nature portfolio

https://doi.org/10.1038/s41564-022-01144-6

Supplementary information

Three families of Asgard archaeal viruses identified in metagenome-assembled genomes

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION

Supplementary text

Enigmatic MGEs of asgardarchaea

Two circular contigs, 7H_11 and 8H_18, were nearly identical and targeted by identical spacers affiliated to Thorarchaeia (Tables S3 and S4), but were assembled from metagenomes originating from samples collected at different depths (59.5 and 68.8 mbsf, respectively). Notably, 7H_42 discovered in our samples from the offshore Shimokita Peninsula, Japan was found to be related to Ga0114923_10000127 and Ga0209976_10000148 originating from the sediment samples from the Sumatra, Indian Ocean (Fig. S1), attesting to the consistency of this emerging group of Asgard MGEs.

The seven MGEs encode diverse proteins involved in DNA replication, repair and metabolism, which are common in MGE and viral genomes but, with the exception of 7H_42, Ga0114923_10000127 and Ga0209976_10000148 which form a group, display little overlap in gene content (see below). Nevertheless, 8H_18, 10H_0 and 7H_42 encode Topo mini-A homologs. Phylogenetic analysis of these proteins showed that, whereas 10H_0 and 7H_42 formed a clade with the Topo mini-A of wyrdvirus WyrdV4, 7H_42 branched among other archaeal sequences (Extended Data 4), suggestive of active exchange of Topo mini-A genes among archaeal viruses and MGEs. The 10H_0 and 7H_42-like MGEs as well as verdandiviruses and wyrdviruses encode multiple non-orthologous Zn-finger proteins, which might be involved in transcription regulation or mediate protein-protein interactions. 10H_0 and 7H_42-like MGEs also share homologs of proliferating cellular nuclear antigen (PCNA) and transcription initiation factor B (TFB) (Fig. S1), both of which have been previously identified in archaeal viruses. For instance, PCNA is encoded by several tailed archaeal viruses infecting halophilic archaea[1,2] and spindle-shaped viruses of Nitrososphaeria[3], whereas TFB homologs are encoded by certain rod-shaped viruses infecting hyperthermophilic Thermoproteota[4]. 10H_0 also encodes Cdc6/Orc1-like origin recognition protein, nucleoside 2-deoxyribosyltransferase, DNA lyase, MazG-like nucleotide pyrophosphohydrolase and bifunctional (p)ppGpp synthase/hydrolase as well as several DNA methyltransferases and nucleases. By contrast, 7H_42-like MGEs encode a DNA primase-superfamily 3 helicase fusion protein that are commonly found in diverse MGEs including diverse varidnaviruses infecting eukaryotes, a Rad51-like recombinase, several nucleases and chromatin-associated proteins containing the HMG domain. The larger elements also encode auxiliary metabolic genes, including PAPS reductase, sulfatase, methylthiotransferase, and enzymes involved in carbohydrate metabolism, which could boost the metabolic activities of the respective hosts. For the smaller contigs, 8H_18 and 8H_67, the vast majority of genes were refractory to functional annotation even using the most sensitive available sequence similarity detection tools, such as HHpred[5].

Auxiliary gene content of asgardarchaeal viruses

By dsDNA virus standards, genomes of verdandiviruses, skuldviruses and wyrdviruses are relatively small (≤20 kb). Thus, the corresponding gene contents are streamlined to include largely the core functions required for virion morphogenesis and genome replication. Nevertheless, some of these viruses encode auxiliary functions, including metabolic genes. In particular, verdandivirus VerdaV1 (and 10H_0 MGE) encode phosphoadenosine phosphosulfate (PAPS) reductase (also known as CysH), an enzyme reducing 3′–phosphoadenylylsulfate to phosphoadenosine-phosphate using thioredoxin as an electron donor. PAPS reductases have been previously identified in certain bacteriophages [6-8] and tailed haloarchaeal viruses [1], where they are thought to confer selective advantage to the host cells through facilitating sulfur metabolism and/or synthesis of sulfur-containing amino acids[7]. PAPS reductase of VerdaV1 might perform a similar function.

Wyrdviruses WyrdV2 and WyrdV6 carry a block of three genes coding for dUTPase, thymidylate synthase X (ThyX) and an uncharacterized protein that is conserved in some phages and is annotated as nucleotide modification associated domain 1 protein (PF07659.13, DUF1599) (Fig. 5). This putative operon is likely to be involved in the biosynthesis of thymidylate from dUTP, to increase the pool of nucleotides available for the synthesis of viral DNA. WyrdV3 encodes a homolog of the nucleoside pyrophosphohydrolase MazG, which in bacteria prevents programmed cell death by degrading the central alarmone, ppGpp[9]. MazG is highly conserved in tailed bacteriophages infecting cyanobacteria[10]. Biochemical characterization of a cyanophage MazG has shown that, instead of degrading ppGpp, it preferentially hydrolyses dGTP and dCTP[11]. Thus, MazG homolog in WyrdV3 might either function in disarming antiviral systems triggered by nucleotide-based alarmones, such as ppGpp, or in adjusting the intracellular nucleotide concentrations for optimal viral genome synthesis. Notably,

MazG homologs are also encoded by asgardarchaeal MGEs 10H_0, 7H_42, Ga0114923_10000127 and Ga0209976_10000148 (Fig. S2).

None of the known archaeal viruses encodes its own RNA polymerase[12]. Nevertheless, various transcription regulators with HTH, Zn-finger or ribbon-helix-helix domains are abundantly encoded in archaeal virus genomes[13]. This is also the case with asgardarchaeal viruses described herein. Verdandiviruses and wyrdviruses encode multiple non-orthologous Zn-finger proteins, whereas skuldviruses encode several proteins with HTH domains (Fig. 4a). In addition, WyrdV1 (as well as 10H_0, 7H_42, Ga0114923_10000127 and Ga0209976_10000148) encodes a transcription initiation factor B (TFB), a homolog of eukaryotic TFIIB, which guides the initiation of RNA transcription[14]. Among archaeal viruses, TFB homologs have been previously identified only in certain rod-shaped viruses infecting hyperthermophilic archaea [4]. Thus, Asgard viruses appear to fully rely on the core transcription machinery of their hosts but encode various transcription factors that could be involved in the recruitment of this machinery for expression of viral genes as well as in the regulation of virus gene transcription. As mentioned above, some of the genes regulated by these transcription factors are likely to encode antidefense proteins.

WyrdV1 and WyrdV3 encode homologs of the carbohydrate-specific 3'-O-methyltransferase[15]. In many archaeal viruses, the structural proteins are glycosylated by either the virus or host encoded glycosyltransferases, although the biological role of this post-translational modification remains unclear. The methyltransferase of WyrdV1 and WyrdV3 could participate in modification of the glycans attached to the virion proteins.

Supplementary references

1 Mizuno, C. M. *et al.* Novel haloarchaeal viruses from Lake Retba infecting Haloferax and Halorubrum species. *Environ Microbiol* **21**, 2129-2147, doi:10.1111/1462-2920.14604 (2019).

2 Raymann, K., Forterre, P., Brochier-Armanet, C. & Gribaldo, S. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol Evol* **6**, 192-212, doi:10.1093/gbe/evu004 (2014).

3 Kim, J. G. *et al.* Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *Proc Natl Acad Sci U S A* **116**, 15645-15650, doi:10.1073/pnas.1905682116 (2019).

4 Baquero, D. P. *et al.* New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *ISME J* **14**, 1821-1833, doi:10.1038/s41396-020-0653-z (2020).

5 Gabler, F. *et al.* Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* **72**, e108, doi:10.1002/cpbi.108 (2020).

6 Mara, P. *et al.* Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *ISME J* **14**, 3079-3092, doi:10.1038/s41396-020-00739-3 (2020).

7 Summer, E. J., Gill, J. J., Upton, C., Gonzalez, C. F. & Young, R. Role of phages in the pathogenesis of Burkholderia, or 'Where are the toxin genes in Burkholderia phages?'. *Curr Opin Microbiol* **10**, 410-417, doi:10.1016/j.mib.2007.05.016 (2007).

8 Farlow, J. *et al.* Genomic characterization of three novel Basilisk-like phages infecting Bacillus anthracis. *BMC Genomics* **19**, 685, doi:10.1186/s12864-018-5056-4 (2018).

9 Gross, M., Marianovsky, I. & Glaser, G. MazG -- a regulator of programmed cell death in Escherichia coli. *Mol Microbiol* **59**, 590-601, doi:10.1111/j.1365-2958.2005.04956.x (2006).

10 Sullivan, M. B. *et al.* Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* **12**, 3035-3056, doi:10.1111/j.1462-2920.2010.02280.x (2010).

11 Rihtman, B. *et al.* Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. *Environ Microbiol Rep* **11**, 448-455, doi:10.1111/1758-2229.12741 (2019).

12 Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D. & Koonin, E. V. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res* **244**, 181-193, doi:10.1016/j.virusres.2017.11.025 (2018).

13 Iranzo, J., Koonin, E. V., Prangishvili, D. & Krupovic, M. Bipartite network analysis of the archaeal virosphere: Evolutionary connections between viruses and capsidless mobile elements. *J Virol* **90**, 11043-11055, doi:10.1128/JVI.01622-16 (2016).

14 Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**, 85-98, doi:10.1038/nrmicro2507 (2011).

15 Bernard, S. M. *et al.* Structural basis of substrate specificity and regiochemistry in the MycF/TylF family of sugar O-methyltransferases. *ACS Chem Biol* **10**, 1340-1351, doi:10.1021/cb5009348 (2015).

Supplementary figures

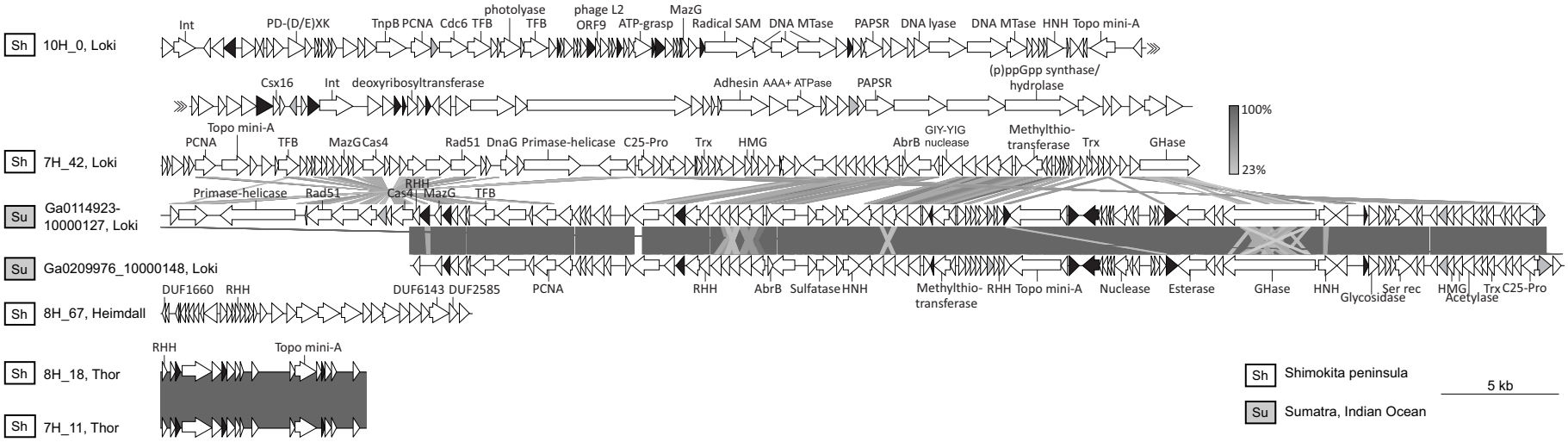
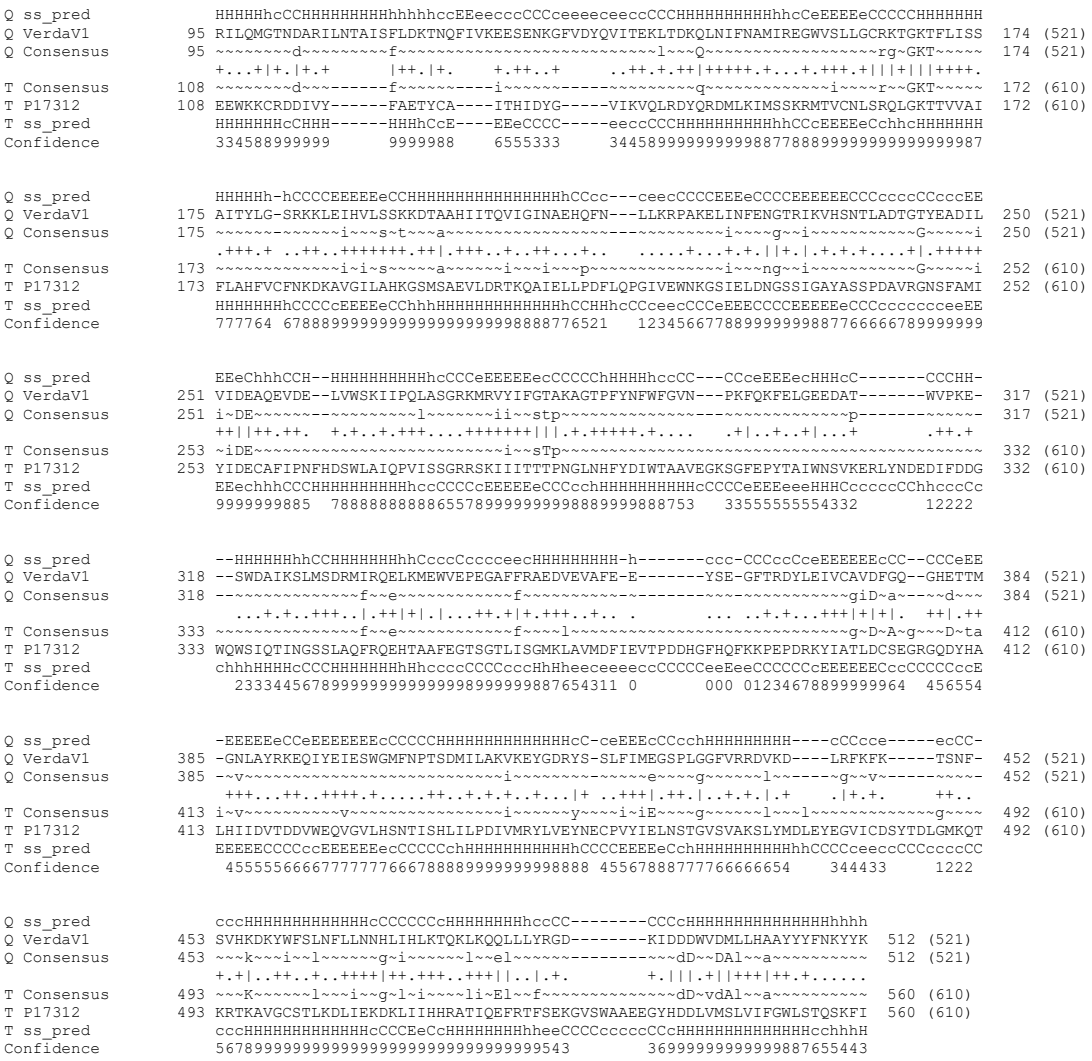


Figure S1. Genome maps of asgardarchaeal CRISPR-targeted MGEs. Due to the lack of known virus hallmark genes, the MGEs are not identified as viruses, but could represent novel virus groups. Genes encoding putative DNA-binding proteins with Zn-binding and helix-turn-helix domains are colored in black and grey, respectively. Grey shading connects genes displaying sequence similarity at the protein level, with the percent of sequence identity depicted with different shades of grey Abbreviations: Int, integrase; PD-(D/E)XK, PD-(D/E)XK family nuclease; PCNA, proliferating cellular nuclear antigen; TFB, transcription initiation factor B; MTase, methyltransferase; HNH, HNH family nuclease; PAPSR, phosphoadenosine phosphosulfate reductase; C25-Pro, C25-family protease; Trx, thioredoxin; GHase, glycoside hydrolase; RHH, ribbon-helix-helix domain-containing DNA binding protein; Ser rec, serine superfamily recombinase; HMG, high mobility group domain-containing chromatin-associated protein.

a

>P17312 TERL_BPT4 Terminase, large subunit OS=Enterobacteria phage T4 OX=10665 GN=17 PE=1 SV=1
Probab=100.00 E-value=4.1e-34 Score=301.28 Aligned_cols=403 Identities=15% Similarity=0.129 Sum_probs=304.0 Template_Neff=10.300



b

>A0A0U5AF03 CAPSD_BPK22 Major capsid protein OS=Pseudomonas phage KPP22 OX=1772250 PE=1 SV=1
Probab=99.97 E-value=3.5e-29 Score=244.70 Aligned_cols=275 Identities=11% Similarity=0.022 Sum_probs=0.0 Template_Neff=7.800

```
Q ss_pred              CcccCCCHHHHHHHHHHHHhhcchhccccCCCCCCCC-ccceEEEeeeeeeceecc--ccCCcccccccc-cceEEEEe
Q VerdaV1          2 VEIHGVSERQWKGIDTQVAQLMREPIVYTRLPSGMQIG-KGKNKAEYFQATDIFYVN--EGNRFSYEDKNT-GARKIRDA   77 (318)
Q Consensus        2 ~~~~~l~~~~~w~~Id~~v~~~~~~~~~L~ar~l~p~v~~~~-~G~~~~~~~~~~~~~d~~~~~~--~g~~~~~~~~~v~~-~~~~~~~~~v   77 (318)
                       ...++++.++|.+||.+|++.++..+.+|+|++ +...  .|+.....++.++..+.+  +++......+.+ ......++
T Consensus       73 ~~~~~ip~~~l~~id~~vi~~~~~~~~~~~~~li~-v~~~g~w~~~~~~~~~~~~e~~G~a~~~~d~~d~p~~~~~~~~~~~~~~v  151 (382)
T A0A0U5AF03     73 TPSIPTPIQFLQTWLPGLVKVMTAARKIDEIIG-IDTVGSWEDQEIVQGIVEPAGTAVEYGDHTNIPLTSWNANFERRTI  151 (382)
T ss_pred              CCCCchHHHHHHHhcchhccccccccHHHHcC-cccCCCceeEEEEEEEecCCceEEEecCCCCCceeeeEeeceeeee
```

```
Q ss_pred              EEEEEEecccccHHHHHHHHhcCCCCchHHHHHHHHHHHHHHHhCeEeecCC---ccCccccceeccccCCCccCCCceeee
Q VerdaV1         78 IITYLTLPIHIDEKDIDASSSEGITPLLVQMEVECRAKMQEKIQKIILHGSA---HTGGRGICTFQQDGDVVEQTYAFEA  154 (318)
Q Consensus       78 ~~i~~~f~i~~~~rel~as~~~g~~~ld~~~a~~A~r~vae~ed~lif~G~~----~~gi~GL~n~~~~~~~~~~~~~g~~~~~~  154 (318)
                       +.++..|.++  ++|++++++.|. +++.+++.+|++.+++++|+++|+|++    ..+++||+|+        ++.+...
T Consensus      152 ~~~~~~~~~~~~~~~~--~~El~~a~~~g~~~l~~~k~~aa~~~~~~~~~~~n~i~f~G~~~~~~~~~~~~~~~GLlN~--------p~l~~~~  220 (382)
T A0A0U5AF03    152 VRGELGMMVG--TLEEGRASAIRL-NSAETKRQQAAIGLEIFRNAIGFYGWQSGLGNRTYGFLND--------PNLPPFQ  220 (382)
T ss_pred              EEEEEEeEec--hHHHHHHHhcCC-CHHHHHHHHHHHHHHHHHHhhEeeeecCCCCCceeecccC--------CCCCccC
```

```
Q ss_pred              c---Cccc--chHHHHHHHHHHHHHH--HHCCCCCC----CEEEEChhHHHHHhcccCCCCCCcHHHHHHHHHHhcCCcc
Q VerdaV1        155 T---LTWT--NNLQFSLEMEIARGLL--RAAHIMPP----YKLWMTPGIRPTLRANVNATTGVSDWKLFQQDWMNSGVIS  223 (318)
Q Consensus      155 ~----~~W~~~t~~~i~~dv~~ai~~L--~~~~g~~~p----~~~L~L~p~~y~~L~~~~~~~~~t~~~~~~~~~i~~~~i~~~~~~~~~~  223 (318)
                       .   .+|.  ++.++++||..++.+|  .+.++++|    ++|+||++.|..|.+  .+..+.    ++++++|+++.
T Consensus      221 ~~~~~~~W~~~T~~eI~~dv~~~~~~~~l~~~s~g~~~p~~~~~~tL~lp~~~~~~~L~~~--------------tvl~~Lk~~~~~--  291 (382)
T A0A0U5AF03    221 TPPSQGWSTADWAGIIGDIREAVRQLRIQSQDQIDPKAEKITLALATSKVDYLSV--TTPYGI----SVSDWIEQTY---  291 (382)
T ss_pred              CCCCCCccccCCHHHHHHHHHHHHHHHHHHcCCCCCCCCCceEEEEeHHHHHHhcc--CCCCCC----cHHHHHHHC---
```

```
Q ss_pred              CCCCCCCCCCchhhceecchh---cCCCCCCCceEEEEEECCC-----------CeEEEEeccCcEecCCCCCCceeece
Q VerdaV1        224 GEKGTEPFIPDINQITETDAL---YNGTLSTATQAFLLWKDDI-----------NCNYIAEAYPIHRAGIPNKEFEGDID  289 (318)
Q Consensus      224 ~~~~~~~~~~~~~~~~~~~~~~i~~~~~~~~~~~--~~~~~~~~~~~~~~~~~~d~~l~~~~~~~~~~~~~~~~~~~~~~--------~~~~i~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~i~  289 (318)
                              |.+. |..++++++.   +.+    +.++++++.+++           +++.+.+|++++.++.+.+++.|.++
T Consensus      292 ------p~i~~I~~~pel~~a~~~g~~------g~~~~~l~~~~~~~~~~~~~~~~g~~~~~~~~~~~~~~~vp~~~~~~~~~~~~~~~~~~~~~~~~~~  360 (382)
T A0A0U5AF03    292 ------PKMR-IVSAPELSGVQMKAQE----PEDALVLFVEDVNAAVDGSTDGGSVFSQLVQSKFITLGVEKRAKSYVED  360 (382)
T ss_pred              ------CCCE-EEECHhhhccccCCC----CccEEEEEECchhcccCCCCCcccceeeeccccceecccEeccceEEEe
```

```
Q ss_pred              EEEeeecEEeecCCCCceEEE
Q VerdaV1        290 YALRWAGCFLPKNPKGAVYV  309 (318)
Q Consensus      290 ~~~~r~gGv~~~~~~~~~~~~~~~~~~~  309 (318)
                       |..+++|+++ |.|..|++.
T Consensus      361 ~~~~t~G~~i~~rP~ai~~~  379 (382)
T A0A0U5AF03    361 FSNGTAGALC-KRPWAVVRY  379 (382)
T ss_pred              eeceeeeEEE-ecccceehhh
```

Figure S2. Results of the HHsearch analysis queried with the putative large terminase subunit (a) and the major capsid proteins (b) of verdandivirus VerdaV1. H(h), α-helix; E(e), β-strand; C(c), coil.

a

>P15794 CAPSD_BPPM2 Major capsid protein P2 OS=Pseudoalteromonas phage PM2 OX=10661 GN=II PE=1 SV=2
Probab=98.32 E-value=0.0012 Score=61.40 Aligned_cols=219 Identities=12% Similarity=0.056 Sum_probs=0.0 Template_Neff=8.200

```
Q ss_pred              eeecCceeeecceeeccEEEEEecCCcccCcchhheeecccceEEEEECCEEEEEE-ecccCcccchHHHHHHHHHHH
Q SkuldV1         15 AESENKHVKIDNTFPIKKIVIIHNPGDFAGGSTGVLIANTILDNLRVFVHGKEVISL-IGDVDVDTAPFAIQLMREANKL   93 (277)
Q Consensus       15 a~~~~~~~~~ti~~tf~ik~I~I~~~~~ga~sGGatG~~VaN~a~~s~rv~~nGk~iI~~-dG~~~vd~~s~gi~lLRE~~~~   93 (277)
                       +.+.+.+++|+.-.....|.+..+           .+...-++.|||.+||+.++.+ +|          +.|..+|++
T Consensus       14 ~~g~~at~~lp~g~tY~~i~l~~t------------~~t~a~I~~I~v~~NG~~i~~~~~~~~~~~------------~~L~~~n~y   71 (269)
T P15794         14 AAGNSCSIKLPIGQTYEVIDLRYS-----------GVTPSQIKNVRVELDGRLLSTYKTL-----------NDLILENTR   71 (269)
T ss_pred              CCCCEEEEEcCCCCCeeEEEEEEEc-----------CCCHHHcCeEEEEEECCEEEEEeCCH-----------HHHHHHHHH
```

```
Q ss_pred              hcccCCccceeEEEcCCccCCC-------------------CeEEEEEEecceeeecCCCCceeeeceEEEEEEeCCCCe
Q SkuldV1         94 RNKVADADEYFEIDFPKAIGRG-------------------HNAYIDFRFNTIANMNDGDRTTYTGATIDIMVEVGTPGR  154 (277)
Q Consensus       94 ~~~VA~~d~y~~I~FP~aip~g-------------------~dvqI~i~~~tsaq~~gGDr~T~ta~t~DI~v~~~~~kG~  154 (277)
                       . +-+..++++.|.|..+--++                   .+.+|++.+++.+.            +-+++.......+..
T Consensus       72 ~-g~~~~~~g~l~i~F~~~~~~~~~~~~~~~~~~~~~~al~t~~~~s~~lev~i~a~~-----------~~p~L~~~a~~~~~~~~~~  140 (269)
T P15794         72 H-KRKIKAGVVSFHFVRPEMKGVNVTDLVQQRMFALGTVGLTTCEIKFDIDEAAA----------GPKLSAIAQKSVGTA  140 (269)
T ss_pred              h-cCCcccCcEEEEecCChhhcCCCchhHheeehhcccCCCccEEEEEEEeCCCCC----------CCEEEEEEEeCCCCC
```

```
Q ss_pred              EEEeeEcceeeccccccCCCccccCCcc--cccceeeeEEEEEecCCCccccceeeEEEeeCceeeecC-cHHHHHHHHccC
Q SkuldV1        155 LSLLRNCGKIVLGANTGRLPEFINPT--EFGYKALNLMVLIEDDGTPSNTAIAKLELRKGSNIIREG-SIPKLQELTKGK  231 (277)
Q Consensus      155 ~ip~~~~g~~s~ga~~G~l~e~lp~~---~plR~~~Lm~it~D~gT~S~tay~~L~Is~g~~~i~DG-siakl~~~~~~~K  231 (277)
                       .-.++....+....-.|.....--|.  ...+|.  +.+-+.|         .++++|+.++..+||. +-+.....+++-
T Consensus      141 ~~~~~~~~~~~~~~~~~~~G~~~~~~lp~~~~~~~i~~--ih~~~~~~~~~~~~~~I~~v~v~~D~~~i~~~~~~~~~~~~~~~~~~~  210 (269)
T P15794        141 PSWLTMRRNFFKQLNNGTTEIADLPRPVGYRIAA--IHIKAAG--------VDAVEFQIDGTKWRDLLKKADNDYILEQY  210 (269)
T ss_pred              CeeEEEEeeeeccCCCcccccccCCCCCcceeEE--EEEecCC--------cCEEEEEECCEEEEecccHHHHHHHHHHh
```

```
Q ss_pred              ceeeccceEEEEEecC--cC--------CHHHeEEEEEeccCccceEEEEEeeccC
Q SkuldV1        232 FSIALGTGFVWIPIKQ--AI--------SASQLKLISQIDSAGTNIEVHWMLTSLK  277 (277)
Q Consensus      232 s~vA~s~G~~~i~fp~--ki--------s~~tLkl~~~~~~tAgt~~e~h~~~~~~~~~  277 (277)
                       .++. -+|++.+.|=.  ..        ..++|+|++..++|++=--+.-.+..++
T Consensus      211 ~~~p-qag~~~~Df~~~~g~~~~~~~~~~~~~~~~~~~~~~l~~~~~~~~~~~~a~~i~i~~E~l~~~~~  265 (269)
T P15794        211 GKAV-LDNTYTIDFMLEGDVYQSVLLDQMIQDLRLKIDSTMDEQAEIIVEYMGVWS  265 (269)
T ss_pred              Cccc-CCCeEEEeCCccCChhhhhhhccccceEEEEEEecCccceEEEEEEecccc
```

b

>3J31_B Major capsid protein; Double jelly roll fold; 4.5A; Sulfolobus turreted icosahedral virus
Probab=97.84 E-value=0.013 Score=58.95 Aligned_cols=242 Identities=13% Similarity=0.060 Sum_probs=0.0 Template_Neff=6.300

```
Q ss_pred             eeEeeecCceeeecceeeccEEEEEecCCccccCcchhhee-----eccccceEEEEECC-EEEEEEeccccCcccchHHH
Q SkuldV1          12 IAFAESENKHVKIDNTFPIKKIVIIHNPGDFAGGSTGVLI-----ANTILDNLRVFVHG-KEVISLIGDVDVDTAPFAIQ   85 (277)
Q Consensus       12 iafa~~~~~~~~ti~~tf~ik~I~I~~~~~ga~sGGatG~~V-----aN~a~~s~rv~~nG-k~iI~~dG~~~vd~~s~gi~   85 (277)
                     .+++.+.++++.|+|.-+|++|.|.......+++++...+     ...+++.|+|..|| .-+.+++|        ...+
T Consensus       13 ~s~~a~~t~tIdLP~~~pIS~I~l~~~~t~~~s~~~~~~~~~~~~~~~p~~~I~kIelv~~Gs~vi~SlsG---------~~l~   84 (345)
T 3J31_B          13 YAWTAGTNIPIKIPRNNFIRKIRVQLIGSISNSGTAAVTLPSAPFPYNLVQTFNLSYEGSKTLYSVSG--------TGLG   84 (345)
T ss_dssp             EECCTTCEEEEEECCSSEEEEEEEEEEEEEECCSSSCEECCCTTTTGGGSCEECEEETTTEESEEEEH--------HHHH
T ss_pred             cccCCCCeEEEEcccCcceeEEEEEEEEeeeCCCCccccCCCCCchHHhceeEEEEECCCceEEEEeH--------HHHH


Q ss_pred             HHHHHHHHhcccCC----------ccceeEEEcC------CccCCCCeEEEEEEecceeeecCCCCceee-eceEEEEEE
Q SkuldV1          86 LMREANKLRNKVAD----------ADEYFEIDFP------KAIGRGHNAYIDFRFNTIANMNDGDRTTYT-GATIDIMVE  148 (277)
Q Consensus       86 lLRE~~~~~~~~VA~----------~d~y~~I~FP------~aip~g~dvqI~i~~~tsaq~~gGDr~T~t-a~t~DI~v~  148 (277)
                     .|-.+++.......           .+..+.+.||      .+..++ +.|++|.++.     .++..++. +.++.++-.
T Consensus       85 aLn~y~~~~~p~~~~~~~~~~~~~~~~~~~~~~~~~IdFG~~~A~d~~~nLqL~It~~~~~------~a~~~s~~~~~~~itv~~~  158 (345)
T 3J31_B          85 ILMYYTTKGQNPAYPAPGTSVPASGSVNLNVMWEFDLARFPATMVQ-NIILSILTGQ-----APSGVSINASFYITITYE  158 (345)
T ss_dssp             HHHHHHTTTCSCEESCTTCEECTTSEEEEEEEEEEEEEEECTTSSS-CEEEEEECCC-----CCTTSEECCEECCEEEEE
T ss_pred             HHHHHHhCCCCCCCCCCCCccCCCCCceeEEEEEEEEccceeccCCC-CeEEEEEECC-----CCCCCccCceeEEEEEEE


Q ss_pred             ----------eCCCCe-------EEEeeEcce-eeccccccCCCccccCCccccccceeeeEEEEEecCC--Cccccceee--
Q SkuldV1         149 ----------VGTPGR-------LSLLRNCGK-IVLGANTGRLPEFINPTEFGYKALNLMVLIEDDG--TPSNTAIAK--   206 (277)
Q Consensus      149 -------------~~~kG~---------~ip~~~~~g~~~s~ga~~G~l~e~lp~~~~plR~~~Lm~it~D~g--T~S~tay~~~--   206 (277)
                               ++.++.       ..+-...-. ++-.+..+.-...-=|...|||+  +++..+|++  -.+++..+.
T Consensus      159 ~~~~~~~~~~~~~~~~~~~~~mp~~~~~~kev~~~~~~~~~~~~~~s~~~~~~idLPtg~~~Rk--lli~~~d~~~g~~~~~~~~~~~~l   236 (345)
T 3J31_B         159 RVTAQEILSEGGLGADGEMPLATVLPKVIEIPTFNVPASSAPIHVAYLQPGQIYKR--QLVYVINSTSGINNTDPTEYEL   236 (345)
T ss_dssp             CCCHHHHHTTTSBCSSCSSBTTCCEECEEEEEEEEEECBCSSCEEEEECCSSSEEEE--EEEEEEBTTTBTCCCCEEEEEE
T ss_pred             EeccccceeeecCCCCcccCCCchhccceEEEeecCCCCCCCCCCceEEEECCCCCeEeE--EEEEEEeCCCCCCCCCcccccce


Q ss_pred             -EEEeeCceeeecCcHHHHHHHHccCce-eeccceEEEEEecC---------c--CCHHHeEEEEEecccCcccceEEE
Q SkuldV1         207 -LELRKGSNIIREGSIPKLQELTKGKFS-IALGTGFVWIPIKQ---------A--ISASQLKLISQIDSAGTNIEVH  270 (277)
Q Consensus      207 -L~Is~g~~~i~DGsiakl~~~~~~~Ks~-vA~s~G~~~i~fp---------k--is~~tLkl~~~~~~tAgt~~e~h  270 (277)
                      |....+. ..++-...-|+++.+..-+ -.+..|++++-|-+         .  -..++++|.+..+.+|.---||
T Consensus      237 ~~~~~~~~-~~~~~~~~~~l~~~~~~~~~~~~~~~~~g~~~iDf~~~~~g~l~l~~~~~~~~~~~~~~L~~~~~~~~~g~~~vi~   312 (345)
T 3J31_B         237 KIVRGVPT-DKIKVSWAALQAENQAEYQVAPYSGASAIIDFRKYFNGDLDLTHAPSDSIEYDLALQNQDNVYSLYVS   312 (345)
T ss_dssp             EECSSSCE-EEEECBHHHHHHHHHHHCSCCSSTTEEEEEGGGTTTSSEECSSSCSSSCEEEEEEESSCEEEEEEEEE
T ss_pred             EEeecCCC-ceeeeCHHHHHHHHHhhhCCCCCCCcEEEEechhhccCccccCCCCCCCCcEEEEEEecCCCCceeEEEE
```

c

>P22535 CAPSD_BPPRD Major capsid protein P3 OS=Enterobacteria phage PRD1 OX=10658 GN=III PE=1 SV=2
Probab=96.16 E-value=1.5 Score=46.49 Aligned_cols=247 Identities=11% Similarity=0.091 Sum_probs=0.0 Template_Neff=5.600

```
Q ss_pred             eeEeee--cCceeeecceeeccEEEEEecCCccccCcchhheeecc-----ccceEEE-EECCEEEEEEeccccCcccchH
Q SkuldV1          12 IAFAES--ENKHVKIDNTFPIKKIVIIHNPGDFAGGSTGVLIANT-----ILDNLRV-FVHGKEVISLIGDVDVDTAPFA   83 (277)
Q Consensus       12 iafa~~--~~~~~~ti~~tf~ik~I~I~~~~~ga~sGGatG~~VaN~-----a~~s~rv-~~nGk~iI~~dG~~~vd~~s~g   83 (277)
                     +.|...  ...+++|+|+--+++|.+..+.......|++|+.-...     +++.||+ -.|+-..++++|..
T Consensus       45 ~~~~~~~~~s~~t~~Lp~~G~L~~l~l~v~~tlt~~~~~t~a~t~~~~~~P~~lI~~I~l~~~gn~~~~~~~sG~~--------  116 (395)
T P22535          45 QTFDPANRSVFDVTPANVGIVKGFLVKVTAAITNNHATEAVALTDFGPANLVQRVIYYDPDNQRHTETSGWH--------  116 (395)
T ss_pred             eeccCcCCcEEEEecCCCceEeeEEEEEEEEEEEcCCCCccccccCCcChhHhcceEEEECCCCCceEEccHHH--------


Q ss_pred             HHHHHHHHHHhcccCCc------------------------------cceeEEEcCCcccCCCC------------eEEEE
Q SkuldV1          84 IQLMREANKLRNKVADA------------------------------DEYFEIDFPKAIGRGH------------NAYID  121 (277)
Q Consensus       84 i~lLRE~~~~~~~~VA~~--------------------------------d~y~~I~FP~aip~g~------------dvqI~  121 (277)
                         |+.++++++.....                                .-.|.+++|-++-++.                +++++
T Consensus      117 ---L~~~n~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~as~~Ag~t~~v~f~l~iPval~~~D~~G~il~qn~~t~l~L~  193 (395)
T P22535         117 ---LHFVNTAKQGAPFLSSMVTDSPIKYGDVMNVIDAPATIAAGATGELTMYYWVPLAYSETDLTGAVLANVPQSKQRLK  193 (395)
T ss_pred             ---HHHHHHHhcCCCeeeeccCCCCccccCCcccccccCcceecCCCeEEEEEEEEeeeeecCCcccceeeecCCCCCccEEEE


Q ss_pred             EEecc---------------eeeecCCCCceeeeceEEEEEEe-----CC-CCeEEE---------eeEcceeeecccccC
Q SkuldV1         122 FRFNT---------------IANMNDGDRTTYTGATIDIMVEV-----GT-PGRLSL---------LRNCGKIVLGANTG  171 (277)
Q Consensus      122 i~~~t---------------saq~~gGDr~T~ta~t~DI~v~~~~~~~~~~~~~~-kG~~ip---------~~~~~g~~s~ga~~G  171 (277)
                     |.+++               .-+..+++..|+..=.-+.|+     -. .+...|         +|..-...-.+..-
T Consensus      194 Vt~at~~~~~~~~~~~~~~~~~a~~y~~~~~~~~t~tl~tv~V~~~~~e~~~~~iPi~~~~~~~~~~P~~dlst~~~l~~~~~~~~~~~~~G~  273 (395)
T P22535         194 LEFANNNTAFAAVGANPLEAIYQGAGAADCEFEEISYTVYQSYLDQLPVGQNGYILPLIDLSTLYNLENSAQAGLTPNVD  273 (395)
T ss_pred             EEEcCccccccccCCCchhHhhcCCCccccceEEeEEEEEEEEEeEeecCCCCCCCccCCCcHHHHHhHhheeeccCCCCce


Q ss_pred             CCccccCCcccccceeeeEEEEEecCCCcccc--cceeeEE-EeeCceeeecCcHHHHHHHHccCceeeccceEEEEEecC-
Q SkuldV1         172 RLPEFINPTEFGYKALNLMVLIEDDGTPSN--TAIAKLE-LRKGSNIIREGSIPKLQELTKGKFSIALGTGFVWIPIKQ-   247 (277)
Q Consensus      172 ~l~e~lp~~~~plR~~~Lm~it~D~gT~S~~--tay~~L~-Is~g~~~i~DGsiakl~~~~~~~Ks~vA~s~G~~~i~fp~-   247 (277)
                     +.-++  |...-||.  |++ ++|.|.+.+  +..+.+. +..+...+++-...-+++.++.+.+.+-.+.|++++.|-.
T Consensus      274 ~~i~L--p~g~~yr~--l~~~v~DnG~~~~~~~~d~~~~~~~l~~n~~di~~~~~~~~~~~~~~~~~~~~yg~~lp~Gvy~~Df~~~   348 (395)
T P22535         274 FVVQY--ANLYRYLS--TIA-VFDNGGSFNAGTDINYLSQRTANFSDTRKLDPKTWAAQTRRRIATDFPKGVYYCDNRDK   348 (395)
T ss_pred             eEEEc--ccccceEE--EEE-EEeCCCccCCCcccccceeeeeCCceeeccCHHHHHHHHHHhhCCCCCCeEEEEeCCCC


Q ss_pred             ---cCCHHHeEEEEEeccCcccceEEEEEee
Q SkuldV1         248 ---AISASQLKLISQIDSAGTNIEVHWMLT  274 (277)
Q Consensus      248 ---kis~~tLkl~~~~~~tAgt~~e~h~~~~~  274 (277)
                        ----+.+.|.+..-+.|.+-.+-+-|.
T Consensus      349 gi~t~~~~~~~~l~~~~~~~~~~~~~~~~~~~~~~~  378 (395)
T P22535         349 PIYTLQYGNVGFVVNPKTVNQNARLLMGYE  378 (395)
T ss_pred             cccccccceeEEEEcccccccccceeEEEEe
```

Figure S3. HHsearch profile-profile comparisons between the putative major capsid protein of skuldvirus SkuldV1 and the double jelly-roll major capsid proteins of (a) corticovirus PM2, (b) turrivirus STIV, and (c) tectivirus PRD1. H(h), α-helix; E(e), β-strand; C(c), coil.

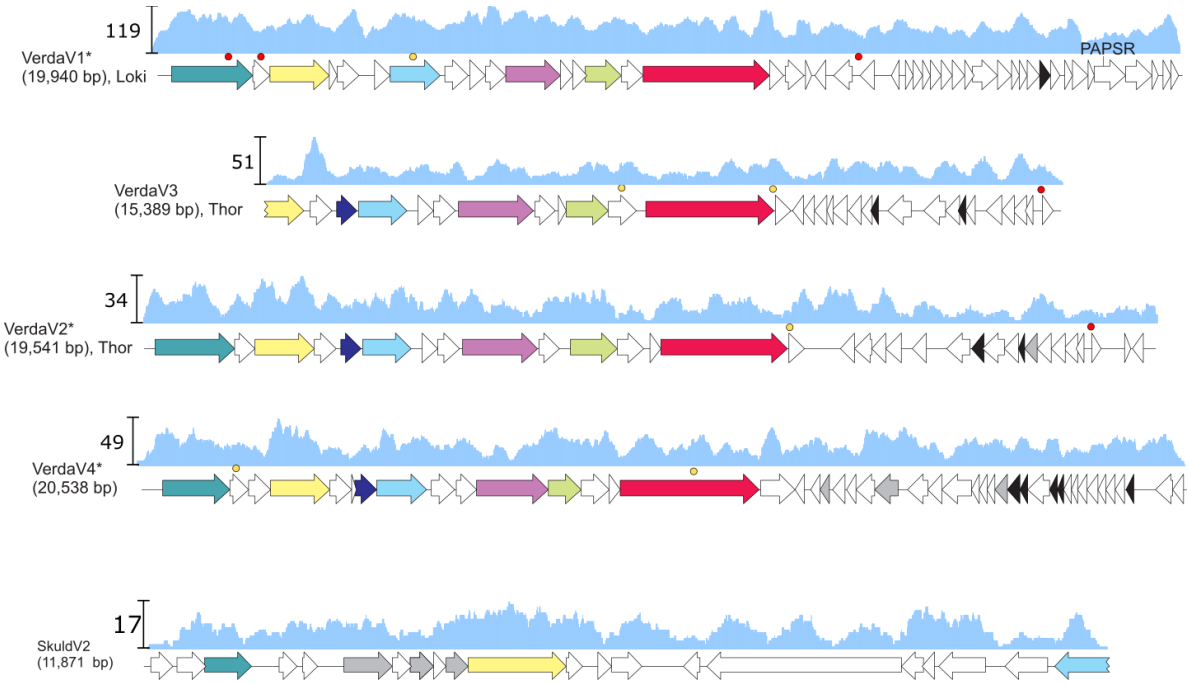


Figure S4. Read depth along viral contigs. The scale on the left shows the maximal coverage for each contig. The open reading frames are colored the same way as in the corresponding Figures 3 and 4 for verdandiviruses and skuldviruses, respectively. Colored circles represent the positions of protospacers targeted by asgardarchaeal CRISPR spacers.

Supplementary Data 1. Output of CRISPRDetect output for CRISPR arrays analyzed in this work.