Using machine learning to map form and function across the virome

Joe Grove

1. Abstract

Machine-learning (ML) is revolutionising molecular virology; enabling the discovery and investigation of novel molecular mechanisms, and paving the way to innovative new treatments for viral disease. We have combined phylogenetics and state-of-the-art ML approaches to predict structures for all viral proteins across the Flaviviridae, a large family of viruses including important human and animal pathogens. We have, thus far, focussed our subsequent analysis on viral glycoproteins, which are responsible for mediating entry of virus particles in to host cells. This has revealed new insights on the origin, evolution and mechanisms of viral entry systems across the Flaviviridae. This approach is highly adaptable and scalable, and we expect it can be applied to investigate diverse functions from across the global virome.