bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ZIKA – How Fast Does This Virus Mutate ?

ABSTRACT

The World Health Organisation has declared the present epidemic of infection with the Zika virus to be a 'Public Health Emergency of International Concern'. The virus appears to have spread from Thailand to French Polynesia in 2013, and has since infected over a million people in the countries of South and Central America. In most cases the infection is mild and transient, but the virus does appear to be strongly neurotropic and cause both birth defects in foetuses and Guillain-Barré syndrome in some adults.

One of the results of the rapidly increasing importance of the Zika virus is that a significant amount of RNA sequence data of the viral genome has appeared in the public domain and is available for scrutiny by scientists and physicians who might not normally consider studying the genetics of a virus. In this paper the techniques and utilities developed in the study of mitochondrial DNA are applied to the Zika virus.

As a result it is possible to show in a simple manner how a phylogenetic tree may be constructed and how the mutation rate of the virus can be measured. The study shows the mutation rate to be about 10 bases a year, in a viral genome of 10,272 bases. This rapid mutation rate will enable the geographic spread of the epidemic to be monitored easily and may also prove useful in assisting the identification of preventative measures that are working, and those which are not.

Whether any of the mutations seen in the present epidemic affect the virulence and behaviour of the virus is uncertain. But there is no clear evidence to show that changes in the viral genome are affecting the pattern of the epidemic. However, it is possible that the foetal damage and Guillain-Barré may be the consequences of the initial viraemia that follows a bite from a mosquito carrying the Zika virus.

Ian Logan M.B., Ch.B., B.A

Exmouth, UK
Email: ianlogan22@btinternet.com

INTRODUCTION

Zika is a RNA virus and was first identified in Uganda in 1947.[1] Over the next 60 years it was the cause of epidemics in several African countries. However, in about 2010 the virus spread to parts of Asia, in particular to Thailand,[2,3] and by 2013 had reached French Polynesia.[4] Since then there has been an explosive epidemic affecting the populations of many countries in both South and Central America. At the present time this epidemic shows no signs of abating.

In the many small epidemics there was no indication of the virus causing anything but mild and transient infections. But in the recent epidemic in Polynesia cases of central nervous system damage were observed and described as being a form of Guillain-Barré disease.[5,6] However, in the current Brazilian epidemic the emphasis has been on the association with birth defects, especially microcephaly resulting from maternal infection with Zika in the first and second trimester of pregnancy.[7] Further cases of Guillain-Barré syndrome have also been seen.

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

On the 1st February 2016 the World Health Organisation declared the emerging Zika epidemic to be a 'Public Health Emergency of International Concern' which shows how this epidemic is now considered to be a major threat to the whole world.[8] Their statement of intent includes the lines:

Appropriate research and development efforts should be intensified for Zika virus vaccines, therapeutics and diagnostics.
National authorities should ensure the rapid and timely reporting and sharing of information of public health importance relevant to this PHEIC.

As a result it can be expected that many research institutions will increase their studies into Zika and related viruses and many new scientific papers will appear in the coming months. At the same time it is to be expected that many more RNA sequences of the virus will appear in the public domain.

VIROLOGY

The Zika virus is a flavivirus carried by mosquitoes and has a host of a monkey or a human. The virus is closely related to the viruses of Yellow Fever, Dengue and West Nile Fever, all of which do cause significant illness and mortality. However there are many other flaviviruses that are less well known and their hosts include horses, sheep, bats, and many other animals. A paper in 1997 listed over 70 viruses in the group and new flaviviruses continue to be identified.[9]
All flaviviruses appear to have much the same structure as each other. The mature virus particles, virions, are about 50 nm in diameter and are icosahedral in shape. Modern electron microscopy can now show the virion in considerable detail.[10,11] The outer part is formed by an *envelope* overlying a phospholipid bi-layer *membrane* and the core contains a single stranded RNA molecule of about 10k bases.

In the mature Zika virion the RNA molecule is termed a *polyprotein* and described as having 10,272 bases, or 3,424 3-base codons for specific amino acids. The translation of bases to functional codons is not perfect, but for analysis purposes it has become accepted to describe the structure of the molecule in this manner:
starting with MKN …. and ending with ….GVL (ie. The codons for: methionine, lysine, asparagine ……. glycine, valine, leucine)

The *polyprotein* is a linear assembly of both *structural* and *non-structural* genes. The structural genes are for the *envelope, membrane and capsid*, and the non-structural genes are usually considered as being NS1, NS2A, NS2B, NS4A, NS4B, and NS5; and for the purposes of this paper this simple explanation will suffice.

