Universität
Bremen

Virus-host interactions during chronic infections

&

Diversity and evolution of Pleolipoviruses

Tomás Alarcón Schumacher

Faculty of Biology
University of Bremen
&
Archaeal Virology group
Max Planck Institute for Marine Microbiology

Academic dissertation in partial fulfilment of the requirements for the degree of

Doctor rerum naturalium **(Dr. Rer. Nat.)**

Gutachten/Reviewers

Gutachterin: Dr. Susanne Erdamnn

Gutachterin: Prof. Dr. Nicole Dubilier

Gutachter: Dr. Mart Krupovic

Prüfer: Prof. Dr Andreas Dotzauer

Prüferin: Prof. Dr. Nicole Dubilier

Prüferin: Dr. Hanna Oksanen

Prüferin: Dr. Susanne Erdamnn

Datum des Promotionskolloquiums: 18.09.2023

Acknowledgements

This work was carried out at the Max Planck institute for Marine Microbiology as part of the International Max Planck Research School of Marine Microbiology (MarMic). MarMic is a joint program between the Max Planck Institute for Marine Microbiology (MPI-MM), the University of Bremen (UniB), the Alfred Wegener Institute - Helmholtz Center for Polar and Marine Research (AWI), and the Constructor University Bremen (CUB).

I would like to express my gratitude and respect to Susanne for being such a motivating and encouraging supervisor. For giving the freedom to pursue and develop my own ideas, as well as sincerely telling me when I should not, your honesty is, and will always be most appreciated I can say that it was a joy to work along someone who genuinely cares about everyone around them.

I would like to express my deep appreciation to all the co-authors and collaborators without whom this work would not have been possible. I would also like to thank to all past and present members of the archaeal virology group, for their unwavering support and all the fruitful discussion along the way. It was always a pleasure to work in such a positive and supporting environment. To our technicians, Danila and Ingrid for their invaluable help in the lab and to all the students that collaborated in our work. To my colleagues and classmates, for their friendship and being always up for some interesting discussions no matter the topic.

I would like to thank my family for their unwavering support, and always being there for me even at a distance. To my parents, for always believing in me, in highs and lows, and for always encouraging me to be better, it took a while, but I think I somehow got it. To my Tata, that did not make it to see this through, but I am pretty sure he would have had the perfect joke lined up for the occasion.

To all my friends, to whom I will be selling low by just saying that they have made this time and experience enjoyable and memorable. Last but in no way least, to Cata for always being there and having my back, words cannot even come close to express how I feel, but it is a real privilege to navigate the intricate paths of life alongside you.

"The only sure way to avoid making mistakes is to have no new ideas."

— Albert Einstein

List of original publications

This thesis is based on the following scientific articles, and structured with each manuscript corresponding to an individual chapter:

- ❑ **Alarcón-Schumacher, Tomas**, Adit Naor, Uri Gophna and Susanne Erdmann. "Isolation of a virus causing a chronic infection in the archaeal model organism *Haloferax volcanii* reveals antiviral activities of a provirus". *Proceedings of the National Academy of Sciences* 119.35 (2022): e2205037119.

- ❑ **Alarcón-Schumacher, Tomas**, Fabian Roeloffs, Julian Schmitz, Alexander Grünberger and Susanne Erdmann. "Understanding the impacts of pleolipovirus chronic infections in host fitness and ecology". In Prep.

- ❑ **Alarcón-Schumacher, Tomas**, Dominik Lücking and Susanne Erdmann. "Revisiting evolutionary trajectories and the organization of the Pleolipoviridae family". *PLOS genetics*, Under review

Additional publications

- ❑ **Alarcón-Schumacher, Tomas**, and Susanne Erdmann. "A trove of Asgard archaeal viruses." *Nature Microbiology* 7.7 (2022): 931-932.

- ❑ Gebhard, L. Johanna, Zlata Vershinin, **Tomás Alarcón-Schumacher**, Jerry Eichler, and Susanne Erdmann. "Influence of N-Glycosylation on Virus–Host Interactions in Halorubrum lacusprofundi." *Viruses* 15, no. 7 (2023): 1469.

- ❑ Extracellular vesicles are the main contributor to the non-viral protected extracellular sequence space". Dominik Lücking , Coraline Mercier , **Tomas Alarcón-Schumacher**, Susanne Erdmann*. *Isme communications*. Under review.

Summary

Viruses are the most abundant biological entities in any given environment. The have the ability to regulate microbial abundances through their lytic cycles, and actively participate in horizontal gene exchange. Intriguingly, not all viruses kill their hosts, as some establish chronic and long-term interactions with their hosts that result in particle production without cell lysis. So far, our knowledge on the impact of chronic viral infections on host metabolism, as well as, on the diversity of viruses performing this life cycle in prokaryotic systems remains scarce.

