

## **Sequence and structural analysis of related proteins in distant viral species**

Saskia Metzler, Silvia Caprari, Olga V. Kalinina

*Max Planck Institute for Informatics, Campus E1 4, 66123, Saarbrücken, Germany, kalinina@mpi-inf.mpg.de*

Unlike cellular organisms, viruses do not constitute a monophyletic group, in which the phylogenetic history can be traced back to a common ancestor. The origin and relatedness of different virus families is currently a subject of active discussion. It is unclear, whether viruses have evolved by reduction of many essential genes from cellular species, descend from mobile elements of other organisms, or whether they precede cellular life and are ancient self-replicating units. Possibly, all these hypotheses are true, for a subset of viral families [1]. The recent discovery of giant viruses [2, 3] revived this discussion with suggestions that a certain clade of them may represent a fourth domain of life [4].

Analysis of evolutionary relationships between distant viral families presents particular difficulties, since the sequence similarity of viral proteins is rarely detectable outside the immediate viral family. We have performed an all-to-all sequence and structural comparison of viral proteins, and focused on cases where similarity is detected between proteins from viruses that use different type of nucleic acid to encode their genome. We can split the corresponding proteins families into families with balanced and unbalanced distribution of viral genome types. For the former category, we recapitulate viral hallmark genes (i.e. genes characteristic to only viruses and present in diverse species [5]) and other known wide-spread viral proteins [6], providing the first comprehensive analysis of these cases. The protein families of the latter category can be often characterized by horizontal gene transfer events. We could not detect any events of horizontal gene transfer between different viruses, however, we have identified several events of horizontal gene transfer from the host to an infecting virus. We have also identified proteins from several protein families that appear in very distant viruses, whose function is likely conserved but whose origin cannot be traced back to a single viral class, which hints at a much more complex network of kinship in the virus world than previously recognized.

1. Dimmock N, Easton A, Leppard K (2009) Introduction to modern virology, 6th edition. Wiley-Blackwell.
2. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306(5700): 1344-1350.
3. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, et al (2013) Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341(6143): 281-286.
4. Nasir A, Kim KM, Caetano-Anolles G (2012) Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC Evol Biol* 12(1): 156.
5. E. Koonin, T. Senkevich, V. Dolja (2006) The ancient virus world and evolution of cells. *Biology Direct*, **1(1)**:29.
6. M. Krupovic, D. H. Bamford (2001) Protein Conservation in Virus Evolution. In *eLS*, John Wiley & Sons, Ltd.