nature microbiology

Supplementary Information for

Unique viruses that infect Archaea related to eukaryotes

Ian M. Rambo, Valerie De Anda, Marguerite V. Langwig and Brett J. Baker

Correspondence to: brett_baker@utexas.edu

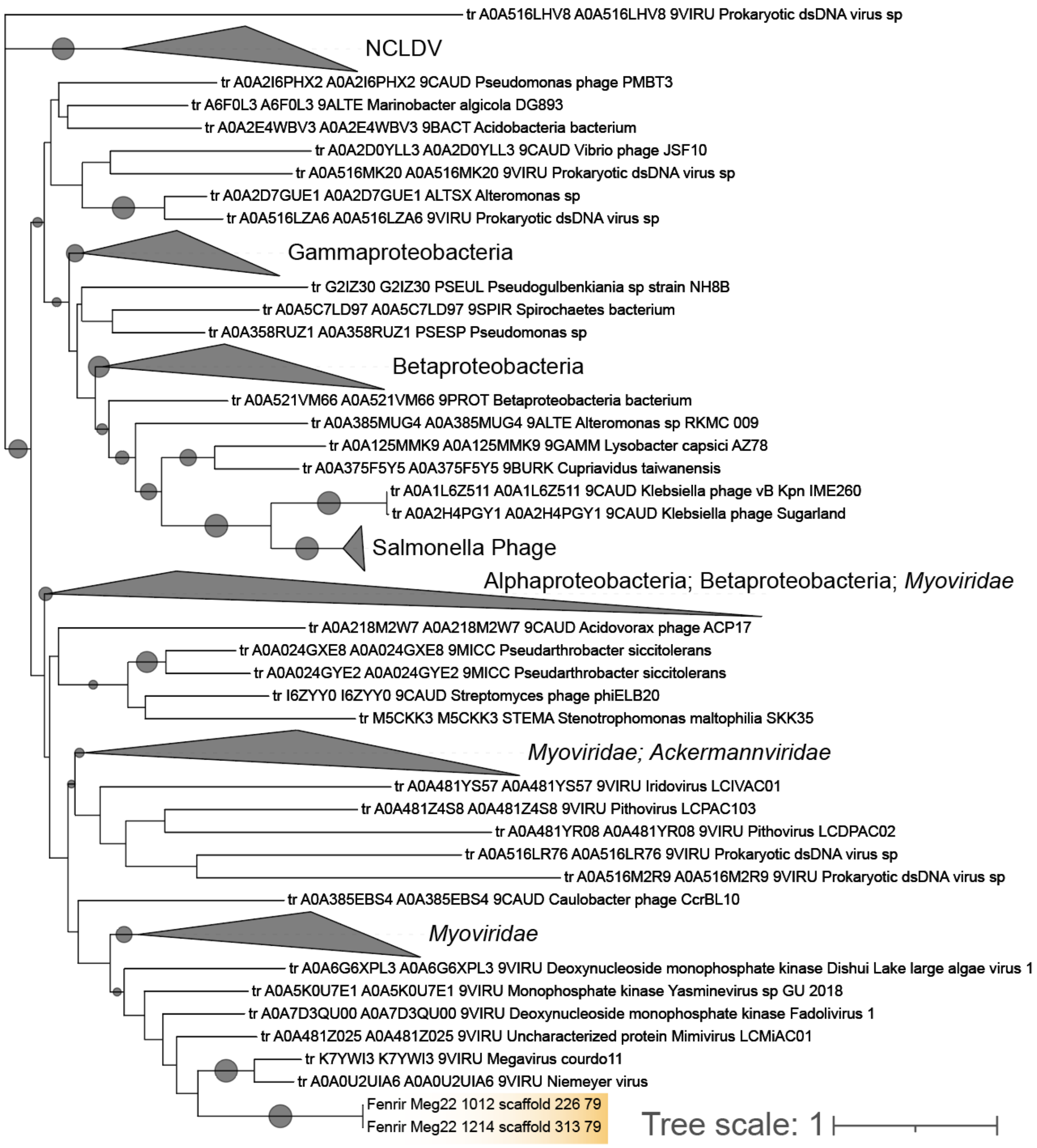
This PDF file includes:

Extended Data Figs. 1-3
Supplementary Text

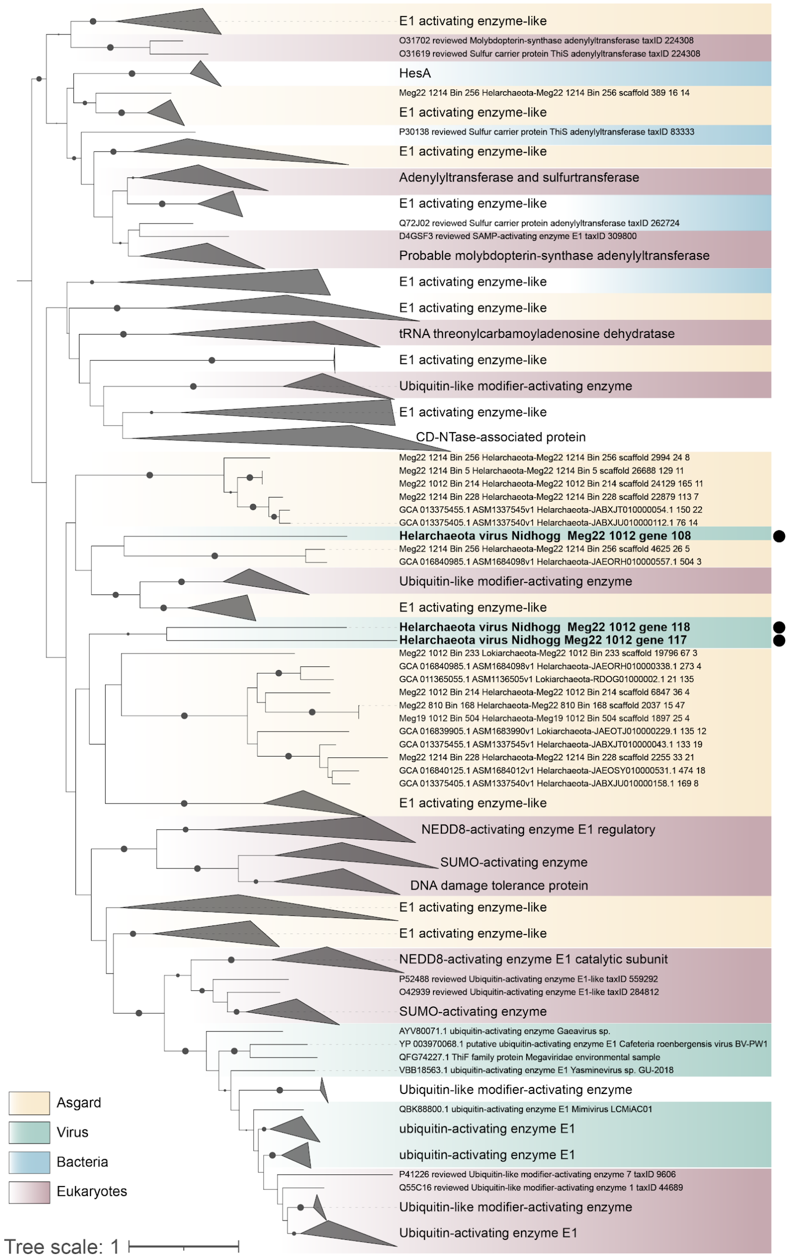
Other Supplementary Information for this manuscript include the following:

Description of Supplementary Datasets 1-7

Extended Data Fig. 1 | Genomic architecture of the complete Helarchaeota virus Nidhogg Meg22_1012. From outside to center: genes described in the main text, genes with homologs not described in the main text, hypothetical proteins, GC content, genome size ruler. Arrows pointing left indicate (-) sense, while those pointing right indicate (+) sense.



Extended Data Fig. 2 | Deoxynucleotide monophosphate kinase phylogeny. A phylogenetic tree of 241 deoxynucleotide/side monophosphate kinase sequences from viruses and bacteria. Circles on branches indicate BOOSTER supports ≥70. Lokiarchaeota virus Fenrir Meg22_1012 and Meg22_1214 sequences are highlighted in gold. The phylogeny was inferred using the LG model with fixed base frequencies and 1000 rapid bootstraps.



Extended Data Fig. 3 | Ubiquitin-activating enzyme phylogeny. A phylogenetic tree of 368 ubiquitin-activating enzyme (E1) protein sequences from archaea, bacteria, eukaryotes, and viruses (taxa are labeled with background colors). Three E1-like protein sequences were identified in Nidhogg viruses, and these are labeled with black circles and bold text. Arched lines show the connections between Nidhogg virus sequences and their Helarchaeota host. This phylogeny was inferred using the LG+R8 model with 1000 ultrafast bootstraps and optimization by nearest neighbor interchange (-bb 1000 -bnni). Circles on tree branches indicate ultrafast bootstrap supports ≥95. The tree is comprised of protein sequences belonging to the NEDD8-activating enzyme E1 catalytic subunit family (n = 11, IPR030468), ubiquitin-activating E1 enzyme (n = 218, IPR035985), viral sequences obtained from NCBI (n = 14), and sequences derived from Lokiarchaeota and Helarchaeota (n = 125).

Supplementary Text

Fig. 2 a-c abbreviations

cas1, CRISPR-associated endonuclease Cas1; *cas2,* CRISPR-associated endoribonuclease Cas2*;* *cas3*, CRISPR-associated endonuclease/helicase Cas3*; cas4*, CRISPR-associated exonuclease Cas4*; cas5*, CRISPR system Cascade subunit Cas5; *cas6,* CRISPR-associated endonuclease Cas6*;* *cas7*, CRISPR-Cas Type I effector complex subunit Cas7*; cas8*, CRISPR-associated protein Cas8*;* *csm2*, Type III CSM-effector complex small subunit Csm2*; csm3*, Type III RAMP superfamily CSM-effector complex Csm3*; csm4*, Type III RAMP superfamily CSM-effector complex Csm4*;* *csm5*, Type III RAMP superfamily CSM-effector complex Csm5*; csx1*, CRISPR system endoribonuclease Csx1; *csx14*, Subtype III-U associated protein Csx14; *cft2*, Cft2 family RNA processing exonuclease; *moaA*, molybdenum cofactor biosynthesis protein MoaA; TA, toxin-antitoxin; NT, nucleotidyl transferase; HEPN, higher eukaryotes and prokaryotes nucleotide-binding domain; kb, kilobase.

Supplementary Dataset 1.

Viral genome overview, Asgard MAG GTDBTk taxonomy, and MAG statistics.

Supplementary Dataset 2.

Minimum Information about an Uncultivated Virus Genome (MiUViG) metadata for viral genomes described in this study.

Supplementary Dataset 3.

Sequences used in the DNA polymerase B phylogeny.

Supplementary Dataset 4.

Viral Protein Family classification membership ratios for Asgard viruses.

Supplementary Dataset 5.

Viral annotations with VIBRANT, DIAMOND, and InterProScan; PhANNs classification; and HHPred results for major capsid proteins predicted with PhANNs.

Supplementary Dataset 6.

InterProScan annotations of Asgard MAGs first detailed in this study, and IMG/M annotations of all MAGs used in this study.

Supplementary Dataset 7.

Sequences used in the ubiquitin-activating enzyme phylogeny.