In this work, we isolated and characterized a novel pleomorphic virus from the *Pleolipoviridae* family - Haloferax pleomorphic virus 1 (HFPV-1) - infecting the archaeon *Haloferax volcanii*. This virus is able to establish a chronic infection in its host that leads to efficient and constant viral release without significant effects on host fitness nor appearance of resistant cells. Remarkably, infection with HFPV-1 prompted a massive remodeling of host transcriptomic program, while specifically inducing the downregulation of host defense mechanisms. Furthermore, we showed that the outcome of the infection is heavily influenced by the cross talk between HFPV-1 and resident proviruses on the host genome. Additionally, we uncovered that viral infection inhibited the transition into the morphology of host cells in a temperature dependent manner, further highlighting the ecological impacts of chronic infecting viruses. Notably, HFPV-1 exhibits a uniquely broad host range among pleomorphic viruses, as it is capable to infect efficiently a large diversity of hosts including members from different families to the host of isolation, which could prove crucial in the development of novel genetic systems for so far non-manipulable haloarchaea. Furthermore, we greatly expanded the diversity of pleolipoviruses by producing and surveying metagenomic dataset, reveling that the diversity of pleolipoviruses is larger than previously thought. Thus, we propose a new organization for the *Pleolipoviridae* family, while also providing further evidence that they likely reshaped their genomes through recombination with different groups of plasmids or other mobile genetic elements.

Our work demonstrates that pleolipoviruses are far more diverse than previously thought and that they can heavily affect host fitness and ecological role, highlight the importance of understanding these chronic-infecting viruses and their impact on host metabolism and ecology in different environments.

Zusammenfassung

Viren sind die am häufigsten vorkommenden biologischen Einheiten in jedem Lebensraum auf diesem Planeten. Sie sind in der Lage, die Abundanz von Mikroorganismen durch ihre lytischen Zyklen zu regulieren, und nehmen aktiv am horizontalen Genaustausch teil. Jedoch, töten nicht alle Viren ihre Wirte, einige gehen chronische und langfristige Interaktionen mit ihren Wirten ein, die zur Partikelproduktion ohne Zelllyse führen. Bislang ist unser Wissen über die Auswirkungen chronischer Virusinfektionen auf den Stoffwechsel des Wirts sowie über die Vielfalt der Viren, die diesen Lebenszyklus in prokaryotischen Systemen vollziehen, noch begrenzt.

In dieser Arbeit haben wir ein neuartiges pleomorphes Virus aus der Familie der *Pleolipoviridae* - Haloferax pleomorphic virus 1 (HFPV-1) - isoliert und charakterisiert, welches das Archaeon *Haloferax volcanii* infiziert. Dieses Virus ist in der Lage, eine chronische Infektion in seinem Wirt zu etablieren, die zu einer effizienten und konstanten Virusfreisetzung führt, ohne dabei signifikante Auswirkungen auf die Fitness des Wirts zu haben, oder das Auftreten von resistenten Zellen zu bewirken. Bemerkenswerterweise führte die Infektion mit HFPV-1 zu einer massiven Umgestaltung des Transkriptoms des Wirts, während gleichzeitig die Herunterregulierung der Abwehrmechanismen des Wirts gezielt induziert wurde. Darüber hinaus konnten wir zeigen, dass das Ergebnis der Infektion stark von Interaktionen zwischen HFPV-1 und residenten Proviren im Wirtsgenom beeinflusst wurde. Weiterhin haben wir herausgefunden, dass während einer Virusinfektion der Übergang in der Morphologie der Wirtszellen in einer temperaturabhängigen Weise gehemmt wird, was die ökologischen Auswirkungen von chronisch infizierenden Viren verdeutlicht.

Des Weitern wies HFPV-1 ein unter den pleomorphen Viren einzigartig breites Wirtsspektrum auf, da es in der Lage ist, eine große Vielfalt von Wirten zu infizieren, darunter Mitglieder verschiedener taxonomischer Familien. Dies könnte sich als entscheidender Faktor, für die Entwicklung neuartiger genetischer Systeme für bisher nicht genetisch manipulierbare Haloarchaea erweisen. Darüber hinaus haben wir die Vielfalt der Pleolipoviren durch die Erstellung und Untersuchung metagenomischer Datensätze erheblich erweitert und dadurch enthüllt, dass die Vielfalt der Pleolipoviren größer ist als bisher angenommen. Daher schlagen wir eine neue Organisation für die Familie der Pleolipoviridae vor und liefern gleichzeitig weitere Beweise dafür, dass sie ihre Genome wahrscheinlich durch Rekombination mit verschiedenen Gruppen von Plasmiden oder anderen mobilen genetischen Elementen umgestaltet haben.

Unsere Arbeiten zeigen, dass Pleolipoviren weitaus vielfältiger sind als bisher angenommen und dass sie die Fitness und die Rolle des Wirts im Ökosystem stark beