# Dynamic molecular epidemiology reveals lineage-associated single-nucleotide variants that alter RNA structure in Chikungunya virus

Thomas Spicher[1], Markus Delitz[1], Adriano de Bernardi Schneider[2], and Michael T. Wolfinger[1,3✉]

[1]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria
[2]AntiViral Research Center, Department of Medicine, University of California San Diego, San Diego, 92103, United States of America
[3]Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währingerstraße 29, 1090 Vienna, Austria

**Chikungunya virus (CHIKV) is an emerging *Alphavirus* which causes millions of human infections every year. Outbreaks have been reported in Africa and Asia since the early 1950s, from three CHIKV lineages: West African, East Central South African, and Asian Urban. As new outbreaks occurred in the Americas, individual strains from the known lineages have evolved, creating new monophyletic groups that generated novel geographic-based lineages. Building on a recently updated phylogeny of CHIKV, we report here the availability of an interactive CHIKV phylodynamics dataset, which is based on more than 900 publicly available CHIKV genomes. We provide an interactive view of CHIKV molecular epidemiology built on Nextstrain, a web-based visualization framework for real-time tracking of pathogen evolution. CHIKV molecular epidemiology reveals single nucleotide variants that change the stability and fold of locally stable RNA structures. We propose alternative RNA structure formation in different CHIKV lineages by predicting more than a dozen RNA elements that are subject to perturbation of the structure ensemble upon variation of a single nucleotide.**

**Chikungunya virus | Molecular epidemiology | Mutation | RNA structure**
**Correspondence: *michael.wolfinger@univie.ac.at***

## Introduction

Chikungunya virus (CHIKV) is an arthropod-borne *Alphavirus* of the family *Togaviridae* that causes millions of human infections every year, particularly in tropic and subtropic regions. CHIKV is the etiological agent of chikungunya fever, an acute febrile illness associated with joint pain, rash, and rarely neurological manifestations (1). CHIKV infection can culminate in chronic arthralgia and arthritis lasting up to several years. CHIKV cycles between vertebrate hosts and hematophagous arthropod vectors, predominantly *Aedes aegypti* and *Aedes albopictus* (2). In Africa CHIKV occurs in an enzootic, sylvatic cycle involving nonhuman primates as hosts, while in Asia CHIKV is mainly maintained in an urban cycle with direct human-mosquito-human transmission (3). The absence of a sylvatic life cycle in Asia suggests that CHIKV originated in Africa and was later carried to Asia (4).

**Geographical spread of Chikungunya virus.** The first documented outbreak of CHIKV was in 1952 on the Makonde Plateau in the Southern Province of Tanganyika (present-day Tanzania) (5, 6). Since then, CHIKV has been emerging in Africa and Asia, with larger outbreaks in the 1960s and 1990s (7). In 2004, CHIKV re-emerged in Kenya and in 2005 an outbreak hit the island of La Réunion, which was the first time that CHIKV occurrences were reported in the southwestern Indian Ocean region (8). The virus, previously assumed to be non-fatal, caused several deaths and also affected neighboring islands including Mayotte, Madagascar, the Seychelles, Comoros, and Mauritius. Following the Indian Ocean islands outbreak, CHIKV spread independently into the Indian Subcontinent and Southeast Asia (9). In the Indian Ocean basin, CHIKV dissemination has been mediated by several mutations, the most prominent being A226V, an amino acid substitution that changes the protein structure of the membrane fusion glycoprotein E1 (10), resulting in increased transmission by *A. albopictus*. As *A. albopictus* is present in temperate regions, this adaptation also has implications on the geographical range of transmission, with CHIKV no longer being bound to tropical and subtropical latitudes (4). Reported cases of CHIKV in Italy, France, Mexico, and the USA, together with the expanding global distribution of *A. albopictus*, facilitated by climate change, raises public health concerns worldwide (11–13).

Early phylogenetic analyses suggested that CHIKV can be separated into three geographically disjoint lineages: West African (WA), East Central South African (ECSA) and Asian Urban Lineage (AUL) (14). Following the 2005 La Réunion outbreak, and the subsequent emergence of CHIKV in India, the existence of a fourth separate lineage, termed Indian Ocean Lineage (IOL), has been proposed (15). While the WA lineage is geographically isolated and shares deep ancestry with the other lineages, investigation of the relationships between taxa belonging to ECSA through phylogenetic inference revealed that this lineage in fact splits into three distinct geographically disjoint epidemic clades (16): The Middle African Lineage (MAL) gave rise to outbreaks in South America (17, 18) and Haiti (19) (South American Lineage, SAL), and the East African Lineage (EAL) gave rise to the IOL. The latter is predominantly found on the Asian continent, except for travel-related cases in which CHIKV has been imported to Europe and North America (3, 20, 21). The third ECSA-derived lineage includes the 1953 Tanganyika strain and encompasses a handful of isolates from Africa and

Asia in a monophyletic group (Africa/Asia Lineage, AAL). The AUL, conversely, is considered a sister clade to all ECSA-derived lineages and has been circulating in Southeast Asia before spreading into Central America and many South American countries from 2013 (3, 22, 23). In 2013, the AUL lineage reached Brazil with the first autochthonous cases being observed in late 2014 (17). Around the same time span 2013/2014 it has emerged in the Caribbean, causing massive spread and nearly three million cases (24). Since then, AUL has spread to multiple regions in America. Interestingly, it is believed that CHIKV was present in the Americas in the 1800s, a courtesy of a spread through navigation, starting in the Caribbean and from there spreading into North and South America, although at the time CHIKV was mislabeled as another febrile disease, dengue (25).

**RNA structure conservation in Chikungunya virus genomes.** CHIKV is a small, spherical, enveloped virus with a single-stranded, (+)-sense RNA genome of approximately 11.8 kb (2, 26) that contains a 5' cap structure and a 3' poly-A tail. The CHIKV genome contains two open reading frames (ORFs) that encode non-structural and structural proteins, respectively, as polyproteins that are post-translationally cleaved (27). The non-structural proteins (nsP1, nsP2, nsP3 and nsP4) constitute the viral replication machinery and are translated from the full-length genome, while the structural proteins (C, E1, E2, E3) form the virus particles and are produced from a subgenomic messenger RNA. The coding sequence is flanked by structured untranslated regions (UTRs) on both ends, which represent the most variable regions of the CHIKV genome (28). This divergence manifests in variable 3'UTR (and thus genome) lengths of individual CHIKV lineages, with ECSA-derived lineages being shorter than WA, and AUL comprising the longest isolates (15). It is plausible to propose patterns of coupled historic mutation and recombination events that eventually resulted in the plasticity observed in present day CHIKV isolates (29). Specific patterns and copy numbers of sequence-level direct repeats in the 3'UTR are characteristic of particular CHIKV lineages and likely represent adaptations of the virus to environmental constraints (30).

Like many other RNA viruses, CHIKV encodes not only viral proteins but also functional RNAs that mediate the viral life cycle. These structured RNAs are found in coding and non-coding regions of the viral genome. A specific fold is often a prerequisite for functional RNAs, and there are selective evolutionary pressures on maintaining these folds, both in coding and non-coding regions (31). As RNA structure is typically conserved at the level of secondary structures, structural homology can be interpreted as a result of evolutionary forces that act on particular RNAs, requiring them to maintain a critical set of structure-determining base pair interactions. In nature, this is achieved by compensatory mutations, i.e. those that maintain base-pair complementarity by a combination of two mutations, e.g. AU → GC, or consistent mutations that change only one pairing partner, e.g. AU → GU. While this kind of structural conservation of functional RNAs, which is also known a covariation, is an ubiquitous trait that is found

in all domains of life, increased mutation rates render viruses particularly interesting in this context.

Although CHIKV is one of the best-studied viruses within the genus *Alphavirus*, knowledge of functional RNA elements and their specific association to viral pathogenesis and replication remains elusive. Unlike other RNA viruses that are characterized by structural conservation of a critical amount of functional RNAs in their UTRs, such as flaviviruses (32) or coronaviruses (33), genus-wide RNA structure conservation does not appear to be prevalent in alphaviruses. In this line, evidence for pervasive RNA structure conservation has not been observed among Sindbis virus (SINV), Venezuelan equine encephalitis virus (VEEV), and CHIKV (34), probably because the genomic location of recognized RNA structure motifs, such as packaging signals, are found at divergent locations in different alphaviruses. However, the apparent absence of covariation patterns among alphavirus species does not exclude the ability of individual species to form highly stable, functional structures. This has been recently affirmed in a genome-wide RNA structure probing study that could confirm known functional RNAs in CHIKV by SHAPE-MaP and characterize several highly structured, potentially functional RNA elements (35).

Likewise, the association between primary sequence and secondary structure in the terminal regions of CHIKV genomes has raised considerable research interest over the last years, mainly motivated by the observation that different CHIKV lineages maintain variable-length 3'UTRs that comprise specific patterns of sequence repeats. While earlier studies identified sequence repeat patterns in the 3'UTR of several alphaviruses (36), recombination by copy-choice mechanisms has been proposed to accelerate CHIKV adaptability, resulting in novel 3'UTR variants (30). In a recent study, we proposed an unambiguous association of sequence repeats in the 3'UTRs of different CHIKV lineages with evolutionarily conserved, structured and unstructured RNA elements (16). Current knowledge about the lack of functional conservation in alphaviruses suggests that potentially functional RNA elements evolved independently in each viral species.

An aspect related to the formation and specificity of RNA structure is the effect of single nucleotide variants (SNVs). These are mutations that can alter the RNA structural ensemble by mediating the base-pairing pattern, potentially resulting in an alternative fold and disrupted functionality. SNVs are sometimes associated with so called riboSNitches, i.e. RNA elements that are subject to perturbation of the structural ensemble resulting in large conformational changes (37, 38). Examples where such events can lead to disease phenotypes in human have been described in the literature (39–42).

**Molecular epidemiology reveals RNA structure-affecting SNVs.** With the availability of large numbers of next generation sequencing data in public databases, multiple efforts to analyze and visualize the spread of infectious diseases have been made (43–45). Nextstrain (https://www.nextstrain.org) (46) is an open source project that makes pathogen phylogenetic data easily accessible to researchers and the interested public, thus facilitating re-

search efforts in the field of pathogen evolution and epidemiology. Nextstrain allows to set up so called community builds that employ the Nextstrain software stack to construct custom real-time phylodynamics resources. Nextstrain community builds have become increasingly popular, and were used, for example, to highlight the genomic epidemiology of the 2018-2020 Ebola virus outbreak in DRC (47), and showcase the mutational dynamics of SARS-CoV-2 superspreading events in Austria (48). We report here the availability of a custom Nextstrain build for CHIKV that encompasses 924 publicly available genomes.

To better understand the evolutionary traits associated with functional RNA conservation among different lineages, we set out to use the CHIKV molecular epidemiology data for studying the impact of lineage-associated sequence variability on viral RNA structure. We were particularly interested in the structural divergence induced by fixed SNVs that are specific to particular lineages, and predict the existence of more than a dozen locally stable RNA elements in the coding regions of the CHIKV genome, whose structural ensemble is substantially altered by lineage-associated SNVs.

## Materials and Methods

**Taxon selection.** We downloaded viral genome and annotation data from the public National Center for Biotechnology Information (NCBI) Genbank database (49) on 23 October 2020. We compiled all temporal and geographic metadata available in these genome records to create data sets for building the Nextstrain instance. Eight sequences of the data set were removed due to missing geographic location metadata or designation as a vaccine sequence. Metadata related to geographic location was associated with the United Nations geoscheme for consistency and labeled in the same way as in de Bernardi Schneider et al. (16). For the temporal analysis we identified two sequences with erroneous sampling dates in the Genbank record. Upon contacting the corresponding authors we were able to correct the dates of the NCBI accessions KX262991.1 and KY435454.1 to 2013 and 2014, respectively.

**Genetic distance.** To calculate the genetic distance within and between Chikungunya lineages we estimated the evolutionary divergence over sequence pairs between groups as implemented in MEGA X under default settings (50, 51). The analyses were conducted using the Maximum Composite Likelihood model (52).

**CHIKV Nextstrain.** We employed the workflow management system Snakemake (53) to build a pipeline for rapid deployment and reproducibility of the CHIKV Nextstrain build. In the first step of the Snakemake workflow, we retrieved metadata including collection date, country, and place of isolation (if available) from the Genbank records of all available CHIKV isolates. Each country was then assigned to one of the following regions: South East Asia, East Asia, South Asia, West Asia, Caribbean, Northern America, Central America, Southern America, Europe, Oceania, Eastern

Africa, Middle Africa, Southern Africa, or Western Africa. In the next step, a file with lineage association for each isolate was created. Isolates with unknown lineage association were assigned to lineages via the the time-resolved phylogenetic tree produced by Nextstrain iteratively. The standard augur pipeline was then applied to construct all relevant data for Nextstrain visualization (46).

**RNA structure modulation via lineage-associated SNVs.** For characterizing SNVs that affect CHIKV RNA structure formation we performed local RNA structure prediction with RNALfold (54) in the reference strain of our Nextstrain build (KT327163.2), limited to sequence lengths of 150 nt and filtered for thermodynamically stable structures. We required a free energy z-score of at least -2 when comparing to 1000 dinucleotide shuffled sequences of the same nucleotide composition, resulting in 138 locally stable candidate structures spread throughout the CHIKV genome. These were then intersected with Nextstrain genome diversity data, yielding a set of 759 candidate RNAs that overlap either one or more variable sites of the CHIKV genome. For each candidate RNA we computed the minimum free energy (MFE) of the non-mutated wild-type (WT) sequence as well as MFEs of all SNV mutants, assessed the base-pair distance between WT and mutant MFE structures with RNAdistance from the ViennaRNA Package (55), employing a base-pair distance cutoff of 15, and filtered for variants that show (almost) complete fixation in one or more lineages. This yielded 14 locally stable RNA elements of the reference strain that overlap a total of 16 lineage-associated SNV, as listed in Table 4. Each SNV was then evaluated for its potential to alter the thermodynamic ensemble of RNA structures with the MutaRNA web server (56) using default parameters.

**Data availability.** The CHIKV Nextstrain instance is available at https://nextstrain.org/community/ViennaRNA/CHIKV. The data can be downloaded by scrolling down to the bottom of the page and clicking on the "Download Data" link.

## Results

**Genetic distance between Chikungunya virus lineages.** Nucleotide divergence analyses based on the average number of base substitutions per site can highlight the proximity among groups of taxa. To get an updated view of the distances within and between CHIKV lineages at the level of nucleotides, we computed the average evolutionary divergence over sequence pairs. Our results show that the nucleotide divergence within each individual lineage is relatively low (< 0.01) with an average of 0.006 (Table 1). The evolutionary divergence between lineages has an average value of 0.073 (Table 2). The WA lineage has a nucleotide divergence to the other lineages that ranges from 0.17-0.19, showing the highest divergence to all other lineages. Also, the major clades encompassing AUL and AUL-Am on one side, and EAL, IOL, MAL, and SAL on the other side present

a nucleotide divergence ranging from 0.065-0.069 between each other. While the genetic divergence between these major clades confirms genotypes that have previously been described in the literature as well-defined lineages, i.e., WA, ECSA, and AUL, lower divergence can be observed between AUL-Am and AUL (0.011), as well as between SAL, MAL, EAL, IOL, AAL, sECSA, ranging from 0.006-0.031.

**Table 1.** Estimates of average evolutionary divergence over sequence pairs within Chikungunya virus lineages. The number of base substitutions per site from averaging over all sequence pairs within each group are shown.

| Lineage | Divergence |
|---------|-----------|
| AUL-Am | 0.0012 |
| AUL | 0.0128 |
| SAL | 0.003 |
| MAL | 0.0107 |
| IOL | 0.0062 |
| EAL | 0.0011 |
| AAL | 0.0066 |
| WA | 0.0102 |

**A Nextstrain build for Chikungunya virus.** In an attempt to provide a publicly available epidemiological and phylogeographical interactive visualization of CHIKV spread, we created a custom Nextstrain build that encompasses all currently available CHIKV genome data. Our community build is available via https://nextstrain.org/community/ViennaRNA/CHIKV and comprises 924 genomes, which represents a substantial increase compared to the 590 genomes considered in the previous most comprehensive study of CHIKV phylogeny (16). The Nextstrain phylogeny is based on a maximum-likelihood tree which is used to infer a timed tree with TreeTime (57) (Figure 1), making available the time of the most recent common ancestor (TMRCA) associated with each individual lineage/major clade of interest (Table 3).

The CHIKV Nextstrain build allows the user to filter data and change the visualization according to preferences, utilizing a set of filters such as date ranges, multiple tree visualization layouts, feature filters and colors. Beside visualizing phylogeography-related characteristics, Nextstrain provides information about the diversity of the underlying sequence data as normalized Shannon entropy in a separate diversity panel. The analysis of nucleotide divergence within the full genome sequences available on the Nextstrain build enabled us to explore lineage-associated genomic variants that change the stability and fold of locally stable RNA structures *in silico*.

**Lineage-specific RNA structures.** To better understand within-species RNA genotype-phenotype associations in viruses, wet set out to assess the impact of lineage-associated, evolutionary fixed SNVs on RNA structure formation in CHIKV. To this end, we performed local RNA secondary structure prediction in the reference strain of our Nextstrain build (KT327163.2, clustering with the AUL-Am lineage),

aiming at characterizing structural elements that show increased thermodynamic stability, as expressed by z-score statistics. We intersected loci that fold into locally stable RNA structures with genome diversity data from Nextstrain to obtain regions of the CHIKV genome that both fold into stable RNA structures and contain one or more single-nucleotide mutations. Each SNV in this set was then characterized in terms of geographic spread and association to specific CHIKV lineages, as well as their base pair distance between non-mutated wild-type MFE structure and mutant MFE structure. Using base-pair distance to pre-filter variants that induce a substantial change in the global fold of the locally stable RNAs, we identified 12 candidate RNAs within the CHIKV coding regions that overlap one SNV, and two candidates that overlap two SNVs each. In total, we have 14 candidate RNAs and 16 fixed SNVs (Table 4).

For each candidate RNA we quantified the effect of mutation-induced changes on the RNA structure ensemble with the `MutaRNA` webserver (56). In addition to comparing the characteristics of wild-type and mutant [1] MFE structures, we assessed the impact of lineage-associated mutations on the entire thermodynamic ensemble of RNA structures by partition function folding.

We pick out two examples that exhibit interesting structural traits: The first example is a variant at nucleotide position 1653 (Figure 2), which has an Adenine (A, wild-type) in the majority of AUL and AUL-Am isolates, while a Guanine (G, mutant) is found at this position in all other lineages. The A1653G variant overlaps a locally stable RNA of 81 nt, whose wild-type sequence folds into a bulged stem-loop structure with an MFE of -23.30 kcal/mol. The mutant folds into a three-way junction structure with an MFE of -28.10 kcal/mol and an equilibrium frequency of 0.3 in the thermodynamic ensemble, suggesting that the mutant is thermodynamically more stable than the wild-type variant. The base pairing potential of wild-type and mutant sequences are depicted as a heat-map dot-plots as well as circular plots in Figure 2. These plots demonstrate the differences in base pairing patterns, highlighting that the SNV considerably modulates the fold of the RNA, thereby enabling the formation of alternatively stacked regions in the mutant that result in thermodynamic stabilization.

Another example encompasses multiple lineage-specific mutations at position 11246, as depicted in Figure 3. While the wild-type (comprising AUL, AUL-Am, AAL, MAL and SAL) has an A at this position, two other variants are clearly associated with different lineages: EAL and IOL have a Cytosine (C) and WA has an Uracil (U) at position 11246. While these variants result in a synonymous mutation at the amino acid level, they induce substantial changes for RNA structure formation. Position 11246 overlaps an RNA element of 80 nt with a wild-type MFE structure of -29.30 kcal/mol and equilibrium frequency of 0.2156. The A11246C and A11246U variants are less stable, with MFE

---

[1] Due to the selection of our reference strain, 'wild-type' refers to isolate KT327163.2, whereas 'mutant' refers to the respective SNV in all candidates listed in Table 4.

**Table 2.** Estimates of evolutionary divergence over sequence pairs between CHIKV lineages. The number of base substitutions per site from averaging over all sequence pairs between groups are shown. Analyses were conducted using the Maximum Composite Likelihood model (52) with 924 CHIKV nucleotide sequences encompassing the lineages/groups observed in this study. Lineages: Asian urban (AUL), AUL-America (AUL-Am), South America (SAL), Middle Africa (MAL), Indian Ocean (IOL), East Africa (EAL), Africa and Asia (AAL), Sister Taxa to ECSA (sECSA), West Africa (WA).

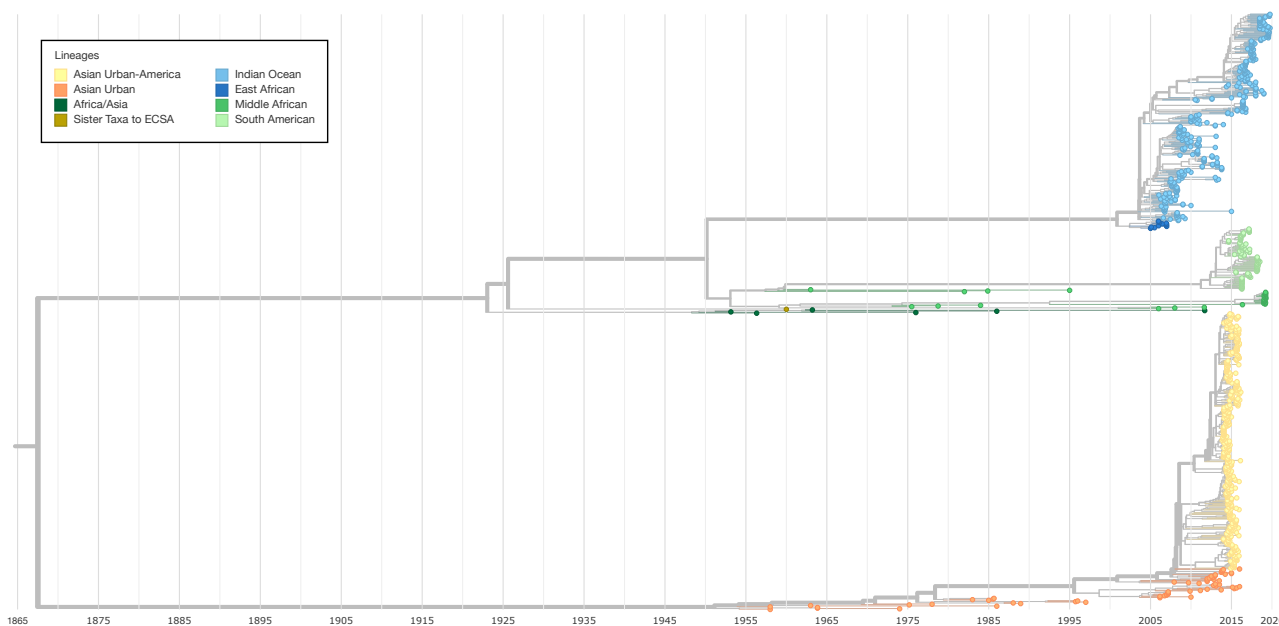| Lineage | AUL-Am | AUL | SAL | MAL | IOL | EAL | AAL | sECSA |
|---|---|---|---|---|---|---|---|---|
| AUL | 0.01108 | | | | | | | |
| SAL | 0.06902 | 0.06622 | | | | | | |
| MAL | 0.06840 | 0.06548 | 0.02432 | | | | | |
| IOL | 0.06988 | 0.06734 | 0.02954 | 0.02631 | | | | |
| EAL | 0.06763 | 0.06533 | 0.02574 | 0.02259 | 0.00584 | | | |
| AAL | 0.06390 | 0.06067 | 0.03145 | 0.02887 | 0.03141 | 0.02807 | | |
| sECSA | 0.06302 | 0.06014 | 0.02957 | 0.02758 | 0.03038 | 0.02690 | 0.01918 | |
| WA | 0.19383 | 0.19228 | 0.17696 | 0.17829 | 0.17917 | 0.17750 | 0.17443 | 0.17505 |



**Fig. 1.** Nextstrain maximum likelihood phylogeny encompassing all CHIKV lineages except the West Africa lineage. The latter is located ancestral to all other clades with an inferred TMRCA date in 1645 (confidence interval 1625–1667, data not shown). Timing has been inferred by TreeTime.

structures of -26.90 kcal/mol (equilibrium frequency 0.0650) and -27.70 kcal/mol (equilibrium frequency 0.1133), respectively. Intriguingly, all variants show varied base pairing patterns in the thermodynamic ensemble, where only 9 stacked base pairs of the closing stem are predicted with high probability.

An interesting observation relates to a C-U mutation at position 10651, which is responsible for the E1-A226V mutation that has been associated with increased CHIKV transmissibility by *A. albopictus*. Although the C10651U mutation overlaps a locally stable region (positions 10594-10682 of the CHIKV genome, z=-3.73), the overall fold of this structure is not altered by the mutation. This suggests that the biological effect is mediated at the protein level rather than the RNA level.

## Discussion

In this contribution we address the question as to what can be learned about CHIKV genotype/phenotype associations by comparative approaches, bringing together different con-

cepts of molecular epidemiology, phylogeny reconstruction, and computational RNA biology. To this end, we build on the Nextstrain (46) framework to provide an interactive phylodynamics resource of CHIKV that reveals spatiotemporal and epidemiological facets of global virus dissemination. Moreover, we employ established tools for RNA structure prediction based on the ViennaRNA (55) Package to infer lineage-specific structural traits.

While the nucleotide divergence within the observed CHIKV lineages is relatively low, this can be explained by the geographical constraint and the reduced collection period for novel lineages. Conversely, the divergence between lineages varies considerably. The highest divergence of the West African lineage when compared to the other CHIKV lineages can be explained by the observation that WA has been an isolated lineage that emerged decades before the more recent outbreaks (14).

However, the divergence between major clades and geographically delineated lineages does raise the question of whether the current nomenclature, as canonized by experts
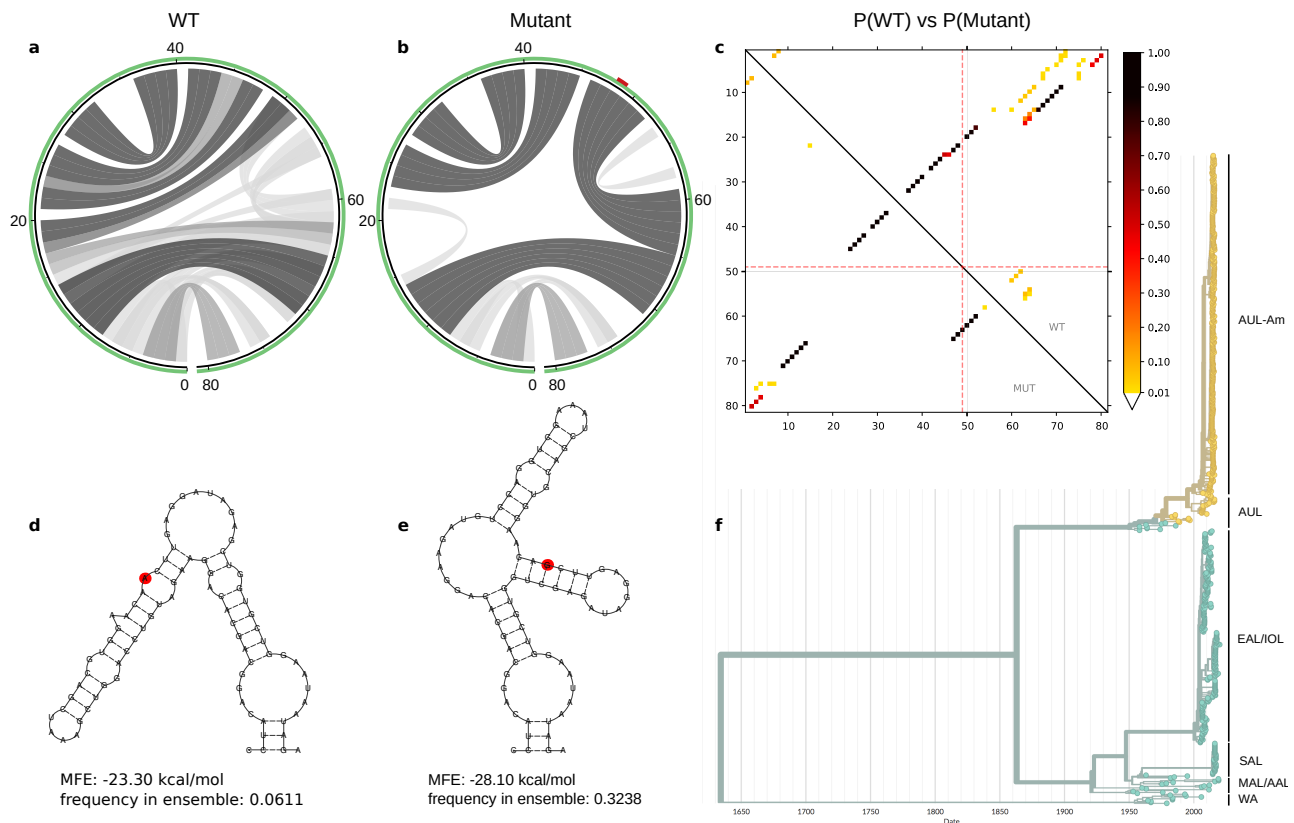
**Fig. 2.** Ensemble properties of the 81ṅt locally stable RNA element that contains the A1653G variant, represented as A49G here. Different patterns of base pair probabilities of the wild type (WT) and mutant are shown as circle plots in **a** and **b**, with darker arcs corresponding to increased base pairing probability. The SNV position in **b** is highlighted in red within the green circle that represents the RNA backbone. The dot plot in **c** shows the differences in base pairing potential of both WT (upper triangle) an mutant (lower triangle), with dark dots indicating high base pairing probability of particular sequence positions. Red dotted lines indicate the mutated position. Minimum free energy structures of the WT and mutant are shown in **d** and **e**, respectively, with the mutated position highlighted in red. The A1653G mutation results in a thermodynamic stabilization by 4.8kcal/mol. This mutation is specific to AUL, and in particular to AUL-Am isolates, highlighted in yellow in the time-resolved phylogenetic tree in **f**.

from the field when first identified, is not misguiding. IOL is an example of a well-established lineage in the literature that presents low divergence from its African origin, but it does introduce the A226V mutation in E1 which mediates the capacity of the virus to replicate in *A. albopictus* (58).

Most viral genotypes/subgenotypes/subclades are based on whole-genome nucleotide divergence of a specific percentage, usually determined by phylogeny showing clades/lineages of a defined 'high' statistical support (>70% bootstrap) (59, 60). Hepatitis B viruses, for example, are classified into genotypes and subgenotypes based on their monophyly, amino acid signatures, and genetic distance (61). In HIV, major distinct clades are classified as major genetic groups, with multiple subtypes within the genetic groups (62).

In the most recent comprehensive phylogeny of CHIKV to date, de Bernardi Schneider et al. (16) analyzed the three major lineages of CHIKV, AUL, IOL, and ECSA. These lineages were broken down based on their monophyly and geographic predominance. Here, we can see that although these strains can still be classified into distinct groups, there should be definite layers, such as lineages/genotypes and sublineages/subgenotypes.

Although we were not aiming at discussing a reclassification of CHIKV into a system independent of geography, we

find the urge to bring to attention that a new coherent system should replace the current taxa classification. Such a new system could assist drug and vaccine development researchers to target specific genotypes or subgenotypes.

In an epidemiological context, the current lineage system, or, respectively, the way the strains are currently grouped, allows the inspection of major outbreak instances and looking at TMRCAs as a way to gauging when a lineage has been introduced in a region, causing outbreaks. From this perspective, our Nextstrain instance provides a reasonable amount of data to investigate deeper the outbreaks that have recently occurred. Importantly, the calculated TMRCA of the major clades is in accordance with previous studies (15). While before December 2013, local CHIKV transmission had not been identified in the Americas (24), our results suggest that the time to the most recent common ancestor for sequences in SAL and AUL-Am were 2011 and 2008, respectively. This result emphasizes the importance of increased surveillance, to identify the virus at the time of introduction, rather than at the time of pandemic (63). The consistency of the dates between Nextstrain TimeTree calculations and previously described TMRCAs in the literature is also reassuring of our ability to provide this additional information on the CHIKV Nexstrain instance.
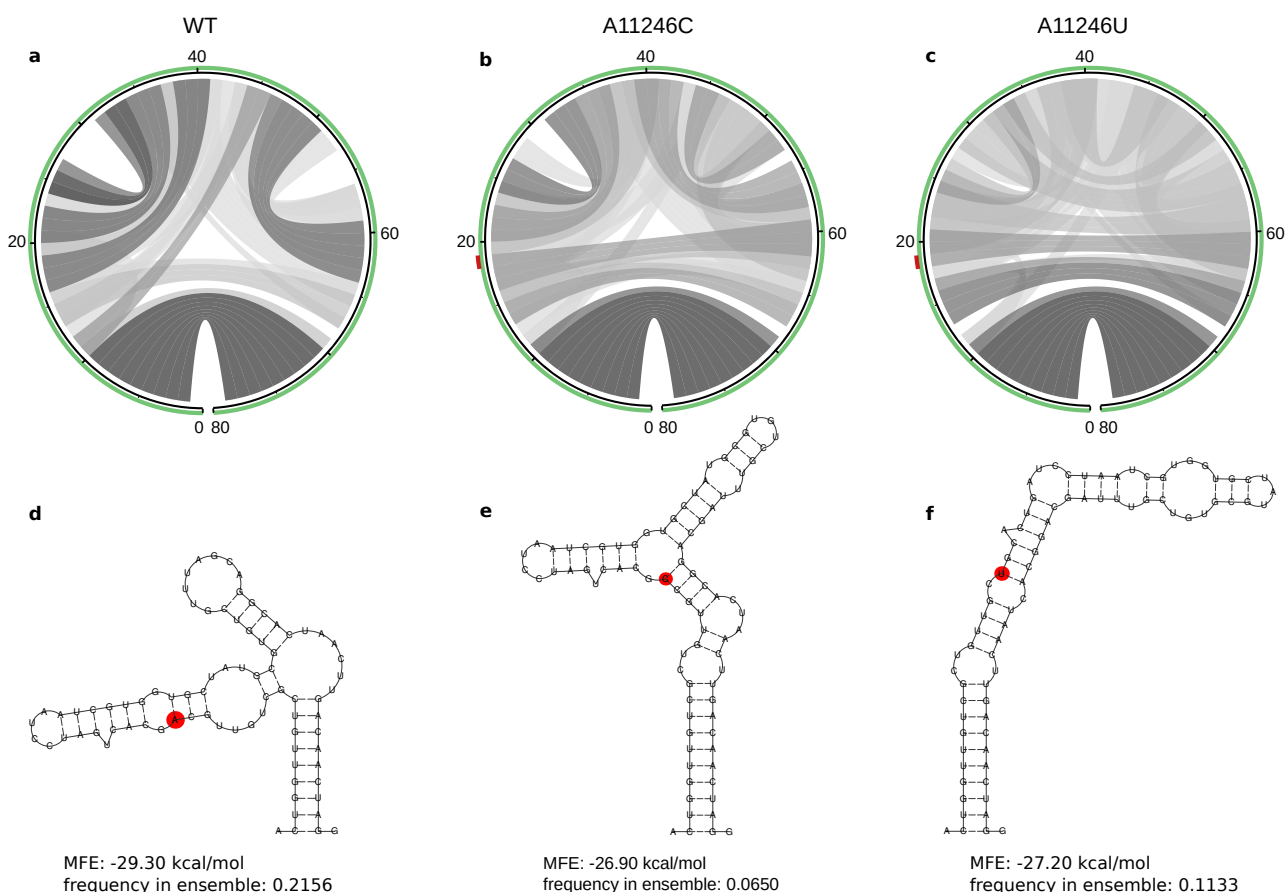
Molecular epidemiology provides a detailed picture of the

**Fig. 3.** Aberration of the thermodynamic ensemble of a locally stable RNA element induced by mutations at position 11446, represented here as A19C and A19U. Structural diversity of the thermodynamic ensembles is shown in circle plots (**a-c**), where hue levels of gray represent pairing probabilities. The closing stem of 9 stacked base pairs is formed in all three variants, while the central parts of the RNA from positions 10–70 show varied base pairing patterns. The SNV at position 19 is highlighted in red on top of the green circles in **b** and **c** as well as in the minimum free energy structures **d-f**.

geographical spread and fixation of RNA variants in viruses and can be used in combination with *in silico* RNA structure prediction to study the structural divergence of different lineages. Owing to the error-prone replication machinery inherent to many RNA viruses, SNVs are created continuously and represent the constitutive driving force behind viral quasispecies (64). Although most of these mutations are considered neutral in an evolutionary context (65), a detailed understanding of the impact and functional associations of lineage-specific SNVs in viruses remain elusive. Single nucleotide mutations can lead to non-synonymous mutations at the amino acid level that result in potentially different protein function. Likewise, nucleotide mutations that culminate in synonymous mutations still have the capacity to alter RNA structure formation, leading to RNA phenotypes that can influence e.g. co-translational protein folding efficiency and thereby mediate viral gene expression patterns (66).

We asked whether Nextstrain can be utilized to infer novel insight into RNA structure-association of individual clades. As Nextstrain is particularly convenient for discriminating characteristics of viral clades/lineages, we set out to expand the sequence-centric approach to computation of lineage-associated structural traits. By combining Nextstrain genome diversity data with RNA structure prediction methods, we could associate sequence variants observed in different CHIKV lineages with alternative RNA structure formation. While we cannot associate lineage-specific SNVs with particular biological functionality at this point, the fixation of nucleotide mutations in certain CHIKV lineages suggests at least that these variants do not have detrimental influence on the viral fitness.

In summary, we show here that the combination of molecular epidemiology data with RNA structure prediction can help to gain insight into hitherto unresolved aspects of genotype-phenotype associations within viral species. On a broader scale, specifically when applied to different viruses, this can augment our understanding of RNA structure evolution.

# Bibliography

1. Nitin Chandak, Rajpal Kashyap, Dinesh Kabra, Poornima Karandikar, Sonali Saha, Shweta Morey, Hemant Purohit, Girdhar Taori, and Hatim Daginawala. Neurological complications of chikungunya virus infection.(original article)(clinical report). *Neurology India*, 57(2):177, 2009. ISSN 0028-3886.
2. NL Forrester, G Palacios, RB Tesh, N Savji, H Guzman, M Sherman, SC Weaver, and WI Lipkin. Genome-scale phylogeny of the alphavirus genus suggests a marine origin. *J Virol*, 86(5):2729–2738, 2011.
3. Scott C Weaver and Naomi L Forrester. Chikungunya: Evolutionary history and recent epidemic spread. *Antivir Res*, 120:32–39, 2015.
4. Zhisheng Her, Yiu-Wing Kam, Raymond T.P Lin, and Lisa F.P Ng. Chikungunya: a bending reality. *Microbes and Infection*, 11(14-15):1165–1176, 2009. ISSN 1286-4579.
5. Marion C. Robinson. An epidemic of virus disease in southern province, tanganyika territory. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 49(1):28–32, 1955.
6. RW Ross. The Newala epidemic: III. The virus: isolation, pathogenic properties and relationship to the epidemic. *Epidemiol Infect*, 54(2):177–191, 1956.

7. Gilles Pialoux, Bernard-Alex Gaüzère, Stéphane Jauréguiberry, and Michel Strobel. Chikungunya, an epidemic arbovirosis. *Lancet Infect Dis*, 7(5):319–327, 2007.

8. P. Renault, J.-L. Solet, D. Sissoko, E. Balleydier, S. Larrieu, L. Filleul, C. Lassalle, J. Thiria, E. Rachou, H. De Valk, D. Ilef, M. Ledrans, I. Quatresous, P. Quenel, and V. Pierre. A major epidemic of chikungunya virus infection on Réunion Island, France, 2005-2006. *American Journal of Tropical Medicine and Hygiene*, 77(4):727–731, 2007. ISSN 00029637.

9. Konstantin A Tsetsarkin, Rubing Chen, and Scott C Weaver. Interspecies transmission and chikungunya virus emergence. *Curr Opin Virol*, 16:143–150, 2016.

10. Isabelle Schuffenecker, Isabelle Iteman, Alain Michault, Séverine Murri, Lionel Frangeul, Marie-Christine Vaney, Rachel Lavenir, Nathalie Pardigon, Jean-Marc Reynes, François Pettinelli, Leon Biscornet, Laure Diancourt, Stéphanie Michel, Stéphane Duquerroy, Ghislaine Guigon, Marie-Pascale Frenkiel, Anne-Claire Bréhin, Nadège Cubito, Philippe Desprès, Frank Kunst, Félix A Rey, Hervé Zeller, and Sylvain Brisse. Genome microevolution of chikungunya viruses causing the indian ocean outbreak (chikungunya virus genome microevolution). *PLoS Medicine*, 3(7):e263, 2006. ISSN 1549-1277.

11. CDC Centers for Disease Control and Prevention. Countries and territories where chikungunya cases have been reported (as of September 17, 2019), 9 2019. https://www.cdc.gov/chikungunya/geo, accessed 06.08.2020.

12. Helene Delatte, Amelie Desvars, Anthony Bouetard, Severine Bord, Geoffrey Gimonneau, Gwenaei Vourc, and Didier Fontenille. Blood-feeding behavior of aedes albopictus, a vector of chikungunya on la reunion.(report). *Vector-Borne and Zoonotic Diseases*, 10(3):249, 2010. ISSN 1530-3667.

13. Ilia Rochlin, Dominick V Ninivaggi, Michael L Hutchinson, and Ary Farajollahi. Climate change and range expansion of the asian tiger mosquito (aedes albopictus) in northeastern usa: implications for public health practitioners. *PloS one*, 8(4):e60874–e60874, 2013. ISSN 1932-6203.

14. Ann M Powers, Aaron C Brault, Robert B Tesh, and Scott C Weaver. Re-emergence of Chikungunya and O'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *J Gen Virol*, 81(2):471–479, 2000.

15. Sara M Volk, Rubing Chen, Konstantin A Tsetsarkin, A Paige Adams, Tzintzuni I Garcia, Amadou A Sall, Farooq Nasar, Amy J Schuh, Edward C Holmes, Stephen Higgs, Payal D Maharaj, Aaron C. Brault, and Scott C. Weaver. Genome-scale phylogenetic analyses of Chikungunya virus reveal independent emergences of recent epidemics and various evolutionary rates. *J Virol*, 84(13):6497–6504, 2010.

16. Adriano de Bernardi Schneider, Roman Ochsenreiter, Reilly Hostager, Ivo L. Hofacker, Daniel Janies, and Michael T Wolfinger. Updated Phylogeny of Chikungunya Virus Suggests Lineage-Specific RNA Architecture. *Viruses*, 11:798, 2019. doi: 10.3390/v11090798. PMCID: PMC6784101.

17. Marcio Roberto Teixeira Nunes, Nuno Rodrigues Faria, Janaina Mota de Vasconcelos, Nick Golding, Moritz UG Kraemer, Layanna Freitas de Oliveira, Raimunda do Socorro da Silva Azevedo, Daisy Elaine Andrade da Silva, Eliana Vieira Pinto da Silva, Sandro Patroca da Silva, et al. Emergence and potential for spread of Chikungunya virus in Brazil. *BMC Med*, 13(1):102, 2015.

18. Maria G Teixeira, Alcina MS Andrade, N Costa Maria da Conceição, Jesuína SM Castro, Francisca LS Oliveira, Cristina SB Goes, Maricelia Maia, Eloisa B Santana, Bruno TD Nunes, and Pedro FC Vasconcelos. East/Central/South African genotype Chikungunya virus, Brazil, 2014. *Emerg Infect Dis*, 21(5):906, 2015.

19. Sarah Keller White, Carla Mavian, Marco Salemi, John Glenn Morris Jr, Maha A Elbadry, Bernard A Okech, John A Lednicky, and James C Dunford. A new "American" subgroup of African-lineage Chikungunya virus detected in and isolated from mosquitoes collected in Haiti, 2016. *PloS One*, 13(5):e0196857, 2018.

20. G Rezza, L Nicoletti, R Angelini, R Romi, AC Finarelli, M Panning, P Cordioli, C Fortuna, S Boros, F Magurano, G Silvi, P Angelini, M Dottori, MG Ciufolini, Gc Majori, and A Cassone. Infection with Chikungunya virus in Italy: an outbreak in a temperate region. *Lancet*, 370(9602):1840–1846, 2007.

21. Robert S. Lanciotti, Olga L. Kosoy, Janeen J. Laven, Amanda J. Panella, Jason O. Velez, Amy J. Lambert, and Grant L. Campbell. Chikungunya virus in US travelers returning from India, 2006. *Emerg Infect Dis*, 13(5):764, 2007. ISSN 1080-6040.

22. X de Lambaillerie, E Leroy, RN Charrel, K Ttsetsarkin, S Higgs, and Ea Gould. Chikungunya virus adapts to tiger mosquito via evolutionary convergence: a sign of things to come? *Virology J*, 5(1), 2008. ISSN 1743422X.

23. Herve Zeller, Wim Van Bortel, and Bertrand Sudre. Chikungunya: its history in Africa and Asia and its spread to new regions in 2013–2014. *J Infect Dis*, 214(suppl_5):S436–S440, 2016.

24. Sergio Yactayo, J Erin Staples, Véronique Millot, Laurence Cibrelus, and Pilar Ramon-Pardo. Epidemiology of chikungunya in the americas. *J Infect Dis*, 214(suppl_5):S441–S445, 2016.

25. Scott B Halstead. Reappearance of chikungunya, formerly called dengue, in the Americas. *Emerg Infect Dis*, 21(4):557, 2015.

26. J H Strauss and E G Strauss. The Alphaviruses: Gene Expression, Replication, and Evolution. *Microbiol Mol Biol R*, 58(3):491–562, 1994.

27. Xiao-Feng Li, Tao Jiang, Yong-Qiang Deng, Hui Zhao, Xue-Dong Yu, Qing Ye, Hong-Jiang Wang, Shun-Ya Zhu, Fu-Chun Zhang, E-De Qin, et al. Complete genome sequence of a Chikungunya virus isolated in Guangdong, China. *J Virol*, 86(16):8904–8905, 2012.

28. Jennifer L Hyde, Rubing Chen, Derek W Trobaugh, Michael S Diamond, Scott C Weaver, William B Klimstra, and Jeffrey Wilusz. The 5' and 3' ends of alphavirus RNAs–non-coding is not non-functional. *Virus Res*, 206:99–107, 2015.

29. Rubing Chen, Eryu Wang, Konstantin A Tsetsarkin, and Scott C Weaver. Chikungunya virus 3' untranslated region: adaptation to mosquitoes and a population bottleneck as major evolutionary forces. *PLoS Pathog*, 9(8):e1003591, 2013.

30. Claudia V Filomatori, Eugenia S Bardossy, Fernando Merwaiss, Yasutsugu Suzuki, Annabelle Henrion, María Carla Saleh, and Diego E Alvarez. RNA recombination at Chikungunya virus 3'UTR as an evolutionary mechanism that provides adaptability. *PLoS Pathog*, 15(4):e1007706, 2019.

31. Michael Kiening, Roman Ochsenreiter, Hans-Jörg Hellinger, Thomas Rattei, Ivo Hofacker, and Dmitrij Frishman. Conserved secondary structures in viral mRNAs. *Viruses*, 11(5):401, 2019.

32. Roman Ochsenreiter, Ivo L Hofacker, and Michael T Wolfinger. Functional RNA Structures in the 3'UTR of Tick-Borne, Insect-Specific and No-Known-Vector Flaviviruses. *Viruses*, 11 (3):298, 2019.

33. Dong Yang and Julian L Leibowitz. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.*, 206:120–133, 2015.

34. Katrina M Kutchko, Emily A Madden, Clayton Morrison, Kenneth S Plante, Wes Sanders, Heather A Vincent, Marta C Cruz Cisneros, Kristin M Long, Nathaniel J Moorman, Mark T Heise, and Alain Laederach. Structural divergence creates new functional features in alphavirus genomes. *Nucleic Acids Res*, 46(7):3657–3670, 2018.

35. Emily A Madden, Kenneth S Plante, Clayton R Morrison, Katrina M Kutchko, Wes Sanders, Kristin M Long, Sharon Taft-Benz, Marta C Cruz Cisneros, Ashlyn Morgan White, Sanjay Sarkar, et al. Using SHAPE-MaP to model RNA secondary structure and identify 3'UTR variation in chikungunya virus. *J Virol*, 2020.

36. Martin Pfeffer, Richard M Kinney, and Oskar-Rüger Kaaden. The alphavirus 3'-nontranslated region: Size heterogeneity and arrangement of repeated sequence elements. *Virology*, 240(1):100–108, 1998.

37. Matthew Halvorsen, Joshua S Martin, Sam Broadaway, and Alain Laederach. Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet*, 6(8):e1001074, 2010.

38. Joshua S Martin, Matthew Halvorsen, Lauren Davis-Neulander, Justin Ritz, Chetna Gopinath, Arthur Beauregard, and Alain Laederach. Structural effects of linkage disequilibrium on the transcriptome. *RNA*, 18(1):77–87, 2012.

39. Yue Wan, Kun Qu, Qiangfeng Cliff Zhang, Ryan A Flynn, Ohad Manor, Zhengqing Ouyang, Jiajing Zhang, Robert C Spitale, Michael P Snyder, Eran Segal, and Howard Y. Chang. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, 505(7485):706–709, 2014.

40. Meredith Corley, Amanda Solem, Kun Qu, Howard Y Chang, and Alain Laederach. Detecting riboSNitches with RNA folding algorithms: a genome-wide benchmark. *Nucleic Acid Res*, 43(3):1859–1868, 2015.

41. Funan He, Ran Wei, Zhan Zhou, Leihuan Huang, Yinan Wang, Jie Tang, Yangyun Zou, Leming Shi, Xun Gu, Melissa J Davis, et al. Integrative Analysis of Somatic Mutations in Non-coding Regions Altering RNA Secondary Structures in Cancer Genomes. *Sci Rep*, 9 (1):1–12, 2019.

42. Jianan Lin, Yang Chen, Yuping Zhang, and Zhengqing Ouyang. Identification and analysis of RNA structural disruptions induced by single nucleotide variants using Riprap and RiboSNitchDB. *NAR Genom Bioinform*, 2(3):lqaa057, 2020.

43. Adriano de Bernardi Schneider, Colby T Ford, Reilly Hostager, John Williams, Michael Cioce, Ümit V Çatalyürek, Joel O Wertheim, and Daniel Janies. Strainhub: A phylogenetic tool to construct pathogen transmission networks. *Bioinformatics*, 36(3):945–947, 2020.

44. Finlay Campbell, Xavier Didelot, Rich Fitzjohn, Neil Ferguson, Anne Cori, and Thibaut Jombart. outbreaker2: a modular platform for outbreak reconstruction. *BMB Bioinf.*, 19(11):1–8, 2018.

45. Nicola De Maio, Colin J Worby, Daniel J Wilson, and Nicole Stoesser. Bayesian reconstruction of transmission within outbreaks using genomic variants. *PLoS Comput. Biol.*, 14(4): e1006117, 2018.

46. James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, 2018.

47. Eddy Kinganda-Lusamaki, Allison Black, Daniel Mukadi, James Hadfield, Placide Mbala-Kingebeni, Catherine Pratt, Amuri Aziza, Moussa Diagne, Bailey White, Nella Bisento, Bibiche Nsunda, Marceline Akonga, Martin Faye, Ousmane Faye, Francois Edidi-Atani, Meris Matondo, Fabrice Mambu, Junior Bulabula, Nicholas Di Paola, Gustavo Palacios, Eric Delaporte, Amadou Alpha Sall, Martine Peeters, Michael Wiley, Steve Ahuka-Mundeke, Trevor Bedford, and Jean-Jaques Muyembe-Tamfum. Operationalizing genomic epidemiology during the Nord-Kivu Ebola outbreak, Democratic Republic of the Congo. *medRxiv*, 2020. doi: 10.1101/2020.06.08.20125567.

48. Alexandra Popa, Jakob-Wendelin Genger, Michael D. Nicholson, Thomas Penz, Daniela Schmid, Stephan W Aberle, Benedikt Agerer, Alexander Lercher, Lukas Endler, Henrique Colaco, Mark Smyth, Michael Schuster, Miguel L. Grau, Francisco Martínez-Jiménez, Oriol Pich, Wegene Borena, Erich Pawelka, Zsofia Keszei, Martin Senekowitsch, Jan Laine, Judith H Aberle, Monika Redlberger-Fritz, Mario Karolyi, Alexander Zoufaly, Sabine Maritschnik, Martin Borkovec, Peter Hufnagl, Manfred Nairz, Günter Weiss, Michael T. Wolfinger, Dorothee von Laer, Giulio Superti-Furga, Nuria Lopez-Bigas, Elisabeth Puchhammer-Stöckl, Franz Allerberger, Franziska Michor, Christoph Bock, and Andreas Bergthaler. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.*, 12, 2020. ISSN 1946-6234. doi: 10.1126/scitranslmed.abe2555. PMID: 33229462.

49. Karen Clark, Ilene Karsch-Mizrachi, David J Lipman, James Ostell, and Eric W Sayers. Genbank. *Nucleic Acids Res*, 44(D1):D67–D72, 2016.

50. Sudhir Kumar, Glen Stecher, and Koichiro Tamura. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*, 33(7):1870–1874, 2016.

51. Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, 35(6): 1547–1549, 2018.

52. Koichiro Tamura, Masatoshi Nei, and Sudhir Kumar. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci USA*, 101(30):11030–11035, 2004.

53. Johannes Köster and Sven Rahmann. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, 08 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts480.

54. Ronny Lorenz and Peter F Stadler. RNA Secondary Structures with Limited Base Pair Span: Exact Backtracking and an Application. *Genes*, 12(14), 2021.

55. Ronny Lorenz, Stephan H Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. ViennaRNA Package 2.0. *Algorithm Mol Biol*, 6(1):26, 2011.

56. Milad Miladi, Martin Raden, Sven Diederichs, and Rolf Backofen. MutaRNA: analysis and visualization of mutation-induced changes in RNA structure. *Nucleic Acids Res*, 37(12):1–5, 2020.

57. Pavel Sagulenko, Vadim Puller, and Richard A Neher. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol*, 4(1):vex042, 2018.

58. Konstantin A Tsetsarkin, Dana L Vanlandingham, Charles E McGee, and Stephen Higgs. A single mutation in Chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog*, 3(12):e201, 2007.

59. Anna L McNaughton, Peter A Revill, Margaret Littlejohn, Philippa C Matthews, and M Azim Ansari. Analysis of genomic-length HBV sequences to determine genotype and subgenotype reference sequences. *J Gen Virol*, pages 1–13, 2020.

60. Nicholas Bbosa, Pontiano Kaleebu, and Deogratius Ssemwanga. Hiv subtype diversity worldwide. *Curr Opin HIV AIDS*, 14(3):153–160, 2019.

61. Adriano de Bernardi Schneider, Carla Osiowy, Reilly Hostager, Henrik Krarup, Malene Borresen, Yasuhito Tanaka, Taylor Morriseau, and Joel O Wertheim. Analysis of Hepatitis B virus genotype D in Greenland suggests the presence of a novel quasi-subgenotype. *Front Microbiol*, 2021. doi: 10.3389/fmicb.2020.602296.

62. David L Robertson, JP Anderson, JA Bradac, JK Carr, B Foley, RK Funkhouser, F Gao, BH Hahn, ML Kalish, C Kuiken, et al. Hiv-1 nomenclature proposal. *Science*, 288(5463): 55–55, 2000.

63. Thiara Manuele Alves Souza, Elzinandes Leal Azeredo, Jessica Badolato-Corrêa, Paulo Vieira Damasco, Carla Santos, Fabienne Petitinga-Paiva, Priscila Conrado Guerra Nunes, Luciana Santos Barbosa, Márcio Costa Cipitelli, Thais Chouin-Carneiro, et al. First report of the East-Central South African genotype of Chikungunya virus in Rio de Janeiro, Brazil. *PLoS Currents*, 9, 2017.

64. Peter Schuster. Quasispecies on fitness landscapes. In Esteban Domingo and Peter Schuster, editors, *Quasispecies: From Theory to Experimental Systems*, pages 61–120. Springer International Publishing, 2016. ISBN 978-3-319-23898-2. doi: 10.1007/82_2015_469.

65. Jemma L Geoghegan and Edward C Holmes. Virus evolution. In Peter M Howley and David M Knipe, editors, *Fields Virology*. Wolters Kluwer, 2021.

66. Guilhem Faure, Aleksey Y Ogurtsov, Svetlana A Shabalina, and Eugene V Koonin. Adaptation of mRNA structure to control protein folding. *RNA Biol*, 14(12):1649–1654, 2017.

**Table 3.** TMRCA estimates of CHIKV lineages, extracted from the Nextstrain instance. Dates are formatted as YYYY-MM-DD.

| Lineage | TMRCA | Date Confidence Interval | Year of First Isolation |
|---------|-------|--------------------------|-------------------------|
| AAL | 1948-04-17 | (1946-10-18,1949-12-14) | 1953 |
| AUL | 1951-02-05 | (1949-09-13,1953-01-04) | 1958 |
| AUL-Am | 2008-03-12 | (2007-10-24,2008-11-10) | 2013 |
| EAL | 2002-05-24 | (2001-02-15,2003-04-20) | 2005 |
| IOL | 2003-08-03 | (2002-10-20,2004-01-14) | 2006 |
| MAL | 1953-01-31 | (1951-05-20,1955-01-14) | 1962 |
| SAL | 2011-03-22 | (2009-06-24,2012-09-15) | 2014 |
| WA | 1954-01-16 | (1952-05-13,1955-09-26) | 1964 |

**Table 4.** Lineage-specific single-nucleotide mutations that modulate the fold of locally stable RNA elements in the CHIKV genome, computed for the reference sequence of our Nextstrain build (accession KT327163.2). The Type column indicates whether a SNV induces a synonymous (S) or non-synonymous (N) mutation at the amino acid level. The Locus column lists global coordinates of the reference strain. Minimum fee energy (MFE) values for wild-type (WT) an mutant (Mut) are given in kcal/mol. Base-pair (BP) distance are computed between MFE structures of WT and Mut. *Subclade. §Isolates collected from 2006 onwards. ‡Isolates collected from 1983 onwards. †Isolates collected from 2019 onwards.

| Variant | Type | Locus | MFE$_{WT}$ | z-score | MFE$_{Mut}$ | BP distance | Lineage association |
|---|---|---|---|---|---|---|---|
| G432A | N | 378–472 | -30.00 | -2.474 | -27.00 | 29 | G: AUL,AUL–Am |
| U810C | S | 783–847 | -21.60 | -2.103 | -18.70 | 42 | U: WA,AUL,AUL–Am,IOL* |
| A1653G | S | 1605–1685 | -23.30 | -2.640 | -28.10 | 15 | A: AUL§,AUL–Am |
| U2122C | S | 2105–2202 | -31.90 | -2.171 | -28.30 | 29 | U: AUL,AUL–Am |
| G2232A | S | 2210–2300 | -31.50 | -2.771 | -27.40 | 36 | G: AUL,AUL–Am |
| C3108U | S | 3093–3192 | -28.20 | -2.141 | -25.60 | 42 | C: WA,AUL,AUL–Am |
| C3682U | S | 3630–3731 | -42.20 | -4.325 | -40.10 | 22 | C: AUL,AUL–Am,IOL* |
| U5508A | N | 5467–5527 | -18.10 | -2.370 | -16.20 | 15 | U: AUL†,AUL–Am |
| G8336C | S | 8312–8395 | -37.10 | -3.023 | -34.40 | 27 | G: AUL†,AUL–Am |
| C8358U | S | 8312–8395 | -37.10 | -3.023 | -36.40 | 27 | C: AUL,AUL–AM,SAL,MAL‡,IOL* |
| G8969A | N | 8918–9019 | -38.60 | -2.447 | -36.10 | 40 | A: AUL†,AUL–Am |
| A9414C | S | 9392–9456 | -17.60 | -2.568 | -15.80 | 18 | A: AUL†,AUL–Am |
| C9197U | S | 9130–9214 | -22.40 | -2.443 | -20.40 | 34 | C: WA,AUL,AUL–AM |
| A10460C | N | 10369–10468 | -31.60 | -2.623 | -31.70 | 18 | A: AUL,AUL–AM |
| A11246C | S | 11228–11307 | -29.30 | -3.049 | -26.90 | 25 | A: AUL,AUL–Am,AAL,SAL,MAL; C: EAL,IOL |
| A11246U | S | 11228–11307 | -29.30 | -3.049 | -27.20 | 36 | A: AUL,AUL–Am,AAL,SAL,MAL; U: WA |