

Review Article

Functional biology and biotechnology of thermophilic viruses

 Ryan K. Doss¹,  Marike Palmer¹,  David A. Mead² and  Brian P. Hedlund^{1,3}

¹School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, Nevada, U.S.A.; ²Varizymes, Middleton, Wisconsin, U.S.A.; ³Nevada Institute of Personalized Medicine, Las Vegas, Nevada, U.S.A.

Correspondence: David A. Mead (dmead@varizymes.com) or Brian P. Hedlund (brian.hedlund@unlv.edu)



Viruses have developed sophisticated biochemical and genetic mechanisms to manipulate and exploit their hosts. Enzymes derived from viruses have been essential research tools since the first days of molecular biology. However, most viral enzymes that have been commercialized are derived from a small number of cultivated viruses, which is remarkable considering the extraordinary diversity and abundance of viruses revealed by metagenomic analysis. Given the explosion of new enzymatic reagents derived from thermophilic prokaryotes over the past 40 years, those obtained from thermophilic viruses should be equally potent tools. This review discusses the still-limited state of the art regarding the functional biology and biotechnology of thermophilic viruses with a focus on DNA polymerases, ligases, endolysins, and coat proteins. Functional analysis of DNA polymerases and primase-polymerases from phages infecting *Thermus*, *Aquificaceae*, and *Nitratiruptor* has revealed new clades of enzymes with strong proofreading and reverse transcriptase capabilities. Thermophilic RNA ligase 1 homologs have been characterized from *Rhodothermus* and *Thermus* phages, with both commercialized for circularization of single-stranded templates. Endolysins from phages infecting *Thermus*, *Meiothermus*, and *Geobacillus* have shown high stability and unusually broad lytic activity against Gram-negative and Gram-positive bacteria, making them targets for commercialization as antimicrobials. Coat proteins from thermophilic viruses infecting *Sulfolobales* and *Thermus* strains have been characterized, with diverse potential applications as molecular shuttles. To gauge the scale of untapped resources for these proteins, we also document over 20,000 genes encoded by uncultivated viral genomes from high-temperature environments that encode DNA polymerase, ligase, endolysin, or coat protein domains.

Introduction

Life in high-temperature environments poses challenges that are met by adaptations that increase the stability of all macromolecules, including proteins. The same properties that make thermophilic proteins vital to their thermophilic hosts—intrinsic stability and activity at high temperatures—also offer important advantages over their mesophilic counterparts for industrial and molecular biology applications. As a classic example, Taq polymerase, isolated from *Thermus aquaticus*, was employed in the 1980s to substitute for *Escherichia coli* DNA polymerase for the polymerase chain reaction (PCR) [1]; the stability of Taq polymerase under conditions required to thermally denature DNA improved the practicality and costs of PCR, and was critical for its rapid expansion as a cornerstone of modern molecular biology, disease diagnostics, forensics, and genetic genealogy, among other technologies [2]. High temperatures also decrease nucleic acid secondary structures, off-target base-pairing, and nonspecific protein–protein and ligand–protein interactions, thereby improving the efficiency and fidelity of a wide variety of biochemical interactions. Thermophily also increases compatibility with a variety of industrial and molecular biology applications, including better performance in viscous solutions, which become more fluid

Received: 31 January 2023
Revised: 28 April 2023
Accepted: 09 May 2023

Version of Record published:
11 August 2023

at higher temperature, and increasing volatility of biofuels [3,4]. In addition, thermophilic enzymes are typically more stable than mesophilic enzymes [5], which can increase shelf-life, enhance stability under a variety of extreme conditions, and simplify purification schemes; for example, heat purification of thermostable enzymes is a simple method for isolating recombinant thermophilic proteins from crude lysates of mesophilic expression systems [6–8].

Even though biotechnology has long relied on enzymes from thermophilic prokaryotes, those from thermophilic viruses that infect them remain remarkably underexplored. All viral genomes encode key enzymes that are necessary for the biology of the virus, including those involved in diversion of host resources for viral genome replication, transcription and translation, evasion of host immunity, packaging, and egress from the host cell [9,10]. Genes encoding these functions occur at much higher frequencies in viral DNA than in host DNA and high recombination rates within viruses promote biochemical innovations. In all, these biochemical innovations provide excellent targets for biotechnological commercialization.

Over the last two decades, rapid progress has been made on the cultivation and molecular biology of thermophilic and hyperthermophilic host–virus pairs, especially among novel archaeal viruses infecting the thermoacidophilic order *Sulfolobales* and to a lesser extent thermophilic *Thermoproteales*, both belonging to the phylum *Thermoproteota* (synonym *Crenarchaeota*) [11,12]. These archaeal viruses belong to the International Committee on Taxonomy of Viruses (ICTV) families of *Lipothrixviridae*, *Rudiviridae*, *Tristomaviridae*, *Turriviridae*, *Ampullaviridae*, *Bicaudaviridae*, *Spiraviridae*, *Fuselloviridae*, *Guttaviridae*, *Clavaviridae*, and *Globuloviridae* [11,13]. Parallel research on thermophilic bacteriophages has largely focused on those infecting the genera *Thermus*, *Meiothermus*, *Geobacillus*, and *Rhodothermus* [14]. Most of these bacteriophages belong to the class *Caudoviricetes* (recently reclassified based on genomic information and not morphology [15]), while others represent novel families not yet placed in higher taxonomic ranks or are largely unclassified (e.g., unclassified myo- and siphoviruses ϕ YS40, G20c, and RM378). These viruses replicate at high temperatures and are often stable at temperatures exceeding the optimal growth temperatures of their host thermophiles or hyperthermophiles [16–18]. However, viral genomes are also universally enriched in poorly annotated genes [19], with tens of thousands of poorly annotated small-gene families being discovered recently [20]. Together, these genes represent a vast, underexplored resource with potential to contribute innumerable advances in biotechnology and biomedicine.

This paper highlights a relatively small number of functionally characterized proteins from thermophilic bacteriophage and archaeal viruses and their potential roles in biotechnology, focusing on proteins with potential applications in DNA synthesis, nucleotide modifications and repair, cell lysis, and nano-trafficking (Figure 1 and Table 1). We also provide an up-to-date accounting of putative proteins of biotechnological interest in uncultivated viral genomes (UViGs) from thermal environments and discuss key opportunities to explore these proteins for biotechnology purposes (Table 2, Supplementary Material).

DNA polymerases

Thermophilic DNA polymerases have been a focal point of development and commercialization for biotechnology companies due to their versatility, with uses ranging from molecular diagnostics to next-generation DNA sequencing. In 1988, Taq polymerase, a family-A DNA polymerase (*PolA*) from *T. aquaticus*, was optimized for PCR due to its thermophily, with the important caveat that it has a high error rate at one mutation per 20,000 base pairs [21,22]. Bacteriophage ϕ 29 DNA polymerase is a mesophilic enzyme with lower error rates [23] that allows for isothermal amplification with proofreading, high processivity, and strand-displacement capabilities, but it only functions at low temperatures. Attempts to identify thermophilic viral DNA polymerases similar to Taq polymerase, but that can proofread, or to identify a suitable thermophilic alternative to ϕ 29 polymerase would improve the capabilities of existing applications. Thus far, some thermophilic viral DNA polymerases have been identified that contain high-fidelity proofreading domains, strand-displacement, or reverse-transcriptase activity [22,24–26].

Several thermophilic viral alternatives to Taq polymerase have been expressed, purified, and characterized (Table 1), including two originating from bacteriophage infecting the genus *Thermus*. Many cultivated *Thermus* phage encode annotated family-A DNA polymerase (*polA*) genes, including vB_Tt72 [27], ϕ YS40 [28], ϕ TMA [29], G20c [30], P7426 [31], P2345 [31], ϕ FA [32], TSP4 [33], and Tth15-6 [34]. Of these, the unclassified myoviruses *Thermus* phage vB_Tt72, phage ϕ YS40, and phage ϕ TMA are closely related and likely represent a novel genus in the *Caudoviricetes* [27]. They have similar *polA* genes (>90% amino acid identity) with low sequence similarity to other polymerases [27]. Biochemical characterization of the vB_Tt72 DNA polymerase revealed proofreading activity even at low dNTP concentrations (0.4 mM), properties that Taq polymerase lacks completely [22]. However, the purified vB_Tt72 polymerase lost function rapidly above 60°C, and was not thermostable enough for PCR [27], limiting potential biotechnology applications.

