Antiviral Research 174 (2020) 104676

Contents lists available at ScienceDirect

Antiviral Research

journal homepage: www.elsevier.com/locate/antiviral

Recombination of B- and T-cell epitope-rich loci from *Aedes*- and *Culex*-borne flaviviruses shapes Zika virus epidemiology

Michael W. Gaunt[a,*], Duane J. Gubler[b], John H.-O. Pettersson[c,d], Goro Kuno[e], Annelies Wilder-Smith[a,f,g], Xavier de Lamballerie[h], Ernest A. Gould[h], Andrew K. Falconar[i]

[a] *London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK*
[b] *Program in Emerging Infectious Diseases, Duke-NUS Medical School, 8 College Rd, 169857, Singapore*
[c] *Department of Medical Biochemistry and Microbiology, Zoonosis Science Center, Uppsala University, Uppsala, Sweden*
[d] *Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, University of Sydney, Sydney, New South Wales, 2006, Australia*
[e] *1648 Collindale Dr, Fort Collins, CO, 80525, USA*
[f] *Department of Public Health and Clinical Medicine, Epidemiology and Public Health, Umeå University, Umeå, Sweden*
[g] *Heidelberg Institute of Global Health, University of Heidelberg, Germany*
[h] *UMR "Unité des Virus Emergents", Aix-Marseille Université-IRD 190, Inserm, 1207-IHU Méditerranée Infection, Marseille, France*
[i] *Departmento de Medicina, Universidad del Norte, Barranquilla, Colombia*

A B S T R A C T

Sporadic human Zika virus (ZIKV) infections have been recorded in Africa and Asia since the 1950s. Major epidemics occurred only after ZIKV emerged in the Pacific islands and spread to the Americas. Specific biological determinants of the explosive epidemic nature of ZIKV have not been identified. Phylogenetic studies revealed incongruence in ZIKV placement in relation to *Aedes*-borne dengue viruses (DENV) and *Culex*-borne flaviviruses. We hypothesized that this incongruence reflects interspecies recombination resulting in ZIKV evasion of cross-protective T-cell immunity. We investigated ZIKV phylogenetic incongruence in relation to: DENV T-cell epitope maps experimentally identified *ex vivo*, published B-cell epitope loci, and $CD8^+$ T-cell epitopes predicted *in silico* for mosquito-borne flaviviruses. Our findings demonstrate that the ZIKV proteome is a hybrid of *Aedes*-borne DENV proteins interspersed amongst *Culex*-borne flavivirus proteins derived through independent interspecies recombination events. These analyses infer that DENV-associated proteins in the ZIKV hybrid proteome generated immunodominant human B-cell responses, whereas ZIKV recombinant derived *Culex*-borne flavivirus-associated proteins generated immunodominant $CD8^+$ and/or $CD4^+$ T-cell responses. *In silico* $CD8^+$ T-cell epitope ZIKV cross-reactive prediction analyses verified this observation. We propose that by acquiring cytotoxic T-cell epitope-rich regions from *Culex*-borne flaviviruses, ZIKV evaded DENV-generated T-cell immune cross-protection. Thus, *Culex*-borne flaviviruses, including West Nile virus and Japanese encephalitis virus, might induce cross-protective T-cell responses against ZIKV. This would explain why explosive ZIKV epidemics occurred in DENV-endemic regions of Micronesia, Polynesia and the Americas where *Culex*-borne flavivirus outbreaks are infrequent and why ZIKV did not cause major epidemics in Asia where *Culex*-borne flaviviruses are widespread.

1. Introduction

Human infections with Zika virus (ZIKV) are known to have occurred in Africa and Asia since at least the 1950s. However, epidemic transmission was not recorded until the virus emerged in Micronesia in 2007 (Duffy et al., 2009), spreading to French Polynesia in 2013 and the American tropics in 2014 (Musso, 2015; Pettersson et al., 2016; Watts et al., 2018; Hamer and Chen, 2019). Although ZIKV dispersal was facilitated by human global travel (Wilder-Smith et al., 2018) and geographic expansion of the mosquito vector, *Aedes aegypti* via commercial transportation (Gould et al., 2017), no specific determinants have been identified to explain the dramatic emergence and dissemination of epidemic ZIKV in human populations where related mosquito-borne flaviviruses (MBFV) were endemic.

Phylogenetic analyses of the genus *Flavivirus* based on the envelope (E) and nonstructural 5 (NS5) genes (Kuno et al., 1998; Gaunt et al., 2001) revealed robust evolutionary associations between viruses, vectors, vertebrate hosts and disease characteristics. These phylogenetic trees divided the MBFV into two clades (groups) according to their primary transmission vector, viz. those transmitted by *Aedes* spp. and those transmitted by *Culex* spp., mosquitoes. In the *Aedes*-associated group, yellow fever virus (YFV), the dengue viruses (DENV), ZIKV and Spondweni virus (SPOV), formed a (paraphyletic) clade, implying that they are direct descendants of a common ancestor (Kuno et al., 2009; Huhtamo et al., 2009; Kolodziejek et al., 2013). The *Culex*-borne flaviviruses formed a monophyletic clade which included West Nile virus

* Corresponding author.
E-mail address: Michael.Gaunt@lshtm.ac.uk (M.W. Gaunt).

https://doi.org/10.1016/j.antiviral.2019.104676
Received 11 June 2019; Received in revised form 6 December 2019; Accepted 8 December 2019

(WNV) and Japanese encephalitis virus (JEV) (Gaunt et al., 2001; Gould et al., 2003; Grard et al., 2010; Liu et al., 2012; Moureau et al., 2015).

Several subsequent whole genome MBFV phylogenies supported this phylo-epidemiological dichotomy (Cook et al., 2012; Huhtamo et al., 2014; Moureau et al., 2015), whilst others placed ZIKV and SPOV with the *Culex*-associated flaviviruses (Huhtamo et al., 2009; Kolodziejek et al., 2013), thus disrupting the phylo-epidemiological division and raising questions concerning the evolutionary relationships of DENV, SPOV and ZIKV (Kuno et al., 2009; Barba-Spaeth et al., 2016). These incongruent phylogenies implied the possibility of genetic recombination which we have now investigated in detail.

The impact of pre-existing DENV B-cell and T-cell immunity against ZIKV infection is complex despite the close homology between their proteomes. Immunodominant cross-reactive humoral (B-cell) responses are generated against the main E-protein of DENV and ZIKV resulting in high (54%) IgG cross-reactivity in sera from infected patients (van Meer et al., 2017; Felix et al., 2017). Cross-reactivity was also demonstrated against the ZIKV nonstructural-1 (NS1) using IgG from populations previously exposed only to DENV (Nurtop et al., 2018). Antibody cross-reactivity between DENV and ZIKV has also been demonstrated by antibody mapping studies (Barba-Spaeth et al., 2016; Priyamvada et al., 2016). Between DENV serotypes, extensive cross-reactive antibody mediated B-cell responses were generated against the prM, E and NS1 glycoproteins and to a limited extent against the NS3 protein (Falconar, 1999, 2007; Lai et al., 2008; da Silva et al., 2009; Costa et al., 2011; Dejnirattisai et al., 2016; Masrinoul et al., 2013).

$CD8^+$ T-cell cross-reactivity and cross-protection between ZIKV and DENV was lower than between DENV serotypes (Weiskopf et al., 2013). For example, distinct ZIKV T-cell responses experimentally identified *ex vivo* were independent of DENV exposure (Grifoni et al., 2018). Part of the explanation is that the majority (55–60%) of ZIKV-specific $CD8^+$ and $CD4^+$ T-cell epitopes were located in the structural C, prM and E proteins (Grifoni et al., 2017), whilst $CD8^+$ and $C4^+$ T-cell cross-reactivity between DENV serotypes was restricted to the C and NS2 to NS5 proteins (Weiskopf et al., 2013). The disparity in cross-reactivity was indicated in ZIKV specific T-cell responses *ex vivo* that lacked cross-reactivity with DENV T-cell epitopes against the NS3 protease S7 (Herrera et al., 2018) and the C protein (Lim et al., 2018).

In this study, we looked for evidence of interspecific recombination amongst the *Aedes*- and *Culex*-associated flaviviruses, where ZIKV presents a phylogenetic tree topology which is either *Aedes*-DENV-associated in the E-protein or *Culex*-MBFV-associated in the NS5-protein. We also addressed the correlation with anomalies in B-cell and/or T-cell cross-reactivity. This analysis extended previous studies which demonstrated disagreement between E− and NS5-protein trees (Kuno et al., 2009; Barba-Spaeth et al., 2016) and hereafter referred to as 'phylogenetic incongruence'. We then assessed whether or not the phylogenetic incongruence of ZIKV correlated with cross-protection from other MBFV and addressed the following questions:

- Is ZIKV phylogenetic incongruence statistically robust and where are the exact amino acid encoded breakpoint positions located within the genome that were responsible for this incongruence?
- Do the regions of the genome causing ZIKV phylogenetic incongruence correlate with:
  - o DENV T-cell epitope maps experimentally identified *ex vivo* (because DENV co-circulate with ZIKV)?
  - o DENV/ZIKV B-cell *ex vivo* and *in vivo* epitope maps?
  - o cross-reactive $CD8^+$ T-cell epitopes predicted *in silico* between ZIKV and other MBFV?
- Accuracy assessment: do *ex vivo* $CD8^+$ T-cell epitopes of MBFV validate the *in silico* epitopes referred to in the final sub-point above concerning *in silico* cross-reactivity?

2. Materials and methods

2.1. Methods overview

For clarity we have summarized the analytical procedures in Appendix A, Fig. S1. The primary approach of this study was to perform viral protein phylogenetic analyses for all available MBFV species and subtypes using amino acid alignments. The rationale for this approach was to understand the differentiation between B-cell and T-cell immunological selection pressure applied to each specific protein in the proteome. Accordingly, we evaluated two amino acid alignments: a 71 taxa alignment based on the recognized MBFV species diversity and a second alignment that included the basal 'insect specific flaviviruses' (ISF) associated with cell fusing agent virus (CFAV) and the flaviviruses associated with the Apoi virus (APOIV) clade (Cook et al., 2012). The 71 taxa MBFV alignment under this evaluation is described in Appendix A, Supplementary data 'Phylogenetic Alignments'.

2.2. Phylogenetic methods

The phylogenetic methodology used the following established approaches in identifying recombination which are described in detail in Appendix A; 'Breakpoint analysis strategy and 'Amino Bootscan', as well as 'All phylogenetic methods and rationale', summarized in Supplementary Information Appendix A, Fig. S1 comprising: 1) a phylogenetic split network analysis (Neighbor-Net) (Bryant and Moulton, 2004) which identifies areas of phylogenetic uncertainty possibly resulting from recombination; 2) an amino acid 'sliding window' approach (VisRD) (Lemey et al., 2009), which identifies recombination breakpoints; 3) parametric bootstrap analysis; 4) amino acid bootscanning ("Amino Bootscan" – see below) and finally, 5) non-parametric bootstrap analysis using 1000 maximum likelihood bootstraps. VisRD is a "sliding window analysis" which produces a noisy phylogenetic signal and therefore all breakpoints were subjected to further confirmatory analysis using amino acid bootscanning.

2.2.1. Custom data pipeline, 'Amino Bootscan'

To achieve this, we constructed a data pipeline called "Amino Bootscan" because of the absence of bootscanning software for amino acid data, essentially a Python (3.7) wrapper for RAxML utilizing itertools (for permuting), Biopython, ETE (Extended Tree Environment), pandas and Matplotlib (available via 'git clone https://github.com/ZikaLab/IEDB_pipeline' and https://genome.guru, password "Recombination9@", under the GNU General Public License v3.0). "Amino Bootscan" is a sliding window phylogenetic data pipeline that assesses the robustness of two incongruent phylogenies and whether they could putatively form a mosaic distribution across a genome. Thus, it minimizes putative mosaics through locus rearrangements. The rationale for the strategy and full description of the method is given in Appendix A, "Breakpoint analysis strategy" and "Amino Bootscan". Here we used a "step size" of 100 amino acids and window size of 600 amino acid residues, which assesses robustness for 500 maximum likelihood bootstraps for every "step" comprising a "quintet tree" (5 taxa tree). The analysis was in two stages, permuting the order of putative incongruent loci within the genome for a given 5 taxa (MBFV) tree and then assessing combinatorial samples of taxa within given MBFV groups. Genome permutations were conducted for TBEV (Neudoerfl strain), YFV (Asibi 17D strain), ZIKV (BeH819015 strain), DENV-1 (75861 strain), and JEV (057434 strain) resulting in 120 different sliding window analyses permutations of 5 loci (Appendix A). The locus genome order was then fixed from the permutations analysis and different MBFV combinations of 5 taxa were assessed that comprised: 1) TBEV (Neudoerfl); 2) YFV (Asibi 17D) and thereafter all combinations of; 3) ZIKV (BeH819015) or SPOV (SM-6-V-1); against 4) DENV-1 (75861), or DENV-2 (16681), or DENV-3 (424); against 5) *Culex* clade I viruses, WNV (unstated strain) or JEV (057434), or SLEV (CbaAr-4005),

or *Culex* clade II virus BSQV (BeAn 4073). All MBFV are described in Supplementary Table 1 Appendix D. The disadvantage of this method is that the computational overhead is very significant for a phylogenetics calculation and requires processor parallelization.

2.2.2. Phylogenetic boostrap analysis

Parametric bootstrap analysis is a maximum likelihood test which identifies potential topological differences between MBFV phylogenies with respect to the placement of ZIKV-SPOV and DENV using a null distribution obtained by amino acid simulation for 100 replications with a JTT amino acid matrix (Swofford-Olsen-Waddell-Hillis test, SOWHat test (Church et al., 2015)) (Appendix A, 'Phylogenetics', 'Parametric bootstrapping').

Non-parametric bootstrap analysis was performed for all amino acid phylogenetic trees using a maximum likelihood approach using RAxML (Stamatakis, 2014). Robustness was determined by bootstrapping using a 4 category gamma (γ) distribution to estimate mutation rate heterogeneity and the analysis included an invariant rate category. Amino acid phylogenies were generated by comparing the likelihood for 'JTT', 'WAG' (Whelan and Goldman, 2001) and 'LG' matrices (Le and Gascuel, 2008) using RAxML. The highest likelihood for the non-parametric bootstrap analysis was the LG matrix. A phylogenetic robustness for a branching order was considered when a non-parametric bootstrap value, denoting the *Aedes* or *Culex* clade, exceeded 80%. Our analysis differed from many others by using amino acid instead of nucleotide analysis, because the third codon position saturates (Zanotto et al., 1996). Modern amino acid matrices, notably LG matrices, provide higher phylogenetic signal than removing the third codon in nucleotide analyses. An identical phylogenetic model was recently used to describe the phylogeny of the global RNA virome (Wolf et al., 2018).

2.3. Mosquito-borne flavivirus NS3 protein domains

The NS3 protein comprises two domains encoding the 1 to 185 amino-terminal serine S7 (S7) protease, and the 186 to 618 amino acid NTPase-helicase protein. The flavivirus NS3 helicase region is a member of two helicase superfamilies with the DEAH/DEAD (Asp [D]-Glu [E]-Ala [A]-His/Asp [H/D]) motif for divalent cation binding and NTPase activity as well as an adjacent C-terminal helicase domain denoted HELICc (Wu et al., 2005).

2.4. Immunological bioinformatic methods

In this analysis, high and low density B-cell epitope loci are defined from the experimentally observed dominant B-cell responses generated against quaternary-, trimer-, dimer-, monomer-specific, conformational or linear, neutralizing or non-neutralizing loci which are cited in the Introduction. Furthermore, in this analysis we define $CD4^+$/$CD8^+$ T-cell epitope maps experimentally identified *ex vivo* as referring to binding affinities of 8-, 9- and 10-mer oligopeptide to $CD8^+$ T-cells from purified peripheral mononuclear blood cells (PMBCs) from convalescent patients.

2.4.1. Custom IEDB data pipeline

To infer putative changes in cross-reactive T-cell epitopes following recombination between ZIKV and *Culex* Clade I viruses (notably WNV and JEV), we constructed *in silico* predicted $CD8^+$ T-cell epitope maps for each given phylogeny using the stand-alone version of MHC class I peptide binding prediction tools (version 2.19.1) from the Immune Epitope Database (IEDB) (Vita et al., 2019). This was automated and parsed using a data pipeline (available via 'git clone https://github.com/ZikaLab/IEDB_pipeline' and https://genome.guru password "Recombination9@", under the GNU General Public License v3.0; Appendix B, T-cell bioinformatics, 'Algorithm') and which is called, the 'IEDB data pipeline'. The pipeline was verified by manually checking 20% of all epitopes, together with internal checks (Appendix B, 'Materials and methods') and verified against MBFV $CD8^+$ T-cell epitopes experimentally identified *ex vivo* for their binding affinities for one recombinant locus (NS2A to NS3 DEAD) (Appendix B, 'Results'; Supplementary Table 2, Appendix D).

All 12 human leucocyte antigen (HLA) alleles used in the analysis were identified using both published (Supplementary Table 2, Appendix D) and selected HLA haplotypes prevalent in the populations in South East (SE) Asia (Appendix B 'SE Asian strains').

2.4.2. IEDB data pipeline criteria

We used *in silico* $CD8^+$ T-cell epitope predictive algorithms based around the artificial neural network (ANN), stabilized matrix method (SMM) and the 'consensus method' (*CM*) (Moutaftsi et al., 2006; González-Galarza et al., 2015); Supplementary Table 2, Appendix D). Putative cross-protective *in silico* $CD8^+$ T-cell epitopes were predicted based on 5 criteria (Supplementary data); an important criterion is the *CM* index which assesses the robustness of a predicted epitope, because previous studies demonstrated the accuracy and reliability of the $CD8^+$ T-cell *in silico* prediction index (Moutaftsi et al., 2006). A *CM* value $<$ 2.5 was used based on the *CM* values of $CD8^+$ T-cell epitopes experimentally verified *ex vivo* and described in Supplementary Table 2 (Appendix D). Thus a *CM* value $<$ 2.5 for ZIKV, as key criteria of epitope prediction, represented a good estimate of its biological validity. All statistical tests were performed using the non-parametric Wilcoxan rank test.

2.4.3. Geographic distribution of HLA haplotypes

To identify appropriate haplotypes in SE Asia, we assessed all available HLA-A and HLA-B haplotype frequency distributions, which included HLA-B*35:01, and correlated them with DENV epitopes, using the Allele Frequency database. Based on country and racial group, we identified HLA haplotypes uniquely dominant in SE Asia and the Pacific Islands, viz. HLA-A*11:01 (each hyperlinked to global distribution map), HLA-A*24:02 (each hyperlinked to global distribution map), and HLA-B*40:01 (hyperlink to global distribution map) (Appendix B, 'SE Asian strains'). HLA-B*35:01 is an example of a haplotype with a moderately increased frequency in Amerindian populations in the Americas.

2.5. Geographic nomenclature

Regarding the geographic division of the Pacific Islands referred to herein, we assigned Papua New Guinea as a member of Melanesia and separate from Australasia, Micronesia as an archipelago within Oceania and French Polynesia as an archipelago within Polynesia.

3. Results

3.1. Phylogenetic analysis

3.1.1. Phylogenetic analysis of the MBFV

An amino acid phylogenetic analysis of the MBFV genomes and the unrooted tree is presented in Fig. 1. All virus abbreviations and source information are listed in Supplementary Table 1 (Appendix D). This network tree depicts a central node of reticulation separating the ZIKV and SPOV sister lineages from the DENV *Aedes* clade and the *Culex* clade viruses. The *Culex* clade then diverges to form clades I and II, the latter separating into Australasian/Melanesian and South American virus groups. In all phylogenies (Fig 1, Figs. 3 and 4), SPOV formed a sister group with ZIKV. *Culex* clade I viruses comprise the medically important MBFV that include WNV and JEV.

3.1.2. ZIKV interspecific recombination breakpoint analysis

A phylogenetic network was observed in the unrooted tree, technically called a 'reticulation node', associated with ZIKV and SPOV (Fig. 1). Reticulation nodes indicate conflicting phylogenetic signals

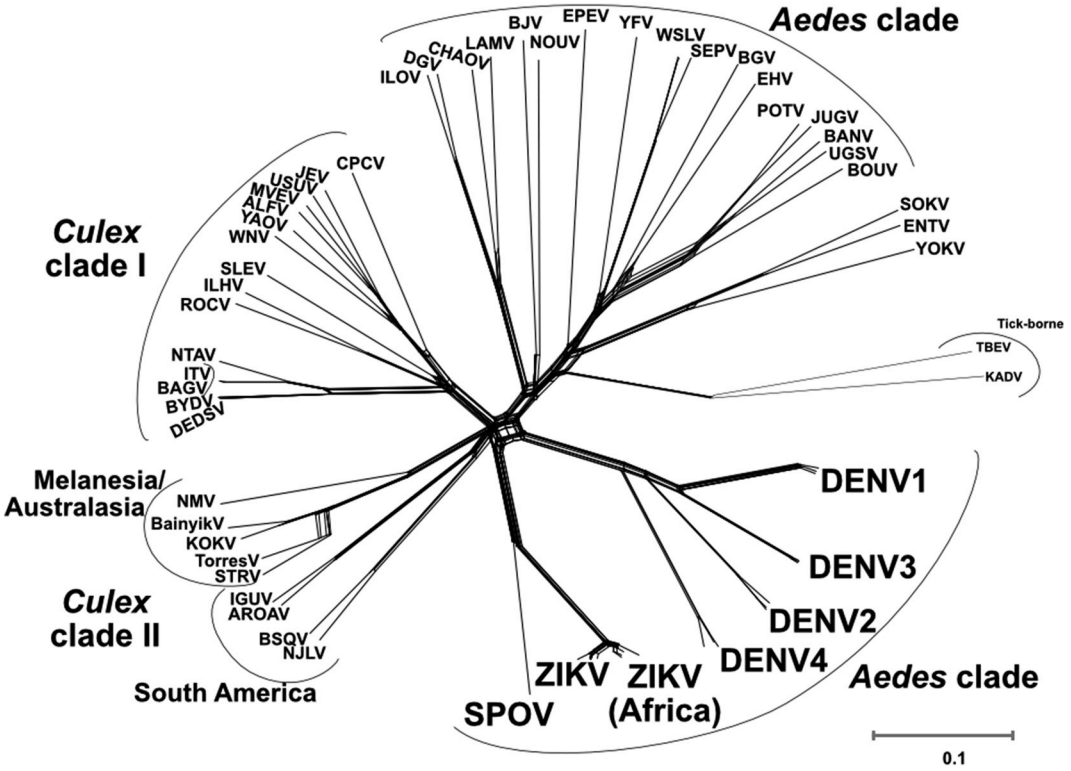


Fig. 1. Network phylogeny (Neighbor-Net) of the MBFV genomes demonstrating the node of reticulation at the centre of the phylogenetic relationship between the ZIKV - SPOV sister group, DENV, and the *Culex* clade MBFV. The tree is unrooted, however in rooted phylogenies (Figs. 3 and 4) tick-borne encephalitis virus (TBEV) and Kadam virus (KADV) were used as outgroup viruses. Virus designations follow the recommendations of the ICTV. Note, that phylogenetic networks represent nodes of phylogenetic uncertainty which may be attributed to recombination (Material and methods).

which could be due to phylogenetic incongruence, and would infer recombination involving ZIKV and SPOV. The node was further analyzed using an amino acid sliding window called VisRD. This method lacks statistical robustness but provides an indication of recombination associated breakpoints (Fig. S2, Appendix C). Putative breakpoints due to phylogenetic incongruence were identified by dramatic shifts in phylogenetic signal on the y-axis of VisRD associated with ZIKV (Fig. S2, Appendix C). The major output of our VisRD analysis, which was distinct from previously published analyses (Kuno et al., 2009; Barba-Spaeth et al., 2016) was a breakpoint within the NS3 HELICc domain, starting 15 amino acids into the NS3 HELICc domain and ending within the C-terminal HELICc domain. Parametric bootstrapping (below) was used to assess the statistical robustness of combining protein sequences between different aligned loci hereafter called “combined loci” (described below and Appendix A), for example assessing the phylogenetic validity of combining two different protein gene alignments into a single alignment (data set).

The robustness of the ZIKV pivotal breakpoints required further confirmation and this was undertaken using a maximum likelihood phylogenetic sliding window analysis written specifically for this investigation, called “Amino Bootscan” (Fig. 2) using a subset of the MBFV and subsequently, maximum likelihood trees for all MBFV (Figs. 3 and 4). Thus all loci designated by breakpoints and where applicable, the combined loci, were examined using “Amino Bootscan” analysis (Fig. 2). Following locus rearrangement analysis, the bootscan distinguished between the B-cell epitope rich loci, where ZIKV-SPOV were monophyletic with the DENV and the T-cell epitope rich loci, where ZIKV-SPOV were monophyletic with the *Culex*-associated MBFV. Moreover all permutations of loci within the genome and all combinations of taxa were assessed. B-cell/T-cell epitope rich loci are identified in Fig. 2. This provided robust confirmation of the breakpoints of VisRD analysis (Fig. S2, Appendix C), and was supported by previous analyses (Kuno et al., 2009; Barba-Spaeth et al., 2016). A complete description of “Amino Bootscan” is given in Appendix A.

