

# Spread of the pandemic Zika virus lineage is associated with NS1 codon usage adaptation in humans

Caio César de Melo Freire<sup>1</sup>, Atila Iamarino<sup>1</sup>, Daniel Ferreira de Lima Neto<sup>1</sup>, Amadou Alpha Sall<sup>2</sup>, and Paolo Marinho de Andrade Zanotto<sup>1\*</sup>

<sup>1</sup>Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Biomedical Sciences Institute, University of Sao Paulo, Sao Paulo, Brazil.

<sup>2</sup>Institut Pasteur de Dakar, Dakar, Senegal.

\*Corresponding author: pzanotto@usp.br.

## ABSTRACT

Zika virus (ZIKV) infections were more common in the zoonotic cycle until the end of the 20<sup>th</sup> century with few human cases in Africa and Southeastern Asia. Recently, the Asian lineage of ZIKV is spreading along human-to-human chains of transmission in the Pacific Islands and in South America. To better understand its recent urban expansion, we compared genetic differences among the lineages. Herein we show that the recent Asian lineage spread is associated with significant NS1 codon usage adaptation to human housekeeping genes, which could facilitate viral replication and increase viral titers. These findings were supported by a significant correlation with growth in Malthusian fitness. Furthermore, we predicted several epitopes in the NS1 protein that are shared between ZIKV and Dengue. Our results imply in a significant dependence of the recent human ZIKV spread on NS1 translational selection.

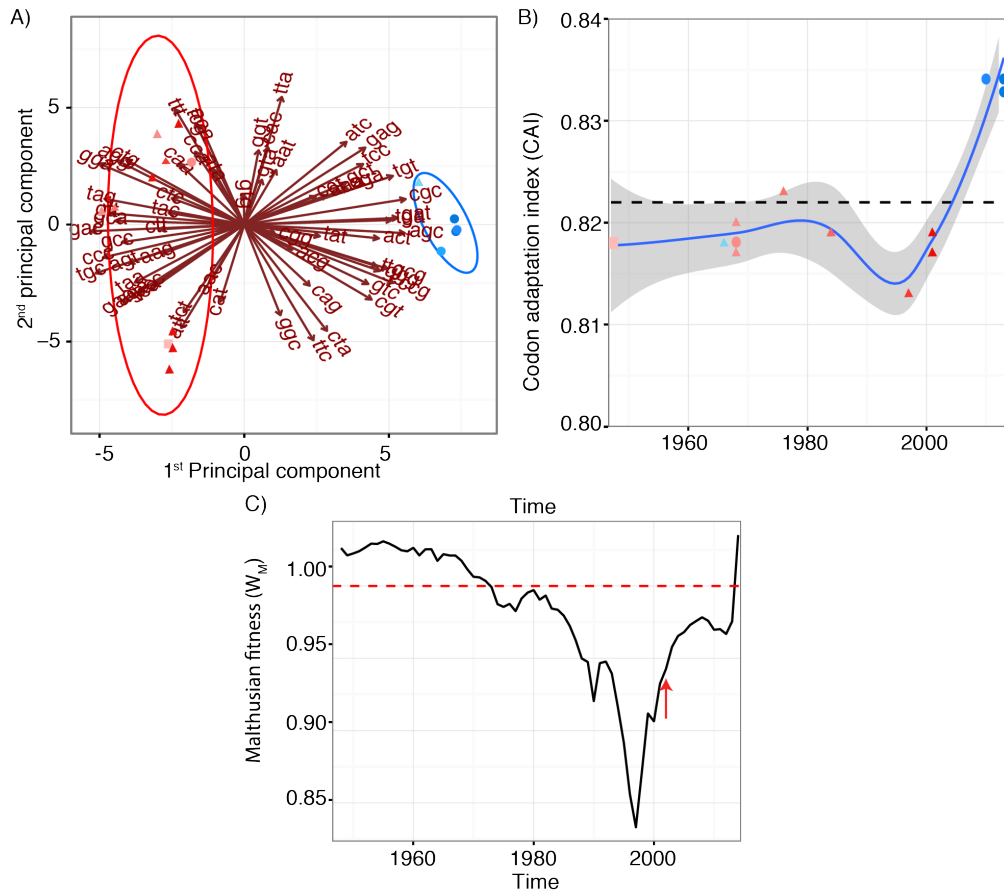
Keywords: Zika virus, emerging diseases, molecular evolution, codon usage adaptation, NS1

