



# HHS Public Access

Author manuscript

*Nat Microbiol.* Author manuscript; available in PMC 2024 June 11.

Published in final edited form as:

*Nat Microbiol.* 2022 July ; 7(7): 962–973. doi:10.1038/s41564-022-01144-6.

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

Sofia Medvedeva<sup>1,2,6</sup>, Jiarui Sun<sup>3</sup>, Natalya Yutin<sup>4</sup>, Eugene V. Koonin<sup>4</sup>, Takuro Nunoura<sup>5,✉</sup>, Christian Rinke<sup>3,✉</sup>, Mart Krupovic<sup>1,✉</sup>

<sup>1</sup>Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, Paris, France

<sup>2</sup>Center of Life Science, Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>3</sup>Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, Australia

<sup>4</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

<sup>5</sup>Research Center for Bioscience and Nanoscience, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Japan

<sup>6</sup>Present address: Institut Pasteur, Université Paris Cité, CNRS UMR6047, Evolutionary Biology of the Microbial Cell Unit, Paris, France

### Abstract

Asgardarchaeota harbour many eukaryotic signature proteins and are widely considered to represent the closest archaeal relatives of eukaryotes. Whether similarities between Asgard archaea and eukaryotes extend to their viromes remains unknown. Here we present 20 metagenome-assembled genomes of Asgardarchaeota from deep-sea sediments of the basin off the Shimokita Peninsula, Japan. By combining a CRISPR spacer search of metagenomic sequences with phylogenomic analysis, we identify three family-level groups of viruses associated with Asgard archaea. The first group, verdandiviruses, includes tailed viruses of the class *Caudoviricetes* (realm *Duplodnaviria*); the second, skuldviruses, consists of viruses with predicted

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**✉Correspondence and requests for materials** should be addressed to Takuro Nunoura, [takuron@jamstec.go.jp](mailto:takuron@jamstec.go.jp), Christian Rinke, [c.rinke@uq.edu.au](mailto:c.rinke@uq.edu.au) or Mart Krupovic, [mart.krupovic@pasteur.fr](mailto:mart.krupovic@pasteur.fr).

**Author contributions**

M.K. initiated the project and designed research. T.N. collected samples and extracted and sequenced DNA. J.S. and C.R. assembled, curated and analysed Asgard archaeal MAGs. S.M. assembled the Asgard archaeal CRISPR and viral datasets. S.M., N.Y., E.V.K. and M.K. analysed viral sequences. S.M. and M.K. wrote the manuscript with input from all authors.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Code availability**

No custom code was used.

**Competing interests**

The authors declare no competing interests.

**Extended data** is available for this paper at <https://doi.org/10.1038/s41564-022-01144-6>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41564-022-01144-6>.

icosahedral capsids of the realm *Varidnaviria*; and the third group, wyrdviruses, is related to spindle-shaped viruses previously identified in other archaea. More than 90% of the proteins encoded by these viruses of Asgard archaea show no sequence similarity to proteins encoded by other known viruses. Nevertheless, all three proposed families consist of viruses typical of prokaryotes, providing no indication of specific evolutionary relationships between viruses infecting Asgard archaea and eukaryotes. Verdandiviruses and skuldviruses are likely to be lytic, whereas wyrdviruses potentially establish chronic infection and are released without host cell lysis. All three groups of viruses are predicted to play important roles in controlling Asgard archaea populations in deep-sea ecosystems.

---

Asgard archaea are an expansive group of metabolically versatile archaea that thrive primarily in anoxic sediments around the globe<sup>1-9</sup>. Based on phylogenomic analyses, Asgard archaea were originally classified into multiple phylum-level lineages, including Lokiarchaeota, Thorarchaeota, Odinararchaeota, Heimdallarchaeota, Helarchaeota, Sifarchaeota, Wukongarchaeota and several others, most of which were named after Norse gods<sup>1,5,7,9-12</sup>. Recently, taxonomic rank normalization using relative evolutionary divergence has suggested that Asgard archaea represent a phylum, tentatively named Asgardarchaeota, including the classes Lokiarchaeia, Thorarchaeia, Odinararchaeia, Heimdallarchaeia, Sifarchaeia, Hermodarchaeia, Sifarchaeia, Baldrarchaeia, Wukongarchaeia and Jordarchaeia, with the other lineages classified as lower-rank taxa within the classes<sup>8,13</sup>. The vast majority of Asgard archaea have been discovered through metagenomics, whereas only one species has been isolated and successfully grown in the laboratory<sup>14</sup>. Asgard archaea gained prominence due to their inferred key role in the origin of eukaryotes<sup>15</sup>. Indeed, Heimdallarchaeia form a sister group to eukaryotes in most phylogenetic analyses<sup>1,5</sup> although alternative phylogenies have also been presented<sup>16,17</sup>. Compared with other archaea, Asgard archaea encode a substantially expanded set of eukaryotic signature proteins, including many proteins implicated in membrane trafficking, vesicle formation and transport, cytoskeleton formation, the ubiquitin network and other processes characteristic of eukaryotes<sup>1,5</sup>. A tantalizing question is whether eukaryotes also inherited viruses and other types of mobile genetic elements (MGEs) from the Asgard archaea.

Viruses infecting archaea are remarkably diverse, in terms of both their genome sequences and virion structures<sup>18-20</sup>. Some archaeal viruses, in particular those with icosahedral virions, are evolutionarily related to bacterial and eukaryotic viruses but the majority of archaeal virus groups are specific to archaea, with no identifiable relatives in the two other domains. Archaea-specific viruses often have odd-shaped virions that resemble lemons, champagne bottles or droplets<sup>19</sup>. Most archaeal viruses have, thus far, been isolated from hyperthermophilic or halophilic hosts, with only a handful of virus species described for methanogenic and ammonia-oxidizing mesophilic archaea<sup>18</sup>. No viruses infecting Asgard archaea have been isolated, primarily due to the inherent difficulty in propagation of Asgard archaeal hosts. Nevertheless, analysis of CRISPR-Cas loci in the genomes of Asgard archaea revealed a remarkable diversity of defence systems in these organisms<sup>21</sup>, implying a rich Asgard archaeal virome. CRISPR arrays are archives of past encounters with viruses and other MGEs, which can be harnessed to uncover the associations between viruses and

their hosts. Indeed, matching the CRISPR spacers from a known organism to viruses with unknown hosts is widely used in host assignment for viruses discovered by metagenomics and, arguably, is the most straightforward and efficient approach to identification of the hosts of viruses infecting prokaryotes<sup>22</sup>.

Here we harness CRISPR spacer sequences from the sequenced Asgard archaeal genomes to search for viruses infecting these organisms, and describe three distinct family-level groups of Asgard-associated viruses, all of which display typical features of viruses infecting bacteria or archaea.

## Results

To search for viruses infecting Asgard archaea, we sequenced 12 metagenomes constructed from total environmental DNA directly extracted from subseafloor sediments originating off the Shimokita Peninsula, Japan (site C9001, water depth 1,180 m)<sup>23</sup> and representing different sediment depths ranging 0.91–363.3 m below the seafloor (mbsf). Asgard archaeal metagenome-assembled genomes (MAGs) were assembled from seven of the 12 samples. A total of 20 Asgardarchaeota MAGs, including 12 Lokiarchaeia, two Thorarchaeia and six Heimdallarchaeia, were analysed in this study (Fig. 1), with an estimated completeness of  $65.95 \pm 17.93$ , estimated quality of  $54.68 \pm 15.78$  and guanine-cytosine content of  $32.78\% \pm 3.99\%$  (Supplementary Table 1).

To assign putative viral genomes to Asgard archaeal hosts, we compiled a dataset of CRISPR spacers from Asgard archaeal MAGs assembled in this study as well as those reported previously (Fig. 2 and Supplementary Table 2). Analysis of Asgard CRISPR arrays allowed us to define Asgard-specific CRISPR repeat sequences, which were used to identify additional CRISPR arrays from the contigs obtained from our sediment samples (Fig. 2a and Methods). In total, our CRISPR spacer dataset (Fig. 2b and Supplementary Data 1) included 2,532 spacers assigned to different Asgardarchaeota lineages, with Lokiarchaeia contributing the highest number (Fig. 2c). All spacers were then used to search for protospacers in the 20 assembled MAGs and putative virus genomes from our dataset, as well as contigs from GenBank and JGI sequence databases. In total, 14 contigs could be assigned to Asgard archaeal hosts based on CRISPR spacer matches with an estimated false positive rate of 0.00276 (Methods and Extended Data Fig. 1). By contrast, none of the contigs was targeted by spacers from the CRISPRHost database (675,911 spacers) and only two, probably false positive, spacers were identified in the CRISPR Spacer Database and Exploration Tool (11,674,395 spacers; Methods).

Eight of the putative viral genomes originated from the metagenomes sequenced in this study (from depths ranging 0.91–87.7 mbsf) whereas six additional genomes were recovered from datasets sequenced previously, including anoxic subseafloor sediment samples from two Pacific Ocean sites (the Hikurangi Subduction Margin (144.3 mbsf)<sup>8</sup> and Cascadia Margin (2 mbsf))<sup>24</sup> and one Indian Ocean site (Sumatra Forearc (1.60–6.07 mbsf; [PRJNA367446](#))). Each genome was targeted (90% identity between spacer and protospacer) by one to six CRISPR spacers assigned to putative Asgard classes Lokiarchaeia, Thorarchaeia and Heimdallarchaeia (Fig. 1 and Supplementary Tables 3 and

4). Based on the conservation of viral hallmark proteins, including major capsid proteins (MCPs), seven of the genomes were unequivocally identified as belonging to viruses from three unrelated groups (Figs. 3–5) whereas the remaining seven could represent either unknown viruses or other types of MGE (Supplementary Fig. 1 and below). The viral dataset was further enriched by searching the JGI and GenBank sequence databases for related contigs. As a result, we collected 21 contigs representing three groups of Asgard archaeal viruses originating from seven geographically remote locations (Fig. 6a). Further analyses using CRISPR-based (SpacePHARER) and kmer-based (WiSH and PHIST) methods confirmed the association of these 21 contigs with Asgard archaeal hosts (Methods and Supplementary Tables 3 and 5). Network analysis of these viral genomes, together with the bacterial and archaeal virus genomes available in RefSeq, showed that the three Asgard virus groups are disconnected from each other as well as from other known viruses (Fig. 6b).

### **Verdandiviruses are tailed viruses of the *Caudoviricetes*.**

Three of the genomes, VerdaV1, VerdaV2 and VerdaV3, assembled from the Shimokita dataset, encode the hallmark proteins specific to viruses of the class *Caudoviricetes*, the most widespread, environmentally abundant and genetically diverse group of viruses<sup>25,26</sup>. Members of the *Caudoviricetes* infect bacteria and archaea and have characteristic virions consisting of an icosahedral capsid and a helical tail attached to one of the capsid vertices. *Caudoviricetes*, together with eukaryotic herpesviruses, form the realm *Duplodnaviria*<sup>27</sup>. Similar to previously characterized bacterial and archaeal *Caudoviricetes*<sup>18</sup>, each of the three virus genomes encodes the HK97-like MCP, a large subunit of the terminase (genome-packaging ATPase-nuclease), the portal protein and several other structural proteins, including tail components (Fig. 3a). Structural modelling of the VerdaV1 MCP using RoseTTAFold<sup>28</sup> yielded a model with the canonical HK97-like fold (Fig. 3b).

The VerdaV1 (19.9 kb) and VerdaV2 (19.5 kb) genomes were assembled as circular contigs (Supplementary Table 3), suggesting that they correspond to complete, terminally redundant virus genomes, whereas VerdaV3 (15.4 kb) probably represents a partial genome. Comparative genomics analysis showed that VerdaV1, VerdaV2 and VerdaV3 belong to the same virus group (Fig. 3a), which we refer to as ‘verdandiviruses’ (for Verdandi, one of the three Norns, the most powerful beings in Norse mythology that govern the lives of gods and mortals). Given that the three genomes were identified in different sediment depths offshore of Shimokita Peninsula (at 18.7, 59.5 and 87.7 mbsf; Supplementary Table 3), we addressed the possibility that related viruses could also be detected in samples from other depths. To this end, we performed BLASTP searches (*E*-value cutoff of  $1 \times 10^{-5}$ ) queried with the corresponding MCP sequences against the assembled sequences from samples retrieved along the depth gradient. The analysis yielded six additional viral contigs and showed that related MCPs (and viruses) are also present in sediment samples from 0.9, 9.3 and 30.8 mbsf, indicating a broad distribution of verdandiviruses through the sediment column. One of the retrieved contigs (VerdaV4, 20.6 kb) was found to be circular and thus also represents a complete virus genome. Furthermore, searches against the GenBank and JGI sequence databases yielded eight hits to virus-like contigs affiliated to Asgard archaea (Supplementary Table 3). Five of the MCPs were encoded within large contigs (>10 kb), including two

(Ga0114925\_10000341 and Ga0114923\_10001063 (VerdaV5 and VerdaV6, respectively)) circular examples. Notably, VerdaV5 was also targeted by a CRISPR spacer from a recently described Asgardarchaeota MAG<sup>1</sup>, further supporting the host assignment for the Verdandivirus group (Fig. 3a and Supplementary Table 3). Finally, we identified a partial provirus related to verdandiviruses integrated in a large genomic contig of Lokiarchaeia (Extended Data Fig. 3). The same contig contained cellular genes unequivocally belonging to Asgardarchaeota (98–99% protein identity), including those encoding ribosomal proteins. Verdandivirus contigs were recovered from four geographically remote sampling sites (Fig. 6a), and maximum-likelihood phylogenetic analysis of the MCPs revealed an overall biogeographic clustering of verdandiviruses (Fig. 3c).