The *envelope* and *membrane* genes define how the outer part of the virion is conformed. This outer part is important as it acts as an antigen for antigen-antibody reactions and also in the interaction between the virus and *entry receptors* as a virion attempts to enter a cell. In consequence, mutations affecting the construction of the *envelope* and *membrane* are probably more significant than mutations in other parts of the genome, and perhaps are of greater influence when it comes to possible changes in virulence.

It remains unclear as to which cells, if not all cells, in the human are susceptible to invasion with Zika., But there does appear to be a preference for cells of the haemopoetic and nervous systems. In all instances the process appears to be the same in that a virion attaches itself to the *entry receptors* on the outer surface of the cell and the virion enters the cell in a process described as endocytosis.[12]

Once in a human cell the *envelope* and *membrane* separate from the core. The virus then hi-jacks the cellular apparatus for its own purposes. The *polyprotein* is copied and cleaved into its constituent parts and the daughter virions produced, each containing its own copy of the *polyprotein*.[13]

At present there are no drugs that prevent replication of the Zika virus, and the older and well-established antiviral drugs, such as amantadine which are active against the influenza virus, are not helpful against the flaviviruses.[14] However a lot of work is being done to find new antiviral drugs.[15]

In relation to the present epidemic the ability of the Zika virus to enter cells of the placenta and the central nervous system is particularly important. However for now it is unclear as to whether or not these invasions are more dependent on the strain of the virus, the genetic make-up of the host, or other factors. But once the virus has crossed the *placental barrier* to enter the foetus or the *blood-brain barrier* to enter the central nervous system it is likely that the usual antigen-antibody reactions

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

are lessened and the virus is able to proliferate more easily. It is also unknown as to how long it might take for the virus to be cleared from the foetus or the central nervous system. Although it does seem likely that virus replication can continue in these areas for many months.

METHODS

The GenBank database

The RNA sequences for the Zika virus in the public domain can be found in the GenBank database of the National Institute of Health.[16]

At present (February 2016) there are 16 complete sequences from virus collected in Africa and Asia before the start of the present epidemic and 14 sequences produced since.

Details of these sequences are given in Table 1.

The corresponding page on the GenBank database for a given sequence can be found by using a URL of the form:

http://www.ncbi.nlm.nih.gov/nuccore/KJ776791

Each page give the amino acid list and the nucleotide base FASTA file for the RNA sequence. However, a GenBank page contains no real explanation as to what each list or file might mean and for this reason the author has developed a pair of *Zika virus utilities* that allows the user to compare one sequence with another.

Table 1: Zika RNA sequences in GenBank database - February 2016

Accession Number	Country of origin	Date of Collection
African sequences:		
1. LC002520	- Uganda	- 1947
2. HQ234499	- Malaysia	- 1966
3. HQ234500	- Nigeria	- 1968
4. KF383116	- Senegal	- 1968
5. KF383115	- CAR	- 1968
6. HQ234501	- Senegal	- 1984
7. KF268948	- CAR	- 1976
8. KF268949	- CAR	- 1980
9. KF268950	- CAR	- 1980
10. KF383117	- Senegal	- 1997
11. KF383118	- Senegal	- 2001
12. KF383119	- Senegal	- 2001
Asian sequences:		
13. EU545988	- Micronesia	- 2007
14. JN860885	- Cambodia	- 2010
15. KU681082	- Philippines	- 2012
16. KU681081	- Thailand	- 2014
Present epidemic:		
17. KJ776791	- Polynesia	- 2013
18. KU365779	- Brazil	- 2015
19. KU365778	- Brazil	- 2015
20. KU365777	- Brazil	- 2015
21. KU365780	- Brazil	- 2015
22. KU312312	- Suriname	- 2015
23. KU501215	- Puerto Rico	- 2015
24. KU509998	- Haiti	- 2014
25. KU321639	- Brazil	- 2015
26. KU527068	- Brazil	- 2015
27. KU647676	- Martinique	- 2015
28. KU501216	- Guatemala	- 2015
29. KU501217	- Guatemala	- 2015
30. KU707826	- Brazil	- 2015

The Zika virus utilities

In conjunction with this paper 2 simple utilities have been prepared and are to be found the author's website. These utilities are in the form of 2 webpages.