The conclusion from the bootscan analysis was also confirmed using all the MBFV in mirror tree analysis (Fig. 3i and 3ii). This depicts the results of non-parametric bootstrap analysis showing: a robust monophyly between ZIKV-SPOV and DENV for the individual protein trees of prM, E, NS4B; and a combined locus comprising the NS1 and partial NS3 protein encompassing most of the HELICc domain, but excluding the C-terminal ‘tip’ (Fig. 3i, Supplementary Table 2, Appendix D).

The trees also depict a robust paraphyly between ZIKV-SPOV and DENV, i.e. a ZIKV-SPOV and *Culex*-associated MBFV monophyly, for a combined locus of the C and NS5 proteins, a locus from the C-terminus of the NS3 HELICc to the NS4A protein inclusively, a locus from the NS2A protein to NS3 DEAD, 15 amino acids of NS3 HELICc (NS2A, NS2B, NS3 comprising S7, DEAD domain, and the N-terminal tip of the HELICc domain) (Fig. 3). This latter region is hereafter referred to as the NS2A to NS3 DEAD locus.

In contrast to the ZIKV-SPOV lineage and Kedougou virus (KEDV) which showed phylogenetic incongruence between DENV B-cell and T-cell rich loci (bootstrap > 75%), we observed robust congruence for all other MBFV groups regardless of locus, viz. between the *Aedes*-YFV group, *Aedes*-DENV group, *Culex* I clade and *Culex* II clade for all MBFV (Figs. 3 and 4).

Parametric bootstrap analysis was used in the initial delineation of phylogenetic loci following the criteria described in Appendix A, “Breakpoint analysis strategy” and demonstrated identical fundamental tree structures between the C and NS5 proteins, both of which comprised a basal DENV clade and a ZIKV-SPOV-*Culex* clade monophyly (Materials and Methods 2.2.2; Appendix A, ‘Phylogenetics’, ‘Parametric bootstrapping’). Parametric bootstrap analysis also demonstrated that the NS1 and NS3 N-terminal HELICc regions were monophyletic for DENV and ZIKV-SPOV (Appendix A, ‘Parametric bootstrapping’). Therefore the C and NS5 proteins and separately, the NS1-NS3 N-terminal HELICc proteins, were analyzed as two distinct alignments for

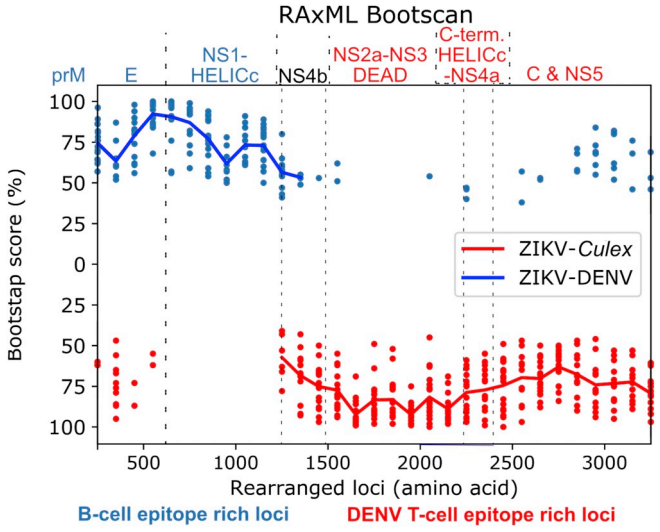


Fig. 2. Maximum likelihood amino acid bootscan, i.e. a bootstrapped sliding window analysis for 500 replications based on "quintet trees' comprising all combinations of 5 taxa from the following 1) TBEV, 2) YFV and thereafter all combinations of 3) ZIKV or SPOV, against 4) DENV; 5) the *Culex* clade I viruses or *Culex* clade II viruses as described in Materials and methods. The step size is 100 and window size of 600 amino acids. The graphic is the automated output of "Amino Bootscan".

Key,

Blue/red lines, bootstrap of all quintet trees for a given "step", i.e. genome position,

ZIKV-DENV (blue line), demonstrates that ZIKV is monophyletic with the DENV,

ZIKV-*Culex* (red line), each quintet tree shows ZIKV is monophyletic with the *Culex*-associated MBFV (*Culex* clade I and *Culex* clade II). For the interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.

non-parametric bootstrap analysis without disrupting the primary phylogenetic signal of interest. However, the parametric bootstrapping also identified KEDV as being incongruent between both the C and NS5 trees and the NS1 and NS3 N-terminal HELICc trees. KEDV was therefore excluded from these data sets (Appendix A, 'Phylogenetics', 'Parametric bootstrapping').

3.1.3. ZIKV and Culex clade virus interspecific recombination

The three loci identified with high-density DENV $CD4^+$ and $C8^+$ T-cell responses, observed *ex vivo,* all placed ZIKV and SPOV within the *Culex*-associated MBFV rather than with the DENV lineages (> 80% maximum likelihood bootstrap support). However, these "T-cell epitope rich trees" were robustly incongruent from each other (with > 80% bootstrap support) as to precisely where ZIKV-SPOV is placed within the *Culex*-associated MBFV clade. Fig. 4 illustrates the individual maximum likelihood phylogenies of the three ZIKV protein loci identified as the C and NS5 ('Tree 1'), the C-terminal HELICc domain and NS4A protein ('Tree 2'), and the NS2A to NS3 DEAD locus ('Tree 3'). Thus, the ZIKV-SPOV lineage for these loci was phylogenetically robust in its association with the *Culex* clade, but 'Trees 1, 2 and 3′ were robustly incongruent when examined against each other with respect to the placement of the ZIKV-SPOV lineage. In 'Tree 1,' the ZIKV-SPOV lineage was monophyletic and basal to the *Culex* clade for the combined C and NS5 protein locus, while in 'Tree 2,' ZIKV-SPOV formed a sister group exclusively with the Australasian/Melanesian *Culex* clade II MBFV for the protein locus of the C-terminal NS3 HELICc to NS4A inclusively. In 'Tree 3,' ZIKV-SPOV formed a sister group with the *Culex* clade I viruses and was thus incongruent with Trees 1 and 2. The bootstrap for the combined C and NS5 locus was highly robust whereas each independent C and NS5 locus had a low bootstrap value. In contrast, the 'ZIKV-SPOV-DENV monophyly' comprising three physically separated loci in the ZIKV-SPOV genome, viz., the prM, E and NS1 proteins, the NS3 HELICc domain and NS4B protein, independently produced identical robust monophyletic phylogenetic trees with respect to ZIKV-SPOV-DENV.

3.2. Immunological properties

3.2.1. MBFV trees and B-cell association

There was a strong correlation between the proteins associated with a ZIKV-SPOV-DENV monophyly, viz. the prM, E and NS1 glycoproteins and the NS3 HELICc helicase excluding the C-terminal domain (Figs. 3 and 4), and their relative B-cell epitope densities. These glycoproteins are released concurrently from infected cells and induce immunodominant antibody (B-cell) responses against both linear and conformational epitopes during human DENV infections. The NS4B gene tree (Figs. 3 and 4), also produced a ZIKV-SPOV-DENV monophyly. However, NS4B protein is not known to generate a strong high level antibody response during DENV infections.

3.2.2. MBFV tree and ex vivo T-cell association

There was a strong correlation between the protein loci which resulted in a ZIKV-SPOV-*Culex* clade monophyly, (i.e. where ZIKV-SPOV is paraphyletic with DENV) with a low density of B-cell epitopes and high density of DENV T-cell epitopes experimentally identified *ex vivo* (i.e. for the C, NS5, NS2A, NS2B and NS3 protein S7 and DEAD domains, the HELICc domain and NS4A protein) (Weiskopf et al., 2013, 2015a, 2016) (Fig. 4). Specifically, proteins with high densities of DENV $CD4^+$ and $CD8^+$ epitopes (C, NS2A, NS3 S7 domain, C-terminal HELICc, NS4A and NS5) were associated with the ZIKV-SPOV-*Culex* clade monophyly for all but one protein, viz. NS4B (Fig. S3, Appendix C). Furthermore, they were not associated with generating immunodominant human B-cell (antibody) responses as was shown with DENV and by finding immunodominant B-cell responses against the E and NS1 glycoproteins through the characterization of anti-ZIKV human monoclonal antibodies (Stettler et al., 2016). Moreover, the MBFV C protein is protected from antibody exposure by a lipid envelope containing the prM and E proteins, whilst the remaining 4 non-structural MBFV proteins (NS2A, NS2B, the NS3 S7 domain, NS4A and NS5 proteins) are involved in viral replication. The NS4B protein was the exception because it contained $CD8^+$ T-cell epitopes but the ZIKV-SPOV-DENV tree is monophyletic for this protein inferring it was not involved in recombination.

Considering the ZIKV-SPOV-*Culex* clade phylogenies (Fig. 4), Tree 1 was strongly associated with $CD4^+$ and $CD8^+$ (C and NS5) T-cell epitopes experimentally identified *ex vivo*, Tree 2 was strongly associated with $CD8^+$ T-cell epitopes (C-terminal NS3 HELICc domain) (Costa et al., 2011; Weiskopf et al., 2013) and Tree 3 was associated with $CD4^+$ and $CD8^+$ T-cell epitopes (NS2A, NS3 S7 & DEAD domains, i.e. the NS2A to NS3 DEAD locus) (Fig. S3). The high numbers of $CD4^+$ and $CD8^+$ T-cell epitopes in DENV were located in the C and NS5 proteins as described in 'Tree 1′, where the ZIKV-SPOV lineage was basal to the *Culex* clade. The dominant $CD8^+$ T-cell epitopes were in the C-terminal region of NS3 HELICc, as described in 'Tree 2', where ZIKV and SPOV were phylogenetically basal to the Australasian/Melanesian MBFV. The remaining $CD4^+$ and $CD8^+$ T-cell epitopes occur within NS2A, NS2B and the NS3 S7, DEAD and N-terminal HELICc domains described in 'Tree 3′, where ZIKV-SPOV was monophyletic to the *Culex* clade I viruses.

3.2.3. Putative T-cell associated protection

HLA haplotypes associated with cytotoxic T-cell mediated protection against DENV infection showed strong variation in human population frequency. The monophyly in 'Tree 2' between ZIKV and Australasian/Melanesian MBFV was striking, because: 1) it involved the most important region of $CD8^+$ T-cell epitopes in DENV-infected