Although virus-encoded homologues could be identified for only 2% (seven out of 298) of verdandivirus proteins by BLASTP searches ( $E$  value =  $1 \times 10^{-5}$ ), functional annotation for some of the other proteins was enabled by sensitive profile–profile comparisons in which HK97-like MCP and a large subunit of the terminase (TerL) MCPs were readily identified with highly significant scores (Supplementary Fig. 2 and Supplementary Table 5). Verdandiviruses display considerable sequence conservation within the capsid formation and genome-packaging modules (Fig. 3a). Notably, upstream of the MCP all viruses carry a gene encoding a putative capsid stabilization protein distantly related to the corresponding protein of marine siphoviruses—for example, TW1 (ref. <sup>29</sup>)—which might be important for maintaining capsid integrity under the high hydrostatic pressures of deep-sea ecosystems. Verdandiviruses encode tail proteins most closely similar to those of siphoviruses, including the major tail protein, tail tape measure protein and a baseplate hub protein with an adhesin domain, suggesting that verdandiviruses possess flexible, non-contractile tails. In most of these viruses, the tape measure protein is relatively short (median length 259 amino acids (aa)), suggesting that the tails themselves are also short. Assuming  $\sim 1.5 \text{ \AA}$  of tail length per amino acid residue of tape measure protein<sup>30,31</sup>, the median tail length of verdandiviruses is predicted to be  $\sim 39 \text{ nm}$ . The genes encoding tail proteins display low sequence conservation, whereas gene contents downstream of the tail modules are distinct in most of the identified viruses (Fig. 3a). Differences in tail modules might correspond to different host ranges of the corresponding viruses. Indeed, whereas VerdaV1 and VerdaV5 were matched by CRISPR spacers from Lokiarchaeia, VerdaV2 and VerdaV3 were predicted to infect Thorarchaeia (Supplementary Table 3). The distinct conservation patterns of the capsid and tail modules probably reflect modular evolution of these Asgard archaeal virus genomes, a common trait in bacterial and archaeal members of the *Caudoviricetes*<sup>32,33</sup>.

Verdandiviruses do not encode any proteins implicated in virus genome replication, and thus can be predicted to be fully reliant on the host for this process. Complete dependence on the host replication machinery is common among archaeal viruses with small and mid-sized genomes<sup>34</sup>, in particular, for the tailed viruses of haloarchaea and methanogens in the families *Saparoviridae*, *Suolaviridae*, *Leisingerviridae* and *Anaerodiviridae*<sup>35</sup>. Verdandiviruses encode predicted DNA-binding proteins with Zn-finger and helix-turn-helix (HTH) motifs, which could participate in recruitment of host replication and transcription machineries. Indeed, in the case of hyperthermophilic archaeal virus SIRV2, a viral HTH protein has been shown to recruit the host DNA sliding clamp protein, a known interaction partner for many other components of the host replisome<sup>36</sup>.

All complete verdandivirus genomes encompass arrays of short genes encoding predicted small, poorly conserved proteins, some of which are adjacent to genes encoding predicted transcriptional regulators (Fig. 3a). Similar to many bacterial viruses of the *Caudoviricetes*<sup>37,38</sup>, some of these small, fast-evolving genes could encode antidefense—in particular, anti-CRISPR proteins.

### Skuldviruses: tailless icosahedral viruses of Asgard archaea.

One of the contigs from Asgard archaeal MAGs assembled from anoxic sediments from the Hikurangi Subduction Margin<sup>8</sup>, SkuldV1, targeted by two Asgard archaeal spacers (Supplementary Table 3), encodes a double jelly-roll (DJR) MCP (Fig. 4a), a hallmark of viruses of the realm *Varidnaviria*, an expansive assemblage of viruses evolutionarily and structurally unrelated to duplodnaviruses<sup>27</sup>. Varidnaviruses are environmentally abundant and infect hosts from all domains of life<sup>27,39</sup>. Sequence searches with the SkuldV1 MCP led to the identification of two additional contigs, one from our dataset (SkuldV2) and the other from the GenBank database ([JABUBK010000319](https://www.ncbi.nlm.nih.gov/nuclseq/JABUBK010000319), hereinafter referred to as SkuldV3). The latter contig was obtained from subseafloor sediments from the Cascadia Margin (Ocean Drilling Programme, site 1244) in the Pacific Ocean<sup>24</sup> and is targeted by four spacers from our dataset. SkuldV1 and SkuldV3 were assembled as circular contigs and thus correspond to complete virus genomes (Fig. 4a). In addition to the DJR MCP, the vast majority of varidnaviruses encode genome-packaging ATPases of the FtsK-HerA superfamily<sup>40</sup>. A homologue of such ATPase was identified in all three SkuldV genomes, indicating that they are bona fide members of the realm *Varidnaviria*. We propose referring to this group of viruses as ‘Skuldviruses’ (for Skuld, another of the Norns).

The vast majority of skuldvirus proteins (97%, 66 of the 68 proteins) show no similarity to proteins encoded by other known viruses. Network analysis using CLANS<sup>41</sup> showed that the skuldvirus MCPs form a cluster separate from the previously characterized<sup>42</sup> groups of DJR MCPs (Fig. 4b). Nevertheless, profile–profile comparisons showed that they are most closely related to the corresponding proteins of prokaryotic viruses of the families *Corticoviridae* (bacteriophage PM2: HHsearch probability of 98.3), *Turriviridae* (archaeal virus STIV: HHsearch probability of 97.8) and *Tectiviridae* (bacteriophage PRD1: HHsearch probability of 96.2) (Supplementary Fig. 3), whereas eukaryotic viruses with DJR MCPs were recovered with considerably lower scores (*Phycodnaviridae*, *Pyramimonas orientalis* virus: HHsearch probability of 83.4). Structural comparison of the SkuldV1 MCP model obtained using RoseTTAFold<sup>28</sup> (Fig. 4b) further showed that it is most similar to the MCP of *Pseudoalteromonas* phage PM2 (ref. <sup>43</sup>), a prototype of the *Corticoviridae*<sup>44</sup> that is widespread in marine ecosystems<sup>45</sup>. Nevertheless, skuldviruses do not share genes with other known viruses other than those encoding the MCP and the genome-packaging ATPase. Corticoviruses, turriviruses and tectiviruses belong to the class *Tectiliviricetes*<sup>27</sup>. We propose that skuldviruses represent a separate virus family within the *Tectiliviricetes*.

Corticoviruses employ a rolling circle mechanism for genome replication and encode characteristic HUH superfamily endonucleases<sup>46</sup>. No such genes or other putative replication genes were identified in skuldviruses. Instead, all three skuldviruses encode a protein related to the A subunit of type IIB topoisomerases, such as topoisomerase VI. The

latter enzyme consists of two distinct subunits, A and B, with the catalytic tyrosine residue responsible for DNA nicking located in the A subunit. Standalone A subunits, dubbed Topo mini-A, have recently been discovered in diverse bacterial and archaeal MGEs<sup>47</sup> but the functions of these proteins remain unknown. In the maximum-likelihood phylogeny of Topo mini-A homologues, skuldvirus proteins form a clade with homologues from methanogenic and ammonia-oxidizing archaea (Extended Data Fig. 4). Skuldviral Topo mini-A might function as a replication protein, possibly initiating the rolling circle replication of the circular skuldvirus genomes in a manner analogous to the HUH endonuclease.

### **Wyrdiviruses are related to spindle-shaped archaeal viruses.**

One of the contigs from Asgard archaeal MAGs assembled from the Hikurangi Subduction Margin<sup>8</sup>, WyrdV1 (15,570 base pairs (bp)), targeted by one CRISPR spacer from our collection, was found to encode two homologues of the MCPs specific to spindle-shaped archaeal viruses<sup>48</sup>. In particular, the closest homologue was found in haloarchaeal virus His1 (family *Halspiviridae*; Extended Data Fig. 5)<sup>49</sup>. His1-like MCPs are ~80 aa in length and contain two hydrophobic segments predicted to be membrane-spanning domains<sup>48</sup> and in all spindle-shaped viruses playing key roles in virion structure and assembly<sup>50</sup>. BLASTP searches using the WyrdV1 MCP as a query against the JGI and GenBank databases identified eight additional contigs (Fig. 5). All these contigs shared several genes encoding a morphogenetic module, including the MCP and receptor-binding adhesin, which in fuselloviruses and halspiviruses is located at one of the pointed ends of the virion<sup>51,52</sup>. In addition, all these viruses encode a AAA<sup>+</sup> ATPase which, in profile-profile comparisons, showed the highest similarity to the morphogenesis (pI) proteins of bacterial filamentous phages (order *Tubulavirales*)<sup>53</sup>. This ATPase is responsible for extrusion of the viral genome through the cellular membrane during virion assembly<sup>54</sup>. We propose referring to this group of viruses as ‘wyrdiviruses’ (for Wyrd (Urðr), the third Norn). The homology between the tubulaviral and wyrdiviral ATPases suggests that virion assembly of wyrdiviruses is mechanistically similar to the extrusion of filamentous bacteriophages. Indeed, spindle-shaped viruses infecting other archaea are released from the host without causing cell lysis through a budding-like mechanism<sup>50,55,56</sup>. Notably, some spindle-shaped viruses encode a single MCP (for example, halspiviruses) whereas others encode two paralogous MCP (for example, fuselloviruses). Similarly, wyrdiviruses encode either one or two MCP paralogs (Fig. 5).

Similar to halspiviruses and thaspiviruses, two of the contigs—Ga0209633\_10003833 and Ga0209976\_10001089, which we refer to as WyrdV2 and WyrdV3, respectively—contained terminal inverted repeats, indicating that these are (nearly) complete genomes. By contrast, contigs JABUBK010000290 and JABUBK010000290, herein referred to as WyrdV4 and WyrdV5, respectively, contained direct terminal repeats, indicative of the completeness and circular structure of the genomes, resembling fuselloviruses. Consistent with the different genome structures, WyrdV2 and WyrdV3 (and halspiviruses) encode protein-primed family B DNA polymerases whereas WyrdV4 and WyrdV5 encode rolling circle replication initiation endonucleases of the HUH superfamily (Fig. 5). Notably WyrdV4, in addition, encodes Topo mini-A. The latter does not cluster with homologues from skuldviruses, but instead forms a clade with the Topo mini-A from the unassigned Asgard archaeal

MGEs 10H\_0 and 7H\_42 (Extended Data Fig. 3). Contigs Ga0209976\_10001631 (WyrdV6) and Ga0209976\_10001236 (WyrdV7) also encode DNA polymerases and are probably linear, nearly complete viral genomes (Fig. 5). Remarkably, the DNA polymerase encoded by WyrdV7 is not orthologous to those of WyrdV2, WyrdV3 and WyrdV6 (Fig. 5). The latter group is most closely related (~30% identity) to the corresponding protein of an uncultured virus (MW522971) associated with *Altiarchaeota*<sup>57</sup>, whereas the former protein is most similar to the DNA polymerase encoded by the spindle-shaped virus infecting marine *Nitrososphaeria*<sup>55</sup> (Supplementary Table 3). Ga0209977\_10002196 and Ga0209976\_10004438 are partial genomes that lack the region encompassing the replication modules. Notably, WyrdV1 lacks either the polymerase or the rolling circle endonuclease gene and, instead, at the equivalent locus contains an array of short genes, some of which might encode replication initiators. Similar dramatic variation in gene content has previously been observed only in haloarchaeal viruses of the family *Pleolipoviridae*, where members of three genera encode non-homologous genome replication proteins<sup>58</sup>. Our present observations further illuminate the remarkable plasticity of the genome replication modules and genome structures in relatively closely related archaeal viruses, in general, and in wyrdviruses in particular.

### Enigmatic MGEs and auxiliary genes of Asgard archaeal viruses.

In addition to the three groups of viruses, we identified seven other contigs (7H\_11, 8H\_18, 8H\_67, 10H\_0, 7H\_42, Ga0114923\_10000127 and Ga0209976\_10000148) targeted by Asgard archaeal CRISPR spacers (Supplementary Table 3). Four of these contigs, 7H\_11, 8H\_18, 10H\_0 and Ga0114923\_10000127, were assembled as circular molecules and probably represent complete MGE genomes of 8,776, 8,776, 84,544 and 58,806 bp, respectively, whereas Ga0209976\_10000148 (48,997 bp), 7H\_42 (44,162 bp) and 8H\_67 (13,282 bp) appeared to be partial. The seven contigs do not encode identifiable homologues of MCPs of known viruses and are likely to represent either unknown viruses or non-viral MGEs, such as plasmids (see Supplementary text for description of Asgard archaeal MGEs). Notably, these MGEs appear to be unrelated to those reported recently in *Heimdallarchaeia*<sup>59</sup>.

Some Asgard archaeal MGEs and viruses encode auxiliary functions, including metabolic genes, such as phosphoadenosine phosphosulfate (PAPS) reductase, which could facilitate sulfur metabolism and/or synthesis of sulfur-containing amino acids<sup>60</sup>; and enzymes involved in nucleotide metabolism, including dUTPase, thymidylate synthase X and nucleoside pyrophosphohydrolase MazG, which might function in disarming antiviral systems triggered by nucleotide-based alarmones, such as ppGpp (Supplementary text).

## Discussion

Here we describe three previously undetected, distinct groups of viruses associated with Asgard archaea of the lineages *Lokiarchaeia* and *Thorarchaeia*. Each group was identified in marine sediment samples from geographically remote sites (Fig. 6), suggesting a wide distribution of these viruses in Asgard archaea-inhabited ecosystems. In addition, we recovered seven CRISPR-targeted MGEs associated with *Lokiarchaeia*, *Thorarchaeia* and

Heimdallarchaea that might represent distinct viruses with structural and morphogenetic proteins unrelated to those of any known viruses or, perhaps more likely, plasmids. Although it is clear that verdandiviruses, skuldviruses and wyrdviruses do not comprise the complete Asgard virome, they provide important insights into the diversity and evolution of the Asgard viruses. All three virus groups are sufficiently distinct from previously characterized viruses to be considered as founding representatives of three previously undescribed families. Wyrdviruses appear to be evolutionarily related to spindle-shaped viruses and thus belong to one of the archaea-specific groups of viruses<sup>19</sup> not known in bacteria or eukaryotes. In archaea, spindle-shaped viruses are widely distributed and infect hyperthermophilic, halophilic and ammonia-oxidizing hosts from different phyla<sup>48,55</sup>. Thus, wyrdviruses further expand the reach of spindle-shaped viruses to Asgard archaea, supporting the notion that this group of viruses was associated with the last archaeal common ancestor (LACA)<sup>61</sup>. By contrast, verdandiviruses and skuldviruses encode HK97-like and DJR MCPs, respectively and, accordingly, at the highest taxonomic level belong to the realms *Duplodnaviria* and *Varidnaviria*. Viruses of both realms have deep evolutionary origins and were proposed to have been present in both LACA and the last universal cellular ancestor<sup>61</sup>. The current work further supports this inference. Although members of both *Duplodnaviria* and *Varidnaviria* also infect eukaryotes, analyses of the verdandivirus and skuldvirus genome and protein sequences unequivocally show that they are more closely, even if distantly, related to their respective prokaryotic relatives. Thus, no putative direct ancestors of eukaryotic viruses were detected. Further exploration of the Asgard archaeal virome is needed to determine whether any of the virus groups associated with extant eukaryotes originate from Asgard viruses.

All prokaryotic members of the realms *Duplodnaviria* and *Varidnaviria* are lytic viruses, which are released from the host cells by lysis (although some alternate lysis with lysogeny)<sup>62–64</sup>. Thus, verdandiviruses and skuldviruses are also likely to kill their hosts at the end of the infection cycle, thereby promoting the turnover of Asgard archaea and nutrient cycling in deep-sea ecosystems. This possibility is consistent with previous results showing that viruses in deep-sea sediments lyse archaea more rapidly compared with bacteria<sup>65</sup>. By contrast, the mechanism of virion release employed by spindle-shaped viruses does not involve cell lysis, with virions continuously released from chronically infected cells<sup>55,56</sup>. Infection of marine Nitrososphaeria with spindle-shaped virus NSV1 resulted in inhibition of host growth and was accompanied by severe reduction in the rate of ammonia oxidation and nitrite reduction<sup>55</sup>. Thus, infection dynamics and the impact of wyrdviruses are likely to be quite different from those of verdandiviruses and skuldviruses. Finally, some Asgard archaeal viruses carry auxiliary metabolic genes, such as the gene encoding PAPS reductase, which might boost the metabolism of infected cells.

The present work is accompanied by two other studies, by Tamarit et al.<sup>66</sup> and Rambo et al.<sup>67</sup>, respectively, describing the identification of other Asgard archaeal viruses. Rambo et al. describe several distinct groups of viruses, all members of the class *Caudoviricetes*. The complete genomes of these viruses are considerably larger than those of verdandiviruses described here and have distinct gene contents, justifying their placements into separate families. The study by Tamarit et al. describes two viruses, Huginnvirus and Muninnvirus, related to icosahedral and spindle-shaped viruses, respectively. Huginnvirus is only distantly

related to skuldviruses described here, with the two groups of viruses forming disconnected clusters in the MCP network (Fig. 4b). Muninnvirus is distantly related to wyrdviruses associated with Lokiarchaeia from our study, but infects a host from the Asgard archaeal class Odinarchaeia. Thus, viruses described in the three studies complement each other by representing distinct virus groups. Collectively, these findings provide valuable insights into the virome of Asgard archaea and open the door for understanding virus–host interactions in deep-sea ecosystems. Undoubtedly, many more Asgard archaeal virus groups remain to be discovered, which should clarify the contribution of these viruses to the evolution of the extant eukaryotic virome.

## Methods

### Site description and sampling.

A total of 365 m of sediment cores were recovered from Hole C9001 C of Site C9001 (water depth 1,180 m) located at the forearc basin off the Shimokita Peninsula, Japan (41° 10.638' N, 142° 12.08' E) by the drilling vessel *Chikyu* during the JAMSTEC CK06-06 cruise in 2006. Coring procedure, subsampling for molecular analyses, profiles of lithology, age model, porewater inorganic chemistry, organic chemistry and cell abundance, and summary of the molecular microbiology in sediments at site C9001 were reported previously<sup>23,68</sup> (and references therein). Subsampled whole round cores (WRCs) for microbiology were stored at –80 °C.

### Sample descriptions.

A total of 12 sediment samples at depths of 0.9, 9.3, 18.5, 30.8, 48.3, 59.5, 68.8, 87.7, 116.4, 154.3, 254.7 and 363.3 mbsf were used in this study. These representatives were chosen based on porewater chemical profiles, lithostratigraphic properties, molecular biology data on core sediments and F430 contents as follows. Note that ten samples—at depths of 0.9, 9.3, 18.5, 30.8, 48.3, 59.5, 116.4, 154.3, 254.3 and 363.3 mbsf—were previously analysed by small subunit ribosomal RNA gene tag sequencing with a 454 FLX Titanium sequencer, and details are reported in ref. <sup>23</sup>.

At 0.9 mbsf, the uppermost section of the core column was chosen; at 9.3 mbsf, the region just beneath the sulfate–methane transition zone was chosen; at 18.5 mbsf, the highest relative abundance of Lokiarchaeia was found in the previous SSU rRNA gene tag sequencing; at 48.3 mbsf, relatively higher abundance of the South African gold mine miscellaneous euryarchaeal group in the archaeal community was observed; at 59.5 mbsf, the predominance of a lineage in Nanoarchaeia (formerly Woesearchaeota) was observed. For sediment, at a depth of 68.8 mbsf the sample harboured anomalously high F430 concentrations, and thus the sample from a depth of 87.7 mbsf was used as reference. Sediment at 116.4 mbsf consisted of ash/pumice whereas other samples used in this study were pelagic clay sediments, and a predominance of Bathyarchaeia was observed. Samples from 154.3, 254.3 and 363.3 mbsf were selected in 100-m depth intervals from the bottom of this drilling hole<sup>23,68</sup>.

### Shotgun metagenomics and SSU rRNA gene tag sequencing.

DNA extraction and shotgun metagenome library construction are described in ref. <sup>69</sup>. Briefly, total environmental DNA was extracted from approximately 5 g of sediment subsampled from the inner part of WRCs using the DNeasy PowerMax Soil Kit (Qiagen), with one minor modification<sup>70</sup>. Next, further purification was performed with NucleoSpin gDNA Clean-up (MACHERY-NAGEL). Purified DNA was used for Illumina sequencing library construction using a KAPA Hyper Prep Kit (for Illumina) (KAPA Biosystems). Sequencing was performed on an Illumina HiSeq 2500 platform (San Diego), and 250-bp paired-end reads were generated.

### Assembly, binning and Asgardarchaeota phylogeny.

Contigs from each sample were assembled using MetaSpades v.0.7.12-r1039 (ref. <sup>71</sup>) and binned with UniteM (<https://github.com/dparks1134/unitem>). MAG sets were generated after dereplication of bins using DAS\_Tool v.1.1.2 (ref. <sup>72</sup>) and quality check with CheckM<sup>73</sup>. MAGs with an estimated quality (completeness – fourfold contamination) <30% were excluded in downstream analysis. Taxonomic assignments of all MAGs were performed using the ‘classify’ function in GTDBtk<sup>74</sup>.

Phylogenetic analysis included 20 Asgardarchaeota MAGs, together with 255 published Asgardarchaeota and 64 non-Asgardarchaeota archaeal genomes. A marker set consisting of 53 ribosomal proteins of the 239 genomes was identified and independently aligned using the ‘identify’ and ‘align’ functions in GTDBtk<sup>74</sup>. Individual multiple sequence alignments were then concatenated and trimmed in GTDBtk using ‘trim\_msa’ (–min\_perc\_aa 0.4). Maximum-likelihood phylogenies of Asgardarchaeota MAGs were initially estimated by FastTree v.2.1.11 (ref. <sup>75</sup>) with default settings, and subsequently inferred using IQ-Tree v.1.6.984 (ref. <sup>76</sup>) under the LG + C10 + F + G + PMSF model with 100 bootstraps. The final consensus tree was visualized and beautified in iTOL<sup>77</sup>.

### Assembly and analysis of viral contigs.

Potential viral and MGE contigs were assembled from metagenomic reads by the MetaViralSPAdes pipeline, with default parameters<sup>78</sup>. Direct and inverted terminal repeats were identified by BLASTN<sup>79</sup>. Open reading frames were identified with Prokka<sup>80</sup> and annotated using HHsearch<sup>81</sup> against Pfam, PDB, SCOPE, CDD and viral protein sequence databases. Minimum information about uncultured virus genomes<sup>82</sup> assembled in the course of this work is provided in Supplementary Table 7. Read depth, along with the assembled viral genomes, is shown in Supplementary Fig. 4. To identify similar contigs in public databases, we searched sequences of MCPs in GenBank and JGI databases (BLASTP, *E*-value cutoff  $1 \times 10^{-5}$ , 70% query coverage). Genomes of assembled viral genomes were compared and visualized using EasyFig<sup>83</sup> with the tBLASTx option. Network analysis of viral genomes was performed using vConTACT v.2.0.9.19 with default parameters against the Viral RefSeq v.201 reference database<sup>84</sup>. The virus network was visualized with Cytoscape<sup>85</sup>.

### Collection of the CRISPR spacer dataset.

CRISPR arrays were detected in metagenomic contigs and 255 published Asgardarchaeota MAGs using minced v.0.4.2 (ref. <sup>86</sup>) and CRISPRDetect<sup>87</sup>. CRISPR arrays from metagenomic contigs were assigned to ‘Asgard’ and ‘non-Asgard’ groups based on CRISPR repeat similarity to previously characterized Asgard CRISPR repeats, with 90% identity cutoff over the full length of the repeat<sup>21</sup>. CRISPR repeat sequences from both metagenomic contigs and Asgard MAGs were then clustered together using an all-against-all BLASTN search (*E*-value cutoff  $1 \times 10^{-5}$ , 90% identity, word size 7), and the result of clustering was visualized in Cytoscape<sup>85</sup>. Consensus sequences for the nine major clusters of CRISPR repeats were constructed using the python package Logomaker<sup>88</sup>.

