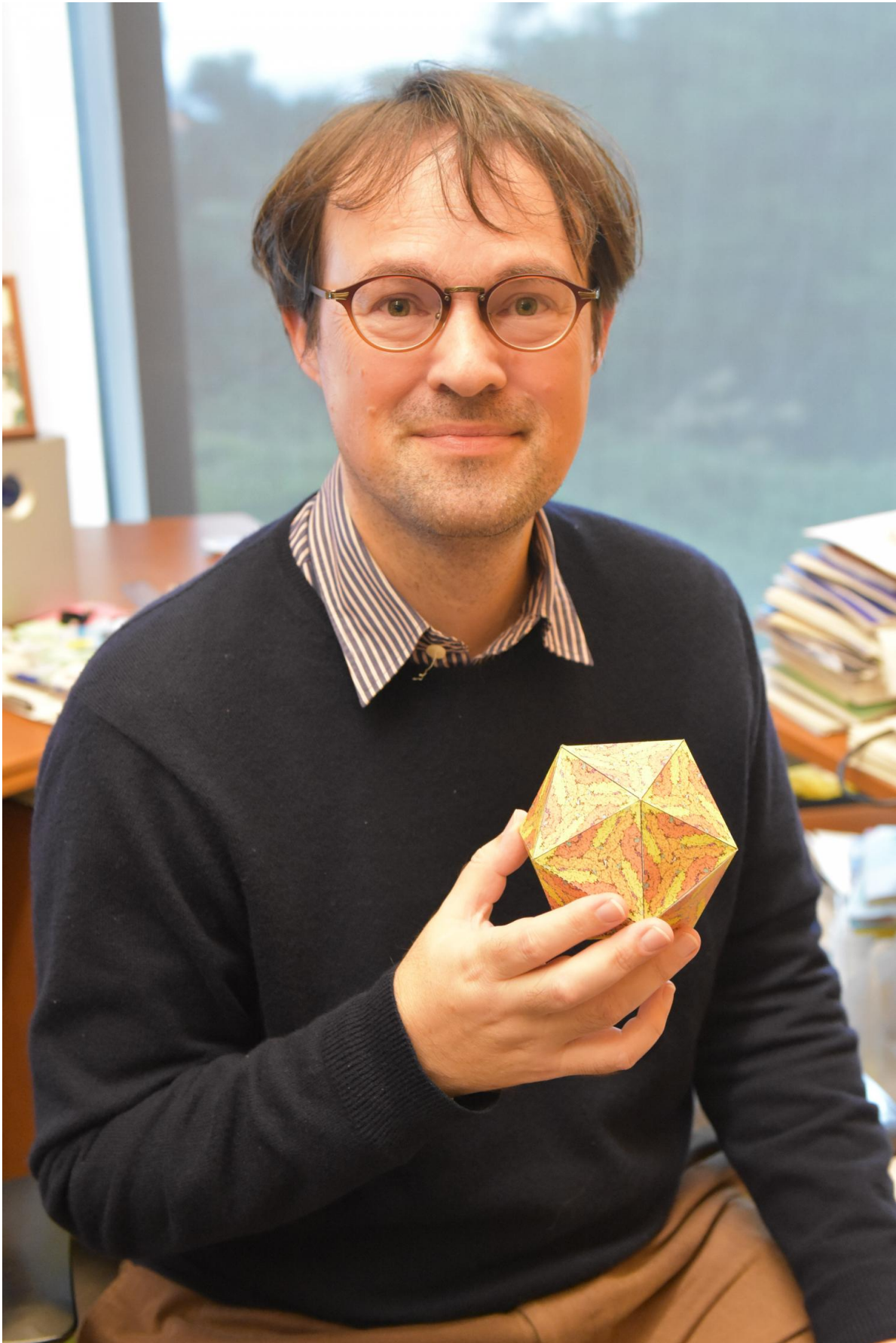# Mathematics supports a new way to classify viruses based on structure

February 23 2017, by Greta Keenan

Professor Robert Sinclair, head of OIST's Mathematical Biology Unit, holds a model of a Zika virus. Credit: Greta Keenan

Professor Robert Sinclair at the Okinawa Institute of Science and Technology Graduate University (OIST) and Professor Dennis Bamford and Dr. Janne Ravantti from the University of Helsinki have found new evidence to support a classification system for viruses based on viral structure.

The team developed a new highly-sensitive computational prototype tool, and used it to detect similarities in the genetic code of viruses with similar outer structures, that conventional tools have failed to detect, suggesting that they share a common ancestor. This is not what would have been expected if similarities in the structure of viruses were due to similar environmental pressures - a phenomenon known as convergence.

The results, published in the *Journal of Virology*, suggest that viral structure could provide a means of categorizing viruses with their close relatives - a potentially superior approach to current classification systems. Application of this new structure-based classification system could make it easier to identify and treat newly emerging viruses that cannot easily be classified with existing classification systems.

Viruses are notoriously difficult to classify due to their enormous diversity, high rates of change and tendency to exchange genetic material. They challenge the very concept of a clear distinction between the living and the dead, with many characteristics resembling those of living things, but lacking the ability to reproduce themselves, without the help of a host cell. As such, they do not fit neatly into the established

biological classification system for cellular organisms.

Existing classification systems are imperfect and often lead to very similar viruses being categorized as entirely different entities. These systems are also unable to account for the fact that viruses are constantly changing.

If scientists could identify something that viruses are unable to change, it could provide a basis for a more meaningful approach to classification and enable the scientific community to tackle emerging viruses, such as HIV, SARS coronavirus and Zika virus, more easily.

Previously observed similarities between the protein shell, or 'capsid', of viruses - that encloses and protects the genetic material - provide a basis for a classification system based on capsid structure, as previously proposed by Prof. Bamford. The few ways in which viruses package themselves are very similar, even between viruses that are likely to have had their common relative more than a billion years ago. Whether this conservation is due to convergence or common descent has been disputed.

For a classification system based on virus capsid structure to be meaningful, the amino acids that provide the building blocks of the capsid proteins should be similar in related viruses. A seeming lack of sufficient amino acid sequence similarity picked up by conventional sequence analysis tools previously undermined capsid structure as a viable way to classify viruses.
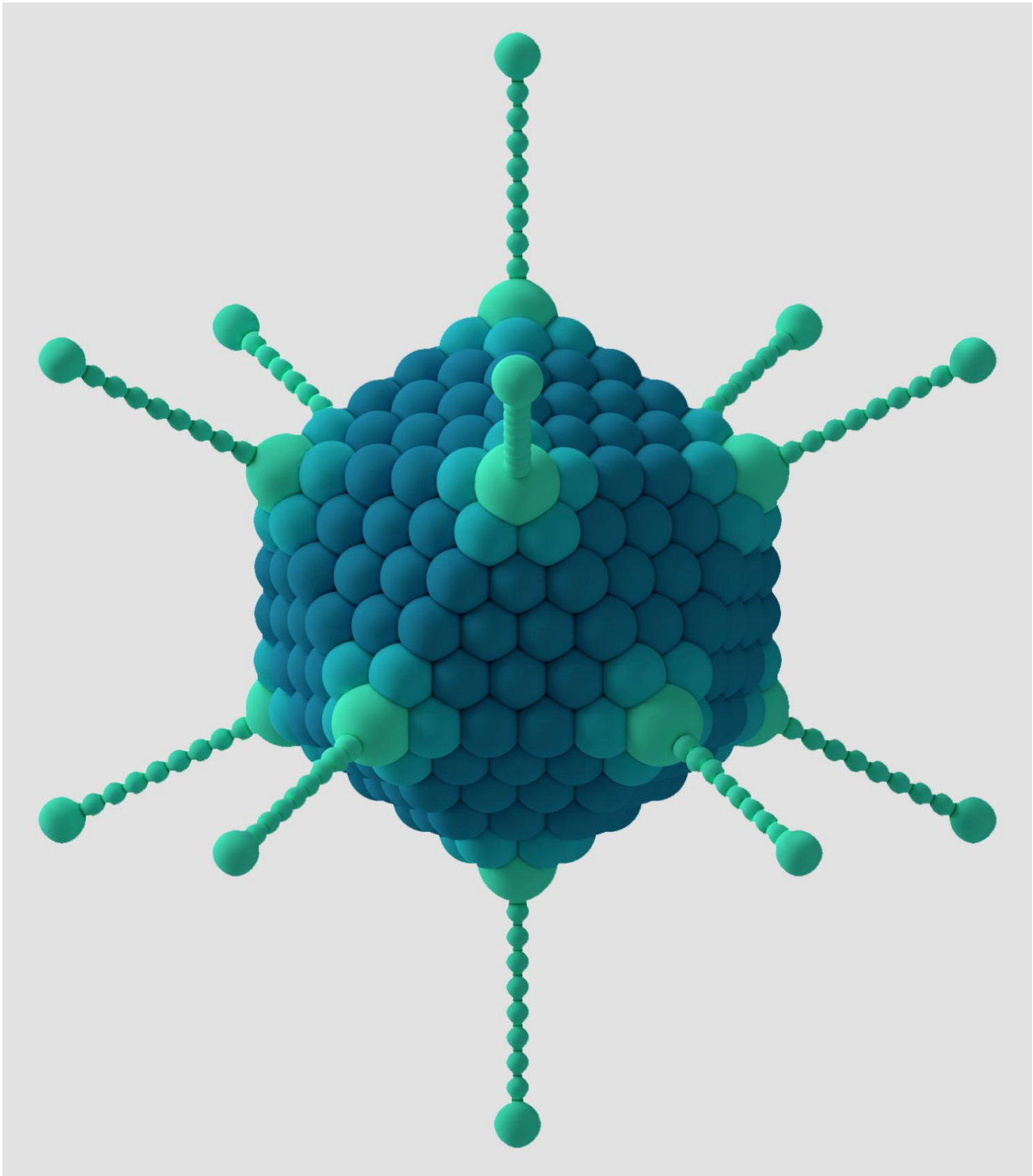
3D schematic of an adenovirus. Adenoviruses are medium-sized viruses with an icosahedral capsid containing a double stranded DNA genome. They cause a range of illnesses from the common cold and to more life-threatening multi-organ diseases in people with weakened immune systems. Credit: Wikimedia Commons

Using ideas from mathematics and computer science, Professor Sinclair from OIST's Mathematical Biology Unit worked with scientists at the University of Helsinki to reinvestigate whether the structure-based classification for viral capsids is in fact supported by previously undetected sequence similarity.

"The conventional tools for detecting sequence similarity are very fast but they can miss things," says Professor Sinclair. "We used a more classical approach that takes longer but is much more sensitive."

The team developed a computational prototype tool called the 'Helsinki Okinawa Sequence Similarity' or HOSS for short, to detect amino acid sequence similarity in viral coat proteins of icosahedral virus capsids - polyhedral capsids with 20 faces. The team also looked at nucleotide sequence similarity.

"By randomly reshuffling the order of amino acids and nucleotides in pairs or triplets of viral sequences, we used statistics to find previously undetected similarities below 17% protein sequence identity, well below what conventional tools are capable of detecting," says Professor Dennis Bamford.

The detection of extremely weak similarities in protein and coding sequences by HOSS suggests that viral capsid similarities are due to common descent, not convergence as previously suspected. This may reflect an aspect of viruses that is extremely difficult to change, and hence provide both a viable approach to classification and a potential therapeutic target.

"Our work is the first to tie structural lineages to sequences so comprehensively," says Professor Sinclair.

The team also demonstrated the power of their method by identifying a candidate capsid gene in the Pandoravirus salinus genome, something which no other team had been able to do.

Now that the researchers have shown that there are similarities between viruses that were previously undetected, further work will focus on finding more efficient methods of data extraction, beyond the HOSS prototype.

"We have also begun shifting our focus to RNA viruses, of which Zika virus and Ebola virus are examples. The genomes of RNA viruses tend to be more highly variable than DNA viruses, and are therefore even more challenging," says Professor Sinclair. "But with a refined method, it could well be possible."

Provided by Okinawa Institute of Science and Technology