The URL for the pages is:
www.ianlogan.co.uk/zikapages/zika.htm
From where the user can choose to use:
either the **Amino Acid Analyser**
or the **Nucleotide Base Analyser**

The **Amino Acid Analyser** has in its source file copies of the amino acid lists for all the complete RNA sequences to be found in the GenBank database and a small Javascript program allows the user to compare any sequence against any other. The results are displayed as a list of amino acid changes.

The **Nucleotide Base Analyser** has in its source file copies of the FASTA files for the complete RNA sequences for the base sequence from Thailand, together with the files from all the sequences for the current epidemic. Again, a small Javascript program enables the user to compare two sequences and find the mutational differences between them.

Although the webpages can be viewed with any of the commonly used web browsers, the author recommends MOZILLA FIREFOX as this browser allows the user to alter the size of the text area, if needed.

It is the author's intention to keep these webpages up-to-date as new Zika RNA sequences appear on the GenBank database.

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

The results from this study can be considered under 3 headings:

1. Non-synonymous amino acid changes observed in the present epidemic

A mutation that causes a non-synonymous change of an amino acid is often considered to be significant. But if the change is between amino acids of similar size and of a similar polarity, there is probably no effective change in the functioning of the target protein.

The amino acid changes shown by the Zika RNA sequences in the present epidemic are listed in Table 2. The table demonstrates that by using this method the sequences can be split into 9 different strains with variously 0 to 6 amino acid changes.

The mutation M2634V is common to all the virus strains that come from countries in South and Central America and is caused by the base mutation A7900G. However, as this mutation is found in the NS4B gene it is unlikely to be of significance as to the virulence or general behaviour of the Zika virus. It is perhaps too early to say that this mutation has absolutely no effect, but for the moment the M2634V mutation can be seen to be a useful marker to the present epidemic.

2. Base mutations in samples collected in the present epidemic

Whereas there are relatively few non-synonymous mutations in the virus strains collected in the present epidemic, there are many more synonymous mutations (i.e. mutations that do not produce a change of amino acid). This means that a meaningful phylogenetic tree can be constructed.

Table 2: Non-synonymous amino acid changes found in sequences from the present epidemic

KJ776791-Polynesia-2013	
KU365778-Brazil-2015 }	M2634V
KU365779-Brazil-2015 }	
KU707826-Brazil-2015 }	
KU365777-Brazil-2015 }	M2634V N2778D
KU365780-Brazil-2015 }	
KU312312-Suriname-2015	M166T T769A M2634V
KU501215-Puerto Rico-2015	I80T A2611V M2634V
KU509998-Haiti-2014 }	Y916H H1857Y I2295M I2445M M2634V
KU321639-Brazil-2015 }	
KU527068-Brazil-2015	K940E T1027A M1143V T2509I M2634V
KU647676-Martinique-2015	D107E R1118W I1226T M2634V T3353A
KU501216-Guatemala-2015 }	V346I G894A M2074L M2634V K2694R R3045C
KU501217-Guatemala-2015 }	

Note: The change M2634V ((Methionine > Valine) occurs in the NS4B gene and is the result of the mutation A7900G which changes the codon from 'ATG' to 'GTG'.

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1 shows the phylogenetic tree produced by using the mutations from the 14 complete Zika sequences currently to be found in the GenBank database. This figure shows the virus samples can now be separated into 12 different strains.

Note that in 3 of the samples there appear to be a *missing* mutation:

C10074T in sequence KU365780-Brazil-2015
T5049C in sequence KU312312-Suriname-2015
G9327A in sequence KU321639-Brazil-2015

and it would seem likely that these omissions result from technical errors.

3. **Making an estimate of the mutation rate for the Zika virus**

The data presented in Figure 1 shows that about 30 mutations (actual range 24-33) have occurred in the strains of Zika virus collected in the course of the present epidemic; and if the duration of the epidemic is take to be about 3 years, the mutation rate is given as about 10 mutations a year.

As the genome of the Zika virus is normally considered to be a *polyprotein* of 10,272 nucleotide bases, the mutation rate can also be considered as being about 0.01% of the RNA mutating each year.

It is not possible, using the data presently available, to give a more accurate value for the mutation rate. But the suggested rate of 10 mutations a year would appear to be a suitable starting point for further studies.

DISCUSSION

The Present Epidemic

The decision by the World Health Organisation to declare a *Public Health Emergency* in February 2016 because of the threat of a pandemic from Zika virus may be thought to have been a pessimistic move. However the evidence appears to indicate that the Zika virus is no longer restricted to localised habitats and largely dependent on the monkey as its host, but now covers a much larger area which includes the countries of South and Central America where it is wholly dependent on the human as its host. Unfortunately, there is no herd immunity against the Zika virus in the populations of these countries, despite the closely related Dengue virus being very prevalent.

The sudden spread of Zika to South and Central America does not appear to have been due any particular change in the mosquito vector or anything to make the Zika virus itself more virulent. But rather from the fact that infected people are now able to fly rapidly from country to country, thereby spreading the disease extremely easily. This means there is little to stop the epidemic continuing to spread to further areas of population which have little, if any, herd immunity against the virus.

The absence of mosquitoes and the low incidence of person-to-person spread of the virus will probably mean the epidemic will not spread in the countries of the Southern and Northern latitudes. But from the evidence obtained so far it would appear likely the epidemic is only at its earliest stage and any suggestions as to what might happen remain speculative.

The Zika phylogentic tree

The 2 utilities prepared for this study show that many distinct strains of the Zika virus now exist, even though the present epidemic is only about 3 years old. When considering just the non-synonymous mutations in the RNA it is now possible to define 9 strains in the present epidemic. However, a more detailed examination looking at the actual mutations of the available sequences distinguishes 12 strains. As more data is made available it is to be expected the number of identifiable strains will increase. The phylogenetic tree as shown in Figure 1 suggests that there is beginning to be a geographical spread of the associated virus strains, with distinct strains now coming from Martinique, Guatemala, Puerto Rico and Suriname, whilst Brazil continues to show a mixture of strains.

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

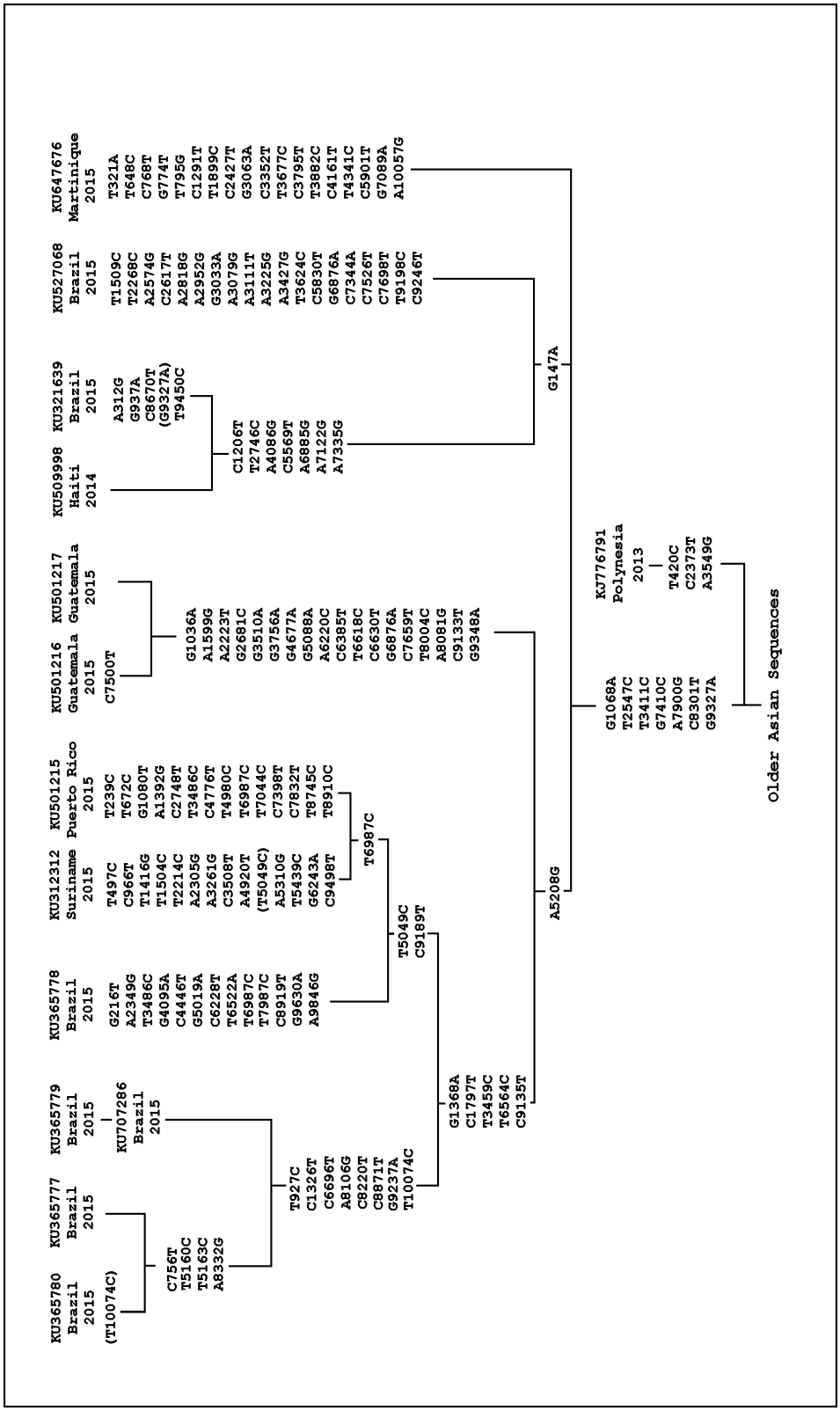

Figure 1

The phylogenetic tree of the 14 Zika RNA sequences from samples collected in the present epidemic.

The samples can now be described as forming 12 different strains.

Note: the 3 missing mutations – indicated by the brackets:

C10074T in sequence KU365780-Brazil-2015, T5049C in sequence KU312312-Suriname-2015, G9327A in sequence KU321639-Brazil-2015

are probably caused by technical errors.

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

An Estimate of the Zika Mutation Rate

The data used in building the phylogenetic tree can also be used to estimate the mutation rate of the Zika virus; and this would appear to be about 10 mutations per year, which is equivalent to about 0.01% of the RNA mutating each year. This rate is very high when compared to the Human DNA mutation rate, where a period of perhaps 250,000 years might be expected.[17] But really it is not appropriate to compare RNA mutations against DNA mutations as DNA replication is a self-correcting process, whereas RNA duplication is liable to many sorts of error.

However, the clear evidence of a high mutation rate in the Zika virus will allow for the present epidemic to be tracked in a fairly simple manner; and also it should be helpful in seeing where a local initiative of mosquito prevention is working and where it is not.

