REPORT

VIRAL GENOMICS

Giant viruses with an expanded complement of translation system components

Frederik Schulz,[1]* Natalya Yutin,[2] Natalia N. Ivanova,[1] Davi R. Ortega,[3]
Tae Kwon Lee,[4]† Julia Vierheilig,[4] Holger Daims,[4] Matthias Horn,[4] Michael Wagner,[4]
Grant J. Jensen,[3,5] Nikos C. Kyrpides,[1] Eugene V. Koonin,[2]* Tanja Woyke[1]*

The discovery of giant viruses blurred the sharp division between viruses and cellular life. Giant virus genomes encode proteins considered as signatures of cellular organisms, particularly translation system components, prompting hypotheses that these viruses derived from a fourth domain of cellular life. Here we report the discovery of a group of giant viruses (Klosneuviruses) in metagenomic data. Compared with other giant viruses, the Klosneuviruses encode an expanded translation machinery, including aminoacyl transfer RNA synthetases with specificities for all 20 amino acids. Notwithstanding the prevalence of translation system components, comprehensive phylogenomic analysis of these genes indicates that Klosneuviruses did not evolve from a cellular ancestor but rather are derived from a much smaller virus through extensive gain of host genes.

Hallmark features of giant viruses are genomes and virions sized in a range previously thought to be characteristic of cellular life (1–3). Giant virus genomes encode homologs of diverse bacterial and eukaryotic genes, including translation system components, reigniting the debate over the origin of these viruses (4–6). It has been proposed that giant viruses constitute remnants of ancient cellular life or are derived from an enigmatic fourth domain of life (3, 7, 8). Alternatively, evidence of signature cellular genes acquired from hosts implies that these viruses evolved from much smaller viral ancestors (6, 9, 10). Discovery of a virus encoding translation machinery that is more complete than in previously identified giant viruses would allow a comprehensive phylogenetic assessment of these signatures of cellular life to discriminate between these two hypotheses.

Analysis of low-complexity metagenomes from a wastewater treatment plant (WWTP) in Klosterneuburg, Austria, revealed clearly separated genomic bins comprising many genes typically found in giant viruses. From these data, a 1.57-Mb genome of a putative virus, which we named Klosneuvirus (KNV), was assembled (Fig. 1, fig. S1, and table S1). The combination of distinct GC distribution and tetranucleotide patterns, the homogenous dispersal of genes exclusively shared with giant viruses, the nearly uniform read coverage (four to seven times) from a deeply sequenced WWTP metagenome, and tight connections between the contigs through duplicated genes and repetitive sequences provide converging lines of evidence that no contaminant contigs were included in the KNV genome bin (figs. S1 and S2). Furthermore, on the basis of WWTP metatranscriptome data (table S2), 15% of the predicted KNV genes were found to be expressed. Electron microscopy of the WWTP biomass revealed ~300-nm particles resembling giant icosahedral viruses (fig. S2).

To detect viruses related to KNV, we screened nearly 7000 environmental metagenomes and discovered three metagenomics bins with high assembly quality and strong overlap in gene content (tables S1 and S3). All three bins were identified as giant virus genomes, ranging from 0.86 Mb (Indivirus) to 1.33 Mb (Hokovirus) to 1.53 Mb (Catovirus) (Fig. 1). The pangenome of KNV and its metagenome-derived relatives (henceforth Klosneuviruses) extends the giant virus gene repertoire by nearly 2500 additional gene families (Fig. 1). The largest gene pool overlap of Klosneuviruses is with the Mimiviridae, with more than 200 shared gene families, but Klosneuviruses also show marked diversity in gene content (Fig. 1). Notably, only 12 of the 355 genes Klosneuviruses share with eukaryotes (but not with other viruses) are found in all four Klosneuvirus genomes (Fig. 1). These findings emphasize the dynamic

evolution of the Klosneuviruses and, under the hypothesis that the origins and abundances of acquired eukaryotic genes are determined by the host lifestyle (11), imply that different Klosneuviruses infect distinct hosts (table S4).

Taxonomic classification of metagenomic 18S ribosomal RNA genes suggests that putative hosts of the Klosneuviruses are from the diverse protist phylum Cercozoa, found in three of the four metagenomes (table S4). With the exception of a single virus isolated from the marine flagellate Cafeteria (12), to date all giant viruses have been recovered in cocultivation with Acanthamoeba (13), which could not be detected together with Klosneuviruses. These results emphasize the key role of metagenomics in the discovery of previously unknown giant viruses independently of their hosts (14, 15).

We mapped all genes of Klosneuviruses to previously established families of orthologous genes of nucleocytoplasmic large DNA viruses (NCLDVs) [nucleocytoplasmic virus orthologous genes (NCVOGs)] (16). Nearly all NCVOGs conserved in the Mimiviridae are present in Klosneuviruses, which indicates completeness of the genome bins (Fig. 2 and fig. S3). The pattern of presence or absence of the core NCLDV genes in Klosneuviruses mostly resembled that in Mimiviruses, which suggests evolutionary affinity. Despite the overall similarity of the core gene representation, Klosneuviruses display distinct features, such as multiple additional paralogs of the major capsid protein (up to 9 in KNV) and the packaging adenosine triphosphatase (Fig. 2). This diversification of the morphogenetic machinery could lead to virion heterogeneity that might play an important role in the virus-host arms race.

To place the Klosneuviruses in a phylogenomic context, we selected five nearly universal NCLDV genes as phylogenetic markers (17). Trees obtained from a concatenated alignment of these genes showed consistent topologies in which Klosneuviruses formed a strongly supported clade positioned between Cafeteria roenbergensis virus (CroV) and the Mimiviruses (Fig. 2 and data S1). On the basis of this strongly supported position and the similarity of the gene repertoires with the Mimiviruses, combined with distinct features, we propose that Klosneuviruses are classified as a new subfamily (provisionally denoted as Klosneuvirinae) in the family Mimiviridae.

To characterize the evolutionary processes that shaped the Klosneuvirus genomes, we identified the most likely path in the evolution of giant viruses, starting from the last common ancestor of the Mimiviridae, a virus with a relatively small genome, to the emergence of at least three distinct lineages: CroV, Mimiviruses, and Klosneuviruses (Fig. 2). We observed gene gain, exceeding the amount of gene loss and leading to substantial genome size increase, in each of these three lineages independently. Lineage-specific gene gain is particularly prevalent in three of the four Klosneuviruses (Fig. 2), resulting in divergent gene complements (Figs. 1 and 2). The reconstructed viral genome evolution follows the accordion model in which phases of preferential gene gain and

Fig. 1. Genome architecture and gene content of the Klosneuviruses. The left panel illustrates genome bins of the Klosneuviruses. From outside to inside: In the first ring, solid circles indicate genes exclusively shared with nucleocytoplasmic large DNA viruses (NCLDVs) (blue), genes specific for Klosneuviruses (white), genes shared with eukaryotes (red), genes shared with Bacteria (green), genes represented in all three domains of cellular life (yellow), and singletons (gray). The second ring displays positions of genes
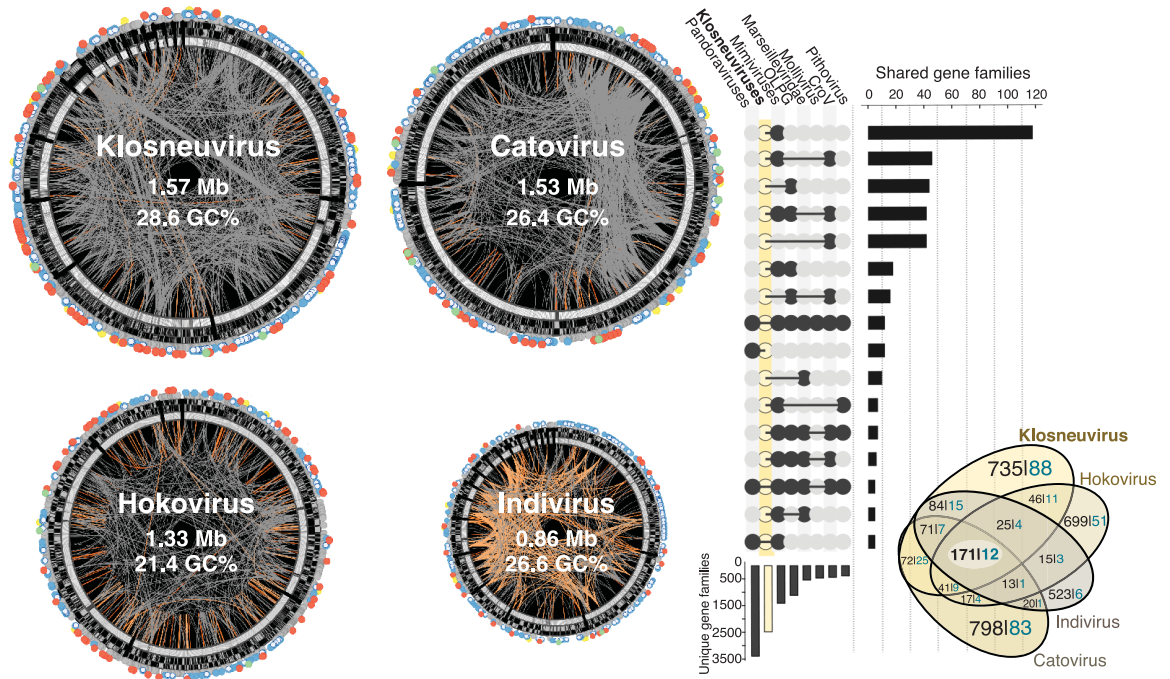

(gray) either on the minus or the plus strand. The next track depicts GC content in shades of gray ranging from 20% (white) to 50% (dark gray). Links connect paralogs (gray) and nearly identical repeats (orange). The middle panel shows the number of gene families shared between and distinct to major NCLDV lineages. Each set of compared lineages is displayed as solid circles connected by horizontal solid lines; the number of shared gene families and total number of distinct gene families in each lineage are shown as bars. OLPG, Organic Lake Phycodnavirus group; CroV, *Cafeteria roenbergensis* virus. The Venn diagram in the right panel shows shared and distinct gene families between the different Klosneuviruses. Numbers in blue indicate gene families shared with eukaryotes but not with other NCLDVs.

Fig. 2. Genome evolution and phylogenetic position of Klosneuviruses. The left panel shows a maximum likelihood tree of the NCLDVs from a concatenated alignment of five core nucleocytoplasmic virus orthologous genes (NCVOGs). The scale bar represents substitutions per site. Branch support values are shown in data S1. Solid circles adjacent to taxon names and on top of internal nodes correlate in diameter with the total number of encoded gene families of the respective node. Indicated at the branches are the number of inferred events of gene gain (blue, not including duplications), loss (red), duplication (purple), and contraction (dark purple) between parent and child nodes. The right panel shows the number of Klosneuvirus genes mapped to ancestral NCVOGs compared with other NCLDVs. The NCVOGs are clustered on the basis of co-occurrence.
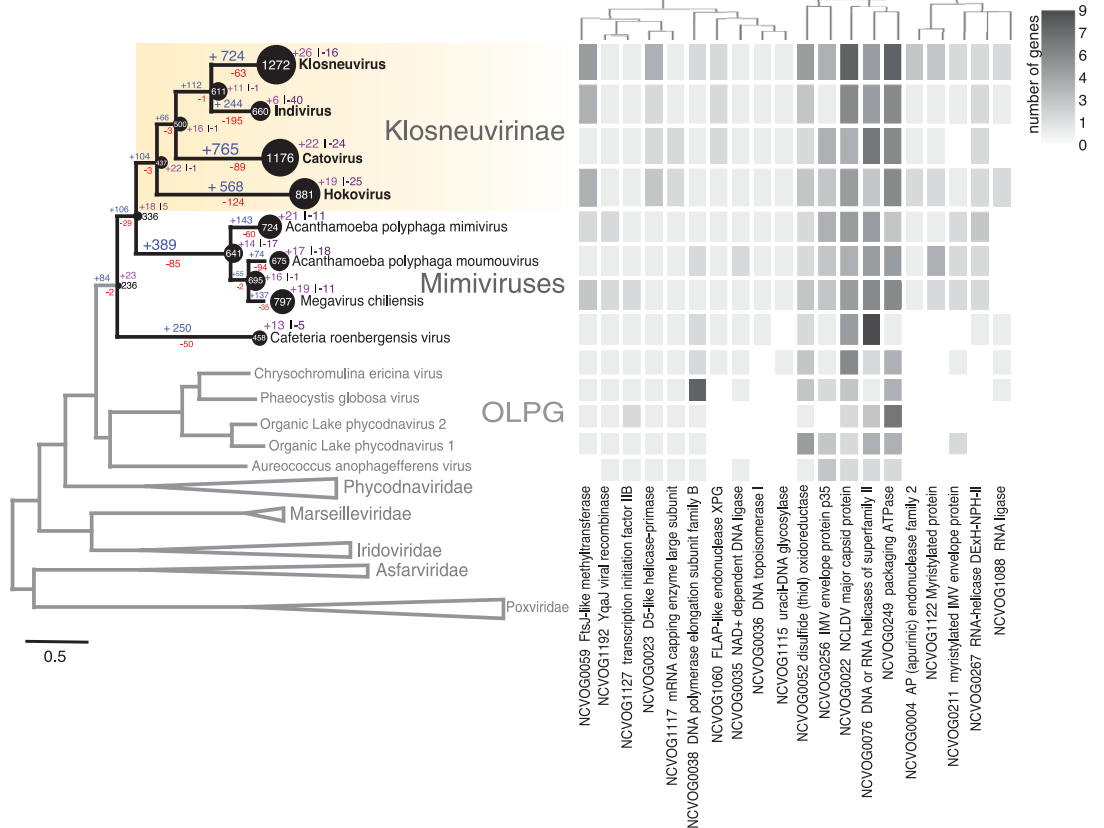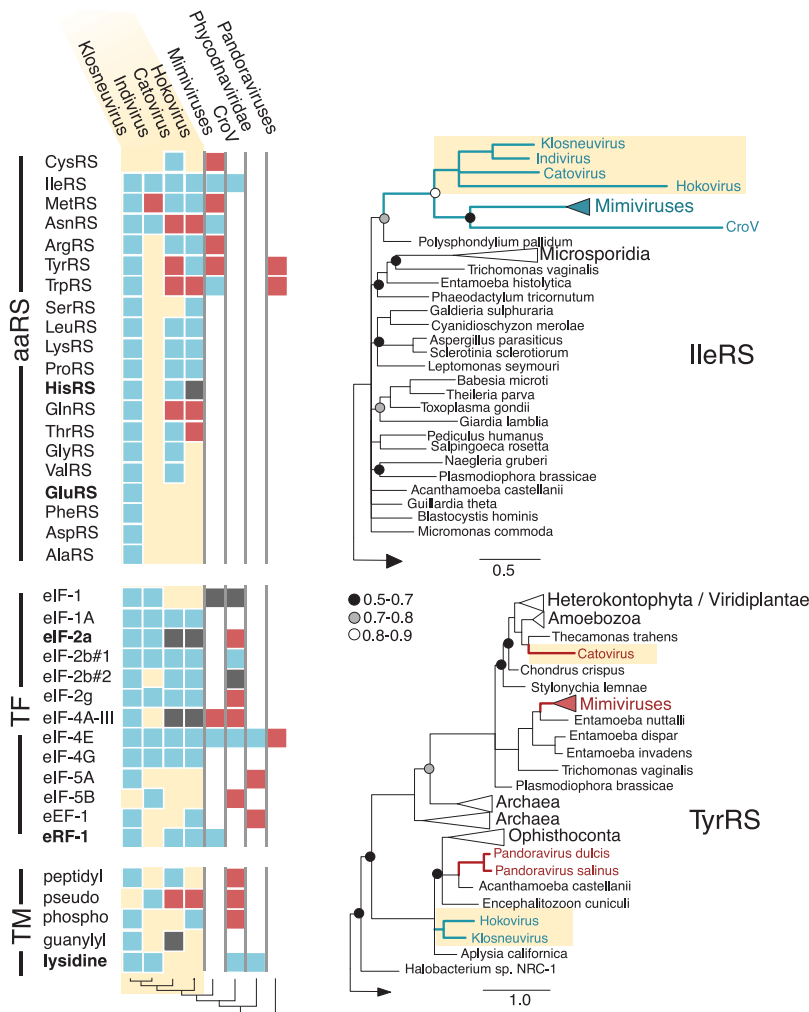
Fig. 3. Evolutionary histories of viral translation system components. The left panel shows the presence and absence of aminoacyl tRNA synthetases (aaRSs), translation factors (TFs), and tRNA modifying enzymes (TMs). Colors indicate whether the respective genes are monophyletic with KNV (blue), were acquired independently (red), or are unresolved in the phylogenetic tree (gray). Gene names in bold indicate potential ancient acquisition or capture from an unknown host. All underlying single phylogenetic trees are shown in data S1. The right panel depicts selected phylogenetic trees of aaRSs: IleRS is an example of the monophyly of the Mimiviridae, whereas TyrRS illustrates independent acquisitions. Branch support values are posterior probabilities, indicated as solid circles. Branches with support below 0.5 are collapsed; scale bars represent substitutions per site.

gene family expansion alternate with gene-loss–dominated phases (18) (Fig. 2). Differences and convergence among viral lineages are likely due to adaptation of different giant viruses to distinct hosts after host switches.

Notably, KNV encodes a large set of translation system components consisting of 25 tRNAs with anticodons for at least 14 different amino acids, as well as more than 40 translation-related proteins, including 19 aminoacyl tRNA synthetases (aaRSs) with distinct amino acid specificities, 11 translation initiation and elongation factors, a peptide chain release factor, and several tRNA modifying enzymes (Fig. 3 and table S5). This expanded translation machinery is partially and differentially shared with other Klosneuviruses (table S5) but does not correlate with genome size.

The wealth of translational genes in Klosneuviruses by far exceeds the that in Mimiviruses, which encode up to seven aaRSs (1).

To elucidate the origins of the translation system components encoded by the Klosneuviruses, we performed comprehensive phylogenomic analysis. The results demonstrate that most Klosnevirus aaRSs and translation factors affiliate with diverse eukaryotes, mainly protists including unicellular algae (Fig. 3 and data S1). These findings are incompatible with the fourth domain hypothesis, which predicts that phylogenetic trees of translation system components would show giant virus branches that are deeply positioned and distant from branches within the eukaryotic subtree, and instead imply piecemeal acquisition of these genes by giant viruses (17). Only two aaRSs

(HisRS and GluRS) were found to branch deeply between Archaea, Bacteria, and eukaryotes, potentially due to accelerated evolution of viral genes (10). Among the seven aaRSs encoded by both Klosneuviruses and Mimiviruses, three (IleRS, AsnRS, and TrpRS) are monophyletic, suggesting early acquisition antedating the split between Klosneuviruses and Mimiviruses. The remaining four are polyphyletic and appear to have been acquired independently (Fig. 3 and data S1). The only aaRS traced to the last common ancestor of the *Mimiviridae* is IleRS.

Virus-encoded translation factors and tRNA modifying enzymes appear to have similar evolutionary patterns. Monophyly with Mimiviruses and/or CroV was apparent for just four of these proteins [eIF-4E, eIF-2b#1, eRF-1, and tRNA(Ile) lysidine-synthase], suggesting their presence in the last common ancestor of the *Mimiviridae* (Fig. 3 and data S1). In summary, the inferred phylogenies of translation-related genes show a mix of monophyly and polyphyly, both within Klosneuviruses and between Klosneuviruses and Mimiviruses. Despite the abundance of translation-related genes in the Klosneuviruses, no virus has yet been found to encode ribosomal RNA or proteins, indicative of selectivity in the capture of translation system components.

The existence of giant viruses with gene complements more similar to the universal genetic makeup of cellular life than that found in the Mimiviruses has been predicted (3, 8). Our report of genomes of giant viruses encoding an expanded complement of translation system components might appear to bring us closer to the hypothetical cellular ancestor of giant viruses. However, the observed evolutionary histories of encoded aaRSs and translation factors are incompatible with the hypothesis of an ancient origin from a cellular ancestor, either a eukaryote or a member of a fourth domain of life. Our results rather imply piecemeal capture of eukaryotic translation machinery components and are most compatible with independent origins of giant viruses from much smaller viruses (17, 19). Although the biological underpinning of the high content of translation-related genes in Klosneuviruses is uncertain, the hosts of these viruses might be particularly efficient in shutting down translation upon virus infection, thus rendering virus-encoded translation system components essential for viral reproduction.

REFERENCES AND NOTES

1. C. Abergel, M. Legendre, J.-M. Claverie, *FEMS Microbiol. Rev.* **39**, 779–796 (2015).
2. B. La Scola *et al.*, *Science* **299**, 2033 (2003).
3. D. Raoult *et al.*, *Science* **306**, 1344–1350 (2004).
4. P. Forterre, *Intervirology* **53**, 362–378 (2010).
5. J. S. Abrahão, R. Araújo, P. Colson, B. La Scola, *PLOS Genet.* **13**, e1006532 (2017).
6. E. V. Koonin, N. Yutin, *Intervirology* **53**, 284–292 (2010).
7. P. Colson, X. de Lamballerie, G. Fournous, D. Raoult, *Intervirology* **55**, 321–332 (2012).
8. J.-M. Claverie, *Genome Biol.* **7**, 110 (2006).
9. T. A. Williams, T. M. Embley, E. Heinz, *PLOS ONE* **6**, e21080 (2011).
10. D. Moreira, P. López-García, *Philos. Trans. R. Soc. London Ser. B* **370**, 20140327 (2015).

11. J. Filée, P. Siguier, M. Chandler, *Trends Genet.* **23**, 10–15 (2007).
12. M. G. Fischer, M. J. Allen, W. H. Wilson, C. A. Suttle, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19508–19513 (2010).
13. J. Y. B. Khalil, J. Andreani, B. La Scola, *Curr. Opin. Microbiol.* **31**, 80–87 (2016).
14. S. Halary, S. Temmam, D. Raoult, C. Desnues, *Curr. Opin. Microbiol.* **31**, 34–43 (2016).
15. P. Simmonds *et al.*, *Nat. Rev. Microbiol.* **15**, 161–168 (2017).
16. N. Yutin, Y. I. Wolf, D. Raoult, E. V. Koonin, *Virol. J.* **6**, 223 (2009).
17. N. Yutin, Y. I. Wolf, E. V. Koonin, *Virology* **466–467**, 38–52 (2014).
18. J. Filée, *Curr. Opin. Virol.* **3**, 595–599 (2013).
19. E. V. Koonin, V. V. Dolja, M. Krupovic, *Virology* **479–480**, 2–25 (2015).