## INTRODUCTION

Changes in nucleotide composition have long been noticed as an important evolutionary mechanism and a telltale of viral adaptation to host (Pepin et al., 2010; Plotkin and Kudla, 2011; Longdon et al., 2014). Codon usage adaptation after a host shift event could be required to fine-tune the interactions between a virus and a new host (Longdon et al., 2014; Bahir et al., 2009). Zika virus (ZIKV) was known as a zoonotic pathogen with sporadic human infections in Africa and latter in Southeastern Asia until the end of the last century (Hayes, 2009). In Africa, it remains in a sylvatic cycle involving mainly monkeys and several *Aedes* mosquitoes (Faye et al., 2014). While its Asian lineage is spreading along long chains human-to-human transmission in the Pacific Islands and in South America, vectored mainly by *Aedes aegypti* (Musso et al., 2015). Crucially, the ZIKV pandemic potential is maximized by being also vectored by *A. albopictus* (Grard et al., 2014), a mosquito that explores higher latitudes and transmitted Chikungunya virus in USA and Europe recently (Kuehn, 2014; Grandadam et al., 2011; Delisle et al., 2015). Additionally, sexual intercourse and perinatal infection may be alternative routes of transmission (Besnard et al., 2014; Foy et al., 2011).

The Asian lineage first caused an outbreak of febrile disease in Yap Island, Federated States of Micronesia, in 2007 (Duffy et al., 2009; Hayes, 2009). In 2013 and 2014, it emerged again and caused a large epidemic in French Polynesia (Cao-Lormeau et al., 2014), spreading to Oceania and arriving in America at Easter Island by 2014 (Musso et al., 2015). Recently, in early 2015, it was reported in several Brazilian provinces (Zanluca et al., 2015; Campos et al., 2015), mainly in the Northeastern region. The intense tourism in this regions promotes a massive traffic of people between Brazil and Europe and could help spread ZIKV further, such as when a traveler returned with Zika fever (ZF) from Bahia state in the Northeastern of Brazil to Italy (Zammarchi et al., 2015). ZF symptoms include lasting arthralgia, headaches and mild fever (Zanluca et al., 2015; Campos et al., 2015). The recent outbreaks of ZIKV infections were also associated with a 20-fold increase in Guillain-Barre syndrome cases in French Polynesia (Musso et al., 2014). The increasing in Guillain-Barre cases was also observed in Bahia state, where ZIKV transmission is concomitant with Dengue (DENV) and Chikungunya viruses (CHIV) and ZF incidence reached 275 cases per 100,000 inhabitants until August 2015 (SESAB, 2015). Worryingly, ZIKV was recently associated to the abrupt increase of newborns with microcephaly in Brazil (Ministério da Saúde, 2015).

## RESULTS



**Figure 1. Zika Virus (ZIKV) codon adaptation and fitness according to lineage.** A) RSCU analysis for the polyprotein coding region shows that the principal component (PCA) agrees with the phylogenetic distinctions between the two ZIKV lineages. The African (red) and Asian (blue) lineages were color-coded according to the isolation dates, lighter colors represent older isolation dates. Shapes represent the isolation host: mosquitoes (triangle), monkey (square) or humans (circles). B) NS1 gene Codon Adaptation Index (CAI) to the human housekeeping genes for the African (red) and Asian (blue) lineages according to the isolation dates. C) Malthusian fitness ( $W_M$ ) estimated for ZIKV since 1947, representing decrease ( $W_M < 1$ ), constant population size ( $W_M = 1$ ), and net growth ( $W_M > 1$ ). The red arrow references the end of African lineage sampling. The Spearman correlation coefficients ( $\rho$ ) between the interpolated CAI values in the Figure 1B and the estimated  $W_M$  in Figure 1C were calculated for three time periods: (i) 1948-1970, (ii) 1971-1992 and (iii) 1992-2014. For the former period, we observed a significant negative correlation ( $\rho = -0.59$  and p-value = 0.004); in the second the correlation was significant and positive ( $\rho = 0.46$  and p-value = 0.04) and in the most recent period we found a significant strong positive correlation ( $\rho = 0.90$  and p-value =  $2.70E - 6$ ).

**Codon preferences of ZIKV lineages are distinct** Because codon preferences can strongly affect gene expression (Plotkin and Kudla, 2011), we estimated the relative synonymous codon usage (RSCU) values (Sharp et al., 1986) for each ZIKV gene sequence (Figure 1A). By means of a principal component analysis (PCA) for RSCU values, we found distinct codon preferences in the African and Asian lineages for the entire polyprotein (Figure 1A) and for each viral gene (Figure S1). The extent of the codon bias was inferred by plotting the effective number of codons versus the proportion of GC-content in the third position for each codon (Wright, 1990). As a consequence, we found significant codon usage bias under purifying selective pressure (Wright, 1990), constraining the codon usage in ZIKV (Figure S2A and Figure S3), as found for other arboviruses (Jenkins and Holmes, 2003). The strong purifying selection, which we found at several codon sites (Table S1), was also observed for mosquito transmitted viruses that cause acute infections, mainly alternating between vectors and vertebrate hosts (Hanada et al., 2004). As expected, high amino acid conservation was observed; e.g. 91.8% of the 353 residues of the 17 NS1 proteins analyzed were identical, which is

indicative of purifying selection.

**Evidences of translational selection to human codon usage in NS1 gene of ZIKV epidemic lineages.** Given the significantly distinct codon preferences between the African and Asian ZIKV lineages shown in Figure 1A, and because only the Asian lineage is associated to massive human outbreaks recently (Musso et al., 2015), we further compared virus codon adaptation to host and vector, human housekeeping and *Aedes aegypti* genes. We found that recent Asian epidemic lineages had stronger codon bias on NS1 and NS4A genes, and were also more adapted to humans. This was suggested by measurements of the codon adaptation index (CAI) (Sharp and Li, 1987) for each ZIKV gene for both lineages, unveiling potential viral adaptation to cellular translation machinery of man and mosquitoes. We show in Figure S2B that all ZIKV strains were significantly adapted to humans (CAI values above threshold for the entire polyprotein) while less adapted to *Aedes aegypti* mosquitoes. Moreover, when CAI values were calculated separately for most genes there was little differences between lineages (Figure S4), as was expected for most of the viral genes (Bahir et al., 2009). Nevertheless, codon adaptation for humans in the NS1 coding region from the recent Asian lineage showed a clear increase in CAI values near the present (Figure 1B), coinciding with its spread to Pacific and America. The strong bias in codon usage observed for NS1 from epidemic strains provided additional evidence of translational selection acting on this gene (solid blue circles in Figure S3D).

**High CAI values for the Asian lineage were correlated with the recent ZIKV growth.** This relevant finding was supported by both, (i) the concurrent Malthusian fitness ( $W_M$ ) values above one (Day and Otto, 2001) and, (ii) the significant strong positive correlation ( $\rho = 0.90$  and p-value =  $2.70E - 6$ ) between  $W_M$  and the interpolated values for CAI in the period from 1990 to 2014 (Figure 1B and 1C). Viral codon usage optimization is critical for fine-tuning the interaction with a given host (Longdon et al., 2014), and the most affected genes are usually those highly expressed (Bahir et al., 2009). Therefore, the high NS1 CAI values that we observed for recent Asian ZIKV were considered as a strong indication of adaptive change, which could be associated to improvement in translational efficiency in humans and increased viremia in patients, as observed for Lassa virus (Andersen et al., 2015). Moreover, the NS1 protein is secreted at high levels by infected cells as hexamers that are implicated in immune evasion strategies (Muller and Young, 2013). We obtained similar results on translational selection on NS4A (Figure S3H and S4H). This may be relevant, since NS4A and NS1 appear to play a role in viral replication (Lindenbach and Rice, 1999), while NS4A may enhance viral survival by preventing cell death by the up-regulation of cell autophagy (McLean et al., 2011).

**ZIKV and DENV could share B-cell epitopes.** Another relevant function of the NS1 protein is in assisting in flavivirus immune evasion (Muller and Young, 2013). NS1-specific antibodies are usually found during secondary infections and there is NS1 cross-reactivity between ZIKV and DENV (Lanciotti et al., 2008; Valdés et al., 2000; Muller and Young, 2013), which could impact on pathogenesis. Because *in silico* epitope prediction have been used extensively to develop peptide-based vaccines and investigate immune responses (He and Zhu, 2015), we inferred the structural similarity between the NS1 of DENV and ZIKV by homology modeling (Figure S5A). We found nine linear and five discontinuous epitopes shared in equivalent positions, despite low sequence identity among them (Table S2). Nevertheless, linear epitopes also shared physicochemical properties (Figure S5B). We further calculated the root mean square deviation (RMSD) and performed the global distance test (GDT) for the shared conformational epitopes and found that they were structurally similar, which reinforce the notion that these epitopes may be shared in these phylogenetically closely related viruses (Kuno et al., 1998). These findings could explain the observation of aggravated health conditions in co-infections or secondary infections by ZIKV on DENV pre-exposed people (Roth et al., 2014).

## DISCUSSION

The differences between the African and Asian lineages could explain the emergence of ZIKV in humans and raises concerns about the consequences of the adaptive genetic changes observed in NS1 (Figure 1B) and the recent increase in viral fitness (Figure 1C) (Pepin et al., 2010; Longdon et al., 2014). Moreover, the limited number of human ZIKV cases in Africa could be associated to low viremia in humans, which was demonstrated by a health-officer volunteer experimentally infected with a virus from the African lineage that failed to infect *A. aegypti* mosquitoes (Bearcroft, 1956). Together, our results suggest that fitness gain is associated with improvement of the NS1 translation in humans by synonymous mutations. Synonymous mutations are a common source of variation, given the constrained non-synonymous substitutions rate imposed to RNA viruses that have to negotiate successful infections, alternating between humans and mosquitoes (Hanada et al., 2004). It remains to be evaluated how the NS1 structural and immunological similarities associate to the aggravated symptoms observed when ZIKV and DENV co-circulate (Roth et al., 2014). For this reason, our findings may also be of considerable relevance for the ongoing development of DENV vaccines (McArthur et al., 2013).

## METHODS

**Sequence datasets.** We investigated all 17 available complete genome sequences of ZIKV from GenBank that had information of year and country of isolation (alignment available in <https://github.com/CaioFreire/CUB>). First, we aligned the coding sequences with MACSE program v0.9 (Ranwez et al., 2011) and curated it with SeaView v4.40 (Gouy et al., 2010). During phylodynamic analyses, we employed the most comprehensive dataset for 51 NS5 gene sequences (also available in <https://github.com/CaioFreire/CUB>) sampled from 1947 to 2014, ranging from 14 countries in Africa, Asia, Pacific Islands and America. Since we previously found evidences of recombination in ZIKV from Africa (Faye et al., 2014) and these events could cause potential errors in phylogenetic inferences (Posada and Crandall, 2002), we screened for recombination in NS5 sequences with the RDP program v4.36 (Martin et al., 2010). Identified recombinants were removed of phylogenetic-based analysis. Codon preferences analyses. We employed the relative synonymous codon usage method (Sharp et al., 1986) with the R-package SeqinR v3.13 (Charif and Lobry, 2007) to estimate the codon preferences for each polyprotein gene sequence. In addition, we employed a principal component analysis (PCA) to assess patterns among RSCU values among viral lineages (Su et al., 2009). We identified the most informative codons, which were informative to discriminate among Asian and African lineages, with a biplot graph for the PCA values with the R-package ggbiplot v0.55 (Vu, 2011), using a group probability of 0.95. The different codon preferences between ZIKV lineages were independently confirmed by high support values (> 80%) obtained from hierarchical clustering analysis, using the R-package Pvcust v1.32 (Suzuki and Shimodaira, 2006).

**Codon usage biases.** We calculated the effective number of codons (ENC) with Emboss v6.60 (Rice et al., 2000) and the proportion of guanine-cytosine content in the third base of the codons (GC3), using Seqin{R} program to evaluate the codon usage bias (CUB). The theoretical curve of ENC x GC3 on the genetic drift was estimated with a Perl script to calculate expected ENC and GC3 values (available in <https://github.com/CaioFreire/CUB>), according to (Wright, 1990).