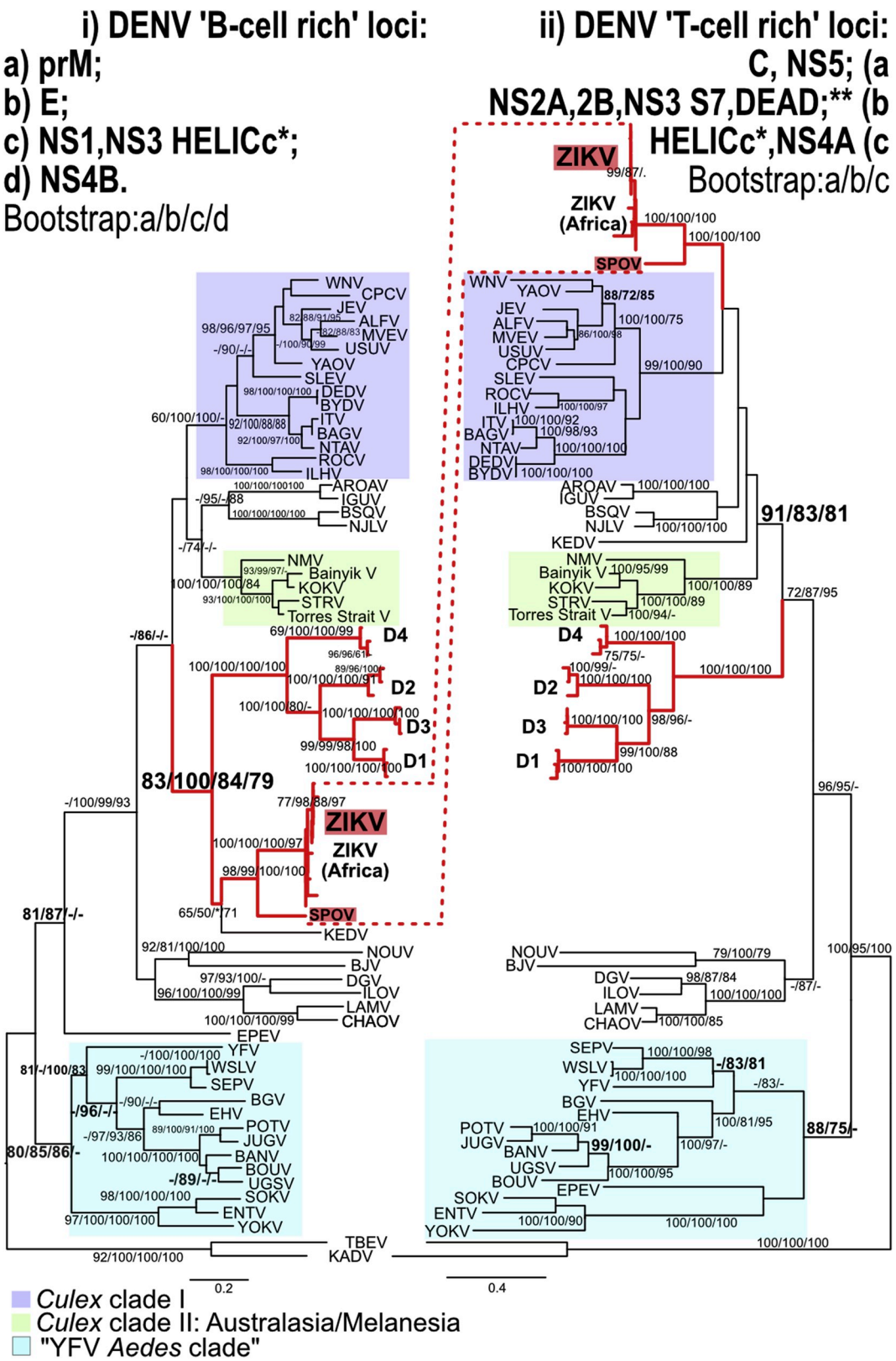


Fig. 3. Amino acid maximum likelihood phylogenies of the predominant B-cell epitope containing proteins (left) and predominant DENV T-cell epitope loci, experimentally observed *ex vivo,* containing proteins (right). The numbers above each node are percentage bootstrap values above 75%. For Fig. 3i the order of the bootstraps along each internal branch represents 4 separate bootstrap protein trees for predominantly B-cell epitope proteins, i.e. a) a prM protein tree, b) an E protein tree, c) a combined locus of NS1 and NS3 HELICc domain protein tree and d) an NS4B protein tree. For Fig. 3ii the order of the bootstraps along each internal branch represents three separate bootstrap protein trees for proteins with high densities of T-cell epitopes experimentally identified *ex vivo*, i.e. a) a combined locus of the C and NS5 protein tree, b) an NS2A to the NS3 DEAD locus protein tree, c) an NS3 C-terminal HELICc region to the NS4A protein tree. Key, * indicates a partial protein domain and ** indicates 15 amino acids of the N-terminal tip of the NS3 HELICc domain. 'B-cell/T-cell rich', refers to B-cell/T-cell epitope rich loci.

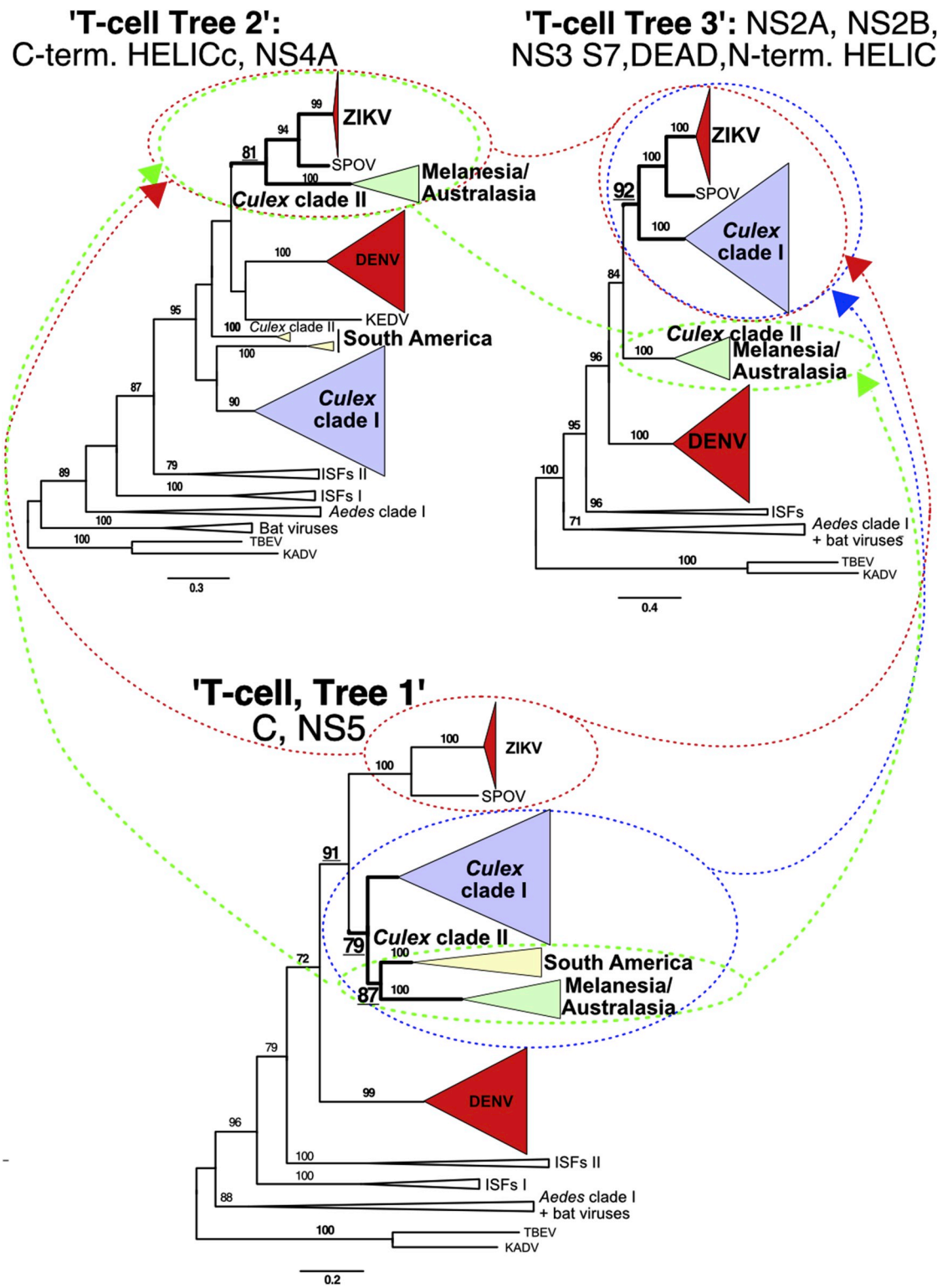
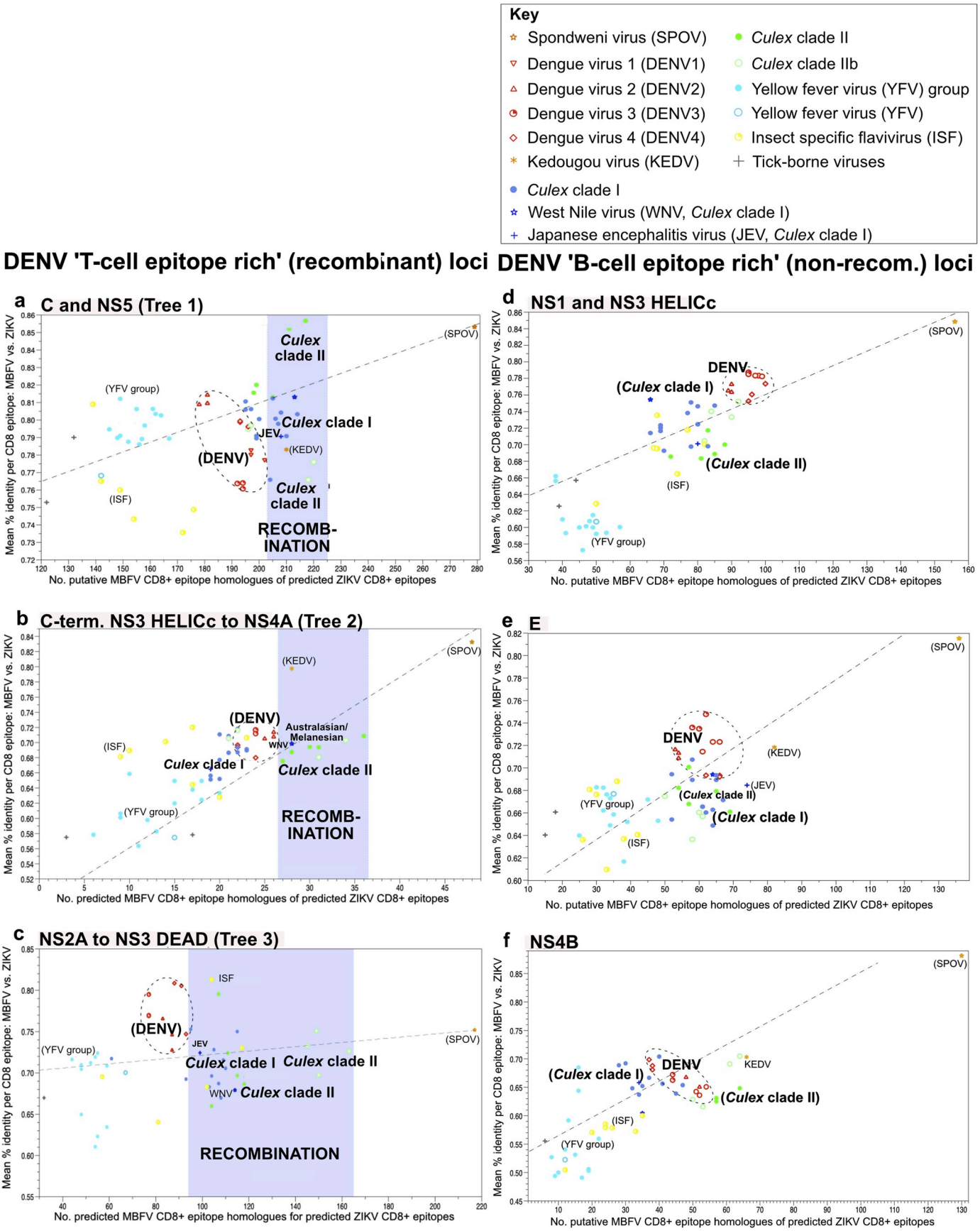


Fig. 4. Amino acid maximum likelihood phylogenies describing the robust incongruence between individual $CD4^+$ and $CD8^+$ T-cell epitope-rich loci experimentally observed *ex vivo* depicted here as 'Tree 1′, 'Tree 2′ and 'Tree 3'. The single numbers above each branch represent bootstrap values above 75%, where * denotes 15 amino acids of N-terminus HELICc.

humans that cluster within 50 amino acids of the N-terminal region of the HELICc domain; 2) 50% of all HLA-B*40:01 $CD8^+$ T-cell epitopes on the DENV genome were located within this small locus, making it a very dense region of $CD8^+$ T-cell epitopes within the DENV translated genome, and 3) HLA-B*40:01 was significantly associated with native Polynesians, New Zealand Maoris (originally from the Cook Islands) and indigenous Taiwanese populations (Edinur et al., 2013). These results are also compatible with $CD8^+$ T-cell responses generated against ZIKV infections in mouse models (Wen et al., 2017). Whether or not the monophyletic relationship between ZIKV and the Melanesian/Australasian viruses infers the influence of T-cell selection pressure will be considered in future investigations.



(caption on next page)

Fig. 5. a to f. Comparison of the *in silico* predicted CD8$^+$ T-cell epitopes in the 'recombinant' against 'non-recombinant' loci between ZIKV and the given MBFV to examine the relationship between total amino acid differences per locus versus the number of shared epitopes for all alignments and MBFV analyzed in Fig. 3. The *in silico* CD8$^+$ T-cell epitope predictions described here were verified using the major published biologically identified epitopes for DENV-1 to 4, JEV and WNV described in Section 3.2.5. The presence of insect specific flaviviruses (ISF) within 5c (NS2A-NS3 DEAD locus) could be due to phylogenetic incongruence within the ISF for this locus (not shown).

