**Molecular evolution of (flavi)viral proteins**

*Černý Jiří (1, 2, 3), Černá Bolfíková Barbora (4), Grubhoffer Libor (1, 2), Růžek Daniel (1, 3)*

(1) Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, České Budějovice, Branišovská 31, 370 05, Czech Republic
(2) Faculty of Science, University of South Bohemia in České Budějovice, České Budějovice, Branišovská 31, 370 05, Czech Republic
(3) Veterinary Research Institute, Brno, Hudcova 296/70, 621 00, Czech Republic
(4) Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Prague 6 - Suchdol, Kamýcká 126, 165 21, Czech Republic

E-mail: cerny@paru.cas.cz

Evolutionary studies on viral proteins are very difficult due to their extreme divergence caused by virus rapid evolution. Due to this phenomenon, most sequence-based methods of molecular phylogenetic are not able to find many suitable homologues for further evolutionary studies or even to reconstruct deep evolutionary relationships among proteins of distantly related viruses or even between viruses and their cellular hosts.
In our study we focused on deciphering evolutionary history of proteins encoded by viruses in genus Flavivirus (family *Flaviviridae*). To identify distant homologues of we employed various highly sensitive algorithms searching for distant sequence (PSI-BLAST, HHblits) and structure homology (DALI, FATCAT). Identified homologues of flaviviral proteins were aligned using structural homology wherever possible. Such structure based alignments were trimmed out from ambiguously aligned amino acid residues and used for reconstruction of evolutionary history of individual proteins.
Final results show that out of roughly 3400 amino acid residues harboring ten major Flavivirus proteins, only NS3, representing 13% of the total genome length is linearly inherited across the whole Flaviviridae family. The remaining 87% of Flavivirus genome are either i) Flavivirus ORFans (C, M, NS1, NS2A, NS2B, NS4A, and NS4B – 46% of Flavivirus genome), ii) genes which have no homologues in other Flaviviridae genera but that have other cellular and viral homologues (E and NS5Met – 22% of the Flavivirus genome), or iii) genes that have homologues in other Flaviviridae genera but even closer homologues in other viruses (NS5Pol – 19% of Flavivirus genome).
Thus, the flavivirus genome is an extremely patchy structure, in which individual genes have a very different evolutionary history. This “patchiness” is most probably a result of multiple recombination events that occurred during the early history of the Flavivirus genome.