Adaptive Diversification between Yellow Fever Virus West African and South American Lineages: A Genome-Wide Study

Yan Li^{1*} and Zexiao Yang²

¹College of Animal Science and Technology, Sichuan Agricultural University, Wenjiang, People's Republic of China;

²College of Veterinary Medicine, Sichuan Agricultural University, Wenjiang, People's Republic of China

Abstract. Yellow fever virus (YFV) has emerged as the causative agent of a vector-borne disease with devastating mortality in the tropics of Africa and the Americas. YFV phylogenies indicate that the isolates collected from West Africa, East and Central Africa, and South America cluster into different lineages and the virus spread into the Americas from Africa. To determine the nature of genetic variation accompanying the intercontinental epidemic, we performed a genome-wide evolutionary study on the West African and South American lineages of YFV. Our results reveal that adaptive genetic diversification has occurred on viral nonstructural protein 5 (NS5), which is crucially required for viral genome replication, in the early epidemic phase of these currently circulating lineages. Furthermore, major amino acid changes relevant to the adaptive diversification generally cluster in different structural regions of NS5 in a lineage-specific manner. These results suggest that YFV has experienced adaptive diversification in the epidemic spread between the continents and shed insights into the genetic determinants of such diversification, which might be beneficial for understanding the emergence and re-emergence of yellow fever as an important global public health issue.

INTRODUCTION

Yellow fever (YF) is an acute hemorrhagic disease caused by arthropod-transmitted yellow fever virus (YFV). The large scale outbreaks of YF in Africa, North and South America, and Europe resulted in devastating mortality from the fifteenth to nineteenth centuries.^{1,2} YF currently remains an important global public health concern. For example, the burden of YF was 84,000–170,000 severe cases and 29,000–60,000 deaths in tropical areas of Africa and South America during 2013.³

YFV possesses a single-stranded, positive-sense RNA genome. The genome consists of two untranslated regions at the 5'- and 3'- ends, and an open reading frame.^{4,5} The open reading frame is translated as a single polypeptide, which is further cleaved into 10 mature proteins by host cellular signal peptidase and virally encoded proteases. These proteins comprise three structural proteins (C, prM, and E) and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5).

YFV is a member of the genus *Flavivirus*, which comprise one of the four genera in the family *Flaviviridae* together with the genera *Hepaciviruses* (e.g., hepatitis C virus), *Pegivirus* (e.g., hepatitis G virus), and *Pestivirus* (e.g., classical swine fever virus).^{6–8} Based on genetic diversity, YFV isolates can be grouped into three lineages: East and Central Africa, West Africa, and South America.^{9–11} It is noteworthy that the lineages present clear geographic structuring—that is, the isolates collected from East and Central Africa, West Africa, and South America cluster into different lineages. Previous studies have revealed that YFV originated in East or Central Africa and later spread to West Africa.⁹ The current South American lineage arose from a common ancestor shared with the West African lineage and was

disseminated to the Americas from West Africa during the slave trade.^{10,11}

The emergence of YF in the Americas provides a powerful historical demonstration of how the microbes rapidly spread through human agency between continents.¹¹ However, the striking phylogeographic profile indicates that, besides the contribution of global trade and transportation in the modern era, the relevance between genetic diversity of YFVs and the intercontinental epidemic should be considered.

The phylogeographic diversity between the viral lineages could reflect an adaptation in the process of epidemic spread. In this context, an important step forward for further understanding the dissemination of YFV between the continents is to describe the quality and quantity of adaptive genetic diversification that accompany the epidemic spread. In this article, to determine the existence of adaptive diversity between the West African and South American lineages and characterize the changes potentially involved in generating the adaptation, we analyzed the genetic variations occurred in the split of these lineages covering the complete polypeptide-coding region.

MATERIALS AND METHODS

Sequence data collection and alignment. Complete genomes and coding sequences of YFV field isolates were collected from the National Center for Biotechnology Information (NCBI) database. Forty-two complete genome or coding sequences were finally used for following evolutionary characterization (Supplemental Table 1). After these sequences were aligned using ClustalW,¹² the codon reading frames of alignment result were then checked manually and the ambiguous codons were removed.

Phylogenetic analysis of YFV isolates. To infer phylogenetic relationships among the YFV field isolates used in this study, we performed maximum likelihood (ML) and neighbor-joining (NJ) analyses in MEGA5 using the complete coding sequences of prM/E and NS5.¹³ For ML analysis, the general time reversible model of evolution allowing

*Address correspondence to Yan Li, College of Animal Science and Technology, Sichuan Agricultural University, Wenjiang, Chengdu, 611130, People's Republic of China. E-mail: liyan@sicau.edu.cn

shape parameter of a gamma distribution plus a proportion of invariant sites was implemented on each of the datasets.¹⁴ For NJ analysis, the Jukes–Cantor model of nucleotide substitution was used.¹⁵ The reliabilities of the reconstructed trees were evaluated by the bootstrap method with 1,000 replications.