### Assignment of CRISPR arrays to Asgardarchaeota.

To assemble the CRISPR spacer database we relied on two categories of CRISPR loci: (1) ‘reference’ sequences of manually curated Asgard CRISPR loci from Makarova et al.<sup>21</sup> and CRISPR loci from Asgard archaeal MAGs; and (2) ‘metagenomic’ CRISPR loci from contigs originating from Shimokita Peninsula subseafloor sediments. CRISPR loci from the latter category were assigned to Asgard archaea if the identity of CRISPR repeat to the reference Asgard archaeal CRISPRs from the former category was >90% and protein sequences encoded on these CRISPR-containing contigs (when present) had the best BLASTP hit to Asgard archaeal proteins. All reference contigs were downloaded from GenBank without annotation. We used prodigal and CRISPR-Cas++ to identify *cas* genes in contigs with CRISPR arrays. Protein sequences were compared to the nr database with BLASTP to confirm Asgard assignment of the contig with CRISPR.

Verdandiviruses were targeted by spacers from five reference and four metagenomic CRISPR loci (Extended Data Fig. 1). The length of reference contigs varied from 3 to 126 kbp. Five reference contigs and one metagenomic contig contained *cas* genes, with the most conserved Cas protein (Cas1) being assigned to Asgard archaea (Lokiarachaeia, Helarchaeia) by the best BLASTP hit with 72–74% identity. Skuldviruses were targeted by spacers from three reference and three metagenomic CRISPR loci, whereas wyrdviruses were targeted by spacers from two reference and three metagenomic CRISPR loci (Extended Data Fig. 1). CRISPR repeats of metagenomic contigs targeting the three groups of viruses were 97% identical to those from the reference category<sup>21</sup>.

### Asgard archaeal CRISPR repeats.

Predicted Asgard archaeal CRISPR arrays contained distinct repeats that could be clustered into nine groups with 90% identity of sequences within groups (Fig. 2a). Comparison of Asgard archaeal CRISPR repeats with those in the public RepeatTyper database produced significant results only for group 9 repeat, which was identified as belonging to CRISPR-Cas type I-E. The closest repeat to group 9 was from *Thermobaculum terrenum*, with two mismatches (27/29). Repeats from group 4 partially matched to *Desulfosarcina* (29/37). Other repeat groups showed no matches in the CRISPR-Cas++ database (<https://crisprcas.i2bc.paris-saclay.fr/>).

### Structural modelling and network analysis of MCPs.

Structural modelling of the representative verdnavivirus and skuldvirus MCPs was performed with RoseTTAFold<sup>28</sup>. The MCPs of skuldviruses were compared with DJR MCPs of other known prokaryotic viruses using CLANS<sup>41</sup>, an implementation of the Fruchterman–Reingold force-directed layout algorithm that treats protein sequences as point masses in a virtual multidimensional space in which they attract or repel each other based on the strength of their pairwise similarities (CLANS *P* values). The reference dataset of DJR MCP was obtained from ref. <sup>42</sup>. Sequences were clustered using CLANS with BLASTP option (*E* value =  $1 \times 10^{-4}$ )<sup>41</sup>.

### Phylogenetic analysis of viral proteins.

Sequences were aligned using MAFFT in ‘Auto’ mode<sup>89</sup>. For phylogenetic analysis, uninformative positions were removed using TrimAl with a gap threshold of 0.2 for verdandiviral MCPs and the gappout option for Topo mini-A<sup>90</sup>. The final alignments contained 323 and 277 positions, respectively. Maximum-likelihood trees were inferred using IQ-TREE v.2 (ref. <sup>76</sup>). The best-fitting substitution models were selected by IQ-TREE, and were LG + G4 and LG + R5 for verdandiviral MCPs and Topo mini-A, respectively. The trees were visualized with iTOL<sup>77</sup>. Phylogenetic trees and underlying alignments in editable format are provided in Supplementary Data 2.

### CRISPR-based virus–host assignments.

Spacer sequences from Asgard archaeal CRISPR arrays were matched to metagenomic contigs and published Asgard MAGs (BLASTN, *E* value =  $1 \times 10^{-5}$ , 90% identity, word size 7), and to viral genomes available on the IGV database (default parameters) at <https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=WorkspaceBlast&page=viralform>. To assess the false positive rate of BLASTN we used a control set of viral contigs from human metagenomes (~400,000 sequences)<sup>91–94</sup>. Seven spacers out of 2,532 (false positive rate = 0.00276) were matched to a control dataset by BLASTN with the same parameters (*E* value =  $1 \times 10^{-5}$ , 90% identity, word size 7). For each spacer we calculated the difference between the identity of best match in the test and control datasets. Spacers with <10% identity difference (*n* = 5) were removed as potential low-complexity spacers. SpacePHARER<sup>95</sup>, with false discovery rate < 0.001, was used for a sensitive search of protospacers with lower nucleotide identity to support the BLASTN results. CRISPRHost and CRISPR Spacer Database and Exploration Tool<sup>96</sup> were used to match large collections of prokaryotic spacers to identified Asgard MGE sequences.

### Kmer-based virus–host assignments.

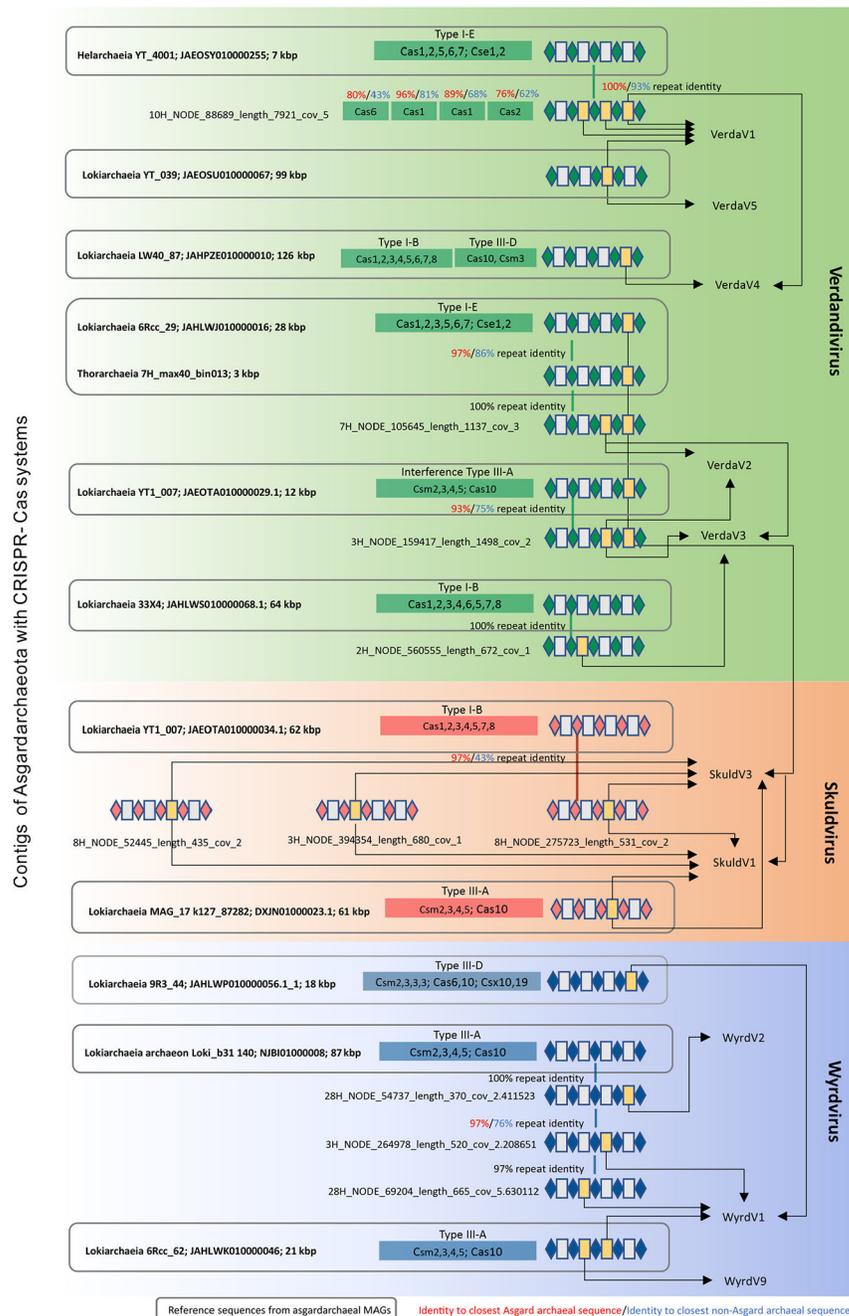
WIsH<sup>97</sup> and PHIST<sup>98</sup>, kmer-based methods that compare 8-mer (WIsH) and 25-mer (PHIST) nucleotide frequencies of viruses and potential hosts, were applied to validate prediction of CRISPR-based virus–host assignment. Unlike other kmer-based methods, PHIST and WIsH can predict virus–host interactions for user-provided host sequences rather than precomputed host databases. For the potential host dataset, we combined 195 MAGs of Asgardarchaeota, 267 non-Asgard MAGs reconstructed from Shimokita Peninsula metagenomes, 2,143 RefSeq archaeal genomes and 1,077 representative bacterial genomes

(Supplementary Table 8). For the virus dataset we used MGE sequences targeted by Asgard spacers. Viral contigs from human metagenomes were used as a negative dataset to calculate null parameters for models of hosts for WISH.

WISH confirmed Asgard archaeal hosts for all MGE sequences targeted by Asgard archaeal spacers, with  $P$  values ranging from  $1.18 \times 10^{-6}$  for WyrV1 to  $1.32 \times 10^{-1}$  for VerdaV2 (Supplementary Table 5). In the 25-mer-based PHIST analysis we considered only those predictions based on  $>10$  kmer. Using these parameters, hosts were predicted for 17 out of 28 (60%) contigs. Of these, 11 (65%) contigs, representing all groups of viruses and MGEs described in this work, were predicted as being associated with Asgard archaea. However, the remaining six, all verdandiviruses, were affiliated to Thermoplasmatota MAG 2H\_mb2\_bin16 assembled from the Shimokita Peninsula dataset. We note that the latter prediction probably resulted from the inclusion of a small (~2 kbp) verdandivirus-like contig in Thermoplasmatota MAG 2H\_mb2\_bin16, which is probably a binning artefact because the full-length verdandivirus contig was assigned to Lokiarachaea MAG. Indeed, we could not confirm Thermoplasmatota assignment as a verdandivirus host with CRISPR spacer analysis.

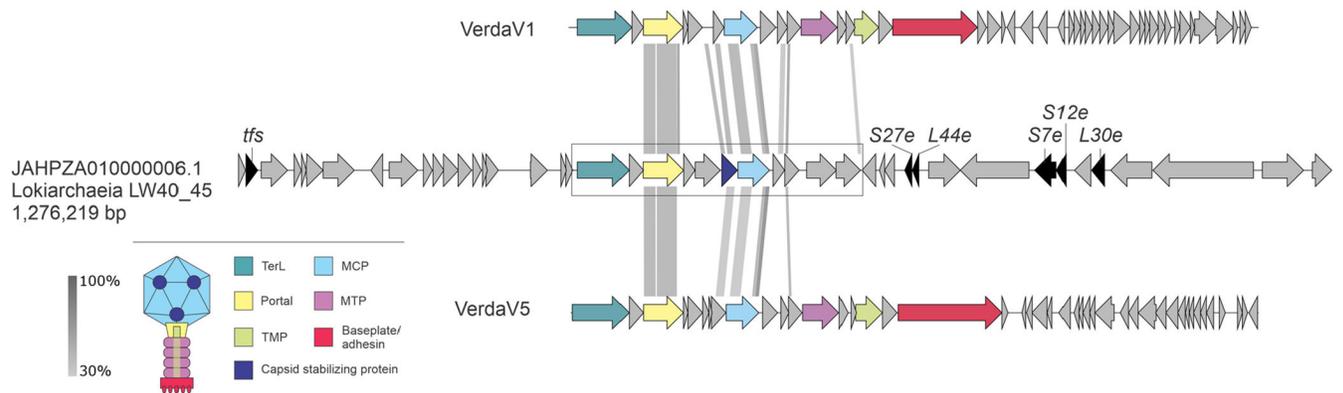
To further verify the validity of our CRISPR matches, we checked the predicted Asgard archaeal viruses and MGEs for the presence of non-Asgard protospacers using large spacer collections of CRISPRHost and CRISPR Spacer Database and Exploration Tool (CRISPR\_SDET). CRISPRHost did not find any protospacers with the default parameters. CRISPR\_SDET search was performed with 90% identity threshold. Two protospacers were found: verdandivirus VerdaV4 was assigned to *Floccifex porci*, a *Firmicutes* bacterium from the pig gut metagenome; two mobile elements from the 'Other' group, JGI-127 and JGI-148, were assigned to *Treponema* sp., a pathogenic bacterium of humans and other mammals. Given that neither of these bacteria resides in deep-sea sediments, the two spacers are likely to represent false positives.

Extended Data

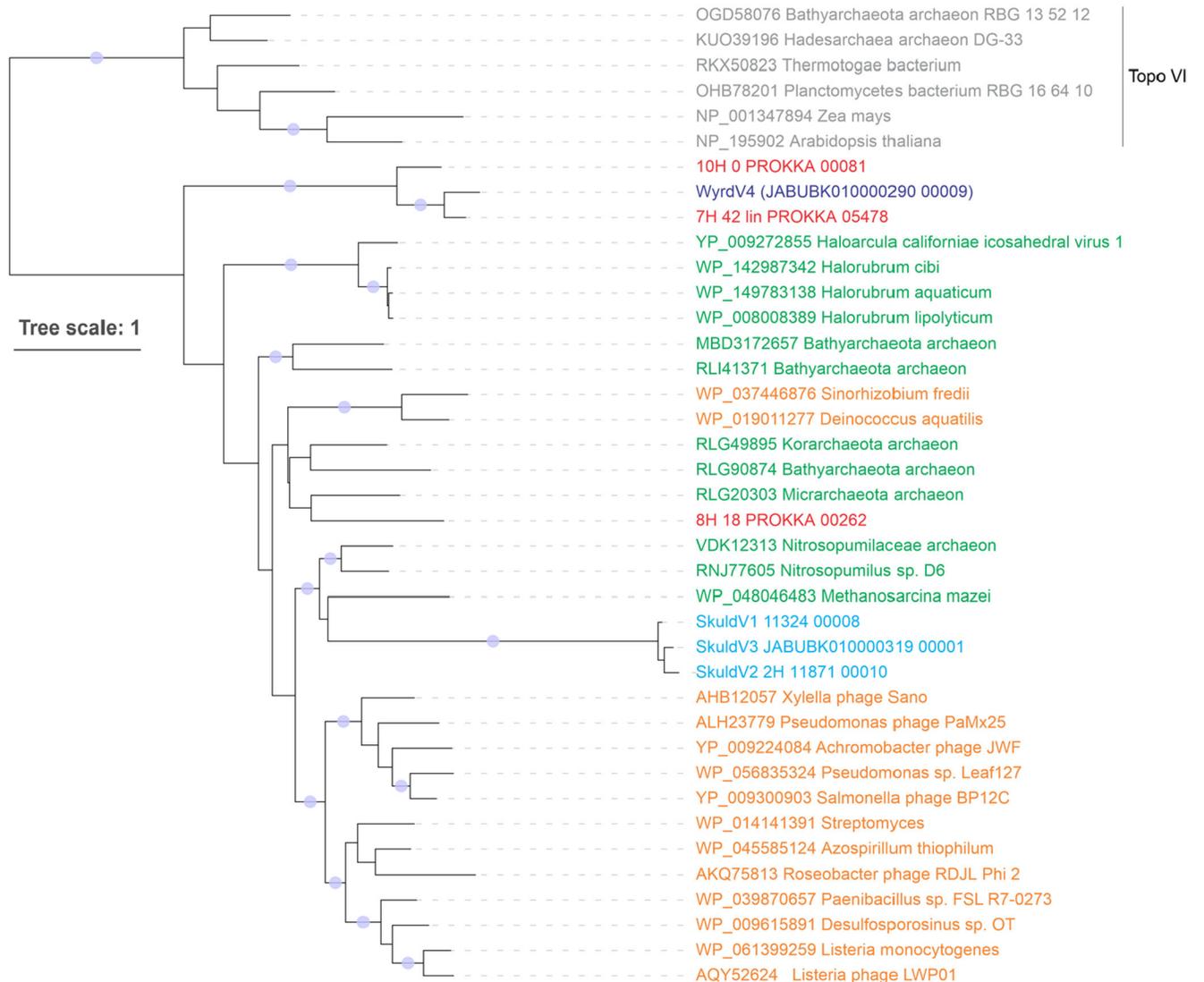


**Extended Data Fig. 1 | Matches between asgardarchaeal CRISPR spacers and viruses.** The figure is divided into three blocks (green, red and blue) corresponding to the three groups of asgardarchaeal viruses, namely, verdandiviruses, skuldviruses and wyrdiviruses. CRISPR repeats and spacers are indicated as diamonds and boxes, respectively. Spacers matching asgardarchaeal viruses are shown in yellow and are connected to the names of

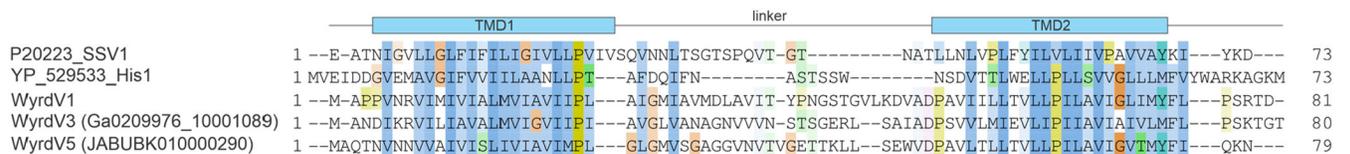
targeted viruses with arrows. Thick vertical lines connect related repeats and the exact pairwise sequence identities are indicated.



**Extended Data Fig. 2 |. Partial provirus integrated within a genomic contig of Lokiarchaea.** The verdandivirus-derived region is boxed. Homologous genes are shown using the same colors and the key is provided on the left of the figure. Housekeeping cellular genes are shown in black and include those encoding transcription factor S (TFS) and ribosomal proteins S7e, S12e, S27e, L30e and L44e. Grey shading connects genes displaying sequence similarity at the protein level, with the percent of sequence identity depicted with different shades of grey.



**Extended Data Fig. 3 |. Maximum likelihood phylogenetic tree of Topo mini-A proteins.** Proteins of skuldviruses and wyrdviruses are shown in cyan and dark blue, respectively, whereas those encoded by other asgardarchaeal MGE are shown in red. Other Topo mini-A homologs encoded by archaea and bacteria (or their corresponding viruses) are shown in green and orange, respectively. Topo VI proteins were used as an outgroup and are colored grey. The tree was constructed using the automatic optimal model selection (LG+R5). The scale bar represents the number of substitutions per site. Circles at the nodes denote aLRT SH-like branch support values larger than 90%.



**Extended Data Fig. 4 |. Sequence alignment of the major capsid proteins of selected wyrdviruses, fusellovirus SSV1 and halspivirus His1.**

When available, proteins are identified with their accession numbers followed by a virus name. TMD, transmembrane domain.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank the crews, technical staff and shipboard scientists of the DV *Chikyu* for the operation and sampling during cruise CK06-06 in 2006. We thank M. Hirai and Y. Takaki for library construction, sequencing and data deposition of seafloor samples off Shimokita. The work in the M.K. laboratory is supported by grants from l'Agence Nationale de la Recherche (nos. ANR-20-CE20-0009-02 and ANR-21-CE11-0001-01) and Ville de Paris (Emergence(s) project MEMREMA). S.M. was supported by the Metchnikov fellowship from Campus France and Russian Science Foundation (grant no. 19-74-20130). N.Y. and E.V.K. are supported by the Intramural Research Program of the National Institutes of Health of the USA (National Library of Medicine). The work of C.R. and J.S. is funded by the Australian Research Council Future Fellow Award (no. FT170100213, to CR). T.N. was partly supported by MEXT KAKENHI (grant nos. JP19H05684 within JP19H05679 (Post-Koch Ecology) and 16H06429, 16K21723 and 16H06437 (NeoVirology)).

## Data availability

The raw reads, as well as assembled virus and MGE genome sequences from the metagenomes described in this study, are available at NCBI under BioProject no. PRJDB12054, BioSample accession nos SAMD00394285–SAMD00394296. Accession numbers of the 20 MAGs assembled in the course of this study are listed in Supplementary Table 1. Accession numbers of the virus and MGE genome sequence are listed in Supplementary Table 3. Source data are provided with this paper.

## References

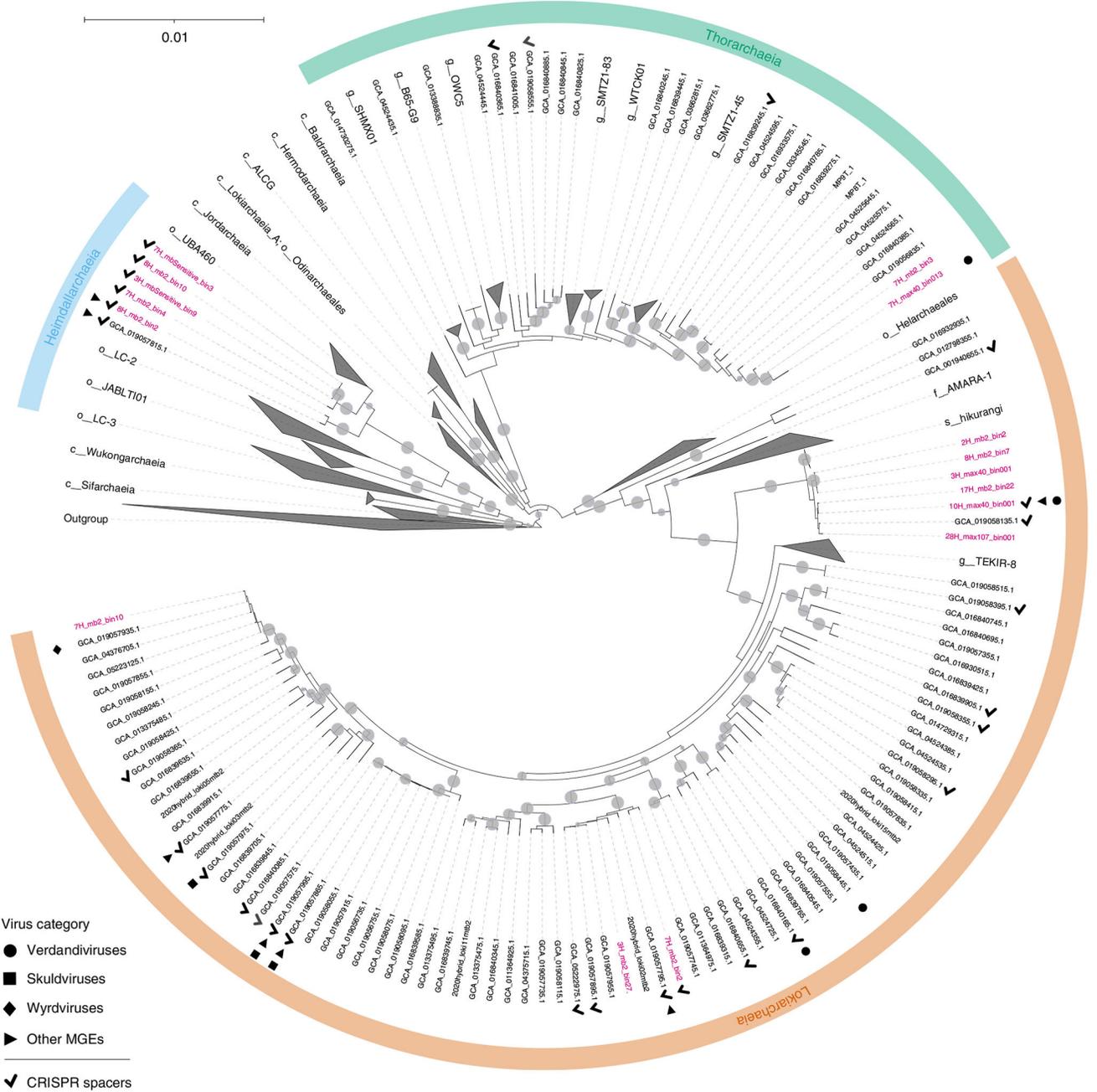
1. Liu Y et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 593, 553–557 (2021). [PubMed: 33911286]
2. Dombrowski N, Teske AP & Baker BJ Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun* 9, 4999 (2018). [PubMed: 30479325]
3. Wong HL et al. Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes. *ISME J.* 12, 2619–2639 (2018). [PubMed: 29980796]
4. Liu Y et al. Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J.* 12, 1021–1031 (2018). [PubMed: 29445130]
5. Zaremba-Niedzwiedzka K et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358 (2017). [PubMed: 28077874]
6. Seitz KW, Lazar CS, Hinrichs KU, Teske AP & Baker BJ Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME J.* 10, 1696–1705 (2016) [PubMed: 26824177]
7. Spang A et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179 (2015). [PubMed: 25945739]
8. Sun J et al. Recoding of stop codons expands the metabolic potential of two novel Asgardarchaeota lineages. *ISME Commun.* 1, 30 (2021). [PubMed: 36739331]
9. Seitz KW et al. Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun* 10, 1822 (2019). [PubMed: 31015394]