3.2.4. Predictive in silico ZIKV-MBFV CD8$^+$ cross-protection

An estimate of the shared MBFV CD8$^+$ epitopes that could induce cross-reactivity in humans with CD8$^+$ T-cell epitopes of ZIKV was performed using the MHC class I peptide binding prediction tools from the Immune Epitope Database (IEDB) (Vita et al., 2019). The accuracy of this epitope prediction method has been experimentally verified (Moutaftsi et al., 2006). We focused on 9 biologically proven MBFV CD8$^+$ T-cell epitopes experimentally identified for the NS2A-NS3 DEAD locus of DENV, JEV and WNV, viz. human leucocyte antigen (HLA) A*01:01, A*02:01, A*26:01, A*30:01, B*07:02, B*08:01, B*15:01, B*35:01, B*38:01 (Supplementary Table 2, Appendix D) and three HLA alleles specifically associated with SE Asia and Australasian/Melanesian populations (Appendix B, 'SE Asian strains'), viz. HLA-A*11:01, A*24:02, and B*40:01. The criteria for cross-reactivity are described in Materials and methods, 2.4.3 and Supplementary Information (Appendix B, 'SE Asian strains'). IEDB analysis was performed with all the MBFV taxa in this analysis (Supplementary Table 1, Appendix D) using all loci. The calculation was automated, parsed, and analyzed in the 'IEDB data pipeline' (Materials and methods 2.2).

Fig. 5 describes the relationship between the average percentage amino acid identity per predicted shared CD8$^+$ epitope and the number of predicted shared epitopes between ZIKV and every other MBFV in the analysis. When compared with the non-recombinant loci (Fig. 5d to f), all three ZIKV - *Culex* clade recombinant protein loci (Fig. 5a to c) had a clearly defined excess of *in silico* predicted, CD8$^+$ T-cell epitopes shared between ZIKV and the *Culex* clade viruses, as opposed to predicted CD8$^+$ T-cell epitopes shared between ZIKV and DENV. This association is statistically significant (Table 1). Thus the C-NS5 (Tree 1), C-terminal NS3 HELICc-NS4A (Tree 2) and NS2A-NS3 DEAD (Tree 3) loci had 7%, 30.2% and 28.2% respectively more shared ZIKV-*Culex* clade CD8$^+$ T-cell epitopes per virus than shared ZIKV-DENV CD8$^+$ epitopes per virus. In contrast, all non-recombinant protein loci showed either a significant decrease in this percentage, notably for NS1-HELICc (−17.2%), or no significant increase or decrease for the prM, E and NS4B protein loci.

Overall, for the *Culex* clade viruses there were 42.8% more putative shared ZIKV CD8$^+$ T-cell epitopes predicted *in silico* in the recombinant loci against their non-recombinant loci ($\frac{213.7_{Culex\,[recombinant]}}{149.6_{Culex\,[non-recombinant]}}$). However, for the DENV there were 13.5% less putative shared ZIKV T-cell CD8$^+$ epitopes predicted *in silico* in the recombinant loci against non-recombinant loci $\frac{192.8_{DENV\,[recombinant]}}{222.8_{DENV\,[non-recombinant]}}$).

The epitope prediction algorithms provided support for the ZIKV recombination events described in Trees 1, 2 and 3 (Fig. 4) correlating with their T-cell epitopes experimentally identified *ex vivo* (see Discussion).

3.2.5. Accuracy analysis of in silico CD8$^+$ prediction

To compare the accuracy of the *in silico* whole genome CD8$^+$ T-cell epitope predictions for each MBFV (section 3.2.4) using the 'IEDB data pipeline' we identified 21 published CD8$^+$ T-cell epitopes derived from *ex vivo* epitope mapping studies within the largest locus in the analysis, viz. the NS2a to NS3 DEAD locus. These published epitopes were identified in patients with prior DENV, JEV or WNV convalescent infections (McMurtrey et al., 2008; Larsen et al., 2010; Weiskopf et al., 2013, 2014, 2015a, 2015b, 2016; Turtle et al., 2016, 2017). The IEDB *in silico* CD8$^+$ T-cell predictions identified all of the 21 experimentally observed epitopes, which is consistent with its strong track-record (Moustaftsi et al., 2006). For example, in a comparison of 18 *in silico* prediction methods, the ANN-based (used in IEDB) "mhcflurry" and "nn_align" outperformed other methods for MHC class I 9-mer and 10-mer binding/non-binding classifications (Zhao and Sher, 2018).

Eleven of these 21 epitopes were defined as putatively capable of inducing cross-reactivity with ZIKV CD8$^+$ T-cell epitopes when analyzed using the IEDB data pipeline (Materials and methods, 2.4.1 and 2.4.2) because these potentially shared epitopes were exclusively within the top 2.5% of predicted epitopes using the comparative method index (*CM*), i.e. resulting in a *CM* value < 2.5 for ZIKV (Material and methods), whilst for the remaining 10 epitopes a *CM* value ≥ 2.5 was estimated for ZIKV (Appendix B, 'Results'). This further verifies the robustness of the IEDB data pipeline and the biological basis of the cross-reactive CD8$^+$ T-cell epitope predictions. IEDB is routinely used to determine the HLA molecules of human subjects whose T-cells were used in *ex vivo* functional epitope mapping studies using long peptides (for example Weiskopf et al., 2013; Turtle et al., 2016).

4. Discussion

Our study provides two novel insights which support and build on previous reports of intraspecific homologous recombination in the MBFV (Holmes et al., 1999; Twiddy and Holmes, 2003; Aaskov et al., 2007; reviewed by Weaver and Vasilakis, 2009; Twiddy and Holmes, 2003; Carney et al., 2012; Faye et al., 2014) and interspecific

Table 1
The number of *in silico* predicted shared CD8$^+$ T-cell epitopes per virus between ZIKV and the *Culex* clade MBFV and predicted shared CD8$^+$ T-cell epitopes per virus between the DENV and ZIKV for each protein locus in the analysis. Here 'percentage increase' is when the number of ZIKV – *Culex* shared CD8$^+$ T-cell epitopes exceeds the number of ZIKV – DENV shared epitopes. Epitope predictions were obtained from the 'IEDB data pipeline'.

	Protein loci	Shared CD8$^+$ T-cell epitopes predicted *in silico* per virus with ZIKV			Wilcoxan test
		Culex	DENV	% increase *Culex*/DENV	
Recombinant loci	C – NS5	205	191.5	+7%	< 0.002*
	NS2A – NS3 DEAD	108.3	84.6	+28%	< 0.004*
	NS3 HELICc-NS4A	30.4	23.3	+30.2%	< 0.0043*
None recombinant loci	prM	19	16	-	< 0.56
	E	61.5	61	–	< 0.03**
	NS1-HELICc	78.4	94.9	−17.2%	< 0.002**
	NS4B	45.7	41.8	–	< 0.055

Key*, significant increase.
**, significant decrease.

heterologous recombination in the MBFV (Cook et al., 2012; Barba-Spaeth et al., 2016). First, we discovered evidence of interspecies recombination in ZIKV and SPOV resulting in hybrid viruses derived from DENV- and *Culex*-associated flaviviruses. Second, we demonstrated that these hybrid genomes contained DENV-related B-cell epitope rich loci interspersed with *Culex* virus-related T-cell epitope rich loci. Thus, we hypothesized that DENV-specific T-cell background immunity would be unlikely to protect populations exposed to the hybrid viruses. Under this hypothesis interspecific recombination provides a rational immunological basis for the explosive emergence of ZIKV in appropriate non-immune populations.

Our first discovery primarily resulted from phylogenetic analysis of translated flavivirus genomes, as opposed to nucleotide sequences which are susceptible to saturation of the third codon position (Zanotto et al., 1996). This produced robust evidence of interspecific recombination within the ZIKV genome which we now realise, is a hybrid of *Aedes*-borne DENV proteins interspersed amongst *Culex*-borne flavivirus proteins. We reached this conclusion by identifying disparities in the placement of ZIKV-SPOV between independent MBFV phylogenetic trees that were attributable to incongruence in specific loci within their genomes. For example, ZIKV and SPOV formed a monophyly with DENV, in trees based on (i) prM, E, NS1 proteins, (ii) the central locus of the NS3 HELICc domain and (iii) the NS4B protein which together span three non-contiguous loci across the genomes. Based on these loci, ZIKV and SPOV are included within the *Aedes*-associated clade of viruses. However, they were placed within the *Culex*-associated clade of viruses in the trees based on (i) the C, NS2A, and NS2B proteins, (ii) the NS3 S7 protease and helicase DEAD domains, and (iii) the C-terminal HELICc domain, the NS4A and NS5 proteins, which together also spanned three non-contiguous loci across their genomes. Moreover, within the genome, these three *Culex*-associated MBFV clade loci alternated with the three *Aedes* clade loci. The corresponding C and NS5 protein loci of ZIKV and SPOV are at opposite ends of the open reading frame of the genome but individual gene trees were uniquely congruent and a combined locus resulted in a large increase in non-parametric bootstrap support across all nodes. Lower, non-parametric bootstrap support for the placement of ZIKV and SPOV within the *Culex* clade using an NS5 tree was previously reported (Kuno et al., 2009; Barba-Spaeth et al., 2016) and bootstrap results for placement of the E-protein were comparable. The flavivirus genome circularizes by intraspecific RNA-RNA pairing between 5′ and 3’ UTRs, connecting C and NS5 loci, as a requirement of replication (Hahn et al., 1987), which is a species-specific RNA-RNA pairing that is maintained during viral RNA polymerase detachment. We therefore confirmed the incongruence first reported by Kuno et al. (2009) and obtained robust bootstrap support for phylogenetic incongruence which violates the epidemiological boundary between the *Aedes* and *Culex*-associated viruses within the MBFV phylogeny identified in our previous work (Gaunt et al., 2001). We recognize that phylogenetic incongruence at deeper levels in the phylogeny is a new concept within MBFV and, as with any evolutionary molecular reconstruction, remains a hypothesis. However, the outcome of our analysis is the identification of a systematic series of *in silico* methods that identify broad shifts in the $CD8^+$ T-cell epitope distribution between viruses, in this case, within the MBFV. Independent indirect evidence in support of this hypothesis derives from the recent discovery that ZIKV was vertically transmissible in *Culex* spp., under laboratory conditions (Phumee et al., 2019) and could possibly be transmitted to *Culex* spp. via viraemic urine, i.e., in an aquatic environment, because similar transmission has been demonstrated for *Aedes* spp. (Du et al., 2019).

We propose that the *Aedes* clade-associated regions of the ZIKV and SPOV genomes are ancestral because these three physically separated loci resulted in identical ZIKV-SPOV-DENV monophyletic trees (Fig. 3i), viz.: 1) the prM, E, and NS1 proteins comprise at least one locus; 2) the NS3 HELICc region is a second locus, excluding its C-terminal domain and; 3) the NS4B protein is a third locus. In common with YFV and other arboviruses, they circulate in *Aedes* and primate transmission cycles in African forest canopies (Gould et al., 2003). In the *Culex*-associated MBFV clade, loci of the ZIKV and SPOV genomes produced three incongruent phylogenies as shown by ‘Trees 1, 2, and 3’ (Fig. 4). These loci therefore represent at least three separate recombination events. The demonstration that ZIKV, SPOV and DENV (and KEDV) fall in the *Aedes* clade (i.e. ancestral *Aedes and* primate transmission) and that ZIKV and SPOV are monophyletic, demonstrating identical recombination patterns throughout the phylogenetically incongruent loci across the genome, implies that a single ZIKV-SPOV ancestor (*Aedes* clade) recombined with three ancestral *Culex* clade viruses. This can be inferred because the ZIKV-SPOV lineages in the three *Culex* clade-associated trees (Fig. 4) are robustly incongruent with respect to the placement of ZIKV-SPOV.

It was previously proposed that the ancestral viruses of DENV, ZIKV and *Culex*-transmitted MBFV were transmitted by mosquitoes (Gould et al., 2003; Huhtamo et al., 2009; Gubler, 2014) and probably had a range of transmission vectors (Gould and Solomon, 2008; Kuno, 2016). Thus, in this or other scenarios ancestral ZIKV versus *Culex*-associated virus recombination events probably occurred during co-infections in primates (*Aedes* spp. vectors) and/or mosquitoes in the sylvatic environment of the African rainforests because extant ZIKV in sylvatic primates predominates in densely forested African bioregions including The Gambia and Uganda (Dick, 1952; Buechler et al., 2017). A detailed series of transmission scenarios is presented in Appendix D, and Supplementary Table 3 (Appendix D). We propose that the resulting recombinant virus reported here would enable ancestral ZIKV and DENV to co-circulate in the same primate hosts and vectors. These results and interpretations infer that the immune selection, described below most likely took place in non-human primates.