Table 1 Functionally characterized proteins from thermophilic viruses and their potential biotechnology applications

| Proteins | Characterized thermophilic viral protein | Source virus (Synonyms) | Functional or structural characterization and biotechnology applications to date | Accession number | References |
|-----------------|--|---|--|---------------------------------|---------------|
| DNA Polymerases | vB_Tt72 PolA | Unclassified myovirus vB_Tt72 (<i>Thermus</i> phage vB_Tt72) | Functional: Verified 3'-5' exonuclease activity, nucleotidyltransferase domain, and performed optimally at 55°C and pH 8.5. | ON714139.1 | [27] |
| | Poll_G20c | Unclassified <i>Oshimavirus</i> G20c (<i>Thermus</i> phage G20c) | Functional: Structurally and functionally characterized confirming 3'-5' exonuclease activity and DNA polymerase activity, with optimal polymerase activity at 70°C at pH 9.1. | KX987127.1 | [30] |
| | 3173 Pol | Metagenomic fragment from 'Pyrovirus' from Octopus Spring, Yellowstone | Functional: Verified 3'-5' exonuclease activity and DNA-dependent DNA polymerase activity and RT activity with optimum of 77°C and half-life of ~11 min. at 94°C. Biotech: One-enzyme RT-PCR. | ADL99605.1 | [26] |
| | OCT 1608-14 Pol | PCR amplified from metagenomic DNA from Octopus Spring, Yellowstone; full-length variant of 3173 Pol from 'Pyrovirus' | Functional: DNA polymerase activity confirmed under conditions similar to those of 3173 Pol. (Contains N-terminal DUF927 domain in addition to 3'-5' exonuclease/DNA polymerase A domain of 3173 Pol.) | KC440900 | [38] |
| | LavaLAMP/ PyroPhage 3173 PolA | Engineered fusion of 'Pyrovirus' 3173 Pol and Sso7d polymerase | Functional: Verified 3'-5' exonuclease activity and DNA-dependent DNA polymerase activity and RT activity with optimum of 85°C. Biotech: RT-PCR, RT-LAMP, cDNA cloning for RNA Seq. Commercialized by Lucigen Co. as LavaLAMP/PyroPhage 3173 PolA. | AFN99414.1 | [25,26,37,40] |
| | Magma DNA polymerase | Chimera of shuffled 'Pyrovirus' Pols with Taq polymerase 5'-3' exonuclease | Functional: Verified higher fidelity over other 3173 Pol variants/relatives (1 error per 10 ⁹ nucleotides), increased primed-template binding affinity. Biotech: Optimized for single-enzyme RT-PCR. | US 2021/0171580 A1 ^a | [40] |
| | NrS-1 primase-polymerase | Unclassified siphovirus NrS-1 (<i>Nitratiruptor</i> phage NrS-1) | Functional: Verified polymerase, primase, and helicase activity at 50°C. Lacks finger and thumb subdomains. Biotech: Primer-free DNA synthesis for whole genome amplification | BAN05337.1 | [42–44] |
| Ligases | RM378 RNA ligase | Unclassified myovirus RM378 (<i>Rhodothermus</i> phage RM378) | Functional: Experimentally determined high specificity (10x higher than T4 RNA ligase), and RNA/ssDNA ligation optimally at 64°C and pH 6–7. Biotech: Patented and commercialized as a competitor to RLM-RACE for cDNA ligation. | NC_004735.1 | [61] |
| | TS2126 RNA ligase | Unclassified virus Ph2119 (<i>Thermus</i> phage Ph2119) | Functional: Experimentally determined high specificity (30x higher than T4 RNA ligase), and RNA/ssDNA ligation optimally at 65°C and pH 7.5. Biotech: Commercialized as CirLigase™ for cDNA ligation and circularization of linear nucleic acids by Epicentre. | COHM52.1 | [62,63] |
| Endolysins | Ts2631 endolysin | Unclassified virus vB_Tsc2631 (<i>Thermus</i> phage vB_Tsc2631) | Functional: <i>In vitro</i> antibacterial activity against Gram-positive and Gram-negative bacteria performing optimally at 40–105°C and pH 7.0–11.0. | KJ561354 | [69,71] |
| | Ph2119 endolysin | Unclassified virus Ph2119 (<i>Thermus</i> Phage Ph2119) | Functional: <i>In vitro</i> antibacterial activity against Gram-negative bacteria performing optimally from 50°C to 105°C and pH 7.5–8.0. | KF408298.1 | [70] |
| | TSPphg endolysin | Unclassified <i>Oshimavirus</i> TSP4 (<i>Thermus</i> phage TSP4) | Functional: <i>In vivo</i> clinical testing in mice infected with multidrug-resistant <i>Staphylococcus aureus</i> reduced infection performing optimally at 40–70°C and pH 7.0–10.0. | QAY18185 | [73,75] |
| | MMPphg endolysin | Unclassified myovirus MMP17 (<i>Meiothermus</i> phage MMP17) | Functional: <i>In vitro</i> antibacterial activity against Gram-positive and Gram-negative bacteria. This enzyme performed optimally from 35°C to 65°C. | QAY18044 | [74,75] |

Continued over

Table 1 Functionally characterized proteins from thermophilic viruses and their potential biotechnology applications (Continued)

| Proteins | Characterized thermophilic viral protein | Source virus (Synonyms) | Functional or structural characterization and biotechnology applications to date | Accession number | References |
|---------------|--|---|---|------------------|------------|
| | MLTphg | Chimera of TSPphg and MMPphg | Functional: <i>In vitro</i> antibacterial activity against Gram-positive and Gram-negative bacteria greater than TSPphg and MMPphg alone optimally from 35 to 40°C. | – | [75] |
| | GVE2 endolysin | Unclassified siphovirus GVE2 (<i>Geobacillus</i> phage GVE2) | Functional: Verified lysis against <i>Geobacillus</i> sp. E263 through interaction with a phage holin and host ABC transporter. This enzyme was active at 60°C. | YP_001285830.1 | [77] |
| | GVE2 _{CAT} -fusions | <i>Geobacillus</i> sp. E263 phage GVE2 catalytic domain fused with several <i>Clostridium perfringens</i> phage cell-wall binding domains | Functional: <i>In vitro</i> analysis revealed antibacterial activity against <i>Clostridium perfringens</i> , performing above 50% activity from 4 to 60°C. Biotech: Patented for commercialization | – | [78] |
| | TP-84 endolysin | <i>Saundersvirus</i> Tp84 (<i>Geobacillus</i> Phage TP-84) | Functional: <i>In vitro</i> analysis revealed biofilm reduction against Gram-positive and Gram-negative bacteria with full activity at 30–75°C. | YP_009600073.1 | [76] |
| Coat Proteins | SIRV2 capsid protein | <i>Iceruvivirus</i> SIRV2 (<i>Sulfolobus islandicus</i> rod-shaped virus 2 SIRV2) | Functional: Experimentally determined stability range of -80°C to 80°C at pH 6. Cryo-EM structure. Biotech related: SIRV2 stability was monitored in different solvents, and attachment sites and ligands identified. | NP_666560.1 | [95,96] |
| | SMV1 coat protein | Unclassified <i>Bicaudaviridae</i> SMV1 (<i>Sulfolobus</i> monocaudavirus 1 SMV1) | Functional: <i>In vivo</i> stability without inflammatory response; passed through simulated gastric fluid, GI tracts of mice, and human intestinal organoids; vector stability and immunogenicity assessed. | YP_009008070.1 | [94] |

^aSequence available in this US patent.

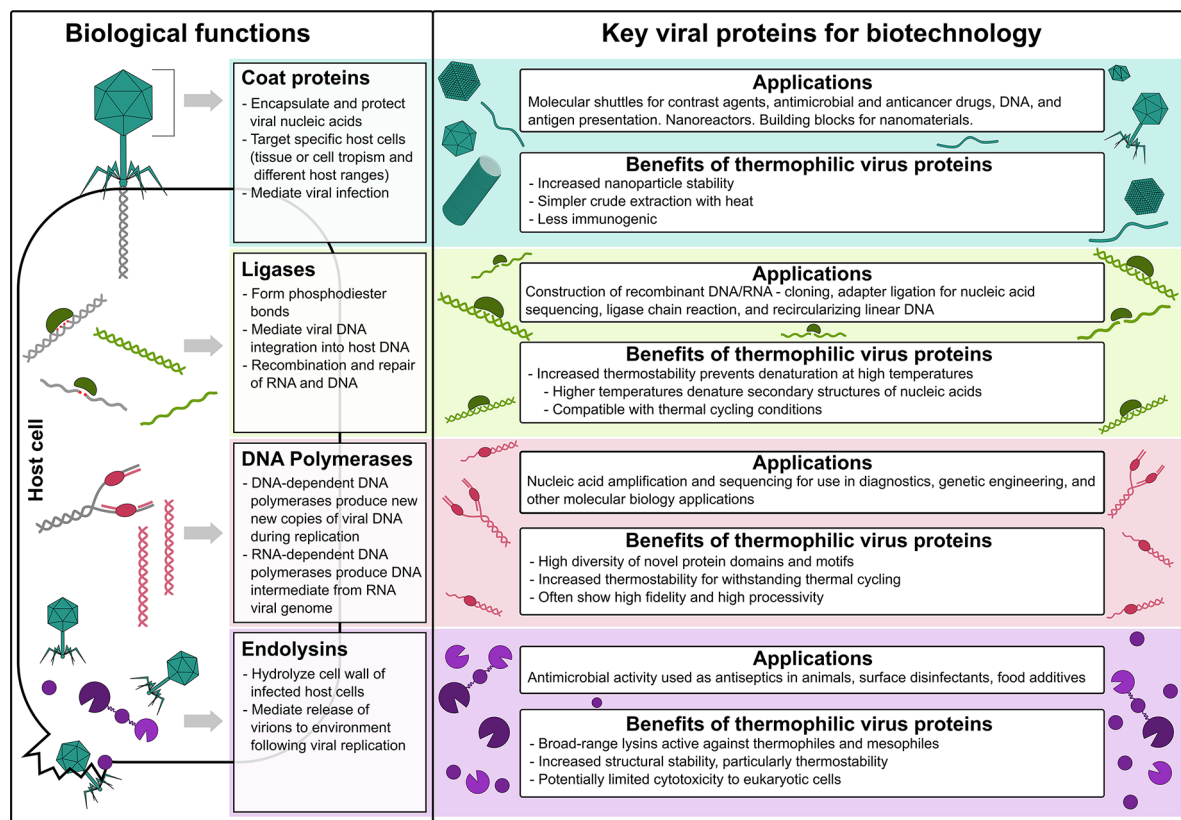


Figure 1. Graphical summary of proteins discussed.

Biological functions and key biotechnological applications of thermophilic viral coat proteins, ligases, DNA polymerases, and endolysins are highlighted.

Table 2 Summary of protein families (Pfam) counts associated with the genes of interest found in UViGs from thermal environments in the IMG/VR v4 database [99] (January 25, 2023)^a

| Protein Category | Pfam queried | Marine thermal systems | | Terrestrial thermal systems | | Total |
|------------------------------------|---|------------------------|-----------------|-------------------------------------|---|---------|
| | | Hydrothermal vents | Marine volcanic | Thermal Springs-Warm/Hot/ (42–90°C) | Other terrestrial geothermal ^b | |
| DNA Polymerases | 00078, 00136, 00476, 03175, 08996, 10391, 14791, 14792, 20286 | 2,889 | 567 | 1,557 | 33 | 5,046 |
| Ligases | 01068, 01653, 03119, 03120, 04675, 04679, 09414, 09511, 11311, 13298, 14743, 18043 | 2,849 | 185 | 1,058 | 19 | 4,111 |
| Endolysins | 00959, 01183, 01476, 01510, 11125, 11860, 18341 | 2,024 | 381 | 2,102 | 6 | 4,513 |
| Coat/Capsid Proteins | 01819, 02305, 03864, 05065, 05356, 05357, 06152, 06673, 07068, 09018, 09063, 09300, 10665, 11651, 12691, 16710, 16855, 16903, 18628, 19199, 19307, 19821, 20036 | 5,264 | 608 | 4,272 | 49 | 10,193 |
| Sum per environment | - | 13,026 | 1,741 | 8,989 | 107 | 23,863 |
| Total UViGs per environment | - | 110,140 | 19,154 | 55,911 | 2,547 | 187,752 |

^aSee **Supplementary Material** for detailed information.

^bSum of the following IMG/VR categories: Thermal Springs-Runoff channels, Sediment-Thermal Springs, and Volcanic Fumaroles.

Thermus phages G20c, P7426, P2345, ϕ FA, and TSP4 were isolated from terrestrial springs on several continents. All belong to the genus *Oshimavirus* [35] and they have similar *polA* genes. The PolA from G20c, PolI_G20c, has been both structurally and functionally characterized [30]. It was shown to be structurally similar to Taq polymerase, consisting of 3'-5' exonuclease, helicase, and PolA domains, including a novel motif named S β α R near the exonuclease domain believed to play a role in substrate binding [30]. Both DNA polymerase and 3'-5' exonuclease activities were experimentally verified, with maximal DNA polymerase activity at 70°C, which is too low for thermocycling-based biotechnology applications.

In addition to these cultivated viruses, a prophage-encoded DNA polymerase within the chromosome of *Thermus antranikianii* was characterized and shown to have strong strand-displacement activity, similar to ϕ 29, but many of the amplification products were highly branched, non-specific DNA molecules. Thus, this polymerase is not suitable as a thermophilic alternative to ϕ 29 [24], but this prophage polymerase shows that thermophilic strand-displacement is possible. A thermophilic polymerase with properties similar to ϕ 29—high fidelity, high processivity, and strong strand-displacement activity—would enable high-fidelity, long-range PCR and would be of considerable biotechnology interest.

Aside from work on cultivated viruses, bioinformatic [36,37] and functional screens [25,26] for DNA polymerases in viral metagenomic DNA from diverse terrestrial hot springs revealed a group of 3'-5' proofreading exonuclease and DNA polymerase (3' exo/pol)-encoding genes within metagenomic fragments and UViGs assigned to the putative genus 'Pyrovirus', which is predicted to infect genera within the *Aquificaceae*. Comparative phylogenetics showed these unusual *polAs* spread by horizontal gene transfer among thermophilic viruses, their *Aquificota* hosts, other diverse bacteria (although only temporarily retained), and the proto-apicoplast that became a symbiotic partner of an ancestor to the eukaryotic phylum *Apicomplexa* [38]; yet, only the viral enzymes encode N-terminal helicase domains (DUF927). The interdomain lateral gene transfers of these large and unique *polAs* suggest they may be associated with dispersal of diversity-generating mechanisms between geothermal and moderate-temperature biomes [38]. An engineered fusion between a 'Pyrovirus' PolA enzyme with the *Sulfolobus solfataricus* Sso7d DNA binding protein [39], called PyroPhage 3173 PolA, was shown to have three novel characteristics, namely (i) reverse transcriptase (RT) activity, (ii) DNA polymerase strand-displacement activity, and (iii) thermostability, which enabled RT-PCR and reverse transcription loop-mediated isothermal amplification (RT-LAMP) [25,26]. PyroPhage 3173 PolA had an error rate approximately 10 times lower than Taq polymerase, and the enzyme was commercialized by Lucigen Corporation (acquired by LCG, Middleton, Wisconsin). Amino acid swapping with *polA* genes from related UViGs, and domain swapping with the Taq polymerase 5'-3' exonuclease resulted in a recombinant enzyme, called Magma DNA polymerase, that was more thermophilic, more accurate (as low as 1 error in 10⁶ nucleotides), and performed better in reverse-transcriptase PCR applications [40].

Viral DNA-directed primase-polymerase-like proteins are predicted to have additional roles in DNA and/or RNA priming, as well as damage-tolerant DNA polymerase activity [41], and at least one such unusual thermophilic enzyme from unclassified deep-sea vent phage NrS-1, infecting *Nitratiruptor* sp. SB155-2 [42], was shown to be functional [43,44]. This enzyme has features found in DNA polymerases (DNA-dependent polymerization), primases (primer-free DNA strand synthesis initiation), helicases (strand displacement), and RNA polymerases (RNA-dependent polymerization), and could be useful in several potential applications, such as primer-free, isothermal whole-genome amplification.

Ligases

DNA and RNA ligases catalyze the formation of phosphodiester bonds between 5'-phosphate and 3'-hydroxyl groups, with activity on DNA or RNA, respectively [45]. These enzymes serve critical functions *in vivo*, including DNA replication and recombination, somatic generation of immune diversity, nucleic acid editing, and DNA/RNA repair (Table 1). Biotechnology applications of ligases include decades-old technologies such as construction of recombinant plasmids or viruses, but also emerging technologies such as library preparation for DNA and microRNA sequencing [46], single nucleotide polymorphism diagnostics using the ligase chain reaction [47], and synthetic gene construction via Gibson assembly, a cornerstone technology of modern synthetic biology [48].

Most biotechnology applications use ligases from bacteriophage T4 [46,47], but T4 DNA and RNA ligases are incompatible with denaturation conditions necessary for ligation chain reaction (90°C) [49–51], are unstable at temperatures used for Gibson assembly (typically 50°C) [48], have limited efficiency due to competition with secondary structures [52], and have low fidelity due to off-target base pairing [52]. More than 25 ligases from archaea have been functionally characterized, with possible biotechnology applications including Gibson assembly [53–55], ligase chain reaction [54,56], 5'-adenylation [57,58], and RNA sequencing [59]. These enzymes have recently been reviewed [60].

Comparatively less work has been done to characterize ligases from thermophilic viruses, and all published work to date has focused on moderately thermophilic ATP-dependent RNA ligase 1 enzymes. In 2003, a thermophilic homolog of T4 RNA ligase 1 from *Rhodothermus* phage RM378, an unclassified myovirus, was shown to have optimal activity at 64°C [19,61]. This enzyme could substitute for T4 RNA ligase 1 in RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) at 60°C and was patented and commercialized for that purpose. In 2005, another thermostable RNA ligase 1 from the unclassified virus *Thermus* phage Ph2119, TS2126 RNA ligase, was also characterized [19,62]. The TS2126 RNA ligase had ~30 times higher specific activity compared with T4 RNA ligase in phosphatase protection assays with a temperature optimum of 70–75°C and it was also more effective with ssDNA ligation. This enzyme complexes with adenylated donors rapidly, with a slower ligation activity, which strongly favors intramolecular ligations. This property has been exploited for 5' preadenylation of DNA oligonucleotide adapters during cDNA library preparation [63] and for circularization of single-stranded DNA or RNA templates for rolling-circle replication or rolling-circle transcription experiments. Kits for the latter produce virtually no linear or circular concatemers and have been trademarked as CircLigase™ ssDNA Ligase and CircLigase™ II ssDNA Ligase by Epicentre (acquired by Illumina, Madison, WI, U.S.A.).

Endolysins

All viruses require mechanisms to escape infected host cells following replication and assembly, and enzyme systems for this purpose are as diverse as the cell envelopes of their host prokaryotes [64–66]. Endolysins of mesophilic bacteriophages are hard to purify, have limited stability and activity under industrial conditions, and are typically highly specific, with many showing activity only against one host species or just a group of strains [67,68]. These shortcomings limit their use as antimicrobials, yet some evidence suggests these limitations may be overcome by their thermophilic counterparts. Currently, several native endolysins from thermophilic viruses and thermophilic recombinant endolysins have been investigated for broad-range applications (Table 1).

