

Fast evolution of a conserved residue of polyomaviruses defines a new mechanism of adaptation that operates by accelerated codon-constrained Val-Ala (COCO-VA) toggling within an intrinsically disordered protein region

Chris Lauber^{1,2}, Siamaque Kazem¹, Alexander A. Kravchenko^{3*}, Els van der Meijden¹, Sander Kooijman¹, Mariet C.W. Feltkamp¹, and Alexander E. Gorbalenya^{1,3,4*}

¹Department of Medical Microbiology, Leiden University Medical Center, 2300-RC, Leiden, The Netherlands, ²Institute for Medical Informatics and Biometry, Technische Universität Dresden, 01307, Dresden, Germany, ³Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119899 Moscow, Russia, ⁴Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119899 Moscow, Russia

It is common knowledge that conserved residues evolve slowly. We challenge generality of this central tenet of molecular biology by describing the fast evolution of a nucleotide position that is among the most conserved in the long overlap of *de novo* and ancestral open reading frames (ORFs) of a large subset of polyomaviruses. The *de novo* ORF is expressed through either the ALTO protein or the Middle T antigen (MT/ALTO), while the ancestral ORF encodes the N-terminal domain of helicase-containing Large T (LT) antigen. In the latter domain the conserved Cys codon of the LXCXE pRB-binding motif constrains codon evolution in the overlapping MT/ALTO ORF to a binary choice between Val and Ala codons, termed here as codon-constrained Val-Ala (COCO-VA) toggling. We found the rate of COCO-VA toggling to approach the speciation rate and to be significantly accelerated compared to the baseline rate of chance substitution in a large monophyletic lineage of MT/ALTO encoding viruses comprising dozens species. We have then extended this analysis to the characterization of the evolution of the COCO-VA site within a single polyomavirus species. To this end, we have analyzed thirteen mostly newly sequenced genomes of Trichodysplasia spinulosa-associated polyomavirus (TSPyV) representing ~40% of reported cases of the Trichodysplasia spinulosa disease in humans world-wide. Only very limited genome variation ($\leq 0.6\%$) was found, with a total of four non-synonymous substitutions (NSS). Three of these affected only MT/ALTO, with one NSS - fixed most early in TSPyV evolution - involving the COCO-VA toggling. Importantly, the COCO-VA site is located in a short linear motif (SLiM) of an intrinsically disordered region, a typical characteristic of adaptive responders. These findings provide evidence that the COCO-VA toggling is under positive selection in TSPyV and many other polyomaviruses that form a monophyletic lineage and infect a wide range of hosts. Thus, the COCO-VA toggling plays a critical role in virus adaptation, which is unprecedented for conserved residues.

*, deceased