The second discovery from this study was that recombination between *Aedes* and *Culex*-borne MBFV protein loci across the ZIKV-SPOV genomes correlated with DENV B-cell and *Culex*-borne MBFV T-cell immunology. Immunodominant B-cell (antibody) reactive proteins were exclusively encoded in *Aedes*-clade-associated protein loci, namely the prM, E, and NS1 proteins and the central region of the NS3 HELICc domain, with the exception of the NS4B protein (Figs. 2 and 3i). DENV T-cell epitopes experimentally observed *ex vivo* were predominantly located in the C, NS2A to the NS3 DEAD domain, C-terminus HELICc to NS4A protein loci inclusively and the NS5 protein (Fig. 2, 3ii and 4), this precisely coincides with the ZIKV loci that had recombined with *Culex*-associated MBFV. The NS4B protein phylogeny was therefore the only exception in the genome (1/10 proteins), and interestingly this protein inhibits antiviral type I α/β interferon (Muñoz-Jordán et al., 2005). Evidence that the NS5 protein also contains B-cell epitopes is equivocal, but may be coincident with the low bootstrap support for the ZIKV-SPOV lineage associating with the *Culex* clade.

Interspecific or trans-species genetic recombination provides a phylogenetic explanation for the observation of T-cell specificity experimentally identified *ex vivo* and B-cell cross-reactivity between DENV and ZIKV. For example, there was a lack of cross-reactivity between $CD4^+$/$CD8^+$ T-cells from DENV and ZIKV infections, for the C protein (Lim et al., 2018), whilst the NS3 protein was cross-reactive. Furthermore, the $CD8^+$ T-cell response of ZIKV NS3 HELICc, excluding its C-terminal domain, showed significant cross-reactivity against DENV-infected cells (Herrera et al., 2018), but in contrast the NS3 protease was highly specific and all these results were highly compatible with our findings. Our discovery of ZIKV recombination is supported by and also partly explains the extensive cross-reactivity of $CD4^+$ T-cells between ZIKV and WNV but low cross-reactivity between ZIKV and YFV (Reynolds et al., 2018). It is also partly supported by the distinct genome specificity of $CD8^+$ T-cells, involving cytotoxicity against ZIKV infected cells, which is independent of DENV exposure in holoendemic geographic regions (Grifoni et al., 2018). However, previous DENV infection does alter the T-cell memory against ZIKV peptides which could result in more efficient control or clearance of the

infection (Grifoni et al., 2017).

We propose that immune-selection to escape cross-protection due to $CD8^+$ and $CD4^+$ cytotoxic T-cells, generated during previous DENV infections, could give the recombinant ZIKV an epidemiological advantage in DENV-endemic areas. This is compatible with current epidemiological observations. The rationale for this hypothesis is that conserved inter-serotypic DENV $CD8^+$ T-cell epitopes contribute to cross-protection against sequential DENV infections with a different serotype (Weiskopf et al., 2013). However, further comparative studies will be required to assess T-cell and antibody responses against ZIKV and DENV for autochthonous infections in SE Asia, the Pacific Islands and the Americas.

The initial support for recombination as a mechanism of ZIKV T-cell immune escape was the correlation between the ZIKV protein loci showing phylogenetic incongruence and separately the established DENV T-cell epitope map experimentally identified, *ex vivo*. The subsequent support for this mechanism was the *in silico* analysis using *in silico* cross-reactive HLA class I $CD8^+$ T-cell epitope prediction algorithms for all of the ZIKV-SPOV recombinant and non-recombinant protein loci using the 'IEDB data pipeline'. All protein loci identified in the ZIKV-SPOV-*Culex* clade recombination events showed an increase of 7.0% (C-NS5 loci), 28.0% (NS2A-NS3 DEAD locus) and 30.3% (C-terminal NS3 HELICc to NS4A loci) of putative shared $CD8^+$ T-cell epitopes when compared with ZIKV-like $CD8^+$ T-cell epitopes in the DENV. The C-terminal NS3 HELICc region to the NS4A protein locus was the most specific recombination event occurring between ZIKV-SPOV and Australasian/Melanesian *Culex* clade II viruses (Tree 2). This represented 50% of all DENV $CD8^+$ HLA B40:01 T-cell epitopes confirmed by *ex vivo* experimental analysis (Weiskopf et al., 2013, 2015b). In summary, ZIKV-SPOV recombinant loci correlated with the DENV-skewed distribution of T-cell epitopes experimentally identified *ex vivo* across the DENV genome. Furthermore, when the recombinant loci were subjected to whole genome *in silico* $CD8^+$ T-cell predictions for putative cross-reactive epitopes between ZIKV and all other MBFV, these recombinant loci showed a significant decrease in the putative cross-reactive $CD8^+$ T-cell responses between DENV and ZIKV than between *Culex*-associated MBFV and ZIKV. We postulate that these results are not a “coincidence” but are the result of recombination that is driven by selection to facilitate the co-circulation of both ZIKV and DENV within the same hosts (primates) and thus reducing the cross-protective memory T-cell responses from prior DENV infections. Admittedly, little is currently known about the T-cell responses of Old World primates. It is noteworthy that all published $CD8^+$ T-cell epitopes within the NS2A-NS3 DEAD locus were identified by the IEDB data pipeline and > 50% (11/21) of these epitopes would be classed as putative shared $CD8^+$ T-cell epitopes between ZIKV and other MBFV.

Thus, ‘Old World’ MBFV including WNV, JEV (*Culex* clade I), *Culex* clade II Stratford virus (STRV) and Kokobera virus (KOKV) could provide cross-protective T-cell responses against ZIKV-infected cells, for example, via $CD8^+$ cytotoxic T-cells generated against the NS2A-NS3 DEAD protein locus. On the other hand, DENV would be far less susceptible to cross-protective $CD8^+$ T-cell responses from WNV, JEV, STRV and KOKV because DENV is not a recombinant with the *Culex* clade MBFV loci. Cross-reactive short-term $CD8^+$ T-cell responses from healthy JEV exposed donors were extensive against DENV and WNVnonstrucutral proteins (Turtle et al., 2016), although in our analysis they appeared to be relatively limited and they were also notably restricted using T-cells from convalescent recipients of the live attenuated SA-14-14-2 JEV vaccine (Turtle et al., 2017).

Early (1950s–1960s) seroprevalence studies performed on human populations in Asia, showed high incidence of protective antibodies against DENV, JEV and WNV, as judged in mouse protection tests (Smithburn et al., 1954; Hammon et al., 1958a,b; Pond, 1963; Rao, 1971). WNV was serologically identified or isolated in most countries from India to Indonesia, the Philippines and China (Chancey et al., 2015) with 16%–58% preponderance in the most specific serological assay through the challenge of human sera treated mice with high (> 50 LD_{50}) doses of ZIKV (Smithburn, 1954; Smithburn et al., 1954; Pond, 1963; Hammon et al. 1958a, 1958b, 1958). More recently WNV was reported to have an apparent 27% seroprevalence in Xinjiang, China (Cao et al., 2017). This contrasts with the very low seroprevalence to WNV and JEV in French Polynesia (1.5% and 1.3% respectively) in 2011–2013, whilst seroprevalence to DENV was 80.3%, supporting the belief that WNV and JEV do not circulate in French Polynesia (Aubry et al., 2015).

Our observation that *in silico*-predicted ZIKV-specific $CD8^+$ T-cell epitopes are predominantly represented by related *Culex* clade flaviviruses, most likely through interspecific recombination, provides a rational explanation for the different epidemiological patterns of ZIKV persistence, characterized by small, localized outbreaks in SE Asia/Australasia/Melanesia, such as 30 cases in Klaten, Central Java, Indonesia (Olson et al. 1981), where WNV, JEV, Murray Valley encephalitis and other related flaviviruses have been endemic for many years. This dramatically contrasts with the explosive outbreaks in Micronesia, French Polynesia and the Americas, despite the presence of endemic DENV (Aubry et al., 2015; Duffy et al., 2009; Musso, 2015; Pettersson et al., 2016, 2018; Gould et al., 2017). Thus, the *Culex*-associated JEV-complex flaviviruses distributed throughout SE Asia and Australasia/Melanesia would provide a background of cross-protective T-cell immunity to ZIKV. However, these viruses are either absent or present at very low prevalence on the Pacific Islands (Aubry et al., 2015). A different scenario pertains in Central/Southern America where *Culex*-associated MBFV, including, BSQV, AROAV, NJLV, IGUV, ROCV, SLEV and WNV circulate primarily via peridomestic and/or sylvan *Culex* spp., rather than domestic urban species. In support of these arguments, DENV, which did not recombine with the Asian *Culex* clade viruses, circulates as an endemic and epidemic virus in the geographic regions where ZIKV also circulates relatively silently.

A parsimonious hypothesis infers that the geographic site for the interspecific recombination of the ZIKV-SPOV ancestor is Africa because ZIKV, SPOV, and KEDV, were originally isolated in sub-Saharan Africa and only ZIKV and very recently SPOV, have emerged out of Africa (White et al., 2018), inferring an African origin for these three viruses. Therefore, because recombination occurred immediately prior to the divergence of two extant African viruses, ZIKV and SPOV, it is the most likely geographic region for the recombination.

There are at least four alternative possibilities for the apparent lack of ZIKV epidemics in Asia. One possibility is explained by the accounts of background immunity of 2–19% neutralizing antibodies against ZIKV in Bangkok (Thailand), Borneo, Laos, Malay Peninsula and western India in seroprevalence surveys (Smithburn, 1954; Smithburn et al., 1954; Pond, 1963; Pastorino et al., 2019), isolated in Malaysia (Marchette et al., 1969) and Indonesia (Olson et al., 1981) (Appendix E). However, there were no detectable ZIKV neutralizing antibodies in northern Viet Nam, or Chiang Mai, Thailand (Pond, 1963). Three further possible scenarios that could explain ZIKV epidemiology in SE Asia are presented in Appendix E and involve ZIKV immunology in relation to other MBFV infections (e.g., Gordon et al., 2019; Masel et al., 2019).

In summary, interspecific recombination could account for why ZIKV did not become epidemic in SE Asia and Australasia/Melanesia, but caused explosive epidemics when it was introduced to the islands of the South Pacific and the Americas, despite widespread prior exposure to DENV but low exposure to *Culex*-borne MBFV. Our findings raise many implications for future studies of MBFV.

Funding

This work was designed, initiated, conducted and financially supported through ZikaPLAN Project number 734584 by the European Commission under Horizon 2020 funding. We are grateful for the full participation of our ZIKAlliance Project partners X de L, JP and EAG, also part of Horizon 2020. We express sincere gratitude to Daniela

Weiskopf and all associated colleagues at the La Jolla Institute for Allergy and Immunology, San Diego, California, USA for supplying the $CD8^+$ and $CD4^+$ T-cell epitope DENV maps and for their comments on the manuscript.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Appendix A. Supplementary data

Supplementary data to this article comprising Appendix A (Fig. S1 and "Phylogenetics"; media component 1), Appendix B ("T-cell bioinformatics"; media component 2), Appendix C (Figs. S2 and S3; media component 3), Appendix D ("Supplementary tables 1, 2 and 3; media component 4) and Appendix E ("Miscellaneous data: ZIKV epidemiology"; media component 5) can be found online at https://doi.org/10.1016/j.antiviral.2019.104676.

References

Aaskov, J., Buzacott, K., Field, E., Lowry, K., Berlioz-Arthaud, A., Holmes, E.C., 2007. Multiple recombinant dengue type 1 viruses in an isolate from a dengue patient. J. Gen. Virol. 88, 3334–3340.

Aubry, M., Finke, J., Teissier, A., Roche, C., Broult, J., Paulous, S., Desprès, P., Cao-Lormeau, V.M., Musso, D., 2015. Seroprevalence of arboviruses among blood donors in French Polynesia, 2011-2013. Int. J. Infect. Dis. 41, 11–12.

Barba-Spaeth, G., Dejnirattisai, W., Rouvinski, A., Vaney, M.C., Medits, I., Sharma, A., Simon-Lorière, E., Sakuntabhai, A., Cao-Lormeau, V.M., Haouz, A., England, P., Stiasny, K., Mongkolsapaya, J., Heinz, F.X., Screaton, G.R., Rey, F.A., 2016. Structural basis of potent Zika-dengue virus antibody cross-neutralization. Nature 536, 48–53.

Bryant, D., Moulton, V., 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21, 255–265.

Buechler, C.R., Bailey, A.L., Weiler, A.M., Barry, G.L., Breitbach, M.E., Stewart, L.M., Jasinska, A.J., Freimer, N.B., Apetrei, C., Phillips-Conroy, J.E., Jolly, C.J., Rogers, J., Friedrich, T.C., O'Connor, D.H., 2017. Seroprevalence of zika virus in wild African green Monkeys and Baboons. mSphere 2 e00392-16.

