

From Mosquitos to Humans: Genetic Evolution of Zika Virus

Lulan Wang,^{1,2,3,10} Stephanie G. Valderramos,^{1,4,10} Aiping Wu,^{2,3} Songying Ouyang,^{1,5} Chunfeng Li,¹ Patricia Brasil,⁶ Myrna Bonaldo,⁷ Thomas Coates,⁸ Karin Nielsen-Saines,⁹ Taijiao Jiang,^{2,3,*} Roghiyh Aliyari,^{1,*} and Genhong Cheng^{1,2,3,*}

¹Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA

²Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China

³Suzhou Institute of Systems Medicine, Suzhou, Jiangsu 215123, China

⁴Department of Obstetrics & Gynecology, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

⁵Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

⁶Instituto Nacional de Infectologia Evandro Chagas, Instituto Oswaldo Cruz – FIOCRUZ, Rio de Janeiro, CEP 21040-900, Brasil

⁷Laboratorio Biologia Molecular de Flavivírus, Instituto Oswaldo Cruz – FIOCRUZ, Rio de Janeiro, CEP 21040-900, Brasil

⁸UCLA Center for World Health, University of California, Los Angeles, Los Angeles, CA 90024, USA

⁹Division of Pediatric Infectious Diseases, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA

¹⁰Co-first author

*Correspondence: jtj@ism.cams.cn (T.J.), aliyarirs@ucla.edu (R.A.), gcheng@mednet.ucla.edu (G.C.)

<http://dx.doi.org/10.1016/j.chom.2016.04.006>

Initially isolated in 1947, Zika virus (ZIKV) has recently emerged as a significant public health concern. Sequence analysis of all 41 known ZIKV RNA open reading frames to date indicates that ZIKV has undergone significant changes in both protein and nucleotide sequences during the past half century.

Zika virus (ZIKV), discovered in 1947, had caused sporadic disease throughout Africa and Asia until the 2007 Micronesia and 2013 French Polynesia outbreaks (Broutet et al., 2016). The rapid expansion of geographic range and increase in severe pathogenicity first noted in the 2015–2016 Brazilian outbreak has raised questions regarding the molecular evolution of this virus. Previously believed to cause only mild disease, mounting evidence points to the capacity of ZIKV to cause neuropathology, including disorders of fetal brain development and Guillain-Barré syndrome (Broutet et al., 2016). In addition to the rise of associated disorders, novel modes of ZIKV transmission have been reported, including maternal-fetal transmission (Brasil et al., 2016; Calvet et al., 2016; Sarno et al., 2016) and sexual transmission (Hills et al., 2016).

ZIKV is a flavivirus closely related to dengue virus (DENV). Its genome is a single-stranded positive-sense RNA molecule of approximately 10,800 base pairs. A single open reading frame (ORF) is flanked by 5' and 3' untranslated regions (UTRs). The resulting single polyprotein is cleaved into the structural proteins capsid (C), pre-membrane protein (prM), and envelope (E) and the non-structural proteins NS1, NS2A, NS2B, NS3, NS4A, 2K, NS4B, and NS5 (Kuno and Chang, 2007). Prior genetic and phylogenetic an-

alyses have identified two main ZIKV lineages, African and Asian, and the recent 21st century epidemics have been traced to the Asian lineage (Faye et al., 2014; Haddow et al., 2012; Lanciotti and Lambert, 2016). Despite circulating throughout Africa and Asia for the latter half of the 20th century, ZIKV infections were not associated with significant human pathology until now. The reasons for this are obscure. It has been hypothesized that the virus may have recently evolved to become more neurotropic, to exhibit increased replicative capacity, and/or to become more transmissible to humans, but causal support for these possibilities is outstanding. To gain a better understanding of the molecular evolution of the virus, we performed detailed phylogenetic and genetic analyses, as well as targeted structural modeling, on all known full-length ORFs of ZIKV available to date (with an emphasis on the recent human strains).

Nucleotide sequences from 41 strains were included in the analysis: 30 human isolates (including two newly reported here), ten mosquito isolates, and one monkey isolate. All sequences greater than 10.1 kb in length were included in the analysis to in order to encompass all complete sequences of ORFs. The strains analyzed, including accession numbers, year, location, and source of isolation,

are listed in Table S1. We first investigated the phylogenetic relationships for all full-length ORF sequences by using maximum likelihood (ML) mapping method with 1,000 replicates. Consistent with prior reports, there were two major lineages of ZIKV, African and Asian (Figure 1A). Interestingly, the African lineage contained eight mosquito isolates, whereas P6-740 (Malaysia/1966) was the sole mosquito isolate in the Asian lineage. All the contemporary human strains share greater sequence homology to P6-740 than Ibh-30656 (Nigeria/1968), suggesting that the ZIKV strains in the recent human outbreak are evolved from the Asian lineage, which is anchored by P6-740 (Figure S1). All of the human strains identified in the 2015–2016 epidemic appear to be more closely related to the H/PF/2013 strain (French Polynesia/2013) than the FSM strain (Micronesia/2007), suggesting that perhaps these two variants have evolved in parallel from a common ancestor. Furthermore, compared to the mosquito strain, 435 and 446 nucleotide changes are evident in FSM and H/PF/2013, respectively, and among them 344 nucleotides are identical. Therefore, the two Asian sub-lineages could have diverged from a common ancestor, arrived in Malaysia, established niches, and later dispersed to South America. It is unclear why the ZIKV strain that already

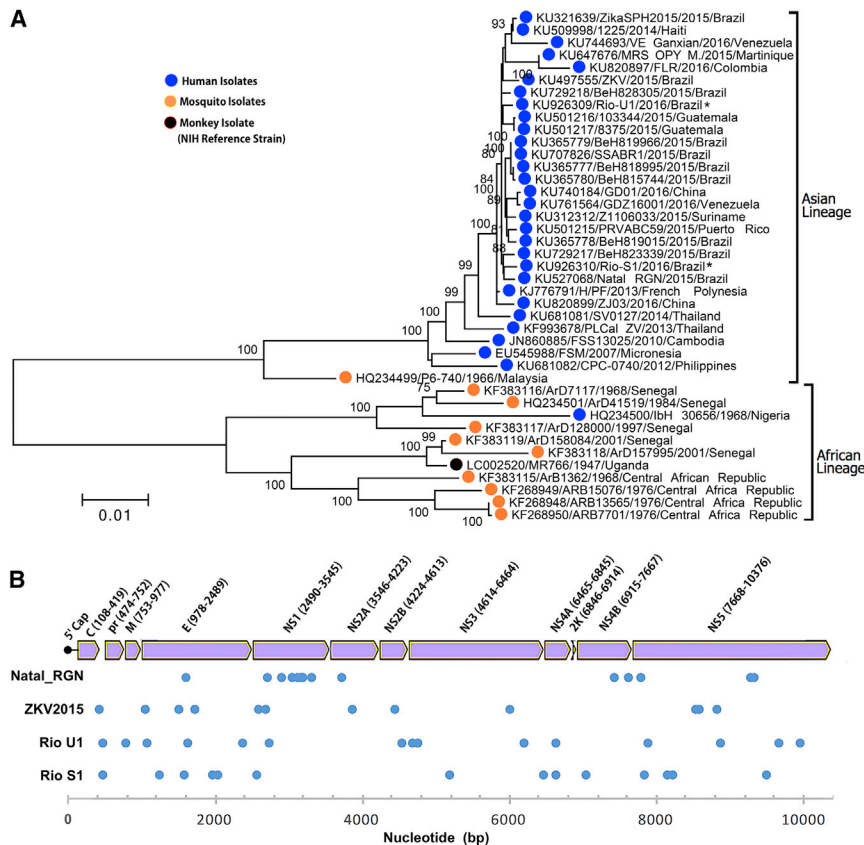


Figure 1. Evolutionary Relationships of Zika Virus

All full-length Zika virus (ZIKV) ORF nucleotide and amino acid sequences (at least 10,100 bp) were obtained from the NIAID Virus Pathogen Database and Analysis Resource (ViPR) (<http://www.viprbrc.org>) and NCBI GenBank (details listed in Table S1).

(A) Phylogenetic tree constructed from nucleotide data from 41 viral complete ORF sequences of ZIKV strains by the maximum-likelihood-method logarithm in MEGA7 based on the Tamura-Nei model. A bootstrap percentage for 1,000 replicates was shown on the left. Branches corresponding to partitions reproduced in less than 70% of bootstrap replicates are not shown. Strains isolated from human, mosquito, and monkey (NIH reference strain) were labeled with blue, orange, and black circles, respectively. The two subtypes were labeled on the right side of the tree. The new strains Rio-U1 and Rio-S1 were highlighted using (*).

(B) Graphical representation of unique nucleotide mutations (blue circle) in Natal_RGN, ZKV2015, Rio-U1, and Rio-S1 strains among all (29 total) current human strains within Asian lineage by using pairwise comparisons. Nucleotide alignments were made using MUSCLE from ViPR. Alignment comparisons were made using Jalview V2.9.

existed in 1966 in Malaysia did not have a significant clinical impact until 50 years later in Oceania. A more rigorous analysis of the potential relationship between the genetic changes and epidemiological topography is required, which will be possible as we gain further sequence information on currently circulating clinical strains and their associated pathology.

We focused further analysis on the nucleotide sequence of four independent human strains with known clinical outcomes. Natal_RGN (KU527068) was isolated from the brain tissue of a fetus with severe microcephaly (Mlakar et al., 2016), and ZKV2015 (KU497555.1) was

isolated from the amniotic fluid of a pregnant patient whose fetus was diagnosed with microcephaly (Calvet et al., 2016). Here, we include two very recent human isolates: Rio-U1 (KU926309.1) and Rio-S1 (KU926310.1) (Bonaldo et al., 2016). Rio-U1 was isolated from the urine of a pregnant woman who presented at 18 weeks' gestation with rash and hand arthralgia, edema, and paresthesias. She recovered acutely from her symptoms, her ultrasound at the time of diagnosis was normal, and she continues to be followed. Rio-S1 was isolated from a man who presented with low grade fever, malaise, rash, conjunctival hyperemia, and

hand and wrist arthralgias, edema, and paresthesias. His illness self-resolved in 10 days. Compared with the 29 other human strains, the number and distribution of unique nucleotide changes are shown in Figure 1B. There were 15, 13, 16, and 15 nucleotide changes in Natal_RGN, ZKV2015, Rio-U1, and Rio-S1, respectively. In pairwise comparisons, there were only three unique amino acid substitutions in Natal_RGN: K940E and T1027A in NS1, and T2509I in NS4B. ZKV2015 had three amino acid substitutions: S550T in E, L1259F in NS2A, and E2831V in NS5. Rio-U1 had only one change: K2039R in NS3. Rio-S1 had three amino acid changes: T625A in E, A2122T in NS4A, and V2688A in NS5.

We were interested in exploring differences in the protein sequence of the African and Asian lineages. Assuming that an African mosquito subtype was the ancestor of the Asian human subtype, we compared the amino acid sequences between eight African strains (seven from mosquitos and one from monkey) and 25 Asian strains isolated from humans. We found that there were 59 amino acid variations located throughout the viral polypeptide sequence that are shared among the individual strains within the African or Asian lineages but are different between these two major lineages (Figure 2A). For comparison, the African human (IbH-30656/Nigeria/1968) and Asian mosquito (P6-740/Malaysia/1966) strains are shown. Our phylogenetic analysis builds upon prior studies by the addition of the most recent human strains to the analysis and further supports the existence of two divergent African and Asian lineages. An important limitation in the analysis of ancestral strain sequences is the potential that these substitutions were adaptations acquired during passages in mouse brains. This is in comparison to modern isolates that were usually sequenced after low passage numbers in monkey or mosquito cell lines. Viral evolution within the murine host is an important question to consider in future experiments, particularly as mouse models are developed to study pathogenesis (Lazear et al., 2016), as well as antiviral and vaccine development. Further studies are needed to elucidate the sequential acquisition of these mutations and their individual contributions to human pathogenesis.

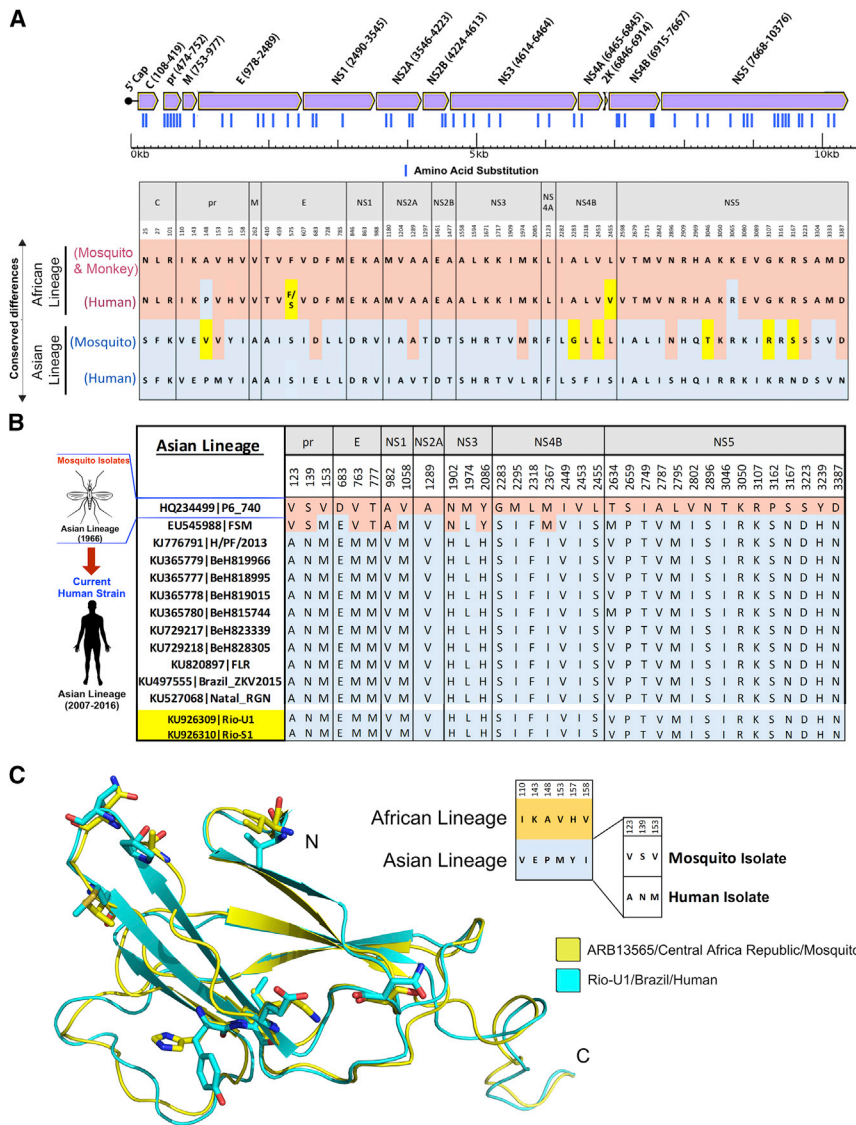


Figure 2. Genetic Evolution of the Asian Lineage and Structural Changes in Pre-Membrane Protein, prM

(A) Graphical illustration of comparison between of Zika virus (ZIKV) from both lineages. Amino acid comparisons were made between African (eight mosquito strains and one monkey strain) and Asian lineages (25 human strains). The sequences are aligned using MUSCLE. Conserved mutations were selected using meta-CATS from ViPR. A graphical map (top bar) of conserved amino acids is shown, represented by blue lines. These sites were in the table below, with the addition of an African human isolate (IbH-30656) and an Asian mosquito isolate (P6-740) for reference. Conserved sites from African and Asian lineages were highlighted in orange and blue, whereas non-related substitutions were highlighted in yellow.

(B) Amino acid substitutions were made from pairwise comparisons of P6-740 against all Asian lineage isolates from humans. All mutations in P6-740 were labeled in orange. Using P6-740 as a reference, identical amino acid substitutions in the human isolates were also highlighted orange, and differences were labeled in blue.

(C) PrM protein of ZIKV shows significant structural alterations. Amino acid substitutions between strains are shown in the table inset. The cartoon represents the predicted overall tertiary structural comparison of ZIKV prM proteins from Rio-U1 (cyan) and ARB13565 (yellow). The automated server CPHmodels-3.0 was employed to build the model according to the homology modeling method. The model was submitted to the SWISS-MODEL Workspace to obtain the 3D structure, then verified using PROCHECK and by Verify3D Structure Evaluation Server and QMEAN. The figure was created in PyMol. The structural templates for the prM protein query sequences were DENV 1 PrM Protein (PDB: 4b03), which shared 48.35% and 50.55% primary sequence identity with the ZIKV PrM proteins from Rio-U1 and ARB13565, respectively. The N and C termini of the structures were labeled with letters. The differences between these two virus strains were shown in sticks.

Due to the recent epidemic and technological advances that now allow rapid and full-length sequencing from direct human isolates, the public bank of human ZIKV sequences has increased from eight at the end of 2014 to 30 as of March 2016. We performed a detailed exploration of the evolution of amino acid polymorphisms of the recent human strains, all of the Asian lineage. Our phylogenetic analysis revealed that all contemporary human isolates share a common ancestor with the P6-740 strain isolated from *Aedes aegypti* mosquito. Comparison of protein sequences using P6-740 as the Asian reference showed that FSM had over 400 variations at the nucleotide level and 26 unique substitutions at the protein level (Figure 2B). Interestingly, when we investigated the sequences of selected human strains identified in more recent epidemics (including FSM, H/PF/2013, and Brazilian strains), we found that all of these strains have acquired changes at an additional eight positions, for a total of 34 amino acid changes compared to P6-740 (Figure 2B). Furthermore, all isolates show identical amino acids at these positions, with the exception of position T2634M/V in the NS5 protein.

Although ZIKV is believed to be primarily transmitted through the mosquito vector, it is interesting to note that no known ZIKV mosquito isolate possesses the same nucleotide sequence as the human strains. One possible explanation is sampling bias, where more recent efforts have focused on isolating the virus from infected humans rather than mosquito arbovirus surveillance. However, it is notable that Duffy et al. were unable to detect ZIKV in mosquitos despite active surveillance during the Micronesia outbreak (Duffy et al., 2009). Alternatively, it is possible that other routes of transmission, such as sexual transmission, may have a greater contribution to the wide spread of ZIKV in the Americas. Intriguingly, it was recently reported that New World strains of *Aedes aegypti* and *albopictus* are poor transmitters of ZIKV (Chouin-Carneiro et al., 2016). Clearly, more studies are urgently needed on natural vector transmission of ZIKV in Asia and the Americas, as well as the possibility of a more prominent contribution of alternative modes of transmission.

In addition to alterations in protein sequences and structures during ZIKV

evolution, nucleotide sequence changes may have an impact on viral genomic stability, replicative efficiency, and thus viral fitness and transmissibility. Strains from the recent epidemic in Brazil showed 14–18 nucleotide mutations compared to the other strains of the Asian lineage isolated from humans. While the nucleotide changes in ZKV2015, Rio-U1, and Rio-S1 are distributed throughout the viral genomic RNA, 50% of the mutations in Natal_RGN, which was isolated from the brain, are located in the NS1 gene. The phenomenon of tissue-specific mutations has been reported for hepatitis C virus, another flavivirus with infectivity to the brain, liver, and blood (Ramachandran et al., 2011). No samples from the brain were taken from the fetus with ZKV2015, and therefore tissue-specific evolution of ZIKV cannot be definitively supported from the available data. However, as more samples isolated from different compartments with known clinical outcomes become available, additional genetic and biochemical assays to determine the potential impact of these changes on viral pathogenesis will be possible.

The pr region of prM protein had the highest percentage variability between the Asian human and the African mosquito subtypes. Six of the 59 (~10%) amino acid variations between these subtypes, namely I110V, K143E, A148P, V153M, H157Y, and V158I, were in the pr region of prM. Furthermore, within the Asian lineage, there were three additional changes in human strains compared to the mosquito strain P6-740: V123A, S139N, and V153M. Structural predictions based on the DENV 1 pr protein (PDB: 4b03) showed significant differences between Rio-U1 and ARB13565 (Figure 2C). Our analysis predicts that A148P could possibly play a critical role in mediating a ten-amino-acid structural change from a loop into a continuous β sheet. This change was only present in human isolates from both lineages, which suggests a potential relevance in human infectivity.

PrM forms a heterodimer with the main viral surface protein, E, in the neutral pH of the lumen of the endoplasmic reticulum (ER). Immature viral particles translocate from the ER to the highly acidic environment of the *trans*-Golgi network where they are packaged into exosomes. The

process of viral maturation takes place in the low-pH environment where viral surface proteins go through a drastic conformational rearrangement due to dissociation of prM-E, formation of E homodimers, and exposure of the prM cleavage site to furin protease. The cleaved pr shields the E protein fusion loop throughout the low-pH condition to prevent secretion of immature particles from the vesicles, and it only dissociates from virions in the extracellular environment (Zhang et al., 2012). The role of prM in viral pathogenesis has been under extensive investigation over the past few years. It has been shown that prM plays a critical role in viral assembly, maturation, heterodimer formation with the E protein, particle secretion, and virulence (Zhang et al., 2012). In our analysis, the six amino acid substitutions in prM (I110V, K143E, A148P, V153M, H157Y, and V158I) resulted in a dramatic predicted structural change of prM between the African and Asian strains. The effect of this structural change on viral function is currently unknown, and further investigations are required to determine whether the observed amino acid changes in prM might have altered the viral pathogenesis of the Asian strain.

It should be noted that Faria et al. have published a recent report detailing a similar phylogenetic analysis on sequences obtained from the Brazilian epidemic (Faria et al., 2016). Our phylogenetic analysis builds upon theirs with additional sequences, as the new strains they reported were available on GenBank when we started this study. However, because detailed clinical information was not available at the time, we did not include these strains in the detailed nucleotide sequence analysis of select strains with known clinical outcomes (Figure 1B). Further, we utilized an alternative approach to structural modeling: rather than mapping, we generated two structural models and overlapped them, leading us to predict the structural change in prM protein.

Our phylogenetic analysis has revealed numerous sequence variations in ZIKV genomes between the African and Asian lineages, as well as among different strains within Asian lineage, as the clinical disease caused by ZIKV has changed from causing only a benign illness to now including severe neuropathology.

Our modeling studies suggest that these sequence variations could mediate specific changes in the prM protein, which could play a role in virulence or improved fitness. In addition, we have narrowed these changes to a reasonable number of amino acid or nucleotide changes that can be tested for their effect on ZIKV infectivity. Future experiments will be required to determine which amino acid or nucleotide substitutions are directly responsible for the possible increased neurotropism, heightened viral fitness, and enhanced transmissibility and infectivity from the mosquito vector to the human host.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one figure and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.chom.2016.04.006>.

ACKNOWLEDGMENTS

We would like to thank Dr. Natalie Quanquin, Dr. Xianzhi Lin, Dr. Elisa Deriu, Sarah Christofersen, and Maxime Chapon for careful reading of the manuscript. This work was supported by the NIH R01, AI069120, AI056154, and AI078389 grants; the Ministry of Science and Technology of China (grants no. 2015ZX09102023 and 2015CB910501); the National Natural Science Foundation of China (81590765); and Institutional Research Fund for Thousand Talents Program at Chinese Academy of Medical Sciences.

REFERENCES

- Bonaldo, M.C., Ribeiro, I.P., Lima, N.S., Santos, A.A.C., Menezes, L.S.R., Cruz, S.O.D., Mello, I.S., Furtado, N.D., Moura, E.E., Damasceno, L., et al. (2016). bioRxiv. Published online March 24, 2016. <http://dx.doi.org/10.1101/045443>.
- Brasil, P., Pereira, J.P., Jr., Raja Gabaglia, C., Damasceno, L., Wakimoto, M., Ribeiro Nogueira, R.M., Carvalho de Sequeira, P., Machado Siqueira, A., Abreu de Carvalho, L.M., Cotrim da Cunha, D., et al. (2016). N. Engl. J. Med. Published online March 4, 2016. <http://dx.doi.org/10.1056/NEJMoa1602412>.
- Broutet, N., Krauer, F., Riesen, M., Khalakhdina, A., Almiron, M., Aldighieri, S., Espinal, M., Low, N., and Dye, C. (2016). N. Engl. J. Med. Published online March 9, 2016. <http://dx.doi.org/10.1056/NEJMp1602708>.
- Calvet, G., Aguiar, R.S., Melo, A.S., Sampaio, S.A., de Filippis, I., Fabri, A., Araujo, E.S., de Sequeira, P.C., de Mendonça, M.C., de Oliveira, L., et al. (2016). Lancet Infect. Dis. Published online February 17, 2016. [http://dx.doi.org/10.1016/S1473-3099\(16\)00095-5](http://dx.doi.org/10.1016/S1473-3099(16)00095-5).
- Chouin-Carneiro, T., Vega-Rua, A., Vazeille, M., Yebakima, A., Girod, R., Goindin, D., Dupont-Rouzeyrol, M., Lourenço-de-Oliveira, R., and Failloux, A.B. (2016). PLoS Negl. Trop. Dis. 10, e0004543.

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., et al. (2009). *N. Engl. J. Med.* **360**, 2536–2543.

Faria, N.R., Azevedo, R.D., Kraemer, M.U.G., Souza, R., Cunha, M.S., Hill, S.C., Thézé, J., Bonsall, M.B., Bowden, T.A., Rissanen, I., et al. (2016). *Science*. Published online March 24, 2016. <http://dx.doi.org/10.1126/science.aaf5036>.

Faye, O., Freire, C.C., Iamarino, A., Faye, O., de Oliveira, J.V., Diallo, M., Zanutto, P.M., and Sall, A.A. (2014). *PLoS Negl. Trop. Dis.* **8**, e2636.

Haddow, A.D., Schuh, A.J., Yasuda, C.Y., Kasper, M.R., Heang, V., Huy, R., Guzman, H., Tesh, R.B.,

and Weaver, S.C. (2012). *PLoS Negl. Trop. Dis.* **6**, e1477.

Hills, S.L., Russell, K., Hennessey, M., Williams, C., Oster, A.M., Fischer, M., and Mead, P. (2016). *MMWR Morb. Mortal. Wkly. Rep.* **65**, 215–216.

Kuno, G., and Chang, G.J. (2007). *Arch. Virol.* **152**, 687–696.

Lanciotti, R.S., and Lambert, A.J. (2016). *Am. J. Trop. Med. Hyg.* **94**, 800–803.

Lazear, H.M., Govero, J., Smith, A.M., Platt, D.J., Fernandez, E., Miner, J.J., and Diamond, M.S. (2016). *Cell Host Microbe*. Published online April 5, 2016. <http://dx.doi.org/10.1016/j.chom.2016.03.010>.

Mlakar, J., Korva, M., Tul, N., Popović, M., Poljšak-Prijatelj, M., Mraz, J., Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T., Fabjan Vodusek, V., et al. (2016). *N. Engl. J. Med.* **374**, 951–958.

Ramachandran, S., Campo, D.S., Dimitrova, Z.E., Xia, G.L., Purdy, M.A., and Khudyakov, Y.E. (2011). *J. Virol.* **85**, 6369–6380.

Sarno, M., Sacramento, G.A., Khouri, R., do Rosário, M.S., Costa, F., Archanjo, G., Santos, L.A., Nery, N., Jr., Vasilakis, N., Ko, A.I., and de Almeida, A.R. (2016). *PLoS Negl. Trop. Dis.* **10**, e0004517.

Zhang, Q., Hunke, C., Yau, Y.H., Seow, V., Lee, S., Tanner, L.B., Guan, X.L., Wenk, M.R., Fibriansah, G., Chew, P.L., et al. (2012). *J. Biol. Chem.* **287**, 40525–40534.