The relationship between mutations and virulence is a very interesting question. But it is probably unlikely that the Zika virus can mutate significantly enough to become as virulent as the Yellow Fever or Dengue viruses.

The Complications of Zika infection

A particular feature of the present epidemic related to the high incidence of foetal abnormalities and cases of Guillain-Barré syndrome. These two complications appear to be very different with the foetal abnormalities being caused by the direct infection of the foetus, whereas in cases of Guillain-Barré syndrome being an exaggerated auto-immune response.

However, in the author's opinion both of these conditions may result from the same underlying cause in which virus gets across the normally impenetrable *placental barrier* and the *blood-brain barrier*. How this happens is unclear, but, perhaps this is just a matter of a person getting a very high initial infection, possibly by having bites from physically large carrier mosquitoes, or maybe bites from several carrier mosquitoes in a very short period of time.

A study using the West Nile virus[18] showed that whilst most of the inoculum from a mosquito bite remains localised in the skin, there is always a significant initial viraemia. In this respect the recent report from Slovenia [7] shows the x-rays of an affected foetus having numerous calcifications in the placenta and brain. Whilst it is unproven it would seem possible that these lesions result from localised 'viral plaque' formation associated with an initial viraemic spread. A similar picture is seen in tuberculosis. Although this disease is caused by a bacterium and not a virus, the resulting x-ray picture of localised calcifications is well recognised and is termed miliary tuberculosis.[19,20]

It is also possible that the risk of developing complications of Zika may reflect the genetic differences between the sufferers and the general population. But at the present stage of our knowledge there is no indication of what particular differences might be important.

Conclusion

This study shows in a simple way how sequencing data from samples of the Zika virus available in the public domain can be collected and analysed. Using this data it is possible to construct a phylogenetic tree and show that in the present epidemic there are already 12 strains of the virus. The data can also be used to show that the Zika virus has a high mutation rate.

This short paper raises as many questions as it tries to answer. The present epidemic is from the Zika virus, but Yellow fever cases are rising in Africa and Dengue affects millions of people each year. So further pandemics from other flaviviruses remain a continuing threat.

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Haddow AJ, Williams MC, Woodall JP, Simpson DI, Goma LK.
TWELVE ISOLATIONS OF ZIKA VIRUS FROM AEDES (STEGOMYIA) AFRICANUS (THEOBALD) TAKEN IN AND ABOVE A UGANDA FOREST.
Bull World Health Organ. 1964;31:57-69.

2. Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, Guzman H, Tesh RB, Weaver SC.
Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage.
PLoS Negl Trop Dis. 2012;6(2):e1477.

3. Fonseca K, Meatherall B, Zarra D, Drebot M, MacDonald J, Pabbaraju K, Wong S, Webster P, Lindsay R, Tellier R.
First case of Zika virus infection in a returning Canadian traveler.
Am J Trop Med Hyg. 2014 Nov;91(5):1035-8.

4. Baronti C, Piorkowski G, Charrel RN, Boubis L, Leparc-Goffart I, de Lamballerie X.
Complete Coding Sequence of Zika Virus from a French Polynesia Outbreak in 2013.
Genome Announc. 2014 Jun 5;2(3).

5. Winer JB.
An Update in Guillain-Barré syndrome.
Autoimmune Dis. 2014; 2014:793024

6. Christian Korff Guillain-Barré syndrome
http://raft.g2hp.net/wp-content/blogs.dir/9/files/2014/04/Guillain-Barré-syndrome.pdf
Hôpitaux Universitaires de Genève 2013
A presentation that shows parts of the original paper:
Guillain G, Barré J, Strohl A.
Sur un syndrome de radiculo-nevrite avec hyperalbuminose du liquide cephalorachidien sans reaction cellulaire. Remarques sur les characteres clinique et graphique des reflexes tendinaux.
Bulletins et Mémoires de la Societe Medicale des Hopitaux de Paris. 1916;40:1462–1470.

7. Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, Kolenc M, Resman Rus K, Vesnaver Vipotnik T, Fabjan Vodušek V, Vizjak A, Pižem J, Petrovec M, Avšič Županc T.
Zika Virus Associated with Microcephaly.
N Engl J Med. 2016 Feb 10.

8. WHO statement on the first meeting of the International Health Regulations (2005) (IHR 2005) Emergency Committee on Zika virus and observed increase in neurological disorders and neonatal malformations
WHO statement 1 February 2016

9. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB.
Phylogeny of the genus Flavivirus
J Virol. 1998 Jan;72(1):73-83.
.

10. Zhang X, Ge P, Yu X, Brannan JM, Bi G, Zhang Q, Schein S, Zhou ZH.
Cryo-EM structure of the mature dengue virus at 3.5-Å resolution.
Nat Struct Mol Biol. 2013 Jan;20(1):105-10

11. Z Hong Zhou
Structures of viral membrane proteins by high-resolution cryoEM
Curr Opin Virol. 2014 April ; 5: 111–119
(of particular relevance is Fig. 3 on page 14)

bioRxiv preprint doi: https://doi.org/10.1101/040303; this version posted February 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12. Manuel Perera-Lecoin, Laurent Meertens, Xavier Carnec, Ali Amara
Flavivirus Entry Receptors: An Update
Viruses 2014, 6, 69-88;

13. Clyde K, Harris E.
RNA secondary structure in the coding region of dengue virus type 2 directs translation start codon selection and is required for viral replication.
J Virol. 2006;80(5):2170–2182.

14. Oxford JS, Logan IS, Potter CW
In vivo selection of an influenza A2 strain resistant to amantadine
Nature 1970 Apr 4;226(5240):82-3.
.

15. Sampath A, Padmanabhan R.
Molecular targets for flavivirus drug discovery.
Antiviral Res. 2009 Jan;81(1):6-15

16. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.
GenBank
Nucleic Acids Res. 2013 Jan;41

17. Logan I
Calculating the Human Mutation Rate by Using a NUMT from the Early Oligocene
http://www.biorxiv.org/content/early/2015/03/13/016428

18. Styer LM, Kent KA, Albright RG, Bennett CJ, Kramer LD, Bernard KA.
Mosquitoes inoculate high doses of West Nile virus as they probe and feed on live hosts.
PLoS Pathog. 2007 Sep 14;3(9):1262-70.

19. Yang M, Zhang JT, Yao Y, Tan QC, Gao T, Tian CL, Huang X, Yu SY.
A clinical study of miliary brain tuberculomas in China.
Jpn J Infect Dis. 2015 Aug 7.
(selected for the x-rays)

20. Khan NA, Akhtar J, Baneen U, Shameem M, Ahmed Z, Bhargava R.
Recurrent pneumothorax: A rare complication of miliary tuberculosis.
N Am J Med Sci. 2011 Sep;3(9):428-30.
(selected for the x-rays)