

Cultivation-independent expansion of the Nucleocytoplasmic Large DNA Viruses

Frederik Schulz^{1*} (fschulz@lbl.gov), Simon Roux¹, David Paez-Espino¹, Sean Jungbluth¹, David Walsh², Vincent J. Denef³, Katherine D. McMahon⁴, Konstantinos T. Konstantinidis⁵, Emiley A. Eloie-Fadrosch¹, Nikos Kyrpides¹, **Tanja Woyke**¹

¹ DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; ² Concordia University, Montréal, Québec, Canada; ³ University of Michigan, Ann Arbor; ⁴ University of Wisconsin-Madison, Madison; ⁵ Georgia Institute of Technology, Atlanta

Project Goals: Employ genome-resolved metagenomics on public environmental sequencing data to assess the diversity and coding potential of giant viruses and predict interactions of these viruses with their eukaryotic hosts.

The discovery of large and giant Nucleocytoplasmic Large DNA Viruses (NCLDV) with genomes in the megabase-range and equipped with a wide variety of features typically associated with cellular organisms was one of the most unexpected, intriguing and spectacular breakthroughs in virology. Current knowledge about the NCLDV is largely derived from viral isolates co-cultivated with protists and algae. Building on the rapidly increasing wealth of publicly available metagenome data, we reconstructed 2,074 NCLDV genomes from 8,535 metagenomes from sampling sites spanning the globe. This led to an 11-fold increase in phylogenetic diversity and a parallel 10-fold expansion in functional diversity. Analyzing 58,023 metagenomic major capsid proteins of large and giant viruses revealed global distribution patterns and underlined their cosmopolitan nature. The discovered viral genomes encoded a wide range of proteins with putative roles in photosynthesis and diverse substrate transport processes, revealing host reprogramming as likely common strategy in the NCLDV. Furthermore, horizontal gene transfer inferences connected viral lineages to diverse eukaryotic hosts. We anticipate that the vast diversity of NCLDV revealed here on a global scale will establish giant viruses as key ecosystem players across Earth's biomes, associated with most major eukaryotic lineages.

References

1. Schulz et al., (2019). Giant virus diversity and host interactions through global metagenomics. *Nature* <https://www.nature.com/articles/s41586-020-1957-x>

This work was conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, under Contract No. DE-AC02-05CH11231 and made use of resources of the National Energy Research Scientific Computing Center, also supported by the DOE Office of Science under Contract No. DE-AC02-05CH11231