10. Farag IF, Zhao R & Biddle JF “Sifarchaeota,” a novel Asgard phylum from Costa Rican sediment capable of polysaccharide degradation and anaerobic methylophily. *Appl. Environ. Microbiol* 87, e02584–20 (2021). [PubMed: 33608286]
11. Zhang JW et al. Newly discovered Asgard archaea Hermodarchaeota potentially degrade alkanes and aromatics via alkyl/benzyl-succinate synthase and benzoyl-CoA pathway. *ISME J.* 15, 1826–1843 (2021). [PubMed: 33452484]
12. Cai M et al. Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic matter degradation. *Sci. China Life Sci* 63, 886–897 (2020). [PubMed: 32201928]
13. Rinke C et al. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat. Microbiol* 6, 946–959 (2021). [PubMed: 34155373]
14. Imachi H et al. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577, 519–525 (2020). [PubMed: 31942073]
15. Lopez-Garcia P & Moreira D The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol* 5, 655–667 (2020). [PubMed: 32341569]
16. Da Cunha V, Gaia M, Nasir A & Forterre P Asgard archaea do not close the debate about the universal tree of life topology. *PLoS Genet.* 14, e1007215 (2018). [PubMed: 29596428]
17. Da Cunha V, Gaia M, Gadelle D, Nasir A & Forterre P Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* 13, e1006810 (2017). [PubMed: 28604769]
18. Baquero DP et al. Structure and assembly of archaeal viruses. *Adv. Virus Res* 108, 127–164 (2020). [PubMed: 33837715]
19. Prangishvili D et al. The enigmatic archaeal virosphere. *Nat. Rev. Microbiol* 15, 724–739 (2017). [PubMed: 29123227]
20. Dellas N, Snyder JC, Bolduc B & Young MJ Archaeal viruses: diversity, replication, and structure. *Annu. Rev. Virol* 1, 399–426 (2014). [PubMed: 26958728]
21. Makarova KS et al. Unprecedented diversity of unique CRISPR-Cas-related systems and Cas1 homologs in Asgard archaea. *CRISPR J.* 3, 156–163 (2020). [PubMed: 33555973]
22. Coclet C & Roux S Global overview and major challenges of host prediction methods for uncultivated phages. *Curr. Opin. Virol* 49, 117–126 (2021). [PubMed: 34126465]
23. Nunoura T et al. Variance and potential niche separation of microbial communities in subseafloor sediments off Shimokita Peninsula, Japan. *Environ. Microbiol* 18, 1889–1906 (2016). [PubMed: 26486095]
24. Glass JB et al. Microbial metabolism and adaptations in Atribacteria-dominated methane hydrate sediments. *Environ. Microbiol* 23, 4646–4660 (2021). [PubMed: 34190392]
25. Dion MB, Oechslin F & Moineau S Phage diversity, genomics and phylogeny. *Nat. Rev. Microbiol* 18, 125–138 (2020). [PubMed: 32015529]
26. Iranzo J, Krupovic M & Koonin EV The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. *mBio* 7, e00978–16 (2016). [PubMed: 27486193]
27. Koonin EV et al. Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev* 84, e00061–19 (2020). [PubMed: 32132243]
28. Baek M et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876 (2021). [PubMed: 34282049]
29. Wang Z et al. Structure of the marine siphovirus TW1: evolution of capsid-stabilizing proteins and tail spikes. *Structure* 26, 238–248 (2018). [PubMed: 29290487]
30. Hendrix RW Tail length determination in double-stranded DNA bacteriophages. *Curr. Top. Microbiol. Immunol* 136, 21–29 (1988). [PubMed: 2967159]
31. Mahony J et al. Functional and structural dissection of the tape measure protein of lactococcal phage TP901-1. *Sci. Rep* 6, 36667 (2016). [PubMed: 27824135]
32. Pope WH et al. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *eLife* 4, e06416 (2015). [PubMed: 25919952]
33. Krupovic M, Forterre P & Bamford DH Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *J. Mol. Biol* 397, 144–160 (2010). [PubMed: 20109464]

34. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D & Koonin EV Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Virus Res.* 244, 181–193 (2018). [PubMed: 29175107]
35. Liu Y et al. Diversity, taxonomy, and evolution of archaeal viruses of the class *Caudoviricetes*. *PLoS Biol.* 19, e3001442 (2021). [PubMed: 34752450]
36. Gardner AF, Bell SD, White MF, Prangishvili D & Krupovic M Protein-protein interactions leading to recruitment of the host DNA sliding clamp by the hyperthermophilic *Sulfolobus islandicus* rod-shaped virus 2. *J. Virol* 88, 7105–7108 (2014). [PubMed: 24696494]
37. Gussow AB et al. Machine-learning approach expands the repertoire of anti-CRISPR protein families. *Nat. Commun* 11, 3784 (2020). [PubMed: 32728052]
38. Li Y & Bondy-Denomy J Anti-CRISPRs go viral: the infection biology of CRISPR-Cas inhibitors. *Cell Host Microbe* 29, 704–714 (2021). [PubMed: 33444542]
39. Krupovic M & Bamford DH Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat. Rev. Microbiol* 6, 941–948 (2008). [PubMed: 19008892]
40. Hong C et al. A structural model of the genome packaging process in a membrane-containing double stranded DNA virus. *PLoS Biol.* 12, e1002024 (2014). [PubMed: 25514469]
41. Frickey T & Lupas A CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20, 3702–3704 (2004). [PubMed: 15284097]
42. Yutin N, Bäckström D, Etema TJG, Krupovic M & Koonin EV Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Virol. J* 15, 67 (2018). [PubMed: 29636073]
43. Abrescia NG et al. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Mol. Cell* 31, 749–761 (2008). [PubMed: 18775333]
44. Oksanen HM, ICTV Report Consortium. ICTV Virus Taxonomy Profile: *Corticoviridae*. *J. Gen. Virol* 98, 888–889 (2017). [PubMed: 28581380]
45. Krupovic M & Bamford DH Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. *BMC Genomics* 8, 236 (2007). [PubMed: 17634101]
46. Kazlauskas D, Varsani A, Koonin EV & Krupovic M Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat. Commun* 10, 3425 (2019). [PubMed: 31366885]
47. Takahashi TS et al. Expanding the type IIB DNA topoisomerase family: identification of new topoisomerase and topoisomerase-like proteins in mobile genetic elements. *NAR Genom. Bioinform* 2, lqz021 (2020). [PubMed: 33575570]
48. Krupovic M, Quemín ER, Bamford DH, Forterre P & Prangishvili D Unification of the globally distributed spindle-shaped viruses of the Archaea. *J. Virol* 88, 2354–2358 (2014). [PubMed: 24335300]
49. Bath C, Cukalac T, Porter K & Dyal-Smith ML His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology* 350, 228–239 (2006). [PubMed: 16530800]
50. Wang F et al. Spindle-shaped archaeal viruses evolved from rod-shaped ancestors to package a larger genome. *Cell* 185, 1297–1307.e11 (2022). [PubMed: 35325592]
51. Hong C et al. Lemon-shaped halo archaeal virus His1 with uniform tail but variable capsid structure. *Proc. Natl Acad. Sci. USA* 112, 2449–2454 (2015). [PubMed: 25675521]
52. Quemín ER et al. *Sulfolobus* spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. *J. Virol* 89, 11681–11691 (2015). [PubMed: 26355093]
53. Roux S et al. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol* 4, 1895–1906 (2019). [PubMed: 31332386]
54. Straus SK & Bo HE Filamentous bacteriophage proteins and assembly. *Subcell. Biochem* 88, 261–279 (2018). [PubMed: 29900501]
55. Kim JG et al. Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *Proc. Natl Acad. Sci. USA* 116, 15645–15650 (2019). [PubMed: 31311861]

56. Quemin ER et al. Eukaryotic-like virus budding in Archaea. *mBio* 7, e01439–16 (2016). [PubMed: 27624130]
57. Rahlff J et al. Lytic archaeal viruses infect abundant primary producers in Earth's crust. *Nat. Commun* 12, 4642 (2020).
58. Bamford DH et al. ICTV Virus Taxonomy Profile: *Pleolipoviridae*. *J. Gen. Virol* 98, 2916–2917 (2017). [PubMed: 29125455]
59. Wu F et al. Unique mobile elements and scalable gene flow at the prokaryote–eukaryote boundary revealed by circularized Asgard archaea genomes. *Nat. Microbiol* 7, 200–212 (2022). [PubMed: 35027677]
60. Summer EJ, Gill JJ, Upton C, Gonzalez CF & 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 (2007). [PubMed: 17719265]
61. Krupovic M, Dolja VV & Koonin EV The LUCA and its complex virome. *Nat. Rev. Microbiol* 18, 661–670 (2020). [PubMed: 32665595]
62. Cahill J & Young R Phage lysis: multiple genes for multiple barriers. *Adv. Virus Res* 103, 33–70 (2019). [PubMed: 30635077]
63. Snyder JC & Young MJ Lytic viruses infecting organisms from the three domains of life. *Biochem. Soc. Trans* 41, 309–313 (2013). [PubMed: 23356303]
64. Krupovic M, Daugelavicius R & Bamford DH A novel lysis system in PM2, a lipid-containing marine double-stranded DNA bacteriophage. *Mol. Microbiol* 64, 1635–1648 (2007). [PubMed: 17555443]
65. Danovaro R et al. Virus-mediated archaeal hecatomb in the deep seafloor. *Sci. Adv* 2, e1600492 (2016). [PubMed: 27757416]
66. Tamarit D et al. A closed *Candidatus* Odinarchaeum chromosome exposes Asgard archaeal viruses. *Nat. Microbiol* 10.1038/s41564-022-01122-y (2022).
67. Rambo IM, de Anda V, Langwig MV & Baker BJ Genomes of six viruses that infect Asgard archaea from deep-sea sediments. *Nat. Microbiol* 10.1038/s41564-022-01150-8 (2022).
68. Kaneko M et al. Insights into the methanogenic population and potential in subsurface marine sediments based on coenzyme F430 as a function-specific compound analysis. *JACS Au* 1, 1743–1751 (2021). [PubMed: 34723277]
69. Hirai M et al. Library construction from subnanogram DNA for pelagic sea water and deep-sea sediments. *Microbes Environ.* 32, 336–343 (2017). [PubMed: 29187708]
70. Hiraoka S et al. Microbial community and geochemical analyses of trans-trench sediments for understanding the roles of hadal environments. *ISME J.* 14, 740–756 (2020). [PubMed: 31827245]
71. Nurk S, Meleshko D, Korobeynikov A & Pevzner PA metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834 (2017). [PubMed: 28298430]
72. Sieber CMK et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol* 3, 836–843 (2018). [PubMed: 29807988]
73. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055 (2015). [PubMed: 25977477]
74. Chaumeil PA, Mussig AJ, Hugenholtz P & Parks DH GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927 (2019). [PubMed: 31730192]
75. Price MN, Dehal PS & Arkin AP FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, e9490 (2010). [PubMed: 20224823]
76. Nguyen LT, Schmidt HA, von Haeseler A & Minh BQ IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol* 32, 268–274 (2015). [PubMed: 25371430]
77. Letunic I & Bork P Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296 (2021). [PubMed: 33885785]
78. Antipov D, Raiko M, Lapidus A & Pevzner PA Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinformatics* 36, 4126–4129 (2020). [PubMed: 32413137]

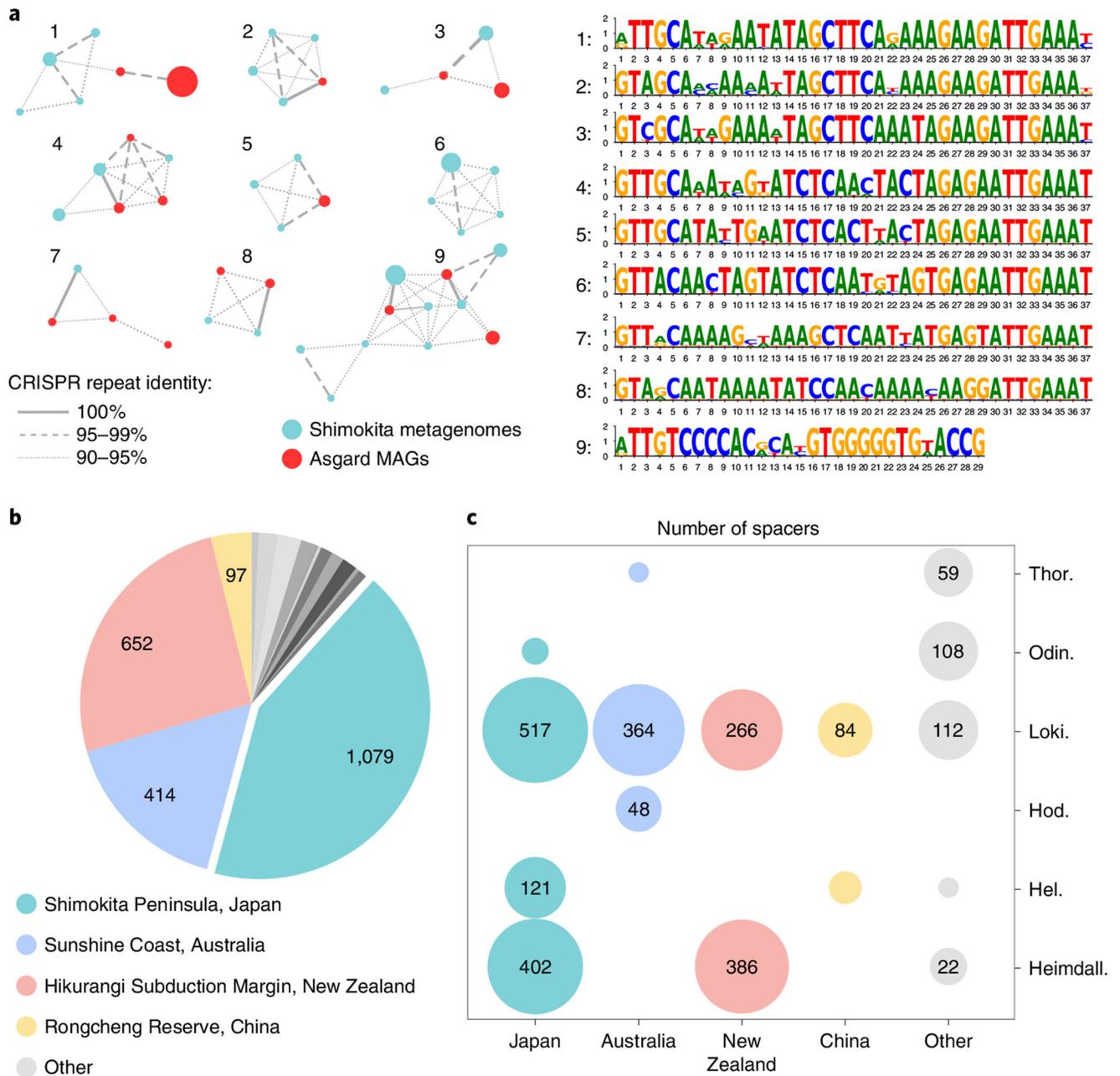
79. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ Basic local alignment search tool. *J. Mol. Biol* 215, 403–410 (1990). [PubMed: 2231712]
80. Seemann T Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069 (2014). [PubMed: 24642063]
81. Steinegger M et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20, 473 (2019). [PubMed: 31521110]
82. Roux S et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol* 37, 29–37 (2019). [PubMed: 30556814]
83. Sullivan MJ, Petty NK & Beatson SA Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010 (2011). [PubMed: 21278367]
84. Bin Jang H et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol* 37, 632–639 (2019). [PubMed: 31061483]
85. Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504 (2003). [PubMed: 14597658]
86. Bland C et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8, 209 (2007). [PubMed: 17577412]
87. Biswas A, Staals RH, Morales SE, Fineran PC & Brown CM CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics* 17, 356 (2016). [PubMed: 27184979]
88. Tareen A & Kinney JB Logomaker: beautiful sequence logos in Python. *Bioinformatics* 36, 2272–2274 (2020). [PubMed: 31821414]
89. Katoh K, Rozewicki J & Yamada KD MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform* 20, 1160–1166 (2019). [PubMed: 28968734]
90. Capella-Gutierrez S, Silla-Martinez JM & Gabaldon T trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973 (2009). [PubMed: 19505945]
91. Nayfach S et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat. Microbiol* 6, 960–970 (2021). [PubMed: 34168315]
92. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD & Lawley TD Massive expansion of human gut bacteriophage diversity. *Cell* 184, 1098–1109 (2021). [PubMed: 33606979]
93. Gregory AC et al. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe* 28, 724–740 (2020). [PubMed: 32841606]
94. Shkoporov AN et al. The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* 26, 527–541 (2019). [PubMed: 31600503]
95. Zhang R et al. SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. *Bioinformatics* 37, 3364–3366 (2021). [PubMed: 33792634]
96. Dion MB et al. Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. *Nucleic Acids Res.* 49, 3127–3138 (2021). [PubMed: 33677572]
97. Galiez C, Siebert M, Enault F, Vincent J & Soding J WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* 33, 3113–3114 (2017). [PubMed: 28957499]
98. Zielezinski A, Deorowicz S & Gudys A PHIST: fast and accurate prediction of prokaryotic hosts from metagenomic viral sequences. *Bioinformatics* 38, 1447–1449 (2021).



**Fig. 1 | Phylogenomic tree of Asgardarchaeota.**

The alignment is based on a concatenated set of 53 protein markers (subsamped to 42 sites each, resulting in a total alignment length of 2,058 sites) from 339 taxa. These taxa encompass 276 Asgardarchaeota MAGs, including 16 recovered in this study (highlighted in pink; MAGs 3H\_mb2\_bin20, 4H\_max40\_bin02,1 4H\_mb2\_bin40 and 7H\_mb2\_bin7 were removed after the alignment trimming step due to low completeness), and 63 non-Asgardarchaeota species representatives from GTDB release RS95 as an outgroup. Maximum-likelihood analysis was performed using IQ-TREE under the LG + C10 + F + G

+ PMSF model. The tree is rooted on the non-Asgardarchaeota group. Nodes with bootstrap statistical support >90% are indicated by grey dots. MAGs containing viral contigs and CRISPR arrays are indicated by different symbols, with the key provided in the bottom left corner. [GCA\\_019058445.1](#) contains a defective verdandivirus-derived provirus (Extended Data Fig. 2).



**Fig. 2 | Description of the Asgardarchaeota CRISPR spacer dataset.**

**a**, Similarity of CRISPR repeats from metagenomic data (teal) and from Asgardarchaeota MAGs (red). Unique CRISPR repeat sequences are shown as nodes in the network. Node diameter is proportional to the number of spacers associated with the repeat sequence in the dataset. Identical CRISPR repeats from both metagenomic data and Asgard MAGs are connected by solid lines; dashed and dotted lines connect CRISPR repeat sequences with 95–99% and 90–95% identities, respectively. Consensus sequences for the nine major clusters are shown on the right. **b**, Sources of Asgardarchaeota CRISPR spacers. The number of spacers in the dataset originating from different geographical sites is indicated on

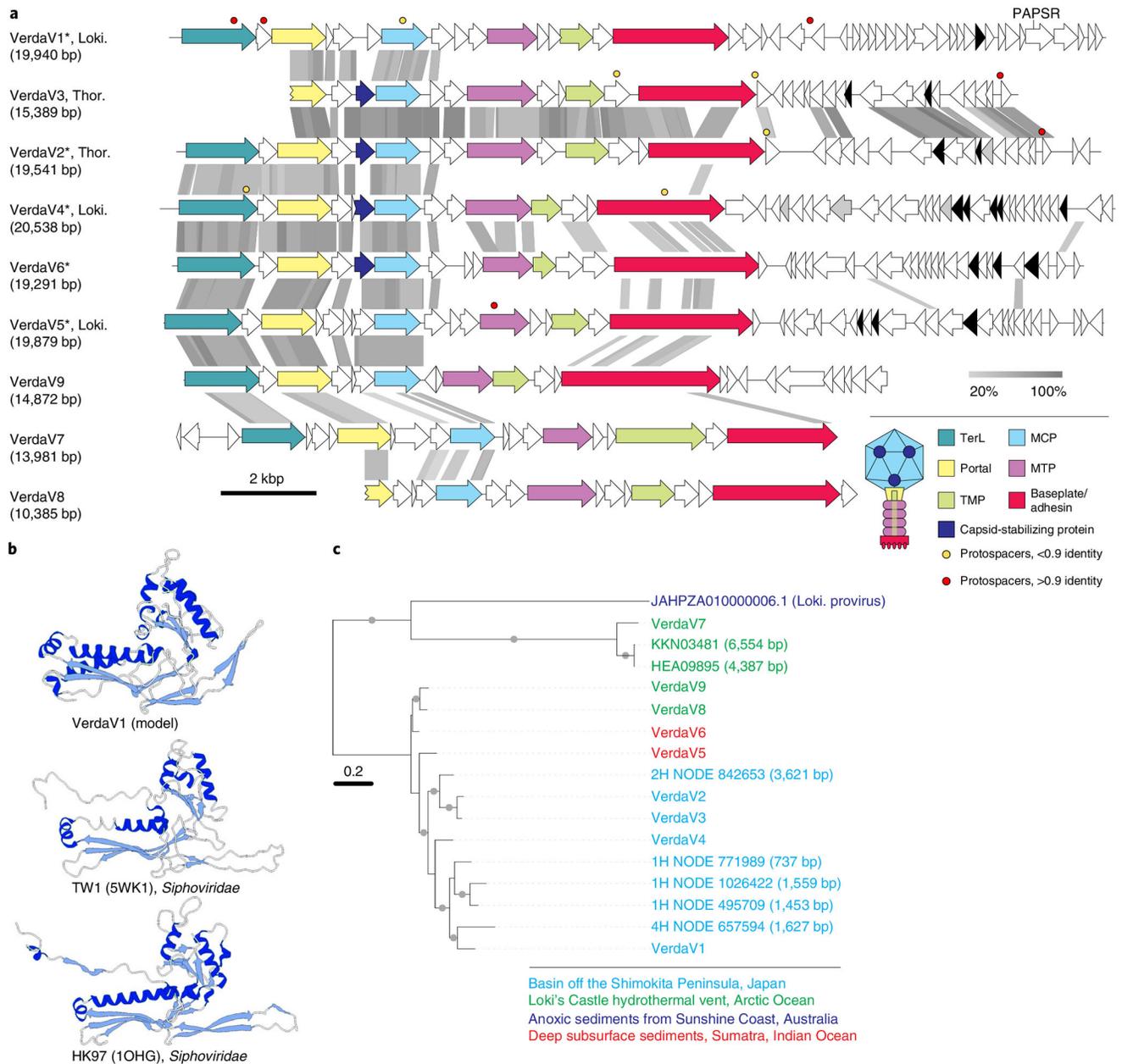
the pie chart. **c.** Taxonomic distribution of Asgardarchaeota MAGs with CRISPR arrays for different isolation sites. Circle size is proportional to the number of spacers in the dataset.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig. 3 |. Diversity of verdandiviruses.**

**a**, Genome maps of verdandiviruses. Homologous genes are shown using the same colours, with the key provided at the bottom. Also shown is the deduced schematic organization of the verdandivirus virion, with colours matching those of genes encoding the corresponding proteins. Genes encoding putative DNA-binding proteins with Zn-binding and HTH domains coloured in black and grey, respectively. Coloured circles indicate the positions of protospacers. Grey shading connects genes displaying sequence similarity at the protein level, with the percentage of sequence identity indicated by different shades of grey (see scale at the bottom). Asterisks denote complete genomes assembled as circular contigs. PAPS<sub>R</sub>, phosphoadenosine phosphosulfate reductase; TMP, tail tape measure protein; MTP,

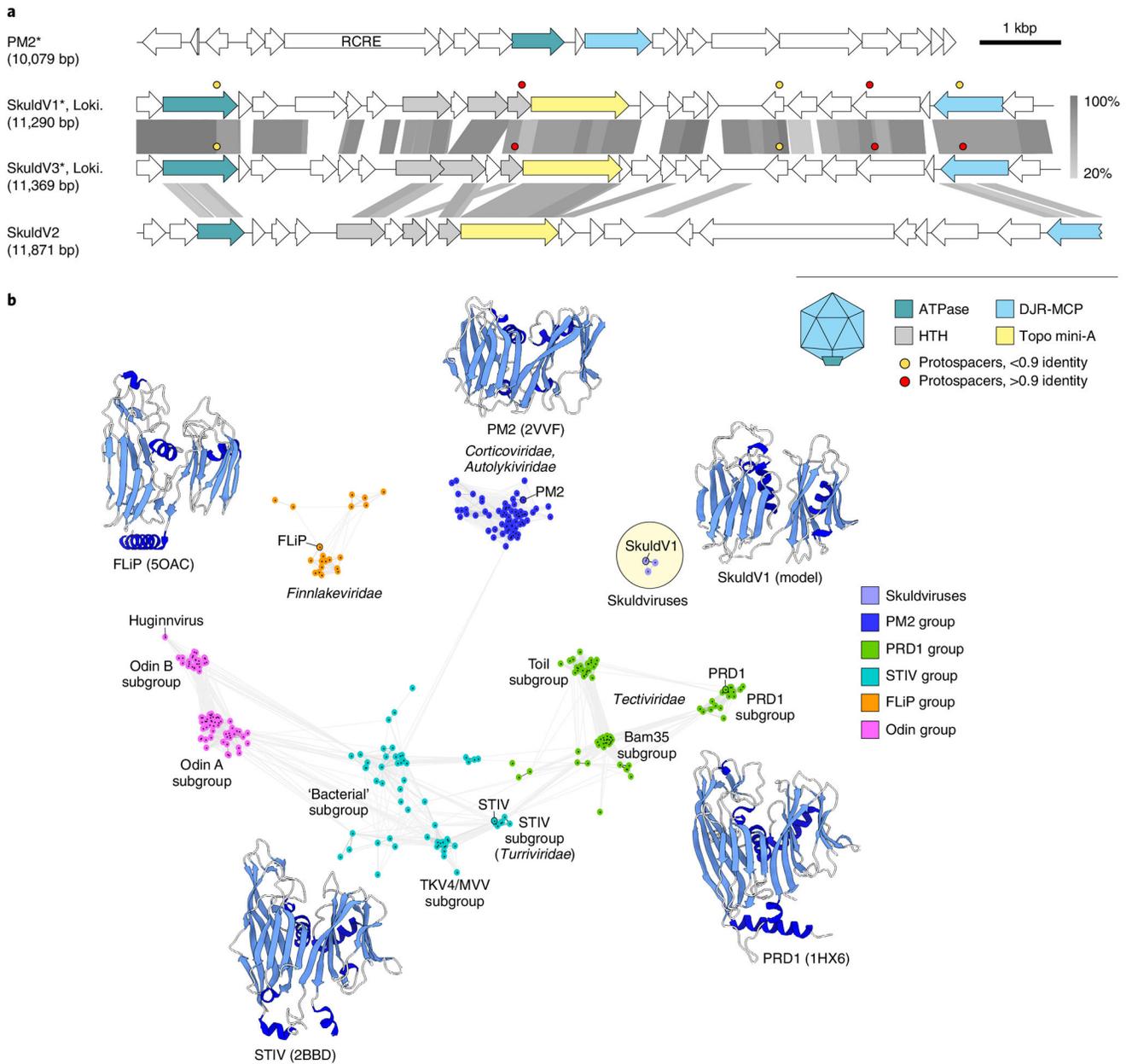
major tail protein. **b**, Comparison of the structural model of the MCP of verdnadivirus AsTV-10H2 with the corresponding structures of siphoviruses TW1 and HK97. The models are coloured according to their secondary structure:  $\alpha$ -helices, dark blue;  $\beta$ -strands, light blue. **c**, Maximum-likelihood phylogeny of MCPs encoded by verdandiviruses. Taxa are coloured based on the source of origin (key at the bottom). The tree was constructed using the automatic optimal model selection (LG + G4) and is midpoint rooted. The scale bar represents the number of substitutions per site. Circles at nodes denote aLRT SH-like branch support values >90%.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

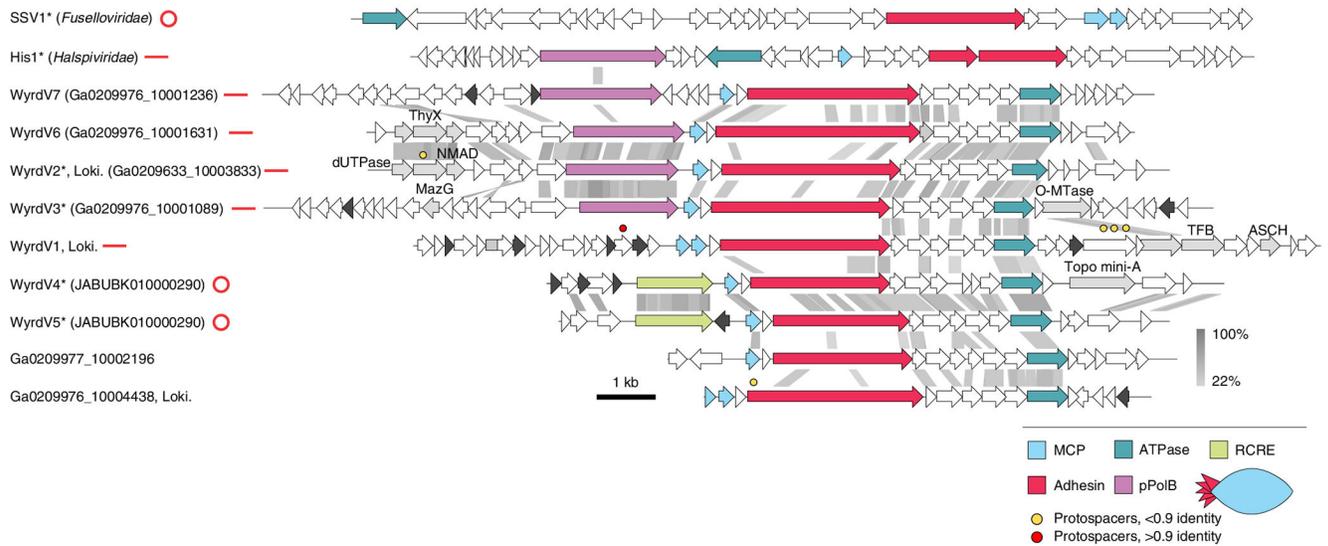


**Fig. 4 | Diversity of skuldviruses.**

**a**, Genome maps of skuldviruses and PM2 virus. Homologous genes are shown using the same colours, with the key provided at the bottom. Also shown is the deduced schematic organization of the skuldvirus virion, with colours matching those of genes encoding the corresponding proteins. Coloured circles indicate the positions of protospacers. Grey shading connects genes displaying sequence similarity at the protein level, with the percentage of sequence identity indicated by different shades of grey (scale on the right). Asterisks denote complete genomes assembled as circular contigs. RCRE, rolling circle replication endonuclease. Genome map of corticovirus PM2 is shown for comparison.

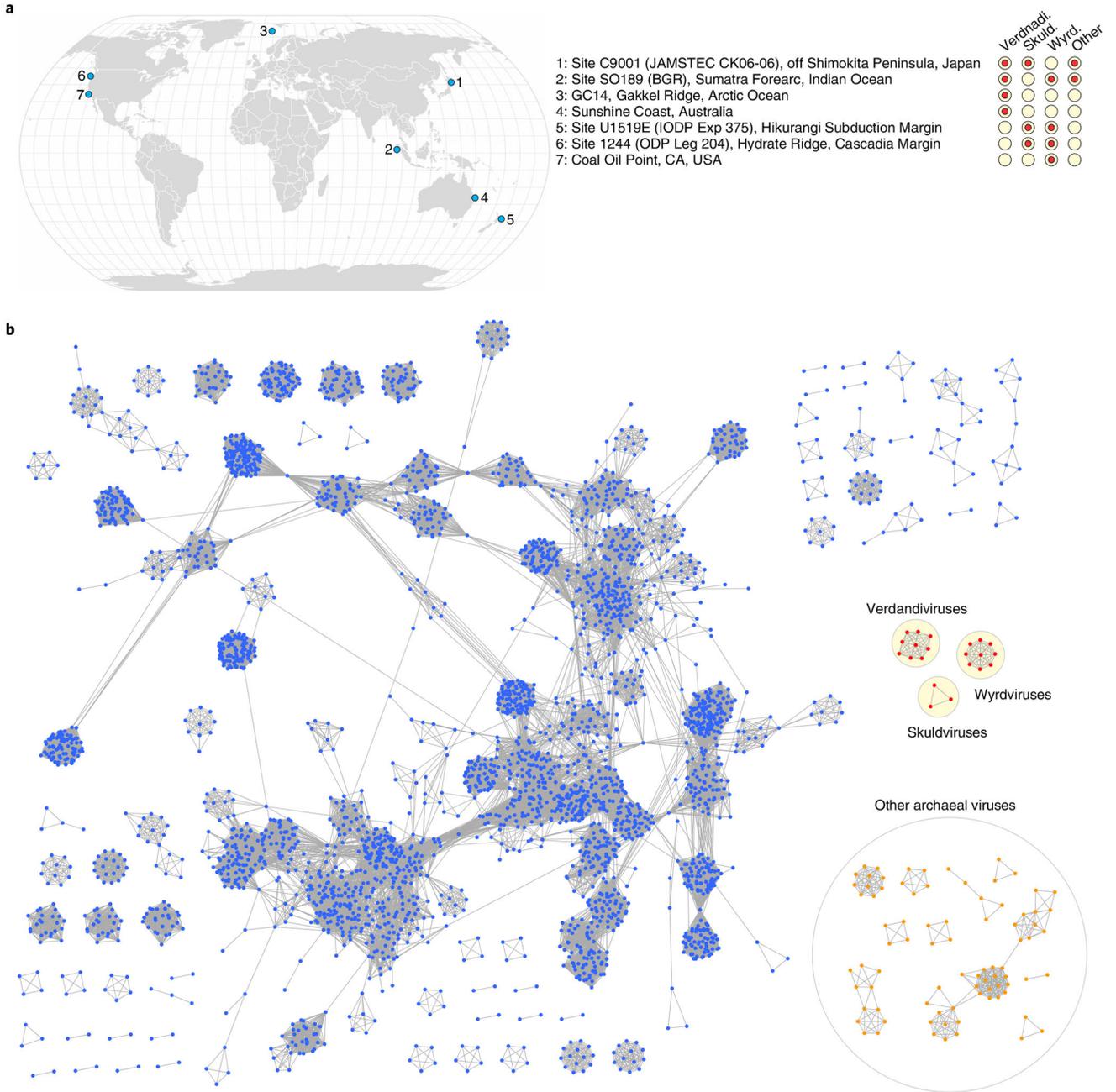
**b**, Sequence similarity network of prokaryotic virus DJR MCPs. Protein sequences were

clustered by pairwise sequence similarity using CLANS. Lines connect sequences with CLANS  $P < 1 \times 10^{-4}$ . CLANS uses  $P$  values of BLASTP comparisons calculated from Poisson distribution of high-scoring segment pairs. Different previously defined groups<sup>42</sup> of DJR MCP are shown as clouds of differentially coloured circles, with the key provided on the right. Note that the Odin group has been named as such previously because some of the MCPs in this cluster originated from *Odinarchaeia* bin<sup>42</sup>. The position of Huginnvirus described by Tamarit et al.<sup>66</sup> is indicated. Subgroups PRD1, Toil and Bam35 are named after corresponding members of the family *Tectiviridae*. Skuldviruses are highlighted within a yellow circle. When available, MCP structures of viruses representing each group are shown next to the corresponding cluster (PDB accession numbers given in parentheses). The Skuldvirus cluster is represented by a structural model of the SkuldV1 MCP. Models are coloured according to their secondary structure:  $\alpha$ -helices, dark blue;  $\beta$ -strands, light blue.



### Fig. 5 | Diversity of wyrdiviruses.

Genome maps of wyrdiviruses, fusellovirus SSV1 and halspivirus His1. Homologous genes are shown using the same colours, with the key provided at the bottom. Also shown is the deduced schematic organization of the skuldvirus virion with colours matching those of the genes encoding the corresponding proteins. Coloured circles indicate the positions of protospacers. Genes encoding putative DNA-binding proteins with Zn-binding and HTH domains are coloured black and grey, respectively. Grey shading connects genes displaying sequence similarity at the protein level, with the percentage of sequence identity indicated by different shades of grey (see scale on the right). Asterisks denote complete genomes assembled as circular contigs or linear contigs with terminal inverted repeats. The topology of each genome is shown next to the corresponding name: circles for circular genomes, lines for linear genomes. ThyX, thymidylate synthase X; NMAD, nucleotide modification-associated domain 1 protein; O-MTase, carbohydrate-specific methyltransferase; TFB, transcription initiation factor B; ASCH, ASC-1 homology domain; pPolB, protein-primed family B DNA polymerase. Genome maps of SSV1 and His1 viruses are shown for comparison.



**Fig. 6 | Asgard archaeal viruses and MGEs.**

**a.** Geographical distribution of Asgard archaeal viruses and MGEs. Filled circles next to geographical locations signify the presence of corresponding virus groups in sediment samples from that location. **b.** Network-based analysis of shared protein clusters among Asgard archaeal viruses and prokaryotic double-stranded DNA viruses. Nodes represent viral genomes and edges represent the strength of connectivity between each genome based on shared protein clusters. Nodes representing genomes of the three groups of Asgard archaeal viruses are shown in red, with the three groups circled within a yellow

background. Nodes corresponding to other bacterial and archaeal viruses are shown in blue and orange, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript