

# Are Transition Mutations That Make CpG Sites Costly to Viruses?



Victoria Caudill, Jasmeen Kaur, Esteban Geo Pineda, Sarina Qin, Ryan Winstead, Pleuni Pennings

## Abstract

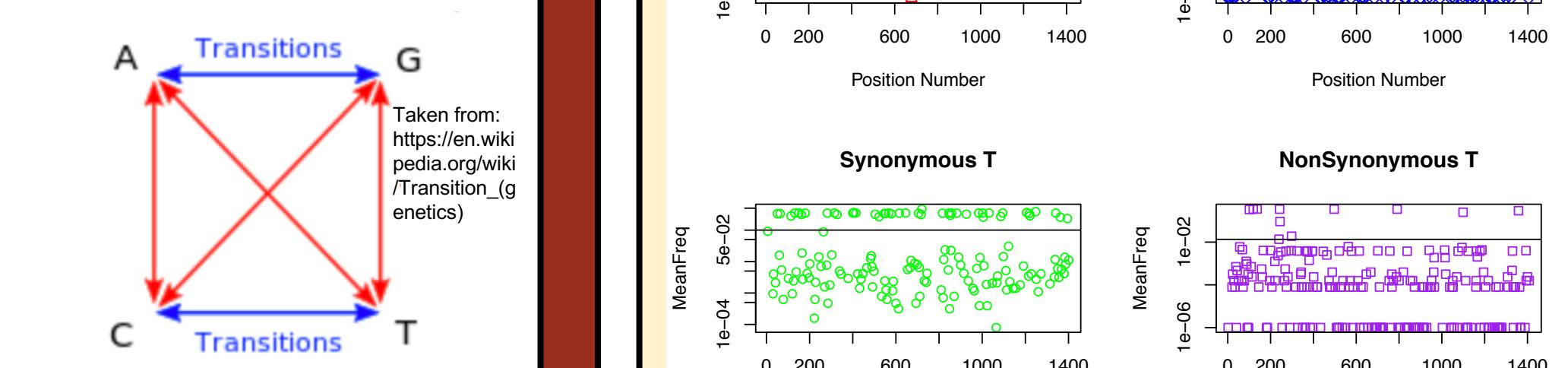
Transition mutations are point mutations that change nucleotide bases Adenine (A) or Thymine (T) to a Cytosine (C) or Guanine (G) or vice versa. When an A or T becomes a C or G, CpG sites can be formed. CpG sites are known to be costly to HIV, but unknown for other viruses. CpG sites are dangerous to the virus because they are more easily recognized as potential pathogens by the host's immune system (Hoelzer et al., Nucleic acid research, 2008). The virus data we used are sequences obtained from GenBank. Through the use of R we graphed and analyzed the virus data. We found that that CpG sites were costly for some viruses, like in Dengue strain 1. This indicates that viruses would want to avoid mutations that create a CpG site. By knowing for which viruses CpG sites are costly, we can further understand host virus interactions.

## CpG and Virus Background

CpG sites are areas of DNA, from 5' to 3', where a Cytosine nucleotide is followed by a Guanine nucleotide, separated by a phosphate group. A synonymous (syn) mutation is a mutation that changes a nucleotide in the DNA, but does not result in an amino acid change, keeping the protein the same. Whereas, a nonsynonymous (nonsyn) mutation is one that alters the amino acid sequence of the protein. CpG sites are formed when there is a transition mutation.

We examined Dengue, Enterovirus, Influenza, and BK.

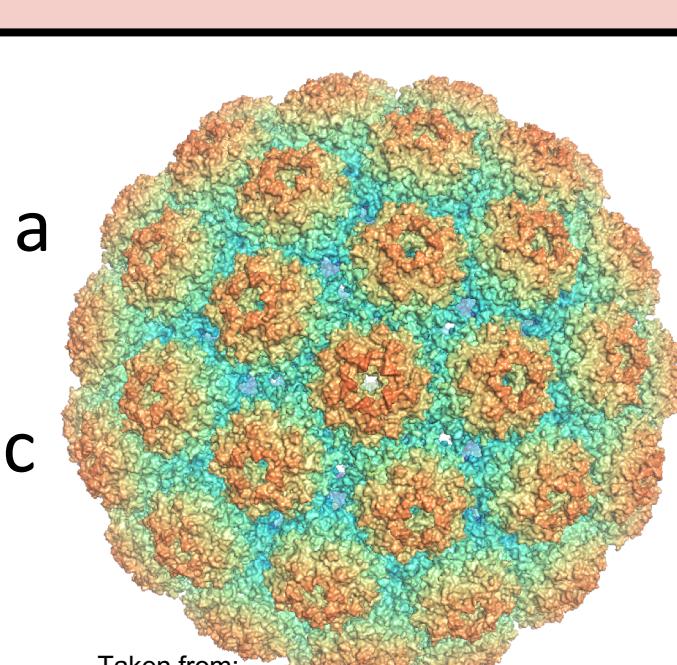
The Dengue virus is transmitted by mosquitoes in tropical areas and has five strains. Enterovirus is a single stranded RNA virus that is transmitted through the intestines. Influenza A, H1N1, is a strain of the influenza virus that is most common in humans. It is the strain that caused the Swine Flu outbreak in 2009 and the Spanish Flu in 1918. The BK virus rarely causes disease and is typically associated with kidney transplant patients.



Taken from: [https://en.wikipedia.org/wiki/Transition\\_\(genetics\)](https://en.wikipedia.org/wiki/Transition_(genetics))

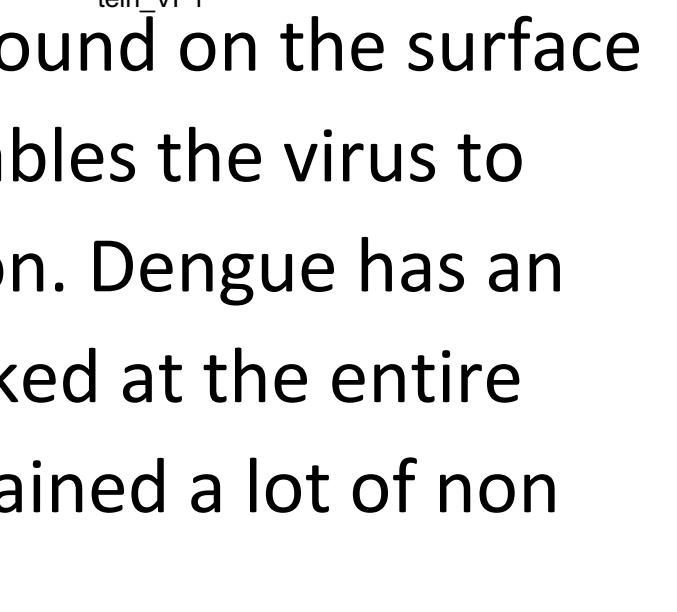
## Protein Background

The VP1 protein is found in Enterovirus and BK, it is a capsid protein needed to begin a viral infection. In order to begin an infection, VP1 surface binding sites interact with specific binding sites on the surfaces of cells.



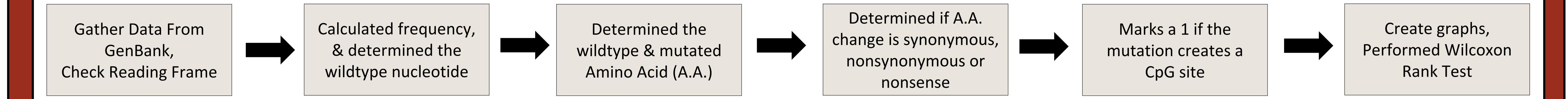
Taken from: [https://en.wikipedia.org/wiki/Major\\_capsid\\_protein\\_VP1](https://en.wikipedia.org/wiki/Major_capsid_protein_VP1)

Neuraminidase is a type of protein that is found on the surface of Influenza viruses that enables the virus to replicate and spread infection. Dengue has an open reading frame. We looked at the entire coding sequence which contained a lot of non structural proteins

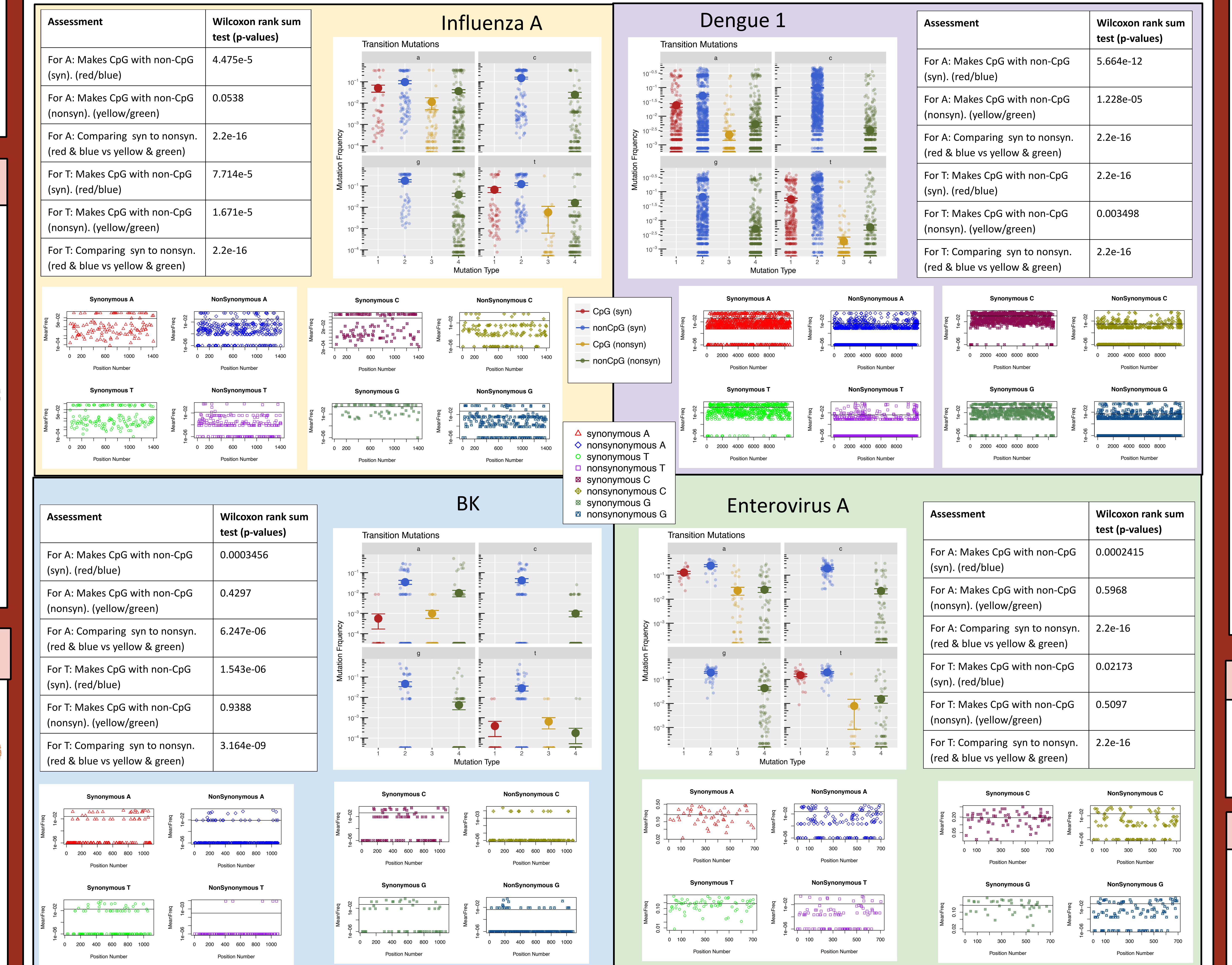


Taken from: <http://pdb101.rcsb.org/motm/113>

## Methods



## Results



## Discussion

### Influenza :

In Influenza, we found that synonymous CpG forming mutations occur less often than synonymous non-CpG forming mutations. The non-synonymous mutations also showed significant differences between CpG forming and non-CpG forming mutations. These results indicate that CpG mutations are costly for the virus.

### Dengue:

Dengue showed similar results to Influenza where all CpG forming mutations were less frequent than non-CpG forming mutations. This data solidifies our hypothesis that CpG forming mutations are costly to the virus.

### BK:

In BK, there was a significant difference between synonymous CpG and non-CpG forming mutations. However, there was no significant difference between nonsynonymous CpG and non-CpG forming mutations. Once again, these results coincide with our hypothesis.

### Enterovirus:

The Enterovirus also showed significance between synonymous CpG and non-CpG forming mutations, but no difference between CpG and non-CpG nonsynonymous mutations. In synonymous mutations CpG sites are costly.

## Future Directions

More research is necessary to fully claim that CpG mutations are costly to all viruses. We would like to further our research by examining more viruses and different proteins within those viruses. With more data, we will be able to prove that CpG sites are indeed more costly to all viruses.

## References

- Bäck, A.T., and Lundkvist, A. (2013). Dengue viruses – an overview. *Infect. Ecol. Epidemiol.* 3.
- Genoni, A., Canducci, F., Rossi, A., Broccolo, F., Chumakov, K., Bono, G., Salerno-Urante, J., Salvatoni, A., Pugliese, A., and Tonoli, A. (2017). Revealing enterovirus infection in chronic human disorders: An integrated diagnostic approach. *Sci. Rep.* 7: 5013.
- Hoelzer, K., Shackelton, L.A., and Parrish, C.R. (2008). Presence and role of cytosine methylation in DNA viruses of animals. *Nucleic Acids Res.* 36: 2825–2837.
- Rani, A., Ranjan, R., McGee, H.S., Metwally, A., Hajjaji, Z., Brennan, D.C., Finn, P.W., and Perkins, D.L. (2016). A diverse virome in kidney transplant patients contains multiple viral subtypes with distinct polymorphisms. *Sci. Rep.* 6: 3327.
- Shtyra, Y.A., Mochalova, L.V., and Bovin, N.V. (2009). Influenza Virus Neuraminidase: Structure and Function. *Acta Naturae* 1, 26–32.

## Acknowledgements

Special thanks to fall 2017 bioinformatics (biol638/738), and Dr. Pleuni Pennings, SFSU Biology Department, Computer Science Department and the College of Science and Engineering (COSE). Funded By NSF