Cao, L., Fu, S., Lv, Z., Tang, C., Cui, S., Li, X., Gao, X., Li, M., Cao, Y., Lei, W., He, Y., Wang, H., Liang, G., 2017. West Nile virus infection in suspected febrile typhoid cases in Xinjiang, China. Emerg. Microb. Infect. 6, e41.

Carney, J., Daly, J.M., Nisalak, A., Solomon, T., 2012. Recombination and positive selection identified in complete genome sequences of Japanese encephalitis virus. Arch. Virol. 157, 75–83.

Chancey, C., Grinev, A., Volkova, E., Rios, M., 2015. The global ecology and epidemiology of West Nile virus. Biol. Med. Res. Int. 376230. https://doi.org/10.1155/2015/376230.

Church, S.H., Ryan, J.F., Dunn, C.W., 2015. Automation and evaluation of the SOWH test with SOWHat. Syst. Biol. 64, 1048–1058.

Cook, S., Moureau, G., Kitchen, A., Gould, E.A., de Lamballerie, X., Holmes, E.C., Harbach, R.E., 2012. Molecular evolution of the insect-specific flaviviruses. J. Gen. Virol. 93, 223–234.

Costa, S.M., Yorio, A.P., Gonçalves, A.J., Vidale, M.M., Costa, E.C., Mohana-Borges, R., Motta, M.A., Freire, M.S., Alves, A.M., 2011. Induction of a protective response in mice by the dengue virus NS3 protein using DNA vaccines. PLoS One 6, e25685.

da Silva, A.N., Nascimento, E.J., Cordeiro, M.T., Gil, L.H., Abath, F.G., Montenegro, S.M., Marques, E.T., 2009. Identification of continuous human B-cell epitopes in the envelope glycoprotein of dengue virus type 3 (DENV-3). PLoS One 4, e742.

Dejnirattisai, W., Supasa, P., Wongwiwat, W., Rouvinski, A., Barba-Spaeth, G., Duangchinda, T., Sakuntabhai, A., Cao-Lormeau, V.M., Malasit, P., Rey, F.A., Mongkolsapaya, J., Screaton, G.R., 2016. Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. Nat. Immunol. 17, 1102–1108.

Dick, G.W., 1952. Epidemiological notes on some viruses isolated in Uganda; yellow fever, Rift valley fever, Bwamba fever, West Nile, Mengo, Semliki forest, Bunyamwera, Ntaya, Uganda S and zika viruses. Trans. R. Soc. Med. Hyg. 47, 13–48.

Du, S., Lui, Y., Lui, J., Zhao, J., Champagne, C., Tong, L., Zhang, R., Zhang, F., Qin, C.-F., Ma, P., Chen, C.-H., Liang, G., Liu, Q., Shi, P.-Y., Cazelles, B., Wang, P., Tian, H., Cheng, G., 2019. *Aedes* mosquitoes acquire and transmit Zika virus by breeding in contaminated aquatic environments. Nat. Commun. 10, 1324.

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., Hayes, E.B., 2009. Zika virus outbreak on Yap island, Federated States OF Micronesia. N. Engl. J. Med. 360, 2536–2543.

Edinur, H.A., Dunn, P.P., Hammond, L., Selwyn, C., Brescia, P., Askar, M., Reville, P., Velickovic, Z.M., Lea, R.A., Chambers, G.K., 2013. HLA and MICA polymorphism in Polynesians and New Zealand Maori: implications for ancestry and health. Hum. Immunol. 74, 1119–1129.

Falconar, A.K., 1999. Identification of an epitope on the dengue virus membrane (M) protein defined by cross-protective monoclonal antibodies: design of an improved epitope sequence based on common determinants present in both envelope (E and M) proteins. Arch. Virol. 144, 2313–2330.

Falconar, A.K., 2007. Antibody responses are generated to immunodominant ELK/KLE-type motifs on the nonstructural-1 glycoprotein during live dengue virus infections in mice and humans: implications for diagnosis, pathogenesis, and vaccine design. Clin. Vaccine Immunol. 14, 493–504.

Faye, O., Freire, C.C., Iamarino, A., Faye, O., de Oliveira, J.V., Diallo, M., Zanotto, P.M., Sall, A.A., 2014. Molecular evolution of Zika virus during its emergence in the 20(th) century. PLoS Neglected Trop. Dis. 8, e2636.

Felix, A.C., Souza, N.C.S., Figueiredo, W.M., Costa, A.A., Inenami, M., da Silva, R.M.G., Levi, J.E., Pannuti, C.S., Romano, C.M., 2017. Cross reactivity of commercial anti-dengue immunoassays in patients with acute Zika virus infection. J. Med. Virol. 89, 1477–1479.

Gaunt, M.W., Sall, A.A., de Lamballerie, X., Falconar, A.K., Dzhivanian, T.I., Gould, E.A., 2001. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J. Gen. Virol. 82, 1867–1876.

González-Galarza, F.F., Takeshita, L.Y., Santos, E.J., Kempson, F., Maia, M.H., da Silva, A.L., Teles e Silva, A.L., Ghattaoraya, G.S., Alfirevic, A., Jones, A.R., Middleton, D., 2015. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Res. 43, 784–788.

Gordon, A., Gresh, L., Ojeda, S., Katzelnick, L.C., Sanchez, N., Mercado, J.C., Chowell, G., Lopez, B., Elizondo, D., Coloma, J., Burger-Calderon, R., Kuan, G., Balmaseda, A., Harris, E., 2019. Prior dengue virus infection and risk of Zika: a pediatric cohort in Nicaragua. PLoS Med. 16, e1002726.

Gould, E.A., Solomon, T., 2008. Pathogenic flaviviruses. Lancet 371, 500–509.

Gould, E.A., de Lamballerie, X., Zanotto, P.M., Holmes, E.C., 2003. Origins, evolution, and vector/host coadaptations within the genus Flavivirus. Adv. Virus Res. 59, 277–314.

Gould, E.A., Pettersson, J., Higgs, S., Charrel, R., de Lamballerie, X., 2017. Emerging arboviruses: why today? One Health 4, 1–13.

Grard, G., Moureau, G., Charrel, R.N., Holmes, E.C., Gould, E.A., de Lamballerie, X., 2010. Genomics and evolution of *Aedes*-borne flaviviruses. J. Gen. Virol. 91, 87–94.

Grifoni, A., Pham, J., Sidney, J., O'Rourke, P.H., Paul, S., Peters, B., Martini, S.R., de Silva, A.D., Ricciardi, M.J., Magnani, D.M., Silveira, C.G.T., Maestri, A., Costa, P.R., de-Oliveira-Pinto, L.M., de Azeredo, E.L., Damasco, P.V., Phillips, E., Mallal, S., de Silva, A.M., Collins, M., Durbin, A., Diehl, S.A., Cerpas, C., Balmaseda, A., Kuan, G., Coloma, J., Harris, E., Crowe, J.E., Stone, M., Norris, P.J., Busch, M., Vivanco-Cid, H., Cox, J., Graham, B.S., Ledgerwood, J.E., Turtle, L., Solomon, T., Kallas, E.G., Watkins, D.I., Weiskopf, D., Sette, A., 2017. Prior Dengue virus exposure shapes T cell immunity to Zika virus in humans. J. Virol. 91 e01469-17.

Grifoni, A., Costa-Ramos, P., Pham, J., Tian, Y., Rosales, S.L., Seumois, G., Sidney, J., de Silva, A.D., Premkumar, L., Collins, M.H., Stone, M., Norris, P.J., Romero, C.M.E., Durbin, A., Ricciardi, M.J., Ledgerwood, J.E., de Silva, A.M., Busch, M., Peters, B., Vijayanand, P., Harris, E., Falconar, A.K., Kallas, E., Weiskopf, D., Sette, A., 2018. Cutting edge: transcriptional profiling reveals multifunctional and cytotoxic antiviral responses of zika virus-specific CD8. J. Immunol. 201, 3487–3491.

Gubler, D.J., 2014. Dengue viruses: their evolution, history and emergence as a global public health problem. In: Gubler, D.J., Ooi, E.E., Vasudevan, S., Farrar, J. (Eds.), Dengue Hemorragic Fever, second ed. Cabi, Wallingford, pp. 1–29.

Hahn, C.S., Hahn, Y.S., Rice, C.M., Lee, E., Dalgarno, L., Strauss, E.G., Strauss, J.H., 1987. Conserved elements in the 3' untranslated region of flavivirus RNAs and potential cyclization sequences. J. Mol. Biol. 198, 33–41.

Hamer, D.H., Chen, L.H., 2019. Zika in Angola and India. J. Travel Med. 26 pii: taz012.

Hammon, W.M., Schrack, W.D., Sather, G.E., 1958a. Serological survey for a arthropod-borne virus infections in the Philippines. Am. J. Trop. Med. Hyg. 7, 323–328.

Hammon, W.M., Tigertt, W.D., Sather, G.E., Berge, T.O., Meiklejohn, G., 1958b. Epidemiologic studies of concurrent virgin epidemics of Japanese B encephalitis and of mumps on Guam, 1947-1948, with subsequent observations including dengue, through 1957. Am. J. Trop. Med. Hyg. 7, 441–467.

Herrera, B.B., Tsai, W.Y., Chang, C.A., Hamel, D.J., Wang, W.K., Lu, Y., Mboup, S., Kanki, P.J., 2018. Sustained specific and cross-reactive T cell responses to Zika and dengue Virus NS3 in West Africa. J. Virol. 92 e01992-17.

Holmes, E.C., Worobey, M., Rambaut, A., 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. 16, 405–409.

Huhtamo, E., Putkuri, N., Kurkela, S., Manni, T., Vaheri, A., Vapalahti, O., Uzcátegui, N.Y., 2009. Characterization of a novel flavivirus from mosquitoes in northern europe that is related to mosquito-borne flaviviruses of the tropics. J. Virol. 83, 9532–9540.

Huhtamo, E., Cook, S., Moureau, G., Uzcátegui, N.Y., Sironen, T., Kuivanen, S., Putkuri, N., Kurkela, S., Harbach, R.E., Firth, A.E., Vapalahti, O., Gould, E.A., de Lamballerie, X., 2014. Novel flaviviruses from mosquitoes: mosquito-specific evolutionary lineages within the phylogenetic group of mosquito-borne flaviviruses. Virology 464–465, 320–329.

Kolodziejek, J., Pachler, K., Bin, H., Mendelson, E., Shulman, L., Orshan, L., Nowotny, N., 2013. Barkedji virus, a novel mosquito-borne flavivirus identified in *Culex perexiguus* mosquitoes, Israel, 2011. J. Gen. Virol. 94, 2449–2457.

Kuno, G., 2016. The boundaries of arboviruses: complexities revealed in their host ranges, virus-host interactions and evolutionary relationships. In: Vasilakis, Nikos, Gubler, Duane J. (Eds.), Arboviruses: Molecular Biology, Evolution and Control. Caister Academic Press, U.K, pp. 219–268.

Kuno, G., Chang, G.J., Tsuchiya, K.R., Karabatsos, N., Cropp, C.B., 1998. Phylogeny of the genus flavivirus. J. Virol. 72, 73–83.

Kuno, G., Chang, G.J.J., Chien, L.L., 2009. Correlations of phylogenetic relation with host range, length of ORF or genes, organisation of conserved sequences in the 3' non-coding region, and viral classification among the members of the genus Flavivirus. In: Feng, Z., Long, M. (Eds.), Viral Genomes: Diversity, Properties Parameters. Nova

Science Publishers, Hauppauge, New York.
Lai, C.Y., Tsai, W.Y., Lin, S.R., Kao, C.L., Hu, H.P., King, C.C., Wu, H.C., Chang, G.J., Wang, W.K., 2008. Antibodies to envelope glycoprotein of dengue virus during the natural course of infection are predominantly cross-reactive and recognize epitopes containing highly conserved residues at the fusion loop of domain II. J. Virol. 82, 6631–6643.
Larsen, M.V., Lelic, A., Parsons, R., Nielsen, M., Hoof, I., Lamberth, K., Loeb, M.B., Buus, S., Bramson, J., Lund, O., 2010. Identification of CD8$^+$ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using NetCTL. PLoS One 5, e12697.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Lemey, P., Lott, M., Martin, D.P., Moulton, V., 2009. Identifying recombinants in human and primate immunodeficiency virus sequence alignments using quartet scanning. BMC Bioinf. 10, 126.
Lim, M.Q., Kumaran, E.A.P., Tan, H.C., Lye, D.C., Leo, Y.S., Ooi, E.E., MacAry, P.A., Bertoletti, A., Rivino, L., 2018. Cross-reactivity and anti-viral function of dengue capsid and NS3-specific memory T cells toward Zika Virus. Front. Immunol. 9, 2225.
Liu, P., Lu, H., Li, S., Moureau, G., Deng, Y.Q., Wang, Y., Zhang, L., Jiang, T., de Lamballerie, X., Qin, C.F., Gould, E.A., Su, J., Gao, G.F., 2012. Genomic and antigenic characterization of the newly emerging Chinese duck egg-drop syndrome flavivirus: genomic comparison with Tembusu and Sitiawan viruses. J. Gen. Virol. 93, 2158–2170.
Marchette, N.J., Garcia, R., Rudnick, A., 1969. Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia. Am. J. Trop. Med. Hyg. 18, 411–415.
Masel, J., McCracken, M.K., Gleeson, T., Morrison, B., Rutherford, G., Imrie, A., Jarman, R.G., Koren, M., Pollett, S., 2019. Does prior dengue virus exposure worsen clinical outcomes of Zika virus infection? A systematic review, pooled analysis and lessons learned. PLoS Neglected Trop. Dis. 13, e0007060.
Masrinoul, P., Omokoko, M.D., Pambudi, S., Ikuta, K., Kurosu, T., 2013. Serotype-specific anti-Dengue virus NS1 mouse antibodies cross-react with prM and are potentially involved in virus production. Viral Immunol. 26, 250–258.
McMurtrey, C.P., Lelic, A., Piazza, P., Chakrabarti, A.K., Yablonsky, E.J., Wahl, A., Bardet, W., Eckerd, A., Cook, R.L., Hess, R., Buchli, R., Loeb, M., Rinaldo, C.R., Bramson, J., Hildebrand, W.H., 2008. Epitope discovery in West Nile virus infection: identification and immune recognition of viral epitopes. Proc. Natl. Acad. Sci. (USA) 105, 2981–2986.
Moureau, G., Cook, S., Lemey, P., Nougairede, A., Forrester, N.L., Khasnatinov, M., Charrel, R.N., Firth, A.E., Gould, E.A., de Lamballerie, X., 2015. New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. PLoS One 10, e0117849.
Moutaftsi, M., Peters, B., Pasquetto, V., Tscharke, D.C., Sidney, J., Bui, H.H., Grey, H., Sette, A., 2006. A consensus epitope prediction approach identifies the breadth of murine T(CD8$^+$)-cell responses to vaccinia virus. Nat. Biotechnol. 24, 817–819.
Muñoz-Jordán, J.L., Laurent-Rolle, M., Ashour, J., Martínez-Sobrido, L., Ashok, M., Lipkin, W.I., García-Sastre, A., 2005. Inhibition of alpha/beta interferon signaling by the NS4B protein of flaviviruses. J. Virol. 79, 8004–8013.
Musso, D., 2015. Zika virus transmission from French Polynesia to Brazil. Emerg. Infect. Dis. 21, 1887.
Nurtop, E., Villarroel, P.M.S., Pastorino, B., Ninove, L., Drexler, J.F., Roca, Y., Gake, B., Dubot-Peres, A., Grard, G., Peyrefitte, C., Priet, S., de Lamballerie, X., Gallian, P., 2018. Combination of ELISA screening and seroneutralisation tests to expedite Zika virus seroprevalence studies. Virol. J. 15, 192.
Olson, J.G., Ksiazek, T., Suhandiman, O., Triwibowo, S., 1981. Zika virus, a cause of fever in Central Java, Indonesia. Trans. Roy. Soc. Trop. Med. Hyg. 75, 389–393.
Pastorino, B., Sengvilaipaseuth, O., Chanthongthip, A., Vongsouvath, M., Souksakhone, C., Mayxay, M., Thirion, L., Newton, P.N., de Lamballerie, X., Dubot-Pérès, A., 2019. Low zika virus seroprevalence in Vientiane, Laos, 2003-2015. Am. J. Trop. Med. Hyg. 100, 639–642.
Pettersson, J.H., Eldholm, V., Seligman, S.J., Lundkvist, Å., Falconar, A.K., Gaunt, M.W., Musso, D., Nougairède, A., Charrel, R., Gould, E.A., de Lamballerie, X., 2016. How did zika virus emerge in the Pacific islands and Latin America. mBio 7 e01239-16.
Pettersson, J.H., Bohlin, J., Dupont-Rouzeyrol, M., Brynildsrud, O.B., Alfsnes, K., Lormeau, Cao, Gaunt, M.W., Falconar, A.K., de Lamballerie, X., Eldholm, V., Musso, D., Gould, E.A., 2018. Re-visiting the evolution, dispersal and epidemiology of Zika virus in Asia. Emerg. Microb. Infect. 7, 79.
Phumee, A., Chompoosri, J., Intayot, P., Boonserm, R., Boonyasuppayakorn, S., Buathong, R., Thavara, U., Tawatsin, A., Joyjinda, Y., Wacharapluesadee, S., Siriyasatien, P., 2019. Vertical transmission of Zika virus in Culex quinquefasciatus Say and Aedes aegypti (L.) mosquitoes. Sci. Rep. 9, 5257.
Pond, W.L., 1963. Arthropod-borne virus antibodies in sera from residents of South-East Asia. Trans. R. Soc. Trop. Med. Hyg. 57, 364–371.
Priyamvada, L., Quicke, K.M., Hudson, W.H., Onlamoon, N., Sewatanon, J., Edupuganti, S., Pattanapanyasat, K., Chokephaibulkit, K., Mulligan, M.J., Wilson, P.C., Ahmed, R., Suthar, M.S., Wrammert, J., 2016. Human antibody responses after dengue virus infection are highly cross-reactive to Zika virus. Proc. Natl. Acad. Sci. (USA) 113, 7852–7857.
Rao, T.R., 1971. Immunological surveys of arbovirus infections in South-East Asia, with special reference to dengue, chikungunya, and Kyasanur Forest disease. Bull. World Health Organ. 44, 585–591.
Reynolds, C.J., Suleyman, O.M., Ortega-Prieto, A.M., Skelton, J.K., Bonnesoeur, P., Blohm, A., Carregaro, V., Silva, J.S., James, E.A., Maillère, B., Dorner, M., Boyton, R.J., Altmann, D.M., 2018. T cell immunity to Zika virus targets immunodominant epitopes that show cross-reactivity with other Flaviviruses. Sci. Rep. 8, 672.
Smithburn, K.C., 1954. Neutralizing antibodies against arthropod-borne viruses in the sera of long-time residents of Malaya and Borneo. Am. J. Hyg. 59, 157–163.
Smithburn, K.C., Kerr, J.A., Gatne, P.B., 1954. Neutralizing antibodies against certain viruses in the sera of residents of India. J. Immunol. 72, 248–257.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogeny. Bioinformatics 30, 1312–1313.
Stettler, K., Beltramello, M., Espinosa, D.A., Graham, V., Cassotta, A., Bianchi, S., Vanzetta, F., Minola, A., Jaconi, S., Mele, F., Foglierini, M., Pedotti, M., Simonelli, L., Dowall, S., Atkinson, B., Percivalle, E., Simmons, C.P., Varani, L., Blum, J., Baldanti, F., Cameroni, E., Hewson, R., Harris, E., Lanzavecchia, A., Sallusto, F., Corti, D., 2016. Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. Science 353, 823–826.
Turtle, L., Bali, T., Buxton, G., Chib, S., Chan, S., Soni, M., Hussain, M., Isenman, H., Fadnis, P., Venkataswamy, M.M., Satishkumar, V., Lewthwaite, P., Kurioka, A., Krishna, S., Shankar, M.V., Ahmed, R., Begum, A., Ravi, V., Desai, A., Yoksan, S., Fernandez, S., Willberg, C.B., Kloverpris, H.N., Conlon, C., Klenerman, P., Satchidanandam, V., Solomon, T., 2016. Human T cell responses to Japanese encephalitis virus in health and disease. J. Exp. Med. 213, 1331–1352.
Turtle, L., Tatullo, F., Bali, T., Ravi, V., Soni, M., Chan, S., Chib, S., Venkataswamy, M.M., Fadnis, P., Yaïch, M., Fernandez, S., Klenerman, P., Satchidanandam, V., Solomon, T., 2017. Cellular immune responses to live attenuated Japanese encephalitis (JE) vaccine SA14-14-2 in adults in a JE/dengue Co-endemic area. PLoS Neglected Trop. Dis. 11, e0005263.
Twiddy, S.S., Holmes, E.C., 2003. The extent of homologous recombination in members of the genus Flavivirus. J. Gen. Virol. 84, 429–440.
van Meer, M.P.A., Mögling, R., Klaasse, J., Chandler, F.D., Pas, S.D., van der Eijk, A.A., Koopmans, M.P.G., Reusken, C.B.E.M., GeurtsvanKessel, C.H., 2017. Re-evaluation of routine dengue virus serology in travellers in the era of Zika virus emergence. J. Clin. Virol. 92, 25–31.
Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., Peters, B., 2019. The immune epitope Database (IEDB): 2018 update. Nucleic Acids Res. 47, 339–343.
Watts, A.G., Huber, C., Bogoch, I.I., Brady, O.J., Kraemer, M.U.G., Khan, K., 2018. Potential Zika virus spread within and beyond India. J. Travel Med. 25 (1). https://doi.org/10.1093/jtm/tay132.
Weaver, S.C., Vasilakis, N., 2009. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect. Genet. Evol. 9, 523–540.
Weiskopf, D., Angelo, M.A., de Azeredo, E.L., Sidney, J., Greenbaum, J.A., Fernando, A.N., Broadwater, A., Kolla, R.V., De Silva, A.D., de Silva, A.M., Mattia, K.A., Doranz, B.J., Grey, H.M., Shresta, S., Peters, B., Sette, A., 2013. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8$^+$ T cells. Proc. Natl. Acad. Sci. (USA) 110, 2046–2053.
Weiskopf, D., Angelo, M.A., Sidney, J., Peters, B., Shresta, S., Sette, A., 2014. Immunodominance changes as a function of the infecting dengue virus serotype and primary versus secondary infection. J. Virol. 88, 11383–11394.
Weiskopf, D., Cerpas, C., Angelo, M.A., Bangs, D.J., Sidney, J., Paul, S., Peters, B., Sanches, F.P., Silvera, C.G., Costa, P.R., Kallas, E.G., Gresh, L., de Silva, A.D., Balmaseda, A., Harris, E., Sette, A., 2015a. Human CD8+ T-Cell responses against the 4 dengue virus serotypes are associated with distinct patterns of protein targets. J. Infect. Dis. 212, 1743–1751.
Weiskopf, D., Angelo, M.A., Bangs, D.J., Sidney, J., Paul, S., Peters, B., de Silva, A.D., Lindow, J.C., Diehl, S.A., Whitehead, S., Durbin, A., Kirkpatrick, B., Sette, A., 2015b. The human CD8$^+$ T cell responses induced by a live attenuated tetravalent dengue vaccine are directed against highly conserved epitopes. J. Virol. 89, 120–128.
Weiskopf, D., Angelo, M.A., Grifoni, A., O'Rourke, P.H., Sidney, J., Paul, S., De Silva, A.D., Phillips, E., Mallal, S., Premawansa, S., Premawansa, G., Wijewickrama, A., Peters, B., Sette, A., 2016. HLA-DRB1 alleles are associated with different magnitudes of dengue virus-specific CD4$^+$ T-cell responses. J. Infect. Dis. 214, 1117–1124.
Wen, J., Tang, W.W., Sheets, N., Ellison, J., Sette, A., Kim, K., Shresta, S., 2017. Identification of Zika virus epitopes reveals immunodominant and protective roles for dengue virus cross-reactive CD8+ T cells. Nat. Microbiol. 2, 17036.
Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699.
White, S.K., Lednicky, J.A., Okech, B.A., Morris, J.G., Dunford, J.C., 2018. Spondweni virus in field-caught *Culex quinquefasciatus* mosquitoes, Haiti, 2016. Emerg. Infect. Dis. 24, 1765–1767.
Wilder-Smith, A., Chang, C.R., Leong, W.Y., 2018. Zika in travellers 1947-2017: a systematic review. J. Travel Med. 25 (1). https://doi.org/10.1093/jtm/tay044.
Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V., Koonin, E.V., 2018. Origins and evolution of the global RNA virome. mBio 9 e02329-18.
Wu, J., Bera, A.K., Kuhn, R.J., Smith, J.L., 2005. Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing. J. Virol. 79, 10268–10277.
Zanotto, P.M., Gould, E.A., Gao, G.F., Harvey, P.H., Holmes, E.C., 1996. Population dynamics of flaviviruses revealed by molecular phylogenies. Proc. Natl. Acad. Sci. (USA) 93, 548–553.
Zhao, W., Sher, X., 2018. Systematically benchmarking peptide-MHC binding predictors: from synthetic to naturally processed epitopes. PLoS Comput. Biol. 14, e1006457.