

Exclusive and Common Subsets of Zika Virus Polyprotein Mutants

Joel K Weltman*

Clinical Professor Emeritus of Medicine, Alpert Medical School, Brown University, USA

Abstract

Two subsets of Zika virus (ZIKV) polyprotein amino acid positions are identified. One subset (Exclusive) consists of mutating amino acid positions which were found only in polyproteins isolated from ZIKV of human origin. A second subset (Common) consists of mutating amino acid positions which were found both in ZIKV polyproteins of human origin and in ZIKV polyproteins of *Aedes* species mosquito origin. The dominance of the Exclusive subset in the polyprotein was found to range from the N-terminus structural proteins until non-structural protein NS3. Although no longer greater than the Common subset, elements of the Exclusive subset existed almost to the C-terminus of non-structural protein NS5. These results are considered in the context of reported epitopic and other biological characteristics of ZIKV.

Keywords: Zika virus; ZIKV; Polyprotein; Information entropy H; Shannon entropy; Exclusive subset; Common subset

Introduction

Zika virus (ZIKV) causes microcephaly, brain abnormalities and other neural diseases in gestating infants who are infected in utero [1, 2]. Because of the highly adverse effects of ZIKV on gestating infants, it is important to develop preventive vaccines and therapeutic agents.

Reported here is a bioinformatics analysis of ZIKV polyprotein that sorts mutations which occurred exclusively in ZIKV isolated from infected human hosts (Exclusive subset) and mutations which occurred in ZIKV isolated both from infected humans and from infected *Aedes* mosquito vectors (Common subset). It is proposed that the exclusive mutations may reflect immunological and other biological processes which occur exclusively in human hosts but not in the mosquito vector. These exclusive mutational objects thus may help guide development of human host-specific targets for vaccines and drugs.

Materials and Methods

Complete sets of full-length ZIKV polyprotein sequences isolated from humans (n=212) and from *Aedes* mosquitos (n=34) were downloaded from the NCBI Zika Virus Resource (<http://www.ncbi.nlm.nih.gov/genome/viruses/variation/Zika>) on 8 April 2017. Polyprotein domains were assigned by alignment with the default MR 766 reference sequence [3]. Polyprotein sequence management was facilitated with Jalview 2.9.0b2 [4].

Computations were performed using Anaconda 2.4.0 (64-bit) with Python 2.7.10, Numpy 1.10.1, Scipy 0.16.0, Sympy 1.0 and Matplotlib 1.4.3. Information entropy (H) was computed by the equation of Shannon [5] and is expressed in bits. Amino acid positions with $H > 0.0$ in the polyprotein were classified and sorted into subsets depending whether the positive H value occurred at amino acid positions in ZIKV polyproteins obtained only from humans (Exclusive subset) or at amino acid positions in ZIKV polyproteins obtained both from humans and from *Aedes* species of mosquitos (Common subset).

The Mann-Whitney nonparametric U test was performed with Scipy stats; a two-tail p-value is reported. Z-tests were performed using 1000 pseudo-random trials and are reported with two-tail probabilities.

Results and Discussion

The basis for the definition of mutational subsets within the set of ZIKV amino acid positions is shown in Figure 1. As previously reported

[6], plots of mutations in ZIKV isolated from humans reveal the mutations which occurred exclusively in viruses isolated from humans (top) and in mutations common to viruses isolated both from humans and from *Aedes* mosquitos (bottom). This sorting of ZIKV mutations into Exclusive and Common subsets is consistent with immunological and metabolic processes which occur in humans but not in mosquitos and with structural processes common to both humans and mosquitos.

There were 219 polyprotein amino acid positions in the human Exclusive subset and 125 polyprotein amino acid positions in the common (human and *Aedes*). subset ($z=4.9475$, $p=7.5178 \times 10^{-07}$). The summed exclusive and common subset counts account for all 344 amino positions where $H > 0.0$ in the complete polyprotein dataset. The mean and standard deviation for H in the Exclusive subset are 0.0745 ± 0.0804 bits. The mean and standard deviation for H in the Common subset are 0.1433 ± 0.1650 bits. In addition to containing fewer subset elements, the average H in the Common subset is significantly greater than the average H in the Exclusive subset (Mann-Whitney $U=6525.5$, $p=7.9471 \times 10^{-17}$).

Distributions of information entropy (H) in ZIKV polyproteins isolated from humans are shown in Figure 2. The elements of the Common subset predominated in the C-terminal half of the polyprotein amino acid chain (Figure 2a), in non-structural regions of the polyprotein (Figure 2d). H summation curves (Figure 2b) were approximated (\rightarrow) by the following two indefinite integrals, where x represents the polyprotein amino acid position:

$$[1]. \quad \sum H(\text{exclusive}) \rightarrow \int H(\text{exclusive}) dx = x^4(2.3142 \times 10^{-13}) - x^3(1.2929 \times 10^{-09}) + x^2(1.4392 \times 10^{-06}) + x(5.5747 \times 10^{-03}) + 6.0842 \times 10^{-01}$$

$$[2]. \quad \sum H(\text{common}) \rightarrow \int H(\text{common}) dx = x^4(2.6759 \times 10^{-13}) - x^3(1.4469 \times 10^{-09}) + x^2(2.6463 \times 10^{-06}) + x(2.2839 \times 10^{-03}) + 3.8470 \times 10^{-01}$$

In equations 1 and 2, the discrete summation curves are used

*Corresponding authors: Joel K Weltman, Clinical Professor Emeritus of Medicine, Alpert Medical School, Brown University Providence, RI 02912, USA, Tel: 4012457588; E-mail: joel_weltman@brown.edu

Received May 02, 2017; Accepted June 01, 2017; Published June 07, 2017

Citation: Weltman JK (2017) Exclusive and Common Subsets of Zika Virus Polyprotein Mutants. J Med Microb Diagn 6: 256. doi:10.4172/2161-0703.1000256

Copyright: © 2017 Weltman JK. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

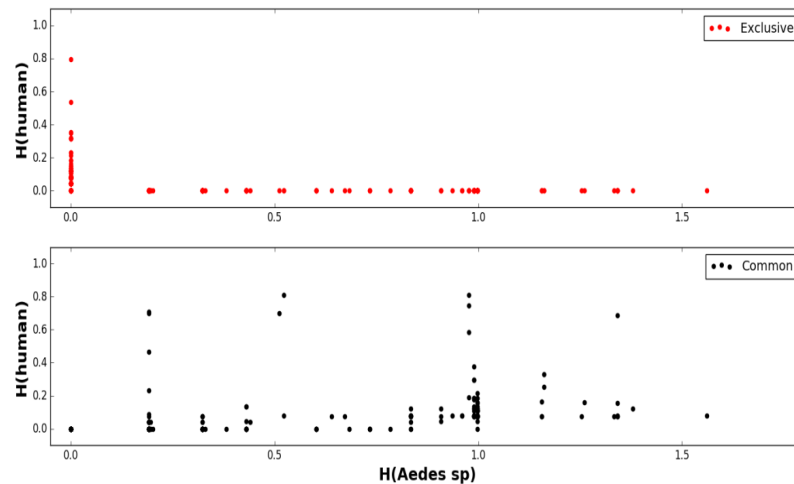


Figure 1: Definition of exclusive and of common subsets of amino acid mutants detected in ZIKAV polyproteins isolated from humans. (top) the positive H-value, objects in the Exclusive subset are represented vertically on the ordinate; (bottom) the positive H values in the Common subset are distributed on the graph surface area, indicating common occurrence of mutant amino acid positions in ZIKV polyproteins isolated from humans and from Aedes mosquitos.

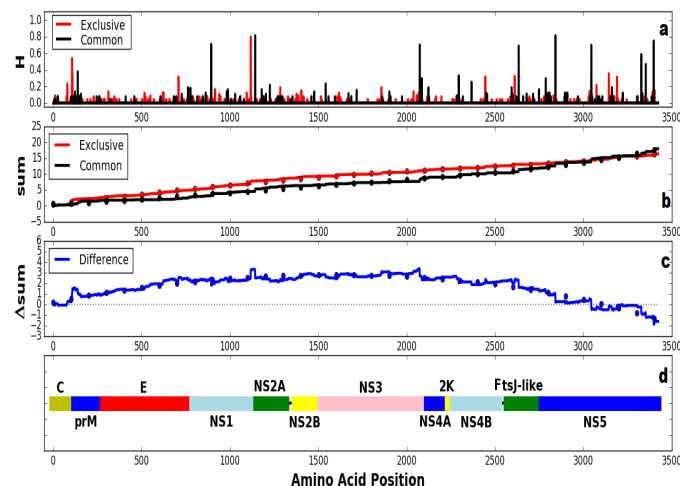


Figure 2: Distribution of information entropy (H) in ZIKV polyprotein subsets. (a) H distributions in exclusive (red) and in common (black) subsets of ZIKV polyprotein amino acid positions; (b) summation curves are shown for the exclusive subset of amino acid positions (red) and for the common subset of amino acid positions (black); (c) differences (blue) between the exclusive H subset summation and the common H subset summation; (d) a reference diagram of the sequential organization protein components of the polyprotein. In graphs (b) and (c), the solid lines denote observed values while closed circles denote values fit according to equation 1 (Exclusive subset) and equation 2 (Common subset). In figure (d), the ZIKV precursor structural protein components of the polyprotein are, with amino acid positions in parentheses: C (capsid protein, 6-122), prM (precursor membrane protein, 128-290) and E (envelope protein, 292-794); the non-structural precursor protein components are NS1 (797-1148), NS2A (1158-1307), NS2B (1377-1502), NS3 serine protease, helicase (1521-2124), NS4A (2125-2242), protein 2K (2243-2265), NS4B (2270-2512), FtsJ-like methyl transferase (2575-2746) and NS5 ribonuclease (2774-3412).

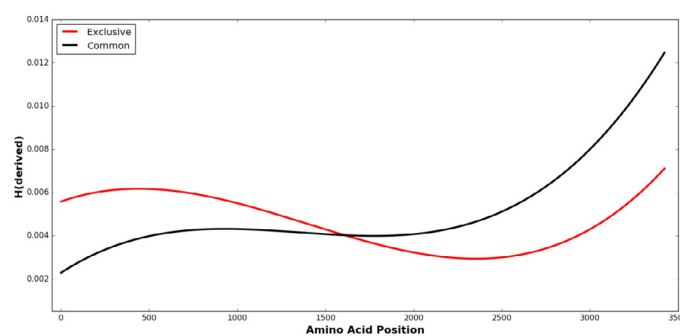


Figure 3: Predicted H distributions in the exclusive and common subsets of ZIKV polyprotein mutating amino acid positions. H values were predicted from first derivatives of equations 1 and 2 by treating those polynomials as indefinite integrals.

to generate approximations as continuous, indefinite integrals. Difference curves (Δ sum, Figure 2c) for the observed summation and approximated polynomials were obtained by subtraction: Δ sum = equation (1) - equation (2). The amino acid position with maximum value in the directly computed Δ sum curve is position 2068, which is within the NS3 non-structural protein domain of the polyprotein (Figure 2d). The amino acid position with maximum value in the Δ sum curve computed from the polynomial approximation is position 1609, which also is within the NS3 non-structural protein domain of the polyprotein. There were 69 amino acid position in the human Exclusive subset, from position 2068 to the C-terminal, with an H sum of 5.4103 bits, which is 33.17% of the Exclusive subset observed total.

Differentiation of the indefinite integrals in Equations 1 and 2 yield the following 2 equations, which were used to produce the H distribution curves in Figure 3:

$$[3]. H(\text{exclusive}) = x^3 (9.2569 \times 10^{-13}) - x^2 (3.8788 \times 10^{-9}) + x (2.8785 \times 10^{-6}) + 0.0056$$

$$[4]. H(\text{common}) = x^3 (1.0704 \times 10^{-12}) - x^2 (4.3407 \times 10^{-9}) + x (5.2926 \times 10^{-6}) + 0.0023$$

The polyprotein position at which $H(\text{exclusive})=H(\text{common})$ in the continuous model of equations 3 and 4 is between amino acid positions 1608 and 1609 which, as noted above, is within the NS3 domain.

The data in Figures 2 and 3 suggest that dominance of the human Exclusive subset of mutations extends beyond the structural protein components of the polyprotein. Precursors of the structural proteins comprise amino acids 1-794 of the polyprotein. The greater mutability of the elements of the human Exclusive subset, in comparison with the Common subset, probably represents functions of processes occurring in the human host but not in the mosquito vector. Such processes may include immunological escape [7], antibody dependent enhancement [8] and other biochemical and biophysical factors present in the host but not in the mosquito vector. For example, viral interactions with interferons are important determinants of outcome of ZIKV infection [9]. There are other, specific immunological human host factors that affect the Zika virus. Neutralizing antibodies against the prM precursor membrane protein (prM) and envelope E protein have been reported [10-12]. The prM and E proteins initially exist as precursors within the structural domain of the polyprotein. Antibody and T-cell responses against the NS1, NS2, NS3, NS4B and NS5 non-structural polyproteins have been reported [10,11,13,14].

The distribution of the elements of the human Exclusive subset is in agreement with the physiological and immunological Zika

virus-host interaction processes discussed above. The activity of the Exclusive subset is detectable well into the non-structural domain of the polyprotein. A more detailed elucidation of the mutational pattern in the human Exclusive subset may increase our understanding of Zika virus biology, thereby facilitating the development of anti-Zika strategies, preventives and therapeutics.

References

1. Reynolds MR, Jones AM, Petersen EE, Lee EH, Rice ME, et al. (2017) Vital signs: Update on Zika virus-associated birth defects and evaluation of all U.S. infants with congenital zika virus exposure-U.S. Zika Pregnancy Registry, 2016. MMWR Morb Mortal Wkly Rep. ePub.
2. Johansson MA, Mier-y-Teran-Romero L, Reefhuis J, Gilboa SM, Hills SL (2016) Zika and the risk of microcephaly. N Engl J Med 375: 1-4.
3. Källberg M, Wang H, Wang S, Peng J, Wang Z, et al. (2012) Template-based protein structure modeling using the RaptorX web server. Nat Protoc 7: 1511-1522.
4. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2: A multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191.
5. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27: 379-423.
6. Weltman JK (2016) sets and subsets of mutating amino acids in Zika Virus polyprotein. J Med Microb Diagn 5: 4.
7. Vossen MT, Westerhout EM, Söderberg-Nauclér C, Wiertz EJ (2002) Viral immune evasion: A masterpiece of evolution. Immunogenetics 54: 527-542
8. Dejnirattisai W, Supasa P, Wongwiwat W, Rouvinski A, Barba-Spaeth G, et al. (2016) Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. Nat Immunol 17: 1102-1108.
9. Winkler CW, Myers LM, Woods TA (2017) Adaptive immune responses to Zika virus are important for Controlling virus infection and preventing infection in brain and testes. J Immunol 198: 3526-3535.
10. Stettler K, Beltramello M, Espinosa DA, Graham V, Cassotta A, et al. (2016) Specificity, cross-reactivity and function of antibodies elicited by Zika virus infection. Science.
11. Xu X, Vaughan K, Weiskopf D, Grifoni A, Diamond MS, et al. (2016) Identifying candidate targets of immune responses in Zika virus based on homology to epitopes in other flavivirus species. PLoS Curr Edition 1.
12. Pardi N, Hogan MJ, Rebecca S, Pelc RS (2017) Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. Nature 543: 248-251.
13. Sironi M, Forni D, Clerici M (2016) Non-structural proteins are preferential positive selection targets in Zika virus and related flaviviruses. PLoS Negl Trop Dis 10: e0004978.
14. Dar H, Zaheer T, Rehman MT, Ali A, Javed A, et al. (2016) Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An *in silico* approach. Asian Pacific J Trop Med 9: 844-850.