Two endolysins from unclassified cultivated phages infecting *T. scotoductus* strain MAT2119, phage vB_Tsc2631 and phage Ph2119, were shown to be homologs of T3 and T7. These endolysins lysed *Thermus* strains, *Deinococcus radiodurans*, and also Gram-negative mesophiles such as *E. coli*, *Salmonella enterica*, *Serratia marcescens*, and *Pseudomonas fluorescens* [69,70]. The Ph2119 endolysin retained 87% activity at 95°C, and the endolysin from vB_Tsc2631 retained 65% activity at 95°C. The endolysin from vB_Tsc2631 has seven charged arginine amino acids near the N-terminus, making the endolysin act like polycationic antibacterial peptides that form pores in cell membranes, allowing for the catalytic center of the endolysin to interact with the peptidoglycan underneath [71]. Arginine is also more thermostable than other positively charged amino acids because its functional group mimics guanidinium [71,72]. The active site of the vB_Tsc2631 is also believed to bind Zn²⁺ ions for catalytic functions and structural stability. These properties are natural advantages necessary for thermophily that overcome many of the obstacles that limit the utility of mesophilic endolysins as antimicrobials.

The endolysin of *Oshimavirus* TSP4 (also known as *Thermus* phage TSP4), TSPphg, was expressed in *E. coli*, purified, and shown to reduce *Staphylococcus aureus* infections in mice, offering promise for clinical treatments for bacterial infections [73]. Further *in vitro* testing showed antimicrobial activity against Gram-negative *S. enterica*, *Klebsiella pneumoniae*, *E. coli*, and Gram-positive *Bacillus subtilis*. The broad antimicrobial activity of TSPphg may be due to strong interactions with peptidoglycan, owing to the six positively charged amino acids near the N-terminus, similar to the endolysin of vB_Tsc2631 [70,73].

Another broad range endolysin, MMPphg, was found in *Meiothermus* phage MMP17, an unclassified myovirus. Purified MMPphg has optimal activity at 65–70°C and lyses both Gram-positive and Gram-negative bacteria, including *E. coli*, *S. aureus*, *S. enterica*, and *Shigella dysenteriae*, and eight different antibiotic-resistant strains of *K. pneumoniae* [74]. The C-terminus of the lysin also contains six positively charged amino acids. The MMP17 endolysin was later artificially fused with TSPphg and the recombinant enzyme, MLTphg, showed higher antimicrobial activity *in vitro* than either individual endolysin [75].

The other two functionally characterized endolysins are from viruses infecting Gram-positive thermophiles. GVE2, an unclassified siphovirus that infects *Geobacillus* sp. E263, encodes an endolysin believed to interact with a host eukaryotic-type ABC transporter to lyse host cells at temperatures from 55 to 90°C [76,77]. Because this endolysin is the first known to interact with ABC transporter proteins, it was further investigated as a potential antimicrobial [78]. The catalytic domain of the GVE2 endolysin was fused to peptidoglycan-binding domains from endolysins of several different *Clostridium perfringens* viruses. The result was chimeric endolysins that could operate up to 70°C, making these enzymes a potential antibiotic treatment for animals that can be added to their heat-sterilized feed [78].

Recently, an endolysin from *Saundersvirus* Tp84 (also known as *Geobacillus* virus TP-84) was investigated as a potential disinfectant for surfaces at high temperatures [79]. This endolysin had activity throughout the temperature range of 30–70°C and inhibited biofilm formation by *Pseudomonas aeruginosa*, *Streptococcus pyogenes*, and *S. aureus*. Extensive human safety testing is recommended for all endolysins to ensure they are safe for consumption if used as additives, and further testing on the long-term stability of these endolysins would be required to evaluate their potential use [78,79].

Phage depolymerases, including both hydrolases and lyases, have been investigated for their ability to degrade polysaccharides or lipids depending on the host's envelope [80–82]. Recent research has suggested that a cocktail of endolysins and envelope depolymerases would produce a greater antimicrobial effect on biofilm-forming bacteria such as *P. aeruginosa* [81,83,84]; however, investigation of thermophilic envelope depolymerases remains sparse. Although several genes from thermophilic phages are annotated as encoding some form of envelope depolymerase [84], evidence for expression of these enzymes are limited to the formation of halos around clear plaques in eight thermophilic *Geobacillus* phages, including TP-84 [79,84]. Due to the relative lack of functional studies performed on the efficacy of thermophilic depolymerases and endolysin cocktails, this presents a notable target for research and development of possible biotechnological applications.

Coat proteins

Viruses consist of nucleic acids encapsulated by protein coats (capsids) that comprise numerous copies of one or more coat protein subunits. As part of the virion, capsids protect nucleic acids [85,86], serve as vehicles for transport that can target specific cells [85], and mediate introduction of nucleic acids into host cells during infection [85,86]. As these particles typically have natural tropism toward certain cell types, and are often resistant to immune defense systems, coat proteins are excellent candidates as nano-traffickers in biomedicine. These coat proteins self-assemble [60,85,87] and inclusion of certain protein domains within them can result in highly specific targeted delivery of compounds [85,88,89].

A variety of molecules can be encapsulated by capsids, resulting in virus or virus-like nanoparticles [19,59,85,90]. These nanoparticles can be used in biomedicine for delivery of contrast agents for medical imaging [85,88,89,91], anticancer or antimicrobial drugs [85,88,89,91], antigen-presenting platforms for vaccines [85,87,91], or engineered shuttles for genetic material in gene therapy [85,91,92]. Additionally, these capsids can also serve as nanoreactors [85,89,90] and can be used in non-medical applications like transport of inorganic compounds and production of nanomaterials [85,93]. However, limitations of capsids from mesophilic viruses infecting animals, plants, or mesophilic microorganisms include their limited chemical and physical stability [85,94], challenges purifying nucleic acid-free capsids from infected host cells or expression hosts [90], and residual immunogenicity [85,89,94], which may result in clearance of nanoparticles from the system before achieving the desired effect [85,94]. Several of these shortcomings can be alleviated by capsids derived from thermophilic viruses (Table 1).

The first thermophilic viral capsids tested for chemical and physical stability in different solvents, and for availability of ligand attachment sites, were those from *Icerudivirus* SIRV2 (also known as *Sulfolobus islandicus* rod-shaped virus 2) [95]. To assess the stability of SIRV2 particles, the structural integrity and infectivity of virions was assessed following incubations in DMSO and ethanol. SIRV2 particles remained intact and infective for 6 days in 20% ethanol, 20% DMSO, or 50% DMSO, respectively, and remained intact in up to 50% ethanol [95]. With its high stability in DMSO, a solvent commonly used for bioconjugation applications, the availability of ligand attachment sites was evaluated through biotinylation of the SIRV2 particles using different compounds to identify reactive carboxylates, carbohydrates, and amines [95]. With amine reactivity found only in the minor coat protein subunits, located at the ends of the rod-shaped virions, and reactive carboxylates and carbohydrates found in both minor and major coat protein subunits, a broad variety of functional groups can be conjugated to these viral nanoparticles, including spatially specific selective bioconjugation to only the minor coat protein subunits [95,96].

More recently, an unclassified virus in the *Bicaudaviridae*, *Sulfolobus* monocaudavirus 1 (SMV1), was tested extensively for potential biotechnology applications, especially as a potential nano-trafficker for biomedical use. SMV1 particles were treated with ethanol, DMSO, simulated gastric fluid, and simulated intestinal fluid solutions to assess stability through *S. islandicus* plaque assays, and particles remained infective for up to 6 days [94]. Subsequently, SMV1 particles were passed through the gastrointestinal tracts of mice or incubated with human intestinal organoids, where limited immune responses were elicited, and no SMV1 particles were detected in off-target organs or tissues [94]. Overall, SMV1 particles fared better than *Inovirus* M13KE, an *E. coli* phage used for comparison, in both the mice and organoids [94], showing great promise as both molecular delivery systems and antigen-presentation platforms.

Expansion of bioprospecting through viral metagenomes

Most research to date has focused on only viruses infecting cultivated thermophilic archaea and bacteria [14,19], thus limiting the overall breadth of our understanding of the thermophilic virome. Given the limited diversity of cultivated thermophilic prokaryotes [97,98] and their viruses [14], and the extremely limited number of biochemically characterized proteins from thermophilic viruses, we propose that UViGs represent a vast resource for the biotechnology sector. To begin to evaluate the potential resource, we searched for pfams that are diagnostic of the four protein groups discussed here—DNA polymerases, ligases, endolysins, and coat proteins—in the Integrated Microbial Genomes/Virus (IMG/VR) v4 database, which contains over 5.5 million high-confidence viral genome contigs from a wide range of biomes [99], focusing on marine and terrestrial geothermal systems (Table 2; Supplementary Material). This search revealed >20,000 potential matches to these four protein groups from >185,000 UViGs, with the largest amount coming from marine hydrothermal systems, followed by high-temperature terrestrial geothermal systems, and the largest protein category being coat proteins, followed by polymerases, endolysins, and ligases. For example, we identified over 5,000 putative polymerases; the vast diversity of polymerase architectures driven by adaptation to thermal environments [88] are ripe for biotechnology exploration.

Despite the vast resources available in UViGs, there are currently some limitations. First, given the genetic and biochemical diversity encoded by UViGs, a vast diversity of biotechnologically useful functions resides in poorly annotated genes that are difficult to bioprospect based on sequence similarity. This hidden resource could be systematically explored using artificial intelligence platforms, including those examining protein folds, which are more highly conserved than primary sequence information [100]. Another limitation is the systematic focus on dsDNA viruses due to predominant library preparation methods used for viral metagenomics, which unfortunately exclude six of the seven groups of the Baltimore classification system [101,102]. This heavy focus on dsDNA viruses ignores many novel architectures of undiscovered viruses and certainly biases our understanding of the thermophilic virome, although known recombination between natural virus populations with different types of genomes may relieve this limitation to a degree [103,104]. Groups such as the RNA Virus Discovery Consortium have deposited many more RNA viral metagenomes into IMG/VR [99] through RNA extraction and reverse transcription prior to or during library preparation, with advancements to date in marine [105,106], sediment [107], and terrestrial ecosystems [108,109], including thermal springs and others [107,110]. Bioinformatic pipelines like VirSorter are also increasing accuracy and providing support for RNA viruses [111]. A separate problem is the identification and classification of metagenomic contigs as UViGs in the first place, which can lead to both false negatives and false positives, although community standards have been developed to improve communication of UViG quality, with UViGs categorized as high quality (>90% completeness), medium quality (50–90% completeness), low quality (<50% completeness), and unsure quality (>120% or no completeness estimate) [99,112]. Despite these challenges, we contend that UViGs provide an immense and poorly explored resource for bioprospecting the global thermophile virome.

In recognition of this resource, projects focused on exploring the sequence coverage of the virosphere are seeing increasing support. This is evident in community-driven sequencing efforts supported by the Joint Genome Institute (e.g., OSTI 1488193, Award 503441), implementation of analysis tools for viruses in collaborative cyberinfrastructure, like CyVerse [113], and the RNA Virus Discovery Consortium [110]. Additionally, in 2016 to 2020, the European Union funded the Virus-X project—Viral Metagenomics for Innovation Value—at €8 million. These projects have expanded the sequence coverage of the global virosphere, expressed and characterized novel proteins [114–116], analyzing crystal structures of expressed genes to aid in functional identification [117], and improved methods to identify and interpret UViGs [111,113], including algorithms to identify host-virus pairs [118] and to improve annotation of uncharacterized viral genes through protein clustering [119,120]. These advancements show not only that thermophilic viral enzymes are an expanding topic of importance for biotechnology, but also that infrastructure and data mining tools are improving to better support the ever-expanding UViG dataset.

Summary

- Viral proteins, particularly polymerases, ligases, endolysins and coat proteins, provide a bountiful but underutilized toolbox for the biotechnology industry to explore.
- Applications involving thermophilic viral proteins provide several benefits that overcome some of the shortcomings of their mesophilic counterparts.

- Bioprospecting of genomes from uncultivated viruses provides a vast and underexplored resource that overcomes the primary impediment of cultivability.

Competing Interests

David Mead is an employee of Varigen Biosciences, doing business as Varizymes, and has worked on commercializing thermophilic enzymes.

Funding

Funding was provided by a grant from the Human Genome Research program at the National Institutes of Health [grant number 1R43HG012181-01]. Additional funding was provided by the UNLV Office of Economic Development and the Troesh Center for Entrepreneurship and Innovation.

Author Contribution

All authors discussed the topic and outline and contributed to the bibliography. All authors contributed to the first draft, which was edited by all authors. R.K.D. and M.P. conducted bioinformatics searches of UViGs from geothermal environments and relevant pfams and compiled data.

Abbreviations

ICTV, International Committee on Taxonomy of Viruses; IMG/VR, Integrated Microbial Genomes/Virus; PCR, polymerase chain reaction; Pfams, protein families; RT-LAMP, reverse transcription loop-mediated isothermal amplification; RLM-RACE, RNA ligase-mediated-rapid amplification of cDNA ends; SIRV2, *Sulfolobus islandicus* rod-shaped virus 2; SMV1, *Sulfolobus* monocaudavirus 1; UViG, uncultivated viral genomes.

References

- Chien, A., Edgar, D.B. and Trela, J.M. (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* **127**, 1550–1557, <https://doi.org/10.1128/jb.127.3.1550-1557.1976>
- Ishino, S. and Ishino, Y. (2014) DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Front. Microbiol.* **5**, 465, <https://doi.org/10.3389/fmicb.2014.00465>
- Barnard, D., Casanueva, A., Tuffin, M. and Cowan, D. (2010) Extremophiles in biofuel synthesis. *Environ. Technol.* **31**, 871–888, <https://doi.org/10.1080/09593331003710236>
- Blumer-Schuetz, S.E., Brown, S.D., Sander, K.B., Bayer, E.A., Kataeva, I., Zurawski, J.V. et al. (2014) Thermophilic lignocellulose deconstruction. *FEMS Microbiol. Rev.* **38**, 393–448, <https://doi.org/10.1111/1574-6976.12044>
- Vielle, C. and Zeikus, G.J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1–43, <https://doi.org/10.1128/MMBR.65.1.1-43.2001>
- Schnell, J. and Kula, M.R. (1989) Investigations of heat treatment to improve the isolation of intracellular enzymes. *Bioprocess Eng.* **4**, 129–138, <https://doi.org/10.1007/BF00369762>
- Olichon, A., Schweizer, D., Muyldermans, S. and de Marco, A. (2007) Heating as a rapid purification method for recovering correctly-folded thermotolerant VH and VHH domains. *BMC Biotechnol.* **7**, 7, <https://doi.org/10.1186/1472-6750-7-7>
- Sakhel, B., Jayanthi, S., Muhoza, D., Okoto, P., Kumar, T.K.S. and Adams, P. (2021) Simplification of the purification of heat stable recombinant low molecular weight proteins and peptides from GST-fusion products. *J. Chromatogr. B.* **1172**, 122627, <https://doi.org/10.1016/j.jchromb.2021.122627>
- Stone, E., Campbell, K., Grant, I. and McAuliffe, O. (2019) Understanding and exploiting phage-host interactions. *Viruses* **11**, 567, <https://doi.org/10.3390/v11060567>
- Pederson, D.M., Welsh, L.C., Marvin, D.A., Sampson, M., Perham, R.N., Yu, M. et al. (2001) The protein capsid of filamentous bacteriophage PH75 from *Thermus thermophilus*. *J. Mol. Biol.* **309**, 401–421, <https://doi.org/10.1006/jmbi.2001.4685>
- Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V. and Krupovic, M. (2017) The enigmatic archaeal virosphere. *Nat. Rev. Microbiol.* **15**, 724–739, <https://doi.org/10.1038/nrmicro.2017.125>
- Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E. et al. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7716–7720, <https://doi.org/10.1073/pnas.0401773101>
- Prangishvili, D. (2013) The wonderful world of archaeal viruses. *Annu. Rev. Microbiol.* **67**, 565–585, <https://doi.org/10.1146/annurev-micro-092412-155633>
- Zablocki, O., van Zyl, L. and Trindade, M. (2018) Biogeography and taxonomic overview of terrestrial hot spring thermophilic phages. *Extremophiles* **22**, 827–837, <https://doi.org/10.1007/s00792-018-1052-5>
- Turner, D., Shkoporov, A.N., Lood, C., Millard, A.D., Dutilh, B.E., Alfenas-Zerbini, P. et al. (2023) Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch. Virol.* **168**, 74, <https://doi.org/10.1007/s00705-022-05694-2>

- 16 Geslin, C., Le Romancer, M., Erauso, G., Gaillard, M., Perrot, G. and Prieur, D. (2003) PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, “*Pyrococcus abyssi*”. *J. Bacteriol.* **185**, 3888–3894, <https://doi.org/10.1128/JB.185.13.3888-3894.2003>
- 17 Mochizuki, T., Yoshida, T., Tanaka, R., Forterre, P., Sako, Y. and Prangishvili, D. (2010) Diversity of viruses of the hyperthermophilic archaeal genus *Aeropyrum*, and isolation of the *Aeropyrum pernix* bacilliform virus 1, APBV1, the first representative of the family *Clavaviridae*. *Virology* **402**, 347–354, <https://doi.org/10.1016/j.virol.2010.03.046>
- 18 Prangishvili, D. and Garrett, R.A. (2005) Viruses of hyperthermophilic *Crenarchaea*. *Trends Microbiol.* **13**, 535–542, <https://doi.org/10.1016/j.tim.2005.08.013>
- 19 Uldahl, K. and Peng, X. (2013) Biology, biodiversity and application of thermophilic viruses. *Thermophilic microbes in environmental and industrial biotechnology*, pp. 271–304, Springer, Dordrecht, https://doi.org/10.1007/978-94-007-5899-5_10
- 20 Fremin, B.J., Bhatt, A.S., Kyrpides, N.C., Small, G.P. and (GP-SmORF), O. R. F. and Consortium (2022) Thousands of small, novel genes predicted in global phage genomes. *Cell Rep.* **39**, 110984, <https://doi.org/10.1016/j.celrep.2022.110984>
- 21 Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T. et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491, <https://doi.org/10.1126/science.2448875>
- 22 Tindall, K.R. and Kunkel, T.A. (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry-US* **27**, 6008–6013, <https://doi.org/10.1021/bi00416a027>
- 23 Dean, F.B., Nelson, J.R., Giesler, T.L. and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–1099, <https://doi.org/10.1101/gr.180501>
- 24 Hjorleifsdottir, S., Blondal, T., Evarsson, A., Fridjónsson, O.H. et al. (2014) Isothermal DNA amplification by a novel and non-ubiquitous *Thermus* polymerase A. *Curr. Biotechnol.* **3**, 76–86, <https://doi.org/10.2174/22115501113026660036>
- 25 Chandler, Y., Koelbl, J., Puckett, J., Moser, M.J., Klingele, A.J., Liles, M.R. et al. (2014) A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (LAMP). *Front. Microbiol.* **5**, 395, <https://doi.org/10.3389/fmicb.2014.00395>
- 26 Moser, M.J., DiFrancesco, R.A., Gowda, K., Klingele, A.J., Sugar, D.R., Stocki, S. et al. (2012) Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme. *PLoS ONE* **7**, e38371, <https://doi.org/10.1371/journal.pone.0038371>
- 27 Dorawa, S., Werbowy, O., Plotka, M., Kaczorowska, A.K., Makowska, J., Kozłowski, L.P. et al. (2022) Molecular characterization of a DNA polymerase from *Thermus thermophilus* MAT72 phage vB.Tt72: A novel Type-A family enzyme with strong proofreading activity. *IJMS* **23**, 7945, <https://doi.org/10.3390/ijms23147945>
- 28 Naryshkina, T., Liu, J., Florens, L., Swanson, S.K., Pavlov, A.R., Pavlova, N.V. et al. (2006) *Thermus thermophilus* bacteriophage ϕ YS40 genome and proteomic characterization of virions. *J. Mol. Biol.* **364**, 667–677, <https://doi.org/10.1016/j.jmb.2006.08.087>
- 29 Tamakoshi, M., Murakami, A., Sugisawa, M., Tsuneizumi, K., Takeda, S., Saheki, T. et al. (2011) Genomic and proteomic characterization of the large *Myoviridae* bacteriophage ϕ TMA of the extreme thermophile *Thermus thermophilus*. *Bacteriophage* **1**, 152–164, <https://doi.org/10.4161/bact.1.3.16712>
- 30 Ahlqvist, J., Linares-Pastén, J.A., Jasilionis, A., Welin, M., Håkansson, M., Svensson, L.A. et al. (2022) Crystal structure of DNA polymerase I from *Thermus* phage G20c. *Acta Crystallogr. D Struct. Biol.* **78**, 1384–1398, <https://doi.org/10.1107/S2059798322009895>
- 31 Minakhin, L., Goel, M., Berdygulova, Z., Ramanculov, E., Florens, L., Glazko, G. et al. (2008) Genome comparison and proteomic characterization of *Thermus thermophilus* bacteriophages P23-45 and P74-26: siphoviruses with triplex-forming sequences and the longest known tails. *J. Mol. Biol.* **378**, 468–480, <https://doi.org/10.1016/j.jmb.2008.02.018>
- 32 Lopatina, A., Medvedeva, S., Artamonova, D., Kolesnik, M., Sitnik, V., Ispatov, Y. et al. (2019) Natural diversity of CRISPR spacers of *Thermus*: evidence of local spacer acquisition and global spacer exchange. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20180092, <https://doi.org/10.1098/rstb.2018.0092>
- 33 Lin, L., Hong, W., Ji, X., Han, J., Huang, L. and Wei, Y. (2010) Isolation and characterization of an extremely long tail *Thermus* bacteriophage from Tengchong hot springs in China. *J. Basic Microbiol.* **50**, 452–456, <https://doi.org/10.1002/jobm.201000116>
- 34 Ahlqvist, J., Linares-Pastén, J.A., Håkansson, M., Jasilionis, A., Kwiatkowska-Semrau, K., Fridjónsson, O.H. et al. (2022) Crystal structure and initial characterization of a novel archaeal-like Holliday junction-resolving enzyme from *Thermus thermophilus* phage Tth15-6. *Acta Crystallogr. D Struct. Biol.* **78**, 212–227, <https://doi.org/10.1107/S2059798321012298>
- 35 Cook, R., Brown, N., Redgwel, T., Rihtman, B., Barnes, M., Clokie, M. et al. (2021) Infrastructure for a PHAge REference database: identification of large-scale biases in the current collection of cultured phage genomes. *Phage* **2**, 214–223, <https://doi.org/10.1089/phage.2021.0007>
- 36 Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M. and Mead, D. (2008) Assembly of viral metagenomes from Yellowstone Hot Springs. *Appl. Environ. Microbiol.* **74**, 4164–4174, <https://doi.org/10.1128/AEM.02598-07>
- 37 Palmer, M., Hedlund, B.P., Roux, S., Tsourkas, P.K., Doss, R.K., Stamereilers, C. et al. (2020) Diversity and distribution of a novel genus of hyperthermophilic *Aquificae* viruses encoding a proof-reading Family-A DNA polymerase. *Front. Microbiol.* **11**, 2809, <https://doi.org/10.3389/fmicb.2020.583361>
- 38 Schoenfeld, T.W., Murugapiran, S.K., Dodsworth, J.A., Floyd, S., Lodes, M., Mead, D.A. et al. (2013) Lateral gene transfer of family A DNA polymerases between thermophilic viruses, aquificae, and apicomplexa. *Mol. Biol. Evol.* **30**, 1653–1664, <https://doi.org/10.1093/molbev/mst078>
- 39 Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M. and Vander Horn, P.B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res.* **32**, 1197–1207, <https://doi.org/10.1093/nar/gkh271>
- 40 Heller, R.C., Chung, S., Crissy, K., Dumas, K., Schuster, D. and Schoenfeld, T.W. (2019) Engineering of a thermostable viral polymerase using metagenome-derived diversity for highly sensitive and specific RT-PCR. *Nucleic Acids Res.* **47**, 3619–3630, <https://doi.org/10.1093/nar/gkz104>
- 41 Guillian, T.A., Jozwiakowski, S.K., Ehlinger, A., Barnes, R.P., Rudd, S.G., Bailey, L.J. et al. (2015) Human PrimPol is a highly error-prone polymerase regulated by single-stranded DNA binding proteins. *Nucleic Acids Res.* **43**, 1056–1068, <https://doi.org/10.1093/nar/gku1321>

- 42 Yoshida-Takashima, Y., Takaki, Y., Shimamura, S., Nunoura, T. and Takai, K. (2013) Genome sequence of a novel deep-sea vent epsilonproteobacterial phage provides new insight into the co-evolution of *Epsilonproteobacteria* and their phages. *Extremophiles* **17**, 405–419, <https://doi.org/10.1007/s00792-013-0529-5>
- 43 Zhu, B., Wang, L., Mitsunobu, H., Lu, X., Hernandez, A.J., Yoshida-Takashima, Y. et al. (2017) Deep-sea vent phage DNA polymerase specifically initiates DNA synthesis in the absence of primers. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2310–E2318, <https://doi.org/10.1073/pnas.1700280114>
- 44 Chen, X., Su, S., Chen, Y., Gao, Y., Li, Y., Shao, Z. et al. (2020) Structural studies reveal a ring-shaped architecture of deep-sea vent phage NfS-1 polymerase. *Nucleic Acids Res.* **48**, 3343–3355, <https://doi.org/10.1093/nar/gkaa071>
- 45 Tomkinson, A.E., Vijayakumar, S., Pascal, J.M. and Ellenberger, T. (2006) DNA ligases: structure, reaction mechanism, and function. *Chem. Rev.* **106**, 687–699, <https://doi.org/10.1021/cr040498d>
- 46 Baran-Gale, J., Kurtz, C.L., Erdos, M.R., Sison, C., Young, A., Fannin, E.E. et al. (2015) Addressing bias in small RNA library preparation for sequencing: A new protocol recovers microRNAs that evade capture by current methods. *Front Genet.* **6**, 352, <https://doi.org/10.3389/fgene.2015.00352>
- 47 Zhang, W., Liu, K., Zhang, P., Cheng, W., Zhang, Y., Li, L. et al. (2021) All-in-one approaches for rapid and highly specific quantification of single nucleotide polymorphisms based on ligase detection reaction using molecular beacons as turn-on probes. *Talanta* **224**, 121717, <https://doi.org/10.1016/j.talanta.2020.121717>
- 48 Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345, <https://doi.org/10.1038/nmeth.1318>
- 49 Becker, A., Lyn, G., Gefter, M. and Hurwitz, J. (1967) The enzymatic repair of DNA, II. Characterization of phage-induced sealase. *Proc. Natl. Acad. Sci. U.S.A.* **58**, 1996–2003, <https://doi.org/10.1073/pnas.58.5.1996>
- 50 Cozzarelli, N.R., Melechen, N.E., Jovin, T.M. and Kornberg, A. (1967) Polynucleotide cellulose as a substrate for a polynucleotide ligase induced by phage T4. *Biochem. Biophys. Res. Co.* **28**, 578–586, [https://doi.org/10.1016/0006-291X\(67\)90353-1](https://doi.org/10.1016/0006-291X(67)90353-1)
- 51 Weiss, B. and Richardson, C.C. (1967) Enzymatic breakage and joining of deoxyribonucleic acid, I. repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* **57**, 1021–1028, <https://doi.org/10.1073/pnas.57.4.1021>
- 52 Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y. and Robb, G.B. (2012) Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* **40**, e54, <https://doi.org/10.1093/nar/gkr1263>
- 53 Rolland, J.-I., Gueguen, Y., Persillon, C., Masson, J.-M. and Dietrich, J. (2004) Characterization of a thermophilic DNA ligase from the archaeon *Thermococcus fumicolans*. *FEMS Microbiol. Lett.* **236**, 267–273, <https://doi.org/10.1111/j.1574-6968.2004.tb09657.x>
- 54 Seo, M.S., Kim, Y.J., Choi, J.J., Lee, M.S., Kim, J.H., Lee, J.-H. et al. (2007) Cloning and expression of a DNA ligase from the hyperthermophilic archaeon *Staphylothermus marinus* and properties of the enzyme. *J. Biotechnol.* **128**, 519–530, <https://doi.org/10.1016/j.jbiotec.2006.09.024>
- 55 Smagin, V.A., Mardanov, A.V., Bonch-Osmolovskaya, E.A. and Ravin, N.V. (2008) Isolation and characteristics of new thermostable DNA ligase from archaea of the genus *Thermococcus*. *Appl. Biochem. Microbiol.* **44**, 473–477, <https://doi.org/10.1134/S0003683808050037>
- 56 Tanabe, M., Ishino, S., Yohda, M., Morikawa, K., Ishino, Y. and Nishida, H. (2012) Structure-based mutational study of an archaeal DNA ligase towards improvement of ligation activity. *ChemBioChem* **13**, 2575–2582, <https://doi.org/10.1002/cbic.201200336>
- 57 Zhelkovsky, A.M. and McReynolds, L.A. (2011) Simple and efficient synthesis of 5' pre-adenylated DNA using thermostable RNA ligase. *Nucleic Acids Res.* **39**, e117, <https://doi.org/10.1093/nar/gkr544>
- 58 Sriskanda, V., Kelman, Z., Hurwitz, J. and Shuman, S. (2000) Characterization of an ATP-dependent DNA ligase from the thermophilic archaeon *Methanobacterium thermoautotrophicum*. *Nucleic Acids Res.* **28**, 2221–2228, <https://doi.org/10.1093/nar/28.11.2221>
- 59 Zhang, L. and Tripathi, A. (2016) Archaeal RNA ligase from *Thermococcus kodakarensis* for template dependent ligation. *RNA Biol.* **14**, 36–44, <https://doi.org/10.1080/15476286.2016.1239688>
- 60 Straub, C.T., Counts, J.A., Nguyen, D.M.N., Wu, C.-H., Zeldes, B.M., Crosby, J.R. et al. (2018) Biotechnology of extremely thermophilic archaea. *FEMS Microbiol. Rev.* **42**, 543–578, <https://doi.org/10.1093/femsre/fuy012>
- 61 Blondal, T., Hjorleifsdottir, S.H., Fridjonsson, O.F., Avarsson, A., Skirnisdottir, S., Hermannsdottir, A.G. et al. (2003) Discovery and characterization of a thermostable bacteriophage RNA ligase homologous to T4 RNA ligase 1. *Nucleic Acids Res.* **31**, 7247–7254, <https://doi.org/10.1093/nar/gkg914>
- 62 Blondal, T., Thorisdottir, A., Unnsteinsdottir, U., Hjorleifsdottir, S., Avarsson, A., Ernstsson, S. et al. (2005) Isolation and characterization of a thermostable RNA ligase 1 from a *Thermus scotoductus* bacteriophage TS2126 with good single-stranded DNA ligation properties. *Nucleic Acids Res.* **33**, 135–142, <https://doi.org/10.1093/nar/gki149>
- 63 Lama, L. and Ryan, K. (2016) Adenylation of small RNA sequencing adapters using the TS2126 RNA ligase I. *RNA* **22**, 155–161, <https://doi.org/10.1261/ra.054999.115>
- 64 Baquero, D.P., Liu, J. and Prangishvili, D. (2021) Egress of archaeal viruses. *Cell. Microbiol.* **23**, <https://doi.org/10.1111/cmi.13394>
- 65 Saier, Jr, M.H. and Reddy, B.L. (2015) Holins in bacteria, eukaryotes, and archaea: multifunctional xenologues with potential biotechnological and biomedical applications. *J. Bacteriol.* **197**, 7–17, <https://doi.org/10.1128/JB.02046-14>
- 66 Cahill, J. and Young, R. (2021) Release of phages from prokaryotic cells. *Encyclopedia of Virology*, pp. 501–518, Elsevier, <https://doi.org/10.1016/B978-0-12-814515-9.00074-6>
- 67 Wong, K.Y., Khair, M.H.M.M., Song, A.A.-L., Masarudin, M.J., Chong, C.M., In, L.L.A. et al. (2022) Endolysins against Streptococci as an antibiotic alternative. *Front Microbiol.* **13**, 935145, <https://doi.org/10.3389/fmicb.2022.935145>
- 68 Blasco, L., Ambroa, A., Trastoy, R., Blierot, I., Moscoso, M., Fernández-García, L. et al. (2020) *In vitro* and *in vivo* efficacy of combinations of colistin and different endolysins against clinical strains of multi-drug resistant pathogens. *Sci. Rep.* **10**, 7163, <https://doi.org/10.1038/s41598-020-64145-7>
- 69 Plotka, M., Kaczorowska, A.-K., Morzywolek, A., Makowska, J., Kozłowski, L.P., Thorisdottir, A. et al. (2015) Biochemical characterization and validation of a catalytic Site of a highly thermostable Ts2631 endolysin from the *Thermus scotoductus* phage vB_Tsc2631. *PLoS ONE* **10**, e0137374, <https://doi.org/10.1371/journal.pone.0137374>

- 70 Plotka, M., Kaczorowska, A.-K., Stefanska, A., Morzywolek, A., Fridjonsson, O.H., Dunin-Horkawicz, S. et al. (2014) Novel highly thermostable endolysin from *Thermus scotoductus* MAT2119 bacteriophage Ph2119 with amino acid sequence similarity to eukaryotic peptidoglycan recognition proteins. *Appl. Environ. Microbiol.* **80**, 886–895, <https://doi.org/10.1128/AEM.03074-13>
- 71 Plotka, M., Sancho-Vaello, E., Dorawa, S., Kaczorowska, A.-K., Kozłowski, L.P., Kaczorowski, T. et al. (2019) Structure and function of the Ts2631 endolysin of *Thermus scotoductus* phage vB_Tsc2631 with unique N-terminal extension used for peptidoglycan binding. *Sci. Rep.* **9**, 1261, <https://doi.org/10.1038/s41598-018-37417-6>
- 72 Sokalingam, S., Raghunathan, G., Soundrarajan, N. and Lee, S.G. (2012) A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE* **7**, e40410, <https://doi.org/10.1371/journal.pone.0040410>
- 73 Wang, F., Ji, X., Li, Q., Zhang, G., Peng, J., Hai, J. et al. (2020) TSPphg lysin from the extremophilic *Thermus* bacteriophage TSP4 as a potential antimicrobial agent against both Gram-negative and Gram-positive pathogenic bacteria. *Viruses* **12**, 192, <https://doi.org/10.3390/v12020192>
- 74 Wang, F., Xiong, Y., Xiao, Y., Han, J., Deng, X. and Lin, L. (2020) MMPphg from the thermophilic *Meiothermus* bacteriophage MMP17 as a potential antimicrobial agent against both Gram-negative and Gram-positive bacteria. *Virology* **17**, 130, <https://doi.org/10.1186/s12985-020-01403-0>
- 75 Wang, F., Liu, X., Deng, Z., Zhang, Y., Ji, X., Xiong, Y. et al. (2020) Design, overproduction and purification of the chimeric phage lysin MLTphg fighting against *Staphylococcus aureus*. *Processes* **8**, 1587, <https://doi.org/10.3390/pr8121587>
- 76 Walmagh, M., Briers, Y., dos Santos, S.B., Azeredo, J. and Lavigne, R. (2012) Characterization of modular bacteriophage endolysins from *Myoviridae* phages OBP_201φ2-1 and PVP-SE1. *PLoS ONE* **7**, e36991, <https://doi.org/10.1371/journal.pone.0036991>
- 77 Jin, M., Ye, T. and Zhang, X. (2013) Roles of bacteriophage GVE2 endolysin in host lysis at high temperatures. *Microbiology+* **159**, 1597–1605, <https://doi.org/10.1099/mic.0.067611-0>
- 78 Swift, S.M., Reid, K.P., Donovan, D.M. and Ramsay, T.G. (2019) Thermophile lytic enzyme fusion proteins that target *Clostridium perfringens*. *Antibiotics* **8**, 214, <https://doi.org/10.3390/antibiotics8040214>
- 79 Żebrowska, J., Żolnierkiewicz, O., Ponikowska, M., Puchalski, M., Krawczun, N., Makowska, J. et al. (2022) Cloning and characterization of a thermostable endolysin of bacteriophage TP-84 as a potential disinfectant and biofilm-removing biological agent. *IJMS* **23**, 7612, <https://doi.org/10.3390/ijms23147612>
- 80 Pires, D.P., Oliveira, H., Melo, L.D., Sillankorva, S. and Azeredo, J. (2016) Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Appl. Microbiol. Biotechnol.* **100**, 2141–2151, <https://doi.org/10.1007/s00253-015-7247-0>
- 81 Drulis-Kawa, Z., Majkowska-Skronek, G. and Maciejewska, B. (2015) Bacteriophages and phage-derived proteins—application approaches. *Curr. Med. Chem.* **22**, 1757–1773, <https://doi.org/10.2174/0929867322666150209152851>
- 82 Yan, J., Mao, J. and Xie, J. (2014) Bacteriophage polysaccharide depolymerases and biomedical applications. *BioDrugs* **28**, 265–274, <https://doi.org/10.1007/s40259-013-0081-y>
- 83 Alkawash, M.A., Sothill, J.S. and Schiller, N.L. (2006) Alginate lyase enhances antibiotic killing of mucoid *Pseudomonas aeruginosa* in biofilms. *APMIS* **114**, 131–138, <https://doi.org/10.1111/j.1600-0463.2006.apm.356.x>
- 84 Lubkowska, B., Jezewska-Frackowiak, J., Sobolewski, I. and Skowron, P.M. (2021) Bacteriophages of thermophilic ‘*Bacillus* group’ bacteria—a review. *Microorganisms* **9**, 1522, <https://doi.org/10.3390/microorganisms9071522>
- 85 Mateu, M.G. (2016) Assembly, engineering and applications of virus-based protein nanoparticles. *Adv. Exp. Med. Biol.* **940**, 83–120, https://doi.org/10.1007/978-3-319-39196-0_5
- 86 Sukenik, L., Mukhamedova, L., Procházková, M., Škubník, K., Plevka, P. and Vácha, R. (2021) Cargo release from nonenveloped viruses and virus-like nanoparticles: capsid rupture or pore formation. *ACS Nano*. **15**, 19233–19243, <https://doi.org/10.1021/acsnano.1c04814>
- 87 Zhai, L., Anderson, D., Bruckner, E. and Tumban, E. (2021) Novel expression of coat proteins from thermophilic bacteriophage φIN93 and evaluation for assembly into virus-like particles. *Protein Expr. Purif.* **187**, 105932, <https://doi.org/10.1016/j.pep.2021.105932>
- 88 Schoenfeld, T., Liles, M., Wommack, K.E., Polson, S.W., Godiska, R. and Mead, D. (2010) Functional viral metagenomics and the next generation of molecular tools. *Trends Microbiol.* **18**, 20–29, <https://doi.org/10.1016/j.tim.2009.10.001>
- 89 Glasgow, J. and Tullman-Ercek, D. (2014) Production and applications of engineered viral capsids. *Appl. Microbiol. Biot.* **98**, 5847–5858, <https://doi.org/10.1007/s00253-014-5787-3>
- 90 Dashti, N.H. and Sainsbury, F. (2020) Virus-derived nanoparticles. *Methods Mol. Biol.* **2073**, 149–162, https://doi.org/10.1007/978-1-4939-9869-2_9
- 91 Sainsbury, F. (2017) Virus-like nanoparticles: emerging tools for targeted cancer diagnostics and therapeutics. *Ther. Deliv.* **8**, 1019–1021, <https://doi.org/10.4155/tde-2017-0098>
- 92 Bryant, D.H., Bashir, A., Sinai, S., Jain, N.K., Ogden, P.J., Riley, P.F. et al. (2021) Deep diversification of an AAV capsid protein by machine learning. *Nat. Biotechnol.* **39**, 691–696, <https://doi.org/10.1038/s41587-020-00793-4>
- 93 Zhang, W., Zhang, X.E. and Li, F. (2018) Virus-based nanoparticles of Simian Virus 40 in the field of nanobiotechnology. *Biotechnol. J.* **13**, e1700619, <https://doi.org/10.1002/biot.201700619>
- 94 Uldahl, K.B., Walk, S.T., Olshefsky, S.C., Young, M.J. and Peng, X. (2017) SMV1, an extremely stable thermophilic virus platform for nanoparticle trafficking in the mammalian GI tract. *J. Appl. Microbiol.* **123**, 1286–1297, <https://doi.org/10.1111/jam.13584>
- 95 Steinmetz, N.F., Bize, A., Findlay, K.C., Lomonosoff, G.P., Manchester, M., Evans, D.J. et al. (2008) Site-specific and spatially controlled addressability of a new viral nanobuilding block: *Sulfolobus islandicus* rod-shaped virus 2. *Adv. Funct. Mater.* **18**, 3478–3486, <https://doi.org/10.1002/adfm.200800711>
- 96 DiMaio, F., Yu, X., Rensen, E., Krupovic, M., Prangishvili, D. and Egelman, E.H. (2015) A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* **348**, 914–917, <https://doi.org/10.1126/science.aaa4181>
- 97 Jiao, J.Y., Liu, L., Hua, Z.S., Fang, B.Z., Zhou, E.M., Salam, N. et al. (2021) Microbial dark matter coming to light: challenges and opportunities. *Natl. Sci. Rev.* **8**, nwa280, <https://doi.org/10.1093/nsr/nwaa280>

- 98 Shu, W.S. and Huang, L.N. (2022) Microbial diversity in extreme environments. *Nat. Rev. Microbiol.* **20**, 219–235, <https://doi.org/10.1038/s41579-021-00648-y>
- 99 Camargo, A.P., Nayfach, S., Chen, I.M.A., Palaniappan, K., Ratner, A., Chu, K. et al. (2022) IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **51**, D733–D743, <https://doi.org/10.1093/nar/gkac1037>
- 100 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O. et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589, <https://doi.org/10.1038/s41586-021-03819-2>
- 101 Kieft, K. and Anantharaman, K. (2022) Virus genomics: what is being overlooked? *Curr. Opin. Virol.* **53**, 101200, <https://doi.org/10.1016/j.coviro.2022.101200>
- 102 Ponsero, A.J. and Hurwitz, B.L. (2019) The promises and pitfalls of machine learning for detecting viruses in aquatic metagenomes. *Front. Microbiol.* **10**, 806, <https://doi.org/10.3389/fmicb.2019.00806>
- 103 Diemer, G.S. and Stedman, K.M. (2012) A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Direct* **7**, 13, <https://doi.org/10.1186/1745-6150-7-13>
- 104 Stedman, K.M. (2015) Deep recombination: RNA and ssDNA virus genes in DNA virus and host genomes. *Ann. Rev. Virol.* **2**, 203–217, <https://doi.org/10.1146/annurev-virology-100114-055127>
- 105 Zeigler Allen, L., McCrow, J.P., Ininbergs, K., Dupont, C.L., Badger, J.H., Hoffman, J.M. et al. (2017) The Baltic Sea virome: diversity and transcriptional activity of DNA and RNA viruses. *mSystems* **2**, e00125–16, <https://doi.org/10.1128/mSystems.00125-16>
- 106 Callanan, J., Stockdale, S.R., Shkoporov, A., Draper, L.A., Ross, R.P. and Hill, C. (2020) Expansion of known ssRNA phage genomes: From tens to over a thousand. *Sci. Adv.* **6**, eaay5981, <https://doi.org/10.1126/sciadv.aay5981>
- 107 Krishnamurthy, S.R., Janowski, A.B., Zhao, G., Barouch, D. and Wang, D. (2016) Hyperexpansion of RNA bacteriophage diversity. *PLoS Biol.* **14**, e1002409, <https://doi.org/10.1371/journal.pbio.1002409>
- 108 Starr, E.P., Nuccio, E.E., Pett-Ridge, J., Banfield, J.F. and Firestone, M.K. (2019) Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25900–25908, <https://doi.org/10.1073/pnas.1908291116>
- 109 Wu, R., Davison, M.R., Gao, Y., Nicora, C.D., McDermott, J.E., Burnum-Johnson, K.E. et al. (2021) Moisture modulates soil reservoirs of active DNA and RNA viruses. *Commun. Biol.* **4**, 992, <https://doi.org/10.1038/s42003-021-02514-2>
- 110 Neri, U., Wolf, Y.I., Roux, S., Camargo, A.P., Lee, B., Kazlauskas, D. et al. (2022) Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell* **185**, 4023e4018–4037e4018, <https://doi.org/10.1016/j.cell.2022.08.023>
- 111 Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O. et al. (2021) VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37, <https://doi.org/10.1186/s40168-020-00990-y>
- 112 Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M. et al. (2019) Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37, <https://doi.org/10.1038/nbt.4306>
- 113 Bolduc, B., Zablocki, O., Guo, J., Zayed, A.A., Vik, D., Dehal, P. et al. (2021) iVirus 2.0: Cyberinfrastructure-supported tools and data to power DNA virus ecology. *ISME Commun.* **1**, 77, <https://doi.org/10.1038/s43705-021-00083-3>
- 114 Freitag-Pohl, S., Jasilionis, A., Håkansson, M., Svensson, L.A., Kovačič, R., Welin, M. et al. (2019) Crystal structures of the *Bacillus subtilis* prophage lytic cassette proteins XepA and YomS. *Acta Crystallogr. D Struct. Biol.* **75**, 1028–1039, <https://doi.org/10.1107/S2059798319013330>
- 115 Liu, Y., Ishino, S., Ishino, Y., Pehau-Arnauudet, G., Krupovic, M. and Prangishvili, D. (2017) A novel type of polyhedral viruses infecting hyperthermophilic archaea. *J. Virol.* **91**, e00589–17, <https://doi.org/10.1128/JVI.00589-17>
- 116 Mizuno, C.M., Guyomar, C., Roux, S., Lavigne, R., Rodriguez-Valera, F., Sullivan, M.B. et al. (2019) Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nat. Commun.* **10**, 752, <https://doi.org/10.1038/s41467-019-08672-6>
- 117 Aevarsson, A., Kaczorowska, A.-K., Adalsteinsson, B.T., Ahlqvist, J., Al-Karadaghi, S., Altenbuchner, J. et al. (2021) Going to extremes – a metagenomic journey into the dark matter of life. *FEMS Microbiol. Lett.* **368**, fnab067, <https://doi.org/10.1093/femsle/fnab067>
- 118 Galiez, C., Siebert, M., Enault, F., Vincent, J. and Söding, J. (2017) WISH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**, 3113–3114, <https://doi.org/10.1093/bioinformatics/btx383>
- 119 Mirdita, M., von den Driesch, L., Galiez, C., Martin, M.J., Söding, J. and Steinegger, M. (2017) Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176, <https://doi.org/10.1093/nar/gkw1081>
- 120 Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J. and Söding, J. (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473, <https://doi.org/10.1186/s12859-019-3019-7>