

The International Committee on Taxonomy of Viruses

Taxonomy Proposal Form, 2024

**Part 1a: Details of taxonomy proposals**

|  |  |  |
| --- | --- | --- |
| **Title:** | Create new phylum, “*Artimaviricota*” in the kingdom *Orthornavirae* (realm *Riboviria*) for classification of a hyperthermophilic RNA virus | |
| **Code assigned:** | 2024.039D |

|  |  |  |  |
| --- | --- | --- | --- |
| **Author(s), affiliation and email address(es):** | | | |
| **Name** | **Affiliation** | **Email address** | **Corresponding author(s)** X |
| Syun-ichi Urayama | Department of Life and Environmental Sciences, Laboratory of Fungal Interaction and Molecular Biology (donated by IFO), University of Tsukuba, Tsukuba, Japan | [urayama.shunichi.gn@u.tsukuba.ac.jp](mailto:urayama.shunichi.gn@u.tsukuba.ac.jp) | X |
| Akihito Fukudome | Howard Hughes Medical Institute,  Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN, USA | [afukudom@iu.edu](mailto:afukudom@iu.edu) |  |
| Eugene V. Koonin | National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. | [koonin@ncbi.nlm.nih.gov](mailto:koonin@ncbi.nlm.nih.gov) |  |
| Takuro Nunoura | Research Center for Bioscience and Nanoscience (CeBN), JAMSTEC, Yokosuka, Japan | [takuron@jamstec.go.jp](mailto:takuron@jamstec.go.jp) |  |
| Mart Krupovic | Institut Pasteur, Université Paris Cité, Archaeal Virology Unit, Paris, France | [mart.krupovic@pasteur.fr](mailto:mart.krupovic@pasteur.fr) | X |

**Part 1b: Taxonomy Proposal Submission**

|  |  |  |  |
| --- | --- | --- | --- |
| **ICTV Subcommittee:** | | | |
| Animal DNA Viruses and Retroviruses |  | Bacterial viruses | **X** |
| Animal minus-strand and dsRNA viruses |  | Fungal and protist viruses |  |
| Animal positive-strand RNA viruses |  | Plant viruses |  |
| Archaeal viruses |  | General - |  |

|  |
| --- |
| **List the ICTV Study Group(s) that have seen or have been involved in creating this proposal:** |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Optional – complete only if formally voted on by an ICTV Study Group:** | | | |
| **Study Group** | **Number of members** | | |
| **Votes in support** | **Votes against** | **No vote** |
|  |  |  |  |
|  |  |  |  |

|  |  |
| --- | --- |
| **Submission date:** | 21/06/2024 |

**Part 1c: Feedback from ICTV Executive Committee (EC) meeting**

|  |  |
| --- | --- |
| **Executive Committee Meeting Decision code:** | **X** |
| A – Accept | **X** |
| Ac – Accept subject to revision by relevant subcommittee chair. No further vote required |  |
| U – Accept without revision but with re-evaluation and email vote by the EC |  |
| Uc – Accept subject to revision and re-evaluation and email vote by the EC |  |
| Ud – Deferred to the next EC meeting, with an invitation to revise based on EC comments |  |
| J - Reject |  |
| W - Withdrawn |  |

|  |
| --- |
| **Comments from the Executive Committee:** |
|  |

**Part 1d: Revised Taxonomy Proposal Submission**

|  |
| --- |
| **Response of proposer:** |
|  |

|  |  |
| --- | --- |
| **Revision date:** | DD/MM/YYYY |

**Part 3:** **TAXONOMIC PROPOSAL**

|  |
| --- |
| **Name of accompanying Excel module:** |
| 2024.039D.A.v1.Artimaviricota\_np.xlsx |

|  |  |  |  |
| --- | --- | --- | --- |
| **Taxonomic changes proposed:** | | | |
| Establish new taxon | **X** | Split taxon |  |
| Abolish taxon |  | Merge taxon |  |
| Move taxon |  | Promote taxon |  |
| Rename taxon |  | Demote taxon |  |
| Move and rename |  |

|  |  |  |
| --- | --- | --- |
| **Is any taxon name used here derived from that of a living person:** | | **N** |
| **Taxon name** | **Person from whom the name is derived** | **Attached X** |
|  |  |  |
|  |  |  |
|  |  |  |