**Codon adaptation of ZIKV genes to humans and *Aedes aegypti* mosquitoes.** In our analysis, CAI is a measure of synonymous codon usage bias based on the codon preference of a viral strain and a codon usage table for a given host (Sharp et al., 1986). To investigate if the codon usage of ZIKV lineages was similar to the hosts in urban settings, regarding humans and *Aedes aegypti*, we calculated the codon adaptation indices (CAI) for each gene from each ZIKV lineage. Since the most pronounced biases are in highly expressed genes (Sharp and Li, 1987; Bahir et al., 2009), we used Emboss to calculate a codon usage table for humans (available in <https://github.com/CaioFreire/CUB>) based on 3803 genes identified as housekeeping (Suzuki and Shimodaira, 2006). Moreover, we calculated CAI for *A. aegypti* using the table available in Codon Usage Database (Nakamura et al., 2000). Importantly, the CAI values obtained with our table based on housekeeping genes were very similar to those found with the table from Codon Usage Database with generic human genes. The CAI values for each sequence from ZIKV genes were calculated with CAIcal program (Puighò et al., 2008). We assessed the confidence of CAI estimates by the calculation of expected CAI values for 500 random sequences with similar GC-content and codon composition for each gene.

**Selection analyses.** We investigated the selection regimens acting on the polyprotein codon sites, calculating the difference ( $\omega$ ) between the estimates of non-synonymous ( $dN$ ) and synonymous ( $dS$ ) substitution rates per codon site. The  $\omega$  values were estimated with single likelihood ancestor counting (SLAC) method with HyPhy program v2.11 (Pond et al., 2005), assuming a significance level ( $\alpha$ ) of 0.05. We employed a maximum likelihood (ML) phylogenetic tree, inferred with GARLI v2.01 (Zwickl, 2006), on NS5 gene alignment without recombinant sequences and the polyprotein gene alignment for the taxa without recombination in the NS5 gene, as input to SLAC. Codon sites under purifying selection were revealed by  $\omega < 0$ , and the opposite is indicative of diversifying selection.

**Phylodynamic analyses.** Using dates of isolation, we were able to estimate a time-scaled Maximum Clade Credibility (MCC) tree for ZIKV NS5 sequences (alignment available in <https://github.com/CaioFreire/CUB>). We used BEAST v1.82 (Drummond et al., 2012), with the evolutionary rate prior ( $\mu$ ) of  $1 \times 10^{-3}$  found previously (Faye et al., 2014). Since purifying selection could underestimate the time to the most recent common ancestor (TMRCA) (Wertheim and Kosakovsky Pond, 2011), we used a substitution model for protein-coding sequences (SRD06) (Shapiro et al., 2006). To infer the demographic history of ZIKV, we employed the Bayesian skyride method (Minin et al., 2008) to estimate the temporal dynamics of effective population size ( $Ne.g$ ) of ZIKV, which approximates the number of infections in time. To reveal the dynamics of viral population size growth, we calculated the Malthusian fitness ( $W_M$ ), which was approximated by the ratio of the population size in sequential time points ( $W_M = Ne.g_t / Ne.g_{t-1} - 1$ ) (Day and Otto, 2001). Moreover, we investigated the correlations between interpolated CAI values for NS1 and  $W_M$ , using the Spearman rank correlation tests in three time intervals: (i) 1947-1969, (ii) 1970-1990, and (iii) 1991-2014.

**Homology modeling and Linear / Discontinuous Epitope Prediction.** These analyses were based on reference sequences, available in GenBank, of NS1 protein from Dengue subtypes 1 to 4 (GenBank accession numbers: AGN94879, AGN94890, ABV03585 and AFX65881) and from ZIKV strains from Senegal and French (GenBank accession numbers: AEN75266.1 and AHZ13508). We aligned the sequences with AliView (Larsson, 2014) and MUSCLE (Edgar, 2004). Sequences with less than 95% identity were selected for each subtype and modeled using YASARA (Krieger and Vriend, 2015) in a BioLinux 8 (Afgan, 2012) with 20 PSI-Blast iterations (e-value = 0.7), considering 6 oligomerization states. 20 templates were downloaded from the Protein Data Bank (PDB - <http://www.rcsb.org/pdb/>) with 5 sequence alignments per template. Modeling was set to low speed with 10 terminal extensions, sampling 50 terminal loops. We checked the produced structures for consistency at the PDBSum server (de Beer et al., 2014) with the Generate option (available at <https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/>) and PROCHECK (Laskowski et al., 1993) stereo-chemical analyses. We calculated relative accessible surface area using the modeled structures with the server GETAREA (Fraczkiewicz and Braun, 1998) (available at <http://curie.utmb.edu/getarea.html>). Linear and discontinuous B Cell epitopes were predicted using the Immune Epitope Database (<http://tools.immuneepitope.org/>). We used the module YASARA view to map on the modeled structures the epitopes found by the IEDB server (Haste Andersen et al., 2006). Linear epitopes were predicted using the Bepipred Linear Epitope Prediction (Larsen et al., 2006). Structural alignments were made to evaluate RMSD and GDT scores between the models for the epitope regions. All results are available in <https://github.com/CaioFreire/CUB>.

### Author contributions

CCMF, AI, DFLN, AAS, and PMAZ designed the experiments and wrote the paper. CCMF and DFLN conducted the experiments. CCMF, AI and DFLN prepared the figures.

### Acknowledgments

We thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for the funding (project #2014/17766-9). CCMF and AI also thank FAPESP for scholarships (#2012/04818-5 and #2014/06090-4). PMAZ holds a CNPq scholarship.

### REFERENCES

- Afgan, E. (2012). Bio-Linux as a tool for bioinformatics training. *Bioinformatics & Bioengineering (BIBE)*.
- Andersen, K. G., Shapiro, B. J., Matranga, C. B., Sealfon, R., Lin, A. E., Moses, L. M., Folarin, O. A., Goba, A., Odia, I., Ehiane, P. E., Momoh, M., England, E. M., Winnicki, S., Branco, L. M., Gire, S. K., Phelan, E., Tariyal, R., Tewhey, R., Omoniwa, O., Fullah, M., Fonnice, R., Fonnice, M., Kanneh, L., Jalloh, S., Gbakie, M., Saffa, S., Karbo, K., Gladden, A. D., Qu, J., Stremlau, M., Nekoui, M., Finucane, H. K., Tabrizi, S., Vitti, J. J., Birren, B., Fitzgerald, M., McCowan, C., Ireland, A., Berlin, A. M., Bochicchio, J., Tazon-Vega, B., Lennon, N. J., Ryan, E. M., Bjornson, Z., Milner Jr., D. A., Lukens, A. K., Broodie, N., Rowland, M., Heinrich, M., Akdag, M., Schieffelin, J. S., Levy, D., Akpan, H., Bausch, D. G., Rubins, K., McCormick, J. B., Lander, E. S., Günther, S., Hensley, L., Okogbenin, S., Schaffner, S. F., Okokhere, P. O., Khan, S. H., Grant, D. S., Akpede, G. O., Asogun, D. A., Gnirke, A., Levin, J. Z., Happi, C. T., Garry, R. F., and Sabeti, P. C. (2015). Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. *Cell*, 162(4):738–750.
- Bahir, I., Fromer, M., Prat, Y., and Linial, M. (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Molecular Systems Biology*, 5(1):311.
- Bearcroft, W. (1956). Zika virus infection experimentally induced in a human volunteer. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 50(5):442–448.
- Besnard, M., Lastère, S., Teissier, A., Cao-Lormeau, V., and Musso, D. (2014). Evidence of perinatal transmission of Zika virus, French Polynesia, December 2013 and February 2014. *Eurosurveillance*, 19(13).
- Campos, G. S., Bandeira, A. C., and Sardi, S. I. (2015). Zika Virus Outbreak, Bahia, Brazil. *Emerging infectious diseases*, 21(10):1885–1886.
- Cao-Lormeau, V.-M., Roche, C., Teissier, A., Robin, E., Berry, A.-L., Mallet, H.-P., Sall, A. A., and Musso, D. (2014). Zika Virus, French Polynesia, South Pacific, 2013. *Emerging infectious diseases*, 20(6):1085–1086.
- Charif, D. and Lobry, J. R. (2007). *Structural approaches to sequence evolution: molecules, networks, populations*. Bastolla.
- Day, T. and Otto, S. P. (2001). *Fitness*. John Wiley & Sons, Ltd, Chichester, UK.
- de Beer, T. A. P., Berka, K., Thornton, J. M., and Laskowski, R. A. (2014). PDBsum additions. *Nucleic Acids Res*, 42(Database issue):D292–6.

- Delisle, E., Rousseau, C., Broche, B., Leparç-Goffart, I., L'Ambert, G., Cochet, A., Prat, C., Foulongne, V., Ferre, J. B., Catelinois, O., Flusin, O., Tchernonog, E., Moussion, I. E., Wiegandt, A., Septfons, A., Mendy, A., Moyano, M. B., Laporte, L., Maurel, J., Jourdain, F., Reynes, J., Paty, M. C., and Golliot, F. (2015). Chikungunya outbreak in Montpellier, France, September to October 2014. *Eurosurveillance*, 20(17).
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8):3–6.
- Duffy, M. R., Chen, T.-H., Hancock, W. T., Powers, A. M., Kool, J. L., Lanciotti, R. S., Pretrick, M., Marfel, M., Holzbauer, S., Dubray, C., Guillaumot, L., Griggs, A., Bel, M., Lambert, A. J., Laven, J., Kosoy, O., Panella, A., Biggerstaff, B. J., Fischer, M., and Hayes, E. B. (2009). Zika Virus Outbreak on Yap Island, Federated States of Micronesia. *New England Journal of Medicine*, 360(24):2536–2543.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–1797.
- Faye, O., Iamarino, A., Freire, C. C. M., Diallo, M., Sall, A. A., Zannotto, P. M. d. A., Faye, O., and de Oliveira, J. V. C. (2014). Molecular evolution of Zika virus during its emergence in the 20(th) century. *PLoS Neglected Tropical Diseases*, 8(1):e2636.
- Foy, B. D., Kobylinski, K. C., Foy, J. L. C., Blitvich, B. J., da Rosa, A. T., Haddow, A. D., Lanciotti, R. S., and Tesh, R. B. (2011). Probable Non-Vector-borne Transmission of Zika Virus, Colorado, USA. *Emerging infectious diseases*, 17(5):880–882.
- Fraczkiewicz, R. and Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of Computational Chemistry*, 19(3).
- Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, 27(2):221–224.
- Grandadam, M., Caro, V., Plumet, S., Thiberge, J.-M., Souarès, Y., Failloux, A.-B., Tolou, H. J., Budelot, M., Cosserat, D., Leparç-Goffart, I., and Desprès, P. (2011). Chikungunya Virus, Southeastern France. *Emerging infectious diseases*, 17(5):910–913.
- Grard, G., Caron, M., Mombo, I. M., Nkoghe, D., Ondo, S. M., Jiolle, D., Fontenille, D., Paupy, C., and Leroy, E. M. (2014). Zika Virus in Gabon (Central Africa) – 2007: A New Threat from *Aedes albopictus*? *PLoS Neglected Tropical Diseases*, 8(2):e2681.
- Hanada, K., Suzuki, Y., and Gojobori, T. (2004). A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Molecular Biology and Evolution*, 21(6):1074–1080.
- Haste Andersen, P., Nielsen, M., and Lund, O. (2006). Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein science : a publication of the Protein Society*, 15(11):2558–2567.
- Hayes, E. B. (2009). Zika Virus Outside Africa. *Emerging infectious diseases*, 15(9):1347–1350.
- He, L. and Zhu, J. (2015). Computational tools for epitope vaccine design and evaluation. *Current Opinion in Virology*, 11:103–112.
- Jenkins, G. M. and Holmes, E. C. (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res*, 92(1):1–7.
- Krieger, E. and Vriend, G. (2015). New ways to boost molecular dynamics simulations. *Journal of computational chemistry*, 36(13):996–1007.
- Kuehn, B. M. (2014). Chikungunya Virus Transmission Found in the United States: US Health Authorities Brace for Wider Spread. *JAMA*, 312(8):776–777.
- Kuno, G., Chang, G. J., Tsuchiya, K. R., Karabatsos, N., and Cropp, C. B. (1998). Phylogeny of the genus *Flavivirus*. *Journal of Virology*, 72(1):73–83.
- Lanciotti, R. S., Kosoy, O. L., Laven, J. J., Velez, J. O., Lambert, A. J., Johnson, A. J., Stanfield, S. M., and Duffy, M. R. (2008). Genetic and Serologic Properties of Zika Virus Associated with an Epidemic, Yap State, Micronesia, 2007. *Emerging infectious diseases*, 14(8):1232–1239.
- Larsen, J. E. P., Lund, O., and Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome research*, 2:2.
- Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22):3276–3278.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2):283–291.
- Lindenbach, B. D. and Rice, C. M. (1999). Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function. *Journal of Virology*, 73(6):4611–4621.

- Longdon, B., Brockhurst, M. a., Welch, J. J., Russell, C. A., Jiggins, F. M., BROCKHURST, M. A., and Longdon, B. (2014). The Evolution and Genetics of Virus Host Shifts. *PLoS Pathogens*, 10.
- Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefevre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics (Oxford, England)*, 26(19):2462–2463.
- McArthur, M. A., Sztejn, M. B., and Edelman, R. (2013). Dengue vaccines: recent developments, ongoing challenges and current candidates. *Expert review of vaccines*, 12(8):933–953.
- McLean, J. E., Wudzinska, A., Datan, E., Quaglino, D., and Zakeri, Z. (2011). Flavivirus NS4A-induced autophagy protects cells against death and enhances virus replication. *The Journal of biological chemistry*, 286(25):22147–22159.
- Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25(7):1459–1471.
- Ministério da Saúde (2015). Ministério divulga boletim epidemiológico. Technical report, Brasília-DF.
- Muller, D. A. and Young, P. R. (2013). The flavivirus NS1 protein: Molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Research*, 98(2):192–208.
- Musso, D., Cao-Lormeau, V.-M., and Gubler, D. J. (2015). Zika virus: following the path of dengue and chikungunya? *The Lancet*, 386(9990):243–244.
- Musso, D., Nilles, E. J., and Cao-Lormeau, V. M. (2014). Rapid spread of emerging Zika virus in the Pacific area. *Clinical Microbiology and Infection*, 20(10):O595–O596.
- Nakamura, Y., Gojobori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research*, 28(1):292–292.
- Pepin, K. M., Lass, S., Pulliam, J. R. C., Read, A. F., and Lloyd-Smith, J. O. (2010). Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nature Reviews Microbiology*, 8(11):802–813.
- Plotkin, J. B. and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, 12(1):32–42.
- Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, 21(5):676–679.
- Posada, D. and Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *Journal of molecular evolution*, 54(3):396–402.
- Puigbò, P., Bravo, I. G., and Garcia-Vallve, S. (2008). CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct*, 3(1):38.
- Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS one*, 6(9):e22594.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG*, 16(6):276–277.
- Roth, A., Mercier, A., Lepers, C., Hoy, D., Duituturaga, S., Benyon, E., Guillaumot, L., and Soares, Y. (2014). Concurrent outbreaks of dengue, chikungunya and Zika virus infections - an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012-2014. *Eurosurveillance*, 19(41).
- SESAB (2015). Situação epidemiológica da dengue, chikungunya e dei/zika. bahia, 2015. Technical report, Salvador.
- Shapiro, B., Rambaut, A., and Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular biology and evolution*, 23(1):7–9.
- Sharp, P. M. and Li, W. H. (1987). The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research*, 15(3):1281–1295.
- Sharp, P. M., Tuohy, T. M. F., and Mosurski, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research*, 14(13):5125–5143.
- Su, M.-W., Lin, H.-M., Yuan, H. S., and Chu, W.-C. (2009). Categorizing host-dependent RNA viruses by principal component analysis of their codon usage preferences. *Journal of computational biology : a journal of computational molecular cell biology*, 16(11):1539–1547.
- Suzuki, R. and Shimodaira, H. (2006). PvcLust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12):1540–1542.
- Valdés, K., Alvarez, M., Pupo, M., Vázquez, S., Rodríguez, R., and Guzmán, M. G. (2000). Human Dengue antibodies against structural and nonstructural proteins. *Clinical and Diagnostic Laboratory Immunology*, 7(5):856–857.
- Vu, V. Q. (2011). *ggbiplot: A ggplot2 based biplot*. R package version 0.55. Available at: <http://github.com/> . . .
- Wertheim, J. O. and Kosakovsky Pond, S. L. (2011). Purifying selection can obscure the ancient age of viral lineages. *Molecular Biology and Evolution*, 28(12):3355–3365.
- Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene*, 87(1):23–29.
- Zammarchi, L., Tappe, D., Fortuna, C., Remoli, M. E., Günther, S., Venturi, G., Bartoloni, A., and Schmidt-Chanasit, J.

- (2015). Zika virus infection in a traveller returning to Europe from Brazil, March 2015. *Eurosurveillance*, 20(23).
- Zanluca, C., Melo, V. C. A. d., Mosimann, A. L. P., Santos, G. I. V. d., Santos, C. N. D. d., Luz, K., Zanluca, C., Melo, V. C. A. d., Mosimann, A. L. P., Santos, G. I. V. d., Santos, C. N. D. d., and Luz, K. (2015). First report of autochthonous transmission of Zika virus in Brazil. *Memórias do Instituto Oswaldo Cruz*, 110(4):569–572.
- Zwickl, D. J. (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD thesis, The University of Texas at Austin, Austin.