Recombination detection. The search for recombination in 42 YFV isolates was conducted with the RDP v4.56 software package (University of Cape Town, Cape Town, South Africa).¹⁶ Recombination events were identified using the BOOTSCAN,¹⁷ the GENECONV,¹⁸ the Maximum Chi Square (MAXCHI),¹⁹ the Recombination Detection Program,²⁰ and the Sister Scanning (SISCAN) methods.²¹ The *P* value cutoff was set to 0.05 and used throughout. Only recombination event detected with three or more methods was taken into consideration.

McDonald–Kreitman test on the coding sequences of all individual proteins. The McDonald–Kreitman (MK) test for protein coding data has been a commonly used method for detecting the presence of adaptive diversification.²² In this test, if the genetic divergence occurs by neutral processes, the ratio of replacement (nonsynonymous) to synonymous divergence between lineages should be roughly equal to the ratio of replacement to synonymous polymorphism within the lineages. By contrast, if some of the replacements have been influenced by positive adaptive selection, the ratio of replacement to synonymous fixed divergence should be greater than the ratio of replacement to synonymous polymorphism.²²

We adopted the MK test to identify adaptive diversification in all 10 individual proteins separately since the viral proteins perform different functions and might have experienced different evolutionary processes (Chambers and others 1990; Rice and others 1985). Because of the relatively small numbers of polymorphisms and divergences in the individual coding sequences, a Fisher's exact test of independence was performed and significant excess of replacement divergence between the lineages can be used to reject the neutral mutation hypothesis in favor of adaptive diversification.²³

Divergence scan. We estimated nucleotide sequence divergence within YFV West African and South American lineages by means of the Simplot program (Johns Hopkins University, Baltimore, MD).²⁴ YFV lineage reference sequences were randomly chosen as follows: West African lineage: JX898873 and U54798; South American lineage: JF912182 and JF912188. The slide window is 200 nt and moved along the entire coding sequence in steps of 20 nt.

Test for the functional divergence on the coding sequences of putative diversified proteins. Because the changes of amino acid physiochemical characteristics at critical sites could bring about the functional diversification as a response to different environmental challenges,²⁵ we then detected type II functional divergence at homologous sites in the split of the lineages from a common ancestor by means of the method Gu-2006 in DIVERGE 3.0 (Iowa State University, Ames, IA).²⁶ Type II functional divergence represents the evolutionary pattern where amino acids at individual sites are conserved in each lineage but their physiochemical characteristics are different. The method Gu-2006 calculates the coefficient of type II functional divergence (θ_{II}) between the generated lineages,²⁶ and the *z* score test is implemented for statistical evaluation based on the estimate of θ_{II} and its

standard error (θ_{II} SE).²⁷ The null hypothesis is $\theta_{II} = 0$, which means that the absence of this type of functional divergence between member lineages. A rejection of the null hypothesis would indicate that the homologous proteins have potentially experienced functional divergence with the split of the lineages that is, changes of functional roles at the sites involved are likely to have occurred. Furthermore, the posterior probability is used to predict the amino acid sites that could be responsible for these functional differences with a probability cutoff.²⁶

RESULTS

Phylogenetic division of YFV isolates. To identify the general relationships of YFV field isolates used in this study, we reconstructed ML and NJ phylogenetic trees with their prM/E and NS5 coding sequences. The clustering structures of the ML trees are generally similar, which comprise three well-supported clades (Figure 1). The clade divisions appear to have strong geographic structuring between and within continents (Figure 1, Supplemental Table 1). Also, the NJ phylogenetic trees for these coding sequences (not shown) have similar relationships to those based on ML method. Because the phylogeographic features are totally consistent with those found in previous studies,^{10,11} these clades were designated West African, East/Central African, and South American lineages as those used by Chang and others.¹⁰ The West African and South American lineages consist of 19 isolates, but the East/Central African lineage just includes four isolates.

Our following analyses focus only on the West African and South American lineages because of the following. 1) Epidemiologically, these lineages have recently split from a common ancestor and exhibit considerable phylogenetic diversity accompanying the intercontinental epidemic. 2) Methodologically, a small amount of the East/Central African YFV sequences counted tends to yield false estimates of genetic variations.²⁸

Signature of adaptive diversity between YFV West African and South American lineages. We performed the MK test to detect the adaptive diversity between the West African and South American lineages. The estimate of this method is based on a key assumption that there is no recombination between the sets of homologous coding sequences from different phylogenetic clades.²² In this study, a premier concern on the intra-species comparison is that the potential recombination between these lineages would invalidate the assumption. Hence, in the first instance, we carried out a recombination analysis covering the entire coding sequence of the 42 YFV genome.¹⁶ The result indicated that only one recombination event may have taken place, and the recombination donors and recombinant belong to the East/Central African lineage, but not the West African and South American lineages (Supplemental Figure 1). This is consistent with the pattern reported by McGee and others that recombination did not act as an important factor in YFV evolution in nature.²⁹ The combined information suggested that the impact of recombination could be ruled out in the following analyses.

We then implemented the MK test to assay the occurrence of adaptive diversification in all individual proteins. The results showed that there are clear differences in evolutionary pattern among these viral proteins (Table 1). For

FIGURE 1. Maximum likelihood phylogenetic trees based on the complete coding sequences of prM/E (**A**) and NS5 (**B**) extracted from the 42 genomes represented in Supplemental Table 1. The tips are labeled with GenBank accession numbers, hosts, countries, and date of collection. The bootstrap values below 90% are not reported. The designation of lineages follows that of previous result.¹⁰

TABLE 1

McDonald–Kreitman test on yellow fever virus individual coding sequences	Fixed	Polymorphic
WA lineage vs. SA lineage		
C		
Replacement	2	24
Synonymous	11	69
P^*		0.513
prM		
Replacement	0	14
Synonymous	22	124
P		0.220
E		
Replacement	6	49
Synonymous	56	389
P		0.831
NS1		
Replacement	7	19
Synonymous	40	251
P		0.083
NS2A		
Replacement	2	33
Synonymous	38	166
P		0.056
NS2B		
Replacement	1	18
Synonymous	14	99
P		0.473
NS3		
Replacement	3	46
Synonymous	73	489
P		0.184
NS4A		
Replacement	3	14
Synonymous	19	112
P		0.720
NS4B		
Replacement	1	26
Synonymous	36	161
P		0.091
NS5		
Replacement	20	75
Synonymous	97	650
P		0.040

SA lineage = South American lineage; WA lineage = West African lineage.

*Probability of obtaining the observed difference, according to the Fisher's exact test (two tailed), under the null hypothesis of McDonald–Kreitman test.

the coding sequence of NS5, the Fisher's exact test of independence indicated that the ratio of replacement differences to synonymous differences that were fixed between lineages was significantly greater than that of intra-lineage replacement versus synonymous polymorphisms ($P < 0.05$). Additionally, the test on that of NS1 exhibited an excess of replacement divergence between lineages with a marginal probability of 0.083. In contrast, the results from the

other protein coding sequences did not display such pattern of evolution.

Of note, because of the high mutation rate of YFV,³⁰ the estimate of adaptive diversification based on MK test is biased downward by the segregation of low-frequency polymorphisms.³¹ Therefore, it is also feasible that the rapid accumulation of polymorphisms on the other proteins might bring about no significant statistic difference between the compared lineages. To put this into perspective, we constructed similarity plots with SimPlot program to determine how related is one protein of isolates within the West African and South American lineages.²⁴ As shown in Supplemental Figure 2, for each lineage, the sequence homology is generally similar throughout genome. In this situation, the presence of the different evolutionary patterns among the viral proteins is difficult to explain under the hypothesis of different mutation distribution in their coding regions. Thus, the combined results suggested that NS5 and possibly NS1 had experienced adaptive diversification with the segregation process of the West African and South American lineages.

Amino acid sites experienced functional divergence between YFV West African and South American lineages. To detect the signature of adaptive diversification at the individual amino acid sites, we implemented the method Gu-2006 in DIVERGE 3 with the NS5 and NS1 coding sequences of the West African and South American lineages.^{26,27} For NS5, the coefficient of type II functional divergence (θ_{II}) value was larger than 0 with significance ($P < 0.01$) (Table 2). Furthermore, a number of amino acid sites had site-specific scores of the posterior probability ratio that rejected a neutral expectation in favor of functional divergence ($P \geq 0.85$). However, for NS1, the estimate of the θ_{II} did not show a similar pattern of evolution and no site has been subjected to functional divergence (Table 2).

Taken together, the accumulated results suggested that NS5, but not the other proteins, had most likely experienced functional divergence with the split of the West African and South American lineages.

DISCUSSION

The dissemination of disease across geographic space could produce different epidemic features. The objective of this study is to identify the genetic signature of YFV underlying the epidemic spread between the continents. We characterized the genome-wide genetic variations between the West African and South American lineages using MK and DIVERGE functional divergence tests. The MK test

TABLE 2

The estimates of the type II functional divergences of NS1 and NS5 coding sequences between yellow fever virus WA and SA lineages

WA lineage vs. SA lineage	Estimate (z score)	Divergence-related sites
NS1	$\theta_{II}^* = 0.003 \pm 0.015$ (0.200)	None (at $P_{II} \dagger \geq 0.85$)
NS5	$\theta_{II} = 0.030 \pm 0.011$ (2.727, $P \dagger < 0.01$)	173, 276, 291, 567, 880 (at $0.95 > P_{II} \geq 0.90$) 107, 177, 178, 229, 313, 322, 413, 526, 527, 563, 734, 801 (at $0.90 > P_{II} \geq 0.85$)

SA lineage = South American lineage; WA lineage = West African lineage.

* θ_{II} , coefficient of type II functional divergence.

†Probability of obtaining the observed coefficient value of type II functional divergence, according to the Z-score test.

‡ P_{II} , the posterior probability of a site to be type II functional divergence-related.

is generally robust to most demographic factors.^{22,32} The exception is that the slightly deleterious nonsynonymous mutations combined with a rapid expansion in population size can decrease the number of nonsynonymous polymorphisms and thereby yield an artifactual signature of adaptive diversity.^{32,33} For YFV, it has been indicated that the South American lineage arose from a common ancestor shared with the West African lineage and was imported to the Americas following the slave trade approximately 300–400 years ago.^{5,6,10,11} These findings raised further concerns about the implication of the epidemic expansion for adaptive diversity evaluation. It is noteworthy that the effect of a demographic factor, such as expansion in population size, is the same for the different genomic regions in the case of no recombination. In this study, however, the results of the MK test on the coding sequences of NS5 showed a clearly different evolutionary pattern compared with the majority of the other coding sequences (Table 1). Thus, the effect of demographic factors could be roughly ruled out—that is, the excess of amino acid segregations on NS5 between the West African and South American lineages is most likely to be a consequence of adaptive diversification rather than an artifact of demographic factors.

The method Gu-2006 in DIVERGE suggested that the type II functional divergence could have contributed to adaptive changes at a number of sites (Table 2). It is of note that the amino acids at these identified sites were highly conserved within each individual lineage. Because the isolates were collected from different countries over the past several decades (Supplemental Table 1), the adaptive changes at these sites have most likely occurred in the early epidemic phase of these currently circulating lineages. These observations further raise the question of which changes may be relevant to the adaptive diversification accompanying the epidemic spread of the South American lineage and that of the West African lineage. Using the East/Central African lineage as an outgroup,^{9,11} we parsimoniously addressed the phylogenetic distribution of the functional divergence-related changes in these lineages. Thirteen out of 17 identified changes were unambiguously mapped onto the lineages, of which six occurred in the South American lineage and seven in the West African lineage (Figure 2, Table 2).

To shed insight into the functional significance of the putative functional divergence-related changes, we looked at the location of the identified sites in three-dimensional structure. It has long been known that the flavivirus NS5 is composed of an N-terminal methyltransferase domain and a C-terminal RNA-dependent RNA polymerase (RdRp) domain.^{34,35} A hallmark of flavivirus genome is the cap structure at the 5'-untranslated region.³⁶ The methyltransferase domain has the ability to perform methylation reactions required for the cap formation,³⁷ and the RdRp domain is responsible for replicating the viral RNA genome in a cap-dependent manner.³⁸ Because the overall structure of YFV protein NS5 is still lacking and only methyltransferase domain (aa 1–268) are well studied,³⁹ the functional divergence-related changes were mapped on the three-dimensional structures of the YFV methyltransferase domain and the full-length flavivirus NS5 from Japanese encephalitis virus.⁴⁰ We note that four out of six South American lineage-specific adaptive changes cluster in the

FIGURE 2. Adaptive amino acid substitutions in the West African and the South American lineages' NS5s. The substitutions occurring in the methyltransferase domain are shown in green and those in the RNA-dependent RNA polymerase domain are indicated in blue. The numbers of amino acid sites represent their positions in the NS5 of yellow fever virus reference sequence NC_002031.

methyltransferase domain and, in contrast, the majority (six out of seven) of the West African lineage-specific and all four phylogenetically ambiguous changes are distributed on the RdRp domain (aa 275–905, correspond to aa 276–905 of Japanese encephalitis virus NS5) (Figures 2 and 3). The inconsistent pattern between the lineages inferred that they underwent quite different evolutionary fates potentially constituting different functional changes.

In terms of the South American lineage-specific changes, four (R107K, A173G, D177E, and N178S) took place on the methyltransferase domain's bottom face adjacent to RdRp domain and two (K313R and Q880K) occurred on the top face of the RdRp domain (Figures 2, 3A and B).^{39,40} Such structural distribution pattern, in combination with the evolutionary co-occurrence within specific lineage at the early epidemic stage, implied coevolution and functional interaction of these identified sites. For the West African lineage-specific changes, except the substitution I229V lied in the methyltransferase domain (Figure 3A), the others are distributed on several disparate regions of the RdRp domain (Figures 2 and 3C). Some of the changes fall in important functional domains. For example, the identified site 322 lies in the β nuclear localization sequence (aa 322–370, correspond to aa 320–368 of dengue virus serotype 2 NS5) which bring YFV NS5 to the nucleus of host cell and site 801 is in the priming loop, which plays a crucial role in flaviviral RNA synthesis.^{41–43} Notably, in classical swine fever virus of the genus *Pestivirus*, it has been reported that adaptive diversification has occurred on the RdRp NS5B between two major groups, which could be relevant to their virulence diversity.⁴⁴ In hepatitis C virus of the genus *Hepacivirus*, a number of adaptive changes have also taken place on the polymerase NS5B with the split of the predominant subtypes 1a and 1b.⁴⁵ Thus, the accumulated results implied that the changes on the NS5 proteins frequently constitute a genetic signature relevant to the biological diversity and (or) large-scale dissemination of the viruses in the family *Flaviviridae*.

Functionally, the observed adaptive diversification on NS5 calls for two scenarios. First, the changes at the identified sites might contribute to the alteration of interacting

FIGURE 3. Stick model showing the locations of divergence-related sites in (A)³⁹ NS5 methyltransferase and (B and C) the full-length flavivirus NS5 from Japanese encephalitis virus.⁴⁰ In (B), five sites identified specifically in the South American lineage are shown in the methyltransferase and RNA-dependent RNA polymerase domains. In (C), six sites identified specifically in the West African lineage and four phylogenetically ambiguous substitution sites are shown in the RNA-dependent RNA polymerase domain. The methyltransferase domain is shown in green and RNA-dependent RNA polymerase domain is indicated in blue. The residues at the identified sites are shown in purple.

behavior between NS5 and putative function-related factors and (or) template. It has been reported that, during viral replication, the methyltransferase domains of the flavivirus NS5 proteins recruit cellular factors kinases (CK1 and PKG),^{46,47} mitochondrial trifunctional protein (MTP),⁴⁸ and PSD-95/Dlg/ZO-1 proteins,^{49,50} while the RdRp domains can interact with the nuclear import receptors importin- α and importin- β , cyclophilin A,⁵² heat shock protein 70 (Hsp70),⁵³ protein kinase G (PKG),⁴⁰ L subunit of human eukaryotic translation initiation factor 3 (eIF3L),⁵⁴ and small nuclear ribonucleoprotein U1A.⁵⁵ These factors can enhance formation of the replication complex, RNA stability, genome translation, and/or posttranslational modification.^{56,57} Considering the remarkable difference in YFV transmission cycles between Africa and South America,^{58,59} the existence of the adaptive changes might favor the interaction between NS5 and cellular factors from different host and (or) vector. The second scenario envisages the arms race between hosts and YFV. The NS5 proteins of different flaviviruses have been shown to antagonize the host antiviral response by blocking the type I interferon (IFN-I or IFN α/β) pathway.^{50,61–63} In these anti-immune strategies, the NS5 proteins can bind to the signal transducer and activator of transcription 1 (STAT1) and/or STAT2, two critical components of the IFN-I signaling pathway, and mediate IFN antagonism which may influence the virulence of flaviviruses. Additionally, with regard to microbial anti-immunology, previous studies have suggested that pathogens have evolved various virulence strategies including blockage of signaling pathways to neutralize host immune responses.⁶⁴ Taken together, in these contexts, the residue changes at the identified sites might influence the interaction of NS5 with cellular factors and (or) the components of host immune defense, then constituting or maintaining functional adaptations of YFV in epidemics.

Epidemiologically, it is noteworthy that, in addition to the striking phylogeographic relationships, currently circulating YFVs exhibit considerably different activity of epidemics in West Africa and South America.^{3,59,65} From the early twentieth century, the majority of YF infections occurred in the former, including some very large epidemics. In contrast, there have been only a small number of cases reported in the latter. Epidemic factors, including climate, distribution and abundance of vectors, and prevention efforts (e.g., vector control and vaccination), which influence the transmission of YFV, have been considered.^{58,59} Unfortunately, there is no clear relationship between these factors and the distinct epidemic activity on the two continents.⁵⁸ For example, the efforts to control the disease via the eradication of its mosquito vector, *Aedes aegypti*, have begun to senesce from the 1960s. The failure to eradicate mosquito resulted in the re-infestation of ranges, where it had never previously occurred.^{58,66} Furthermore, some countries in South America have estimated YF vaccination coverage well below the 80% threshold, which is required for a population to effectively prevent an epidemic.⁶⁷ Consequently, about 400 million people live in the Americas at risk, which are similar to that (~500 million people) in Africa.³ Taken together, these data suggested that the cause of the epidemic discrepancy between West Africa and South America needs to be elucidated with approaches from multiple disciplines. The adaptive changes identified in this study dropped helpful hints for the decisive genetic factors underlying the dissemination and the diversity of epidemic activity of YFV between the continents.

In summary, our analyses provided the evidence that adaptive diversification has occurred on the viral protein NS5 between the West African and South American

lineages. In addition, the majority of lineage-specific changes presented different distribution patterns on NS5 structural regions. Certainly, future structural and functional studies will be pivotal to understand the impacts of these adaptive genetic changes. If the inferences are experimentally supported, it might extend our knowledge regarding the molecular basis for the emergence and re-emergence of YF.

Received August 25, 2016. Accepted for publication November 10, 2016.

Published online January 2, 2017.

Note: Supplemental table and figures appear at www.ajtmh.org.

Authors' addresses: Yan Li, College of Animal Science and Technology, Sichuan Agricultural University, Wenjiang, People's Republic of China, E-mail: liyan@sicau.edu.cn. Zexiao Yang, College of Veterinary Medicine, Sichuan Agricultural University, Wenjiang, People's Republic of China, E-mail: yzxyang2003@126.com.

REFERENCES

1. Scott HH, 1939. *A History of Tropical Medicine*. Based on the Fitzpatrick Lectures Delivered Before the Royal College of Physicians of London 1937–38. London, UK: Edward Arnold and Co.
2. Monath TP, 1988. *The Arboviruses: Epidemiology and Ecology*, Vol. V. Boca Raton, FL: CRC Press.
3. WHO, 2016. *Yellow Fever*. Geneva, Switzerland: World Health Organization. Available at: <http://www.who.int/mediacentre/factsheets/fs100/en/>. Accessed May 1, 2016.
4. Chambers TJ, Hahn CS, Galler R, Rice CM, 1990. Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol* 44: 649–688.
5. Rice CM, Lenches EM, Eddy SR, Shin SJ, Sheets RL, Strauss JH, 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* 229: 726–733.
6. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, 2012. *Family - Flaviviridae*. *Virus Taxonomy*. San Diego, CA: Elsevier, 1003–1020.
7. Adams MJ, Lefkowitz EJ, King AM, Carstens EB, 2013. Recently agreed changes to the international code of virus classification and nomenclature. *Arch Virol* 158: 2633–2639.
8. Stapleton JT, Fong S, Muerhoff AS, Bukh J, Simmonds P, 2011. The GB viruses: A review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *J Gen Virol* 92: 233–246.
9. Mutebi JP, Wang H, Li L, Bryant JE, Barrett AD, 2001. Phylogenetic and evolutionary relationships among yellow fever virus isolates in Africa. *J Virol* 75: 6999–7008.
10. Chang GJ, Cropp BC, Kinney RM, Trent DW, Gubler DJ, 1995. Nucleotide sequence variation of the envelope protein gene identifies two distinct genotypes of yellow fever virus. *J Virol* 69: 5773–5780.
11. Bryant JE, Holmes EC, Barrett AD, 2007. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog* 3: e75.
12. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.
13. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
14. Lanave C, Preparata G, Saccone C, Serio G, 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol* 20: 86–93.
15. Jukes TH, Cantor CR, 1969. Chapter 24, *Evolution of Protein Molecules*. Munro HN, ed. *Mammalian Protein Metabolism*. New York, NY: Academic Press, 21–132.
16. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B, 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1: 1–5.
17. Martin DP, Posada D, Crandall KA, Williamson C, 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* 21: 98–102.
18. Padidam M, Sawyer S, Fauquet CM, 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218–225.
19. Smith JM, 1992. Analyzing the mosaic structure of genes. *J Mol Evol* 34: 126–129.
20. Martin D, Rybicki E, 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16: 562–563.
21. Gibbs MJ, Armstrong JS, Gibbs AJ, 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16: 573–582.
22. McDonald JH, Kreitman M, 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351: 652–654.
23. McDonald JH, 2009. *Handbook of Biological Statistics*. Baltimore, MD: Sparky House Publishing.
24. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC, 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73: 152–160.
25. Li W-H, 1983. Evolution of duplicate genes and pseudo genes. Nei M, Koehn RK, eds. *Evolution of Genes and Proteins*. Sunderland, MA: Sinauer Associates, 14–37.
26. Gu X, 2006. A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol Biol Evol* 23: 1937–1945.
27. Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z, Zeng Y, 2013. An update of DIVERGE software for functional divergence analysis of protein family. *Mol Biol Evol* 30: 1713–1719.
28. Li W-H, 1997. *Molecular Evolution*. Sunderland, MA: Sinauer Associates.
29. McGee CE, Tsetsarkin KA, Guy B, Lang J, Plante K, Vanlandingham DL, Higgs S, 2011. Stability of yellow fever virus under recombinatory pressure as compared with chikungunya virus. *PLoS One* 6: e23247.
30. Jenkins GM, Rambaut A, Pybus OG, Holmes EC, 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol* 54: 156–165.
31. Fay JC, Wyckoff GJ, Wu CI, 2001. Positive and negative selection on the human genome. *Genetics* 158: 1227–1234.
32. Eyre-Walker A, 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol* 21: 569–575.
33. Nielsen R, 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity (Edinb)* 86: 641–647.
34. Koonin EV, 1993. Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and lambda 2 protein of reovirus. *J Gen Virol* 74: 733–740.
35. Koonin EV, 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72: 2197–2206.
36. Cleaves GR, Dubin DT, 1979. Methylation status of intracellular dengue type 2 40 S RNA. *Virology* 96: 159–165.
37. Davidson AD, 2009. Chapter 2. New insights into flavivirus nonstructural protein 5. *Adv Virus Res* 74: 41–101.
38. Chu PW, Westaway EG, 1987. Characterization of Kunjin virus RNA-dependent RNA polymerase: reinitiation of synthesis in vitro. *Virology* 157: 330–337.
39. Geiss BJ, Thompson AA, Andrews AJ, Sons RL, Gari HH, Keenan SM, Peersen OB, 2009. Analysis of flavivirus NS5 methyltransferase cap binding. *J Mol Biol* 385: 1643–1654.
40. Lu G, Gong P, 2013. Crystal structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog* 9: e1003549.
41. Malet H, Egloff MP, Selisko B, Butcher RE, Wright PJ, Roberts M, Gruez A, Sulzenbacher G, Vonrhein C, Bricogne G, Mackenzie JM, Khromykh AA, Davidson AD, Canard B, 2007. Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. *J Biol Chem* 282: 10678–10689.

42. Yap TL, Xu T, Chen YL, Malet H, Egloff MP, Canard B, Vasudevan SG, Lescar J, 2007. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J Virol* 81: 4753–4765.
43. Buckley A, Gaidamovich S, Turchinskaya A, Gould EA, 1992. Monoclonal antibodies identify the NS5 yellow fever virus non-structural protein in the nuclei of infected cells. *J Gen Virol* 73: 1125–1130.
44. Li Y, Yang Z, 2015. Episodic adaptive diversification of classical swine fever virus RNA-dependent RNA polymerase NS5B. *Can J Microbiol* 61: 948–954.
45. Li Y, Wang R, Du X, Zhang M, Xie M, 2016. Genome-wide analysis for identification of adaptive diversification between hepatitis C virus subtypes 1a and 1b. *Can J Microbiol* 62: 608–616.
46. Bhattacharya D, Mayuri, Best SM, Perera R, Kuhn RJ, Striker R, 2009. Protein kinase G phosphorylates mosquito-borne flavivirus NS5. *J Virol* 83: 9195–9205.
47. Bhattacharya D, Ansari IH, Striker R, 2009. The flaviviral methyltransferase is a substrate of Casein Kinase 1. *Virus Res* 141: 101–104.
48. Kao YT, Chang BL, Liang JJ, Tsai HJ, Lee YL, Lin RJ, Lin YL, 2015. Japanese encephalitis virus nonstructural protein NS5 interacts with mitochondrial trifunctional protein and impairs fatty acid beta-oxidation. *PLoS Pathog* 11: e1004750.
49. Melik W, Ellencrona K, Wigerius M, Hedstrom C, Elvang A, Johansson M, 2012. Two PDZ binding motifs within NS5 have roles in Tick-borne encephalitis virus replication. *Virus Res* 169: 54–62.
50. Werne K, Wigerius M, Johansson M, 2008. Tick-borne encephalitis virus NS5 associates with membrane protein scribble and impairs interferon-stimulated JAK-STAT signalling. *Cell Microbiol* 10: 696–712.
51. Johansson M, Brooks AJ, Jans DA, Vasudevan SG, 2001. A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-beta and the viral helicase, NS3. *J Gen Virol* 82: 735–745.
52. Qing M, Yang F, Zhang B, Zou G, Robida JM, Yuan Z, Tang H, Shi PY, 2009. Cyclosporine inhibits flavivirus replication through blocking the interaction between host cyclophilins and viral NS5 protein. *Antimicrob Agents Chemother* 53: 3226–3235.
53. Ye J, Chen Z, Zhang B, Miao H, Zohaib A, Xu Q, Chen H, Cao S, 2013. Heat shock protein 70 is associated with replicase complex of Japanese encephalitis virus and positively regulates viral genome replication. *PLoS One* 8: e75188.
54. Ana Ts M, Ana Cb T, Danilo Vb D, Roberta Vm B, Maria Cfs M, Gavioli AF, Laura Hvg G, Oliveira AG, Zanelli CF, Valentini SR, 2013. The eukaryotic translation initiation factor 3 subunit L protein interacts with flavivirus NS5 and may modulate yellow fever virus replication. *Virology* 10: 503–509.
55. Bronzoni RV, Madrid MC, Duarte DV, Pellegrini VO, Pacca CC, Carmo AC, Zanelli CF, Valentini SR, Santacruz-Perez C, Barbosa JA, Lutz CS, Rahal P, Nogueira ML, 2011. The small nuclear ribonucleoprotein U1A interacts with NS5 from yellow fever virus. *Arch Virol* 156: 931–938.
56. Selisko B, Wang C, Harris E, Canard B, 2014. Regulation of flavivirus RNA synthesis and replication. *Curr Opin Virol* 9: 74–83.
57. Brinton MA, 2013. Replication cycle and molecular biology of the West Nile virus. *Viruses* 6: 13–53.
58. Hanley KA, Monath TP, Weaver SC, Rossi SL, Richman RL, Vasilakis N, 2013. Fever versus fever: the role of host and vector susceptibility and interspecific competition in shaping the current and future distributions of the sylvatic cycles of dengue virus and yellow fever virus. *Infect Genet Evol* 19: 292–311.
59. Carrington CV, Auguste AJ, 2013. Evolutionary and ecological factors underlying the tempo and distribution of yellow fever virus activity. *Infect Genet Evol* 13: 198–210.
60. Ashour J, Laurent-Rolle M, Shi PY, Garcia-Sastre A, 2009. NS5 of dengue virus mediates STAT2 binding and degradation. *J Virol* 83: 5408–5418.
61. Pisanelli G, Laurent-Rolle M, Morrison J, Garcia-Sastre A, 2015. ID: 121: STAT2 is a determinant of yellow fever virus host tropism. *Cytokine* 76: 88.
62. Lin RJ, Chang BL, Yu HP, Liao CL, Lin YL, 2006. Blocking of interferon-induced Jak-Stat signaling by Japanese encephalitis virus NS5 through a protein tyrosine phosphatase-mediated mechanism. *J Virol* 80: 5908–5918.
63. Laurent-Rolle M, Boer EF, Lubick KJ, Wolfenbarger JB, Carmody AB, Rockx B, Liu W, Ashour J, Shupert WL, Holbrook MR, Barrett AD, Mason PW, Bloom ME, Garcia-Sastre A, Khromykh AA, Best SM, 2010. The NS5 protein of the virulent West Nile virus NY99 strain is a potent antagonist of type I interferon-mediated JAK-STAT signaling. *J Virol* 84: 3503–3515.
64. Finlay BB, McFadden G, 2006. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* 124: 767–782.
65. Pan AHO, 2005. *Control of Yellow Fever: Field Guide*. Washington, DC: Paho Scientific and Technical Publication.
66. Gubler DJ, 2004. The changing epidemiology of yellow fever and dengue, 1900 to 2003: full circle? *Comp Immunol Microbiol Infect Dis* 27: 319–330.
67. WHO, 2015. *Immunization Coverage*. Geneva, Switzerland: World Health Organization. Available at: http://www.who.int/immunization/monitoring_surveillance/routine/coverage/en/index4.html. Accessed July 15, 2016.