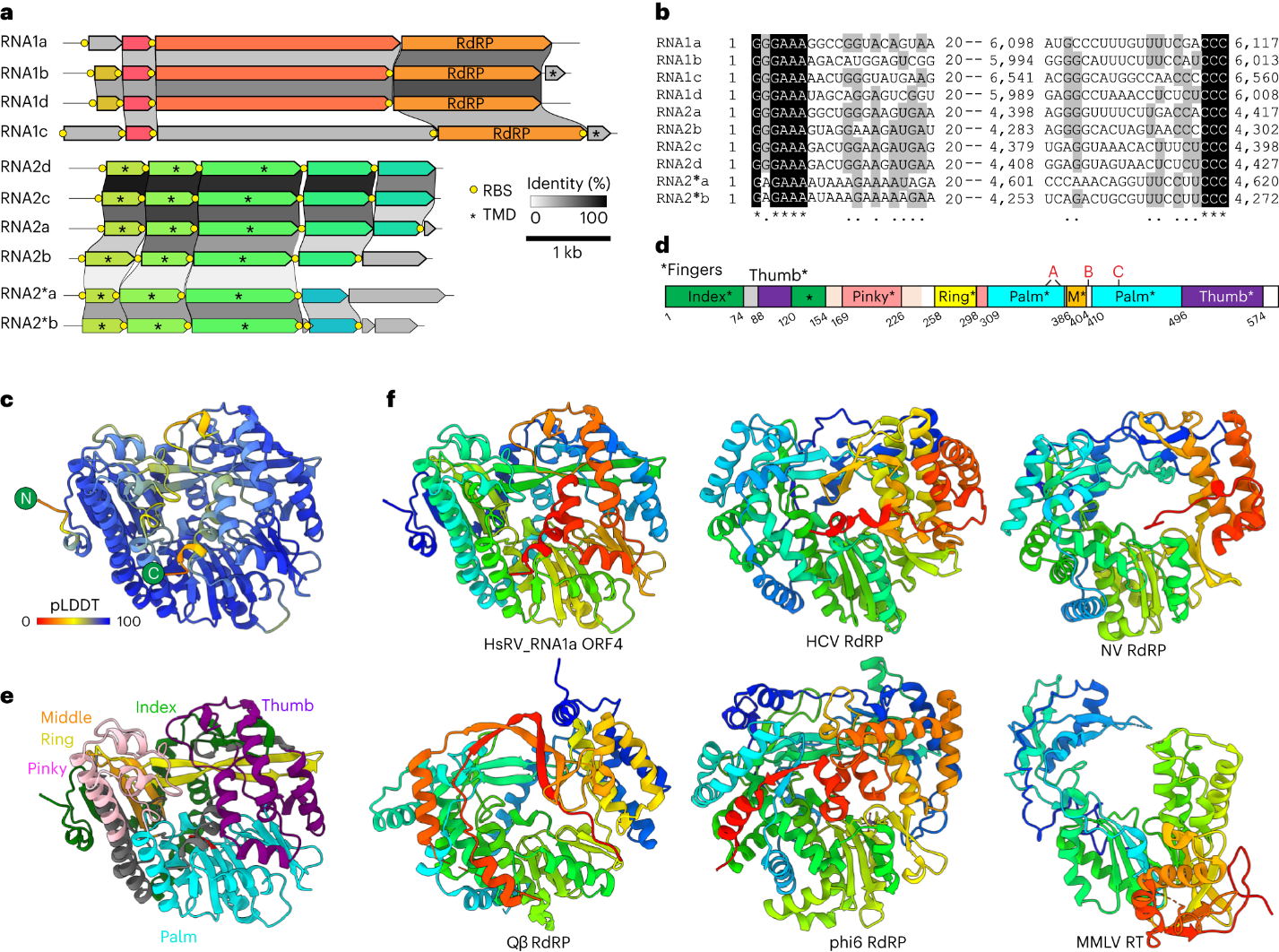
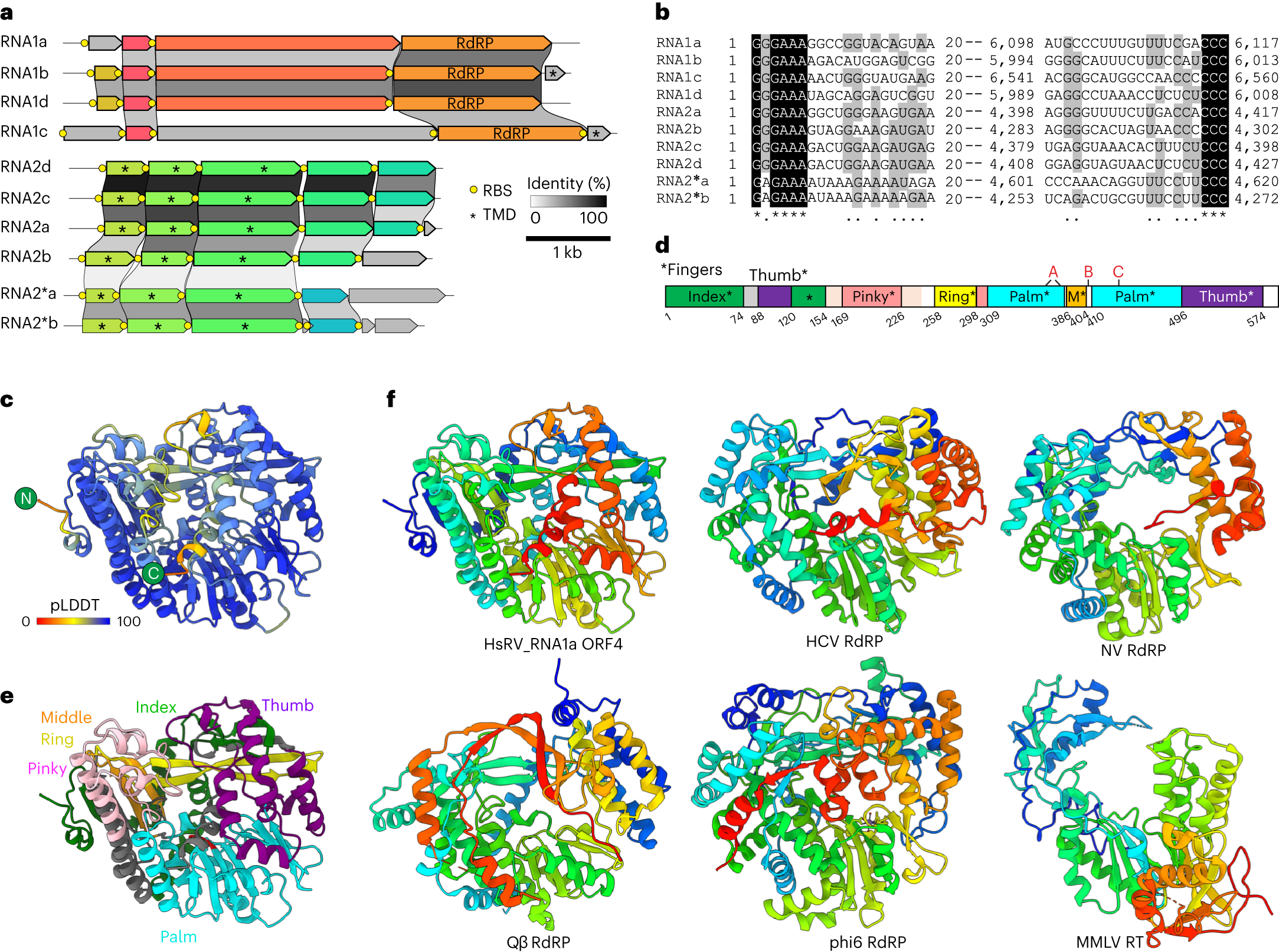
|  |
| --- |
| **Abstract of Taxonomy Proposal:** |
| *Taxonomic rank(s) affected*: *Riboviria*, *Orthornavirae*  *Description of current taxonomy*:  Realm *Riboviria* includes two kingdoms, *Orthornavirae* and *Pararnavirae*, which include highly diverse viruses that encode RNA dependent RNA polymerases (RdRP) and reverse transcriptases (RT), respectively. Kingdom *Orthornavirae* includes six phyla which were established based on phylogenetic analysis of the RdRP and comparative analysis of the viral genomes and proteins.  *Proposed* *taxonomic change(s):*  We propose to create a new phylum in the kingdom *Orthornavirae* for classification of a group of RNA viruses discovered in hot springs that are characterized by unusual RdRPs.  *Justification*:  The RdRPs of HsRV1 and its relatives seem to deviate from the RdRP consensus farther than any of the other recently discovered putative phyla, with none of which they appear to be affiliated, and possess unusual structural features that appear to link them to viral RTs. |

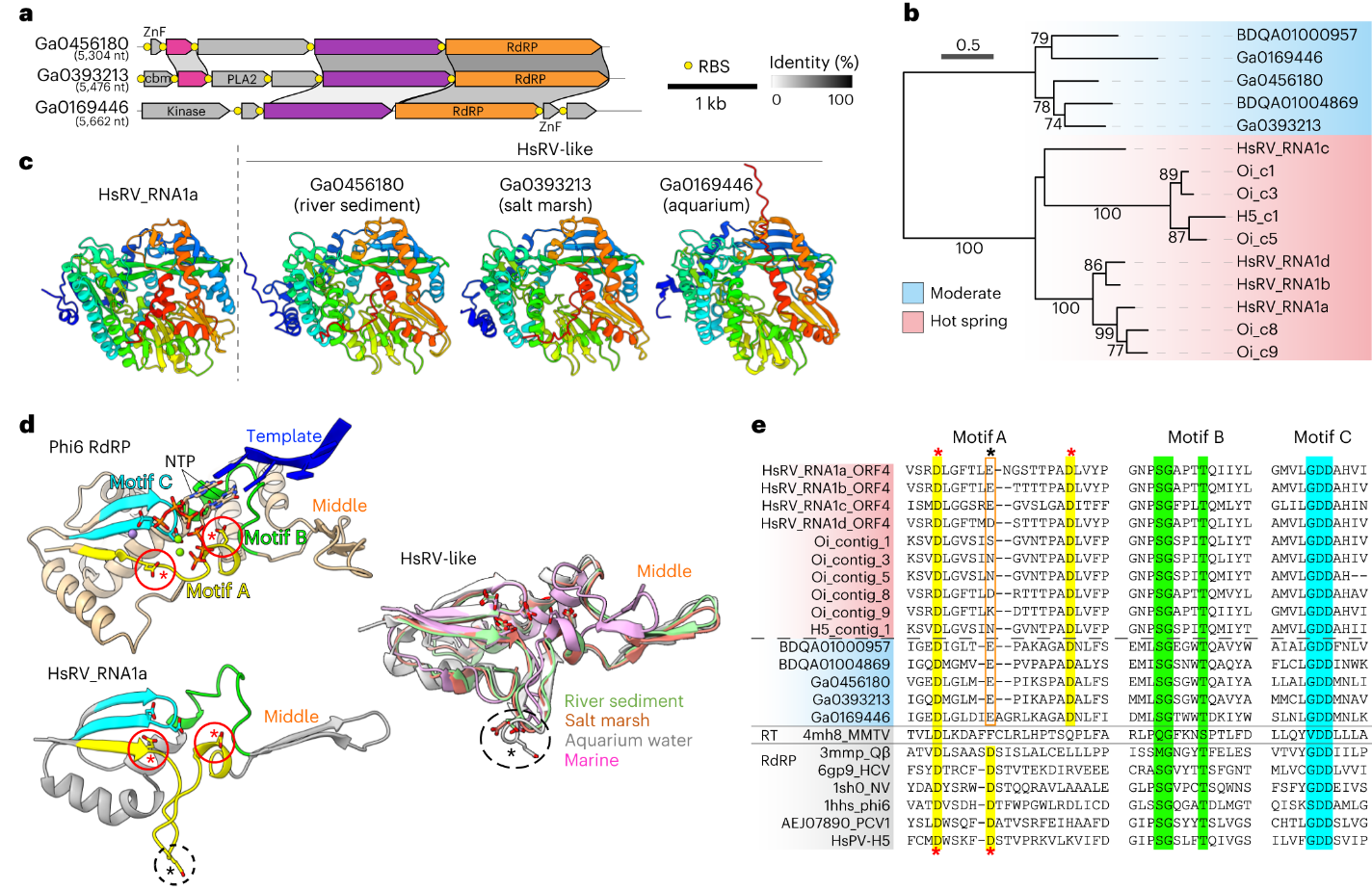
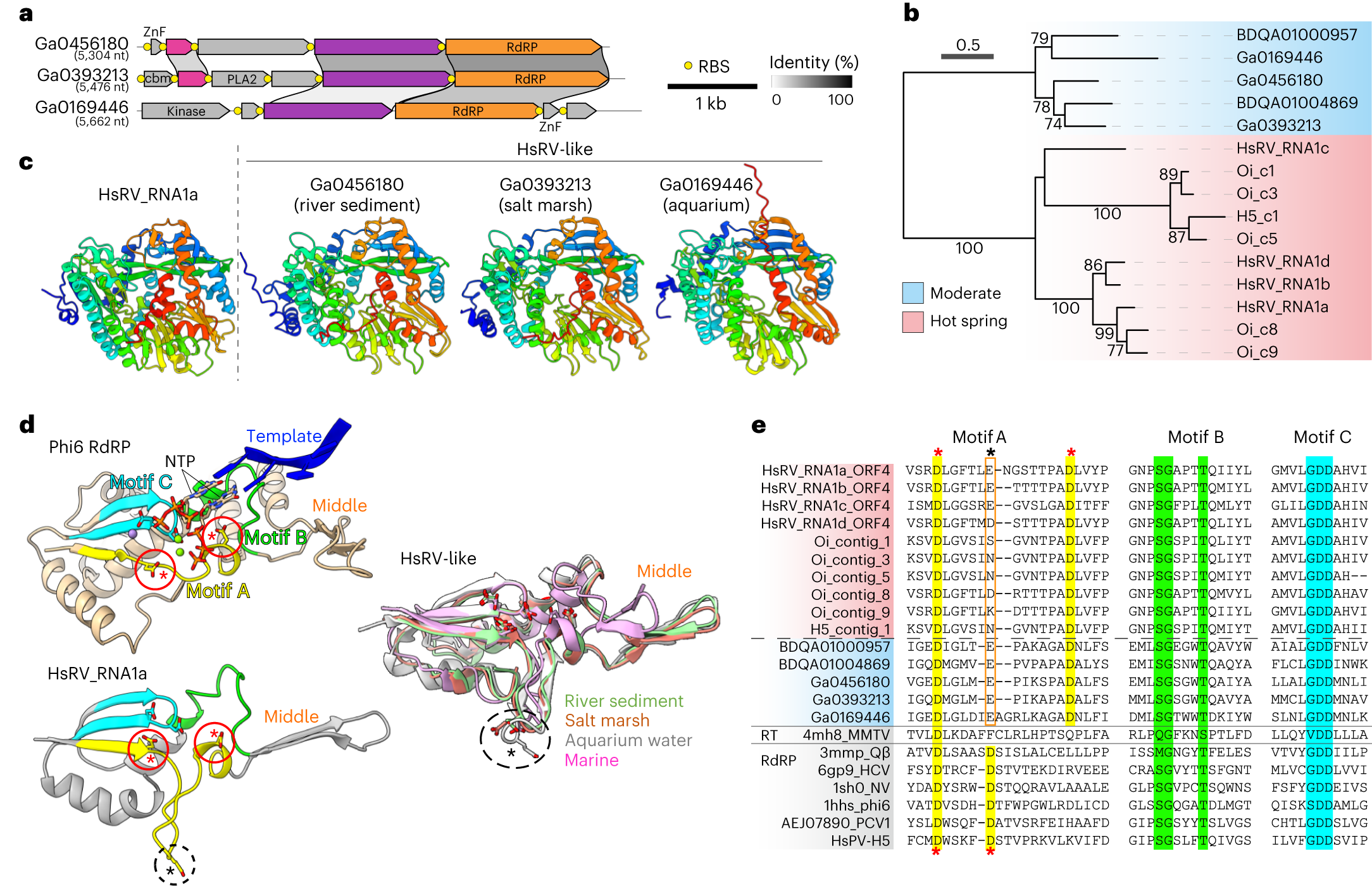
|  |
| --- |
| * **Text of Taxonomy proposal:** |
| *Taxonomic rank(s) affected*: *Riboviria*, *Orthornavirae*  *Description of current taxonomy*:  Realm *Riboviria* consists of two kingdoms, *Orthornavirae* and *Pararnavirae*, which include highly diverse viruses that encode RNA dependent RNA polymerases (RdRP) and reverse transcriptases (RT), respectively. RdRPs and RT enzymes are homologous and share the same structural fold based on the palm domain which encompasses the catalytic residues distributed across three conserved motifs, known as A, B and C. Kingdom *Orthornavirae* includes six phyla which were established based on phylogenetic analysis of the RdRP and comparative analysis of the viral genomes and proteins.  *Proposed* *taxonomic change(s)*:  We propose to create a new phylum in the kingdom *Orthornavirae* for classification of a group of RNA viruses discovered in hot springs that are characterized by an unusual RdRP.  *Demarcation criteria:*  The main demarcation criterion for inclusion into the phylum is the presence of an RdRP showing sequence or structural similarity to the RdRPs encoded by other members of this taxon. The specific signatures of this RdRP are explained in the section below. Currently, only one species is proposed to be established. Thus, appropriate species demarcation criteria will be established in the future once more members of this group are discovered.  *Justification*:  **Discovery and diversity of HsRV-like viruses**  Fragmented and primer-Ligated DsRNA Sequencing (FLDS) of hot spring water samples (79.3 °C, pH 2.2) from the Unzen area, Japan, yielded three populations of contigs (Fig. 1A) which collectively recruited ∼50% of the clean FLDS reads (Urayama et al., 2024). Among these contigs, we identified closely similar 5′- and 3′-terminal sequences (Fig. 1B), a characteristic feature of segmented RNA viruses. On the basis of the similarity of the 5′- and 3′-terminal sequences, lengths of the segments and gene content, we concluded that two sets of contigs constituted genomes of a distinct group of bipartite RNA viruses. The segments were denoted RNA1, RNA2 and RNA2\*. In total, we obtained complete sequences for 4, 4 and 2 divergent variants of segments RNA1, RNA2 and RNA2\*, respectively (Fig. 1A). The similarity between the termini of the segments precluded assignment of all sets of segments to particular virus strains. However, segments RNA1a and RNA2a were most abundant and had longer conserved terminal sequences and were thus assigned to the same virus strain with a bisegmented genome. This strain was denoted Hot spring RNA virus 1 (HsRV1).  RNA1, RNA2 and RNA2\* harbored 4–5, 5–6 and 5–7 open reading frames (ORFs), respectively (Fig. 1A). None of the predicted proteins encoded by these RNAs showed significant similarity (BLASTP E-value = 5 × 10−03) to any protein sequences in public databases. Even the most sensitive profile–profile searches using HHpred yielded no significant (HHpred probability >90%) hits for any of the predicted proteins. However, HHpred searches queried with the amino acid sequence of ORF4 from the RNA1 segment produced a partial hit to several RdRPs. Although the hits were not significant (HHpred probability <90%) and encompassed only a small region of the RdRP (∼15% of the target profile), the aligned region covered the diagnostic RdRP motifs B (SGxxxT, x – any amino acid) and C (GDD).  Using the multiple sequence alignment that included the identified RNA1 ORF4 homologues, a high-quality (average per-residue Local Distance Difference Test (pLDDT) = 90.7) AlphaFold2 model of the putative RdRP was obtained (Fig. 2C). Examination of this model revealed a topology typical of the palm-domain polymerases, with readily discernible ‘Fingers’, ‘Palm’ and ‘Thumb’ subdomains (Fig. 2D,E) and overall architecture similar to that of viral RdRPs (Fig. 2F), albeit with notable unique structural features. In particular, the RNA1 ORF4 model displayed an extended and highly ordered ‘Fingers’ subdomain, with the ‘fingertips’ forming a 5-stranded β-sheet that is missing in other RdRPs and interacts with the ‘Thumb’ subdomain. The conserved motifs B and C identified by HHpred were located within the Palm subdomain, at positions equivalent to those in other RdRPs. Structural superposition of the Palm subdomains from different RdRPs allowed identification of the third core motif, A, in RNA1 ORF4 (see below). Thus, we concluded that RNA1 ORF4 encodes an RdRP. The four RdRPs encoded by the complete RNA1 segments shared 37 to 75% pairwise amino acid sequence identity and thus appear to represent four distinct virus species (or even higher taxa). Further searches of the FLDS datasets revealed several additional contigs with a high (>90%) identity to HsRV\_RNA1b RdRP. In addition, several contigs with moderate (>60%) identity to HsRV RNA1a or RNA1b were detected, suggesting a considerable diversity of HsRV-like viruses in hot springs.  The sequence profile of the HsRV RdRP was used to search the previously described FLDS sequence data from coastal seawater samples19, leading to the identification of two additional contigs (GenBank accessions: BDQA01000957 and BDQA01004869) encoding incomplete HsRV-like RdRPs. Searches against the IMG/VR database queried with these RdRPs yielded significant hits (E-value ≤ 1 × 10−05) to three additional putative RdRPs encoded by apparently complete or near-complete 5.3–5.6-kb-long genome segments: Ga0456180\_000042, Ga0393213\_00017, Ga0169446\_00510 (Fig. 2a). The three segments originate from floodplain (river sediments), salt marsh and aquarium samples, respectively. Phylogenetic analysis of HsRV-like RdRPs showed clear separation between viruses from the hot spring and those from moderate aquatic environments (Fig. 3b). Collectively, these results indicate that HsRV-like viruses are broadly distributed in both hot springs and non-extreme aquatic ecosystems.  **Structural similarities between HsRV-like RdRPs and RTs**  AF2 models of the three HsRV-like RdRPs from moderate ecosystems showed clear structural similarity with the HsRV RdRP, including the extended ‘Fingers’ subdomain (Fig. 2c). Another signature feature of these proteins is an unusual, extended RdRP motif A. In the canonical motif A, the two conserved Asp residues involved in catalysis and substrate discrimination, respectively, are separated by 4–5 residues and bracket the catalytic GDD residues of motif C (Fig. 2d,e). By contrast, in HsRV-like RdRPs, the second Asp residue of motif A is not conserved, and the corresponding residue is located in a loop facing perpendicularly away from motif C, suggesting that it cannot perform the same function. However, all analyzed HsRV-like RdRPs contain an Asp (Asp\*) which is located 12–14 residues away from the first Asp of motif A (Fig. 2e). Despite the extended spacing in the protein sequence, Asp\* occupies a position equivalent to that of the second Asp of the canonical motif A (Fig. 2d,e) and is likely to be its counterpart involved in substrate discrimination.  We next performed structural clustering on the basis of the pairwise DALI Z-scores of the HsRV-like RdRPs together with selected RdRPs of other riboviruses, including putative phyla of RNA phages identified in recent metatrascriptome analyses and RTs encoded by eukaryotic viruses of the order *Ortervirales* as well as non-viral RTs from bacteria and eukaryotes (Fig. 3a). The HsRV-like RdRPs from both hot springs and moderate aquatic ecosystems formed a tight cluster, underscoring their relatedness despite high sequence divergence. All previously known viral RdRPs formed a clade in the structure-based dendrogram, but the HsRV-like RdRPs remained separated from those (Fig. 3a). The two viral RdRP clusters were interspersed with the RTs, such that the viral RTs were the closest structural neighbours of the HsRV-like RdRPs. This result confirms the extreme divergence of the HsRV-like RdRPs and might reflect a closer relationship to viral RTs. This unexpected link was strengthened by the comparison of the ‘Palm’ subdomain of HsRV-like RdRPs with homologues from other riboviruses as well as viral and non-viral RTs. In RdRPs of riboviruses from the 5 established phyla, the first β-strand (blue in Fig. 3b) containing motif A and the motif B-containing α-helix are separated by a characteristic helix-turn-helix (HTH) region followed by a β-hairpin corresponding to the ‘Middle’ finger subdomain (Fig. 1d,e). However, the HTH motif is absent in both the HsRV-like RdRPs and viral RTs. Notably, non-viral RTs, such as those from group II introns or retrons, contain the HTH motif but lack the β-hairpin region, which is compatible with the intermediate position of RTs between the two clades of viral RdRPs. Thus, the HsRV-like RdRPs might comprise an evolutionary intermediate between viral RdRPs and RTs. A BLASTN search against the metagenomic DNA sequences obtained from the hot springs did not detect HsRV-like sequences, suggesting that HsRV-like viruses are bona fide riboviruses that lack a DNA intermediate stage.  **HsRV-like viruses probably infect prokaryotic hosts**  All samples in which HsRVs were detected nearly exclusively contained rRNA sequences from prokaryotes, with eukaryotic presence being below 1%. This is consistent with eukaryotes being unable to thrive in polyextremophilic conditions combining high temperatures and acidic pH. The microbial communities in the HsRV1-containing sample were dominated by bacteria. Thus, HsRV1 most probably infects bacteria. Consistently, nearly every gene in HsRVs is preceded by a Shine-Dalgarno (SD) motif, which is essential for translation initiation in many prokaryotes, and their conservation is a diagnostic feature of prokaryotic genes that has been used to assign bacterial hosts to several groups of RNA viruses, namely, picobirnaviruses and partitiviruses (Krishnamurthy and Wang, 2018; Neri et al., 2022). The lack of eukaryotes in the hot spring samples, contrasted by the dominance of bacteria, together with the presence of typical prokaryotic SD motifs upstream of the predicted virus genes and the polycistronic organization of the viral genomes, strongly suggest that HsRV are viruses of thermophilic bacteria.  **Proposed taxonomy**  The RdRPs of HsRV1 and its relatives seem to deviate from the RdRP consensus farther than any of the other recently discovered putative phyla, with none of which they appear to be affiliated, and possess unusual (predicted) structural features that appear to link them to viral RTs. Whether this connection reflects an intermediate position of the HsRV-like viruses between the kingdoms *Orthornavirae* and *Pararnavirae*, or results from convergent evolution, remains uncertain and should be clarified by sequencing and structural analysis of additional members of this group of viruses. Regardless, HsRV-like viruses cannot be included in any of the existing phyla and thus should be considered as representatives of a separate phylum in the kingdom *Orthornavirae* or even a third kingdom within the realm *Riboviria*. Currently, for classification of HsRV1, we propose to create a new phylum within *Orthornavirae* and name it “*Artimaviricota*”. The phylum would include the following new intermediate taxa: class “*Furtirnaviricetes”*, order “*Divaquavirales*”, family “*Hakuzoviridae*”, genus “*Atsuirnavirus*” and species “*Atsuirnavirus caloris*”.  **Etymology**  “*Artimaviricota*”, after the potential link to viral RTs (arti) and ‘artima’ which means ‘close’ in Lithuanian.  “*Furtirnaviricetes”*, after ‘*furtivus*’, Latin for *hidden*, *secret*, and RNA, referring to the genome type.  “*Divaquavirales*”, after diverse aquatic environments in which these viruses were discovered.  “*Hakuzoviridae*”, after Hakuzo, the White elephant which is a signature of Fugen Bosatsu (Samantabhadra in Buddhism), referring to the Fugen-dake, the largest volcano of the Mt. Unzen (an area where the virus was discovered).  “*Atsuirnavirus*”after ‘*atsui*’, hot in Japanese, and RNA, referring to the genome type.  *“caloris*”, after Latin calor (sing. gen.): of the heat. |

|  |
| --- |
| **References:** |
| Krishnamurthy SR, Wang D. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. Virology. 2018; 516:108-114. doi: 10.1016/j.virol.2018.01.006. PMID: 29346073  Neri U, et al. Expansion of the global RNA virome reveals diverse clades of bacteriophages.  Cell. 2022; 185(21):4023-4037.e18. doi: 10.1016/j.cell.2022.08.023. PMID: 36174579  Urayama SI, Fukudome A, Hirai M, Okumura T, Nishimura Y, Takaki Y, Kurosawa N, Koonin EV, Krupovic M, Nunoura T. Double-stranded RNA sequencing reveals distinct riboviruses associated with thermoacidophilic bacteria from hot springs in Japan. Nat Microbiol. 2024; 9(2):514-523. doi: 10.1038/s41564-023-01579-5. PMID: 38233646 |

|  |
| --- |
| **Tables, Figures:** |

<Start here>

 **Figure 1.** Unusual bipartite RNA virus genomes from the Oi hot spring. A, Genome organization and conservation of the three genomic segments (RNA1, RNA2 and RNA2\*) of HsRV. ORFs encoding homologous proteins are shown as arrows with identical colours. Yellow circles represent predicted Shine-Dalgarno ribosome binding sequences (RBS). Asterisks denote putative genes encoding predicted transmembrane domain (TMD)-containing proteins. B, MSA of the 5ʹ- and 3ʹ-terminal regions of the coding strands of reconstructed genome segments. Black shading, 100% nucleotide identity; grey shading, >50% nucleotide identity. C, Quality assessment of the AlphaFold2 model of the HsRV RdRP. The structural model is coloured on the basis of the pLDDT scores (average pLDDT = 90.7), with the colour key shown at the bottom left corner. D,E, Domain organization of the HsRV RdRP. D, Schematic representation of the domain organization, with exact coordinates of each subdomain, including the five ‘Fingers’, indicated. M, middle finger. The positions of the motifs A, B and C are indicated. E, The structural model of HsRV RdRP coloured using the same scheme as in D. F, Comparison of the HsRV RdRP with homologues from other RNA viruses, including hepatitis C virus (HCV; PDB: 6GP9), Norwalk virus (NV; PDB: 1SH0), Qβ (PDB: 3MMP), phi6 (PDB: 1HHS) as well as RT from Moloney murine leukaemia virus (MMLV; PDB: 4MH8). The structures are coloured using the rainbow scheme, from blue N terminus to red C terminus.



**Figure 2.** HsRV-like viruses from moderate environments. a, RdRP-encoding segments of HsRV-like viruses from non-extreme aquatic ecosystems. ORFs encoding homologous proteins are shown as arrows with identical colours. b, Maximum-likelihood phylogeny of the HsRV-like RdRPs encoded by viruses from extreme (pink) and moderate (blue) ecosystems. Node support was assessed using the SH-aLRT, with the corresponding values (%) shown on the branches. The scale bar represents the number of substitutions per site. c, Comparison of the HsRV RdRP with the homologues encoded by viruses from moderate aquatic ecosystems. The model was produced using AlphaFold2. The models are coloured using the rainbow scheme, from blue N terminus to red C terminus. d, Comparison of the catalytic cores encompassing the conserved RdRP motifs A (yellow), B (green) and C (cyan). Top: the structure of bacteriophage phi6 RdRP with the substrate nucleoside triphosphates (NTP) and template RNA strand (blue ribbon). Bottom: the HsRV RdRP. Middle: structurally superposed HsRV-like RdRPs from moderate ecosystems. The NTP and active site residues of motifs A and C are shown using the stick representation. The conserved aspartate residues of motif A are circled, with structurally equivalent residues indicated with red asterisks, whereas the non-conserved residue located in the loop facing away from the motif C in HsRV and related RdRP is indicated with the black asterisk. e, Sequence alignment of the conserved motifs of HsRV-like RdRPs from extreme (red shading) and moderate (blue shading) ecosystems with the corresponding regions from RdRPs and RT from other viruses (grey shading), including Moloney murine leukaemia virus (MMLV), hepatitis C virus (HCV), Norwalk virus (NV), PCV1 and hot spring partiti-like virus H5 (HsPV-H5). The sequences are indicated with the PDB or GenBank accession numbers. The conserved residues are shaded yellow, green and cyan, respectively, matching those in d. The conserved aspartate residues of motif A are highlighted in yellow, with structurally equivalent residues indicated with red asterisks, whereas the non-conserved residue in HsRV-like RdRPs located at the equivalent position as the second aspartate in other RdRPs is indicated with the black asterisk.

A collage of images of a human body

Description automatically generated

**Figure 3.** Structural relationships between RdRPs and RTs. a, Matrix and cluster dendrogram were constructed on the basis of the pairwise Z-score comparisons calculated using DALI. Different protein groups are highlighted with different background colours on the dendrogram: green, RdRPs from previously characterized viruses; blue, viral and non-viral RTs; red, HsRV-like RdRPs. The colour scale indicates the corresponding Z-scores. hPBV, human picobirnavirus; FMDV, foot-and-mouth disease virus; ReoV, reovirus; LACV, La Crosse virus; HIV-1, human immunodeficiency virus 1; TERT, telomerase RT; non-LTR R2 retroel., non-long terminal repeat R2 retroelement; AF2, AlphaFold2 model. For experimentally determined structures, the corresponding PDB accession numbers are indicated at the bottom of the matrix. b, Structural comparison of the core domain of RdRPs and RT encompassing the conserved motifs A–C. The structures are coloured using the rainbow scheme, from blue N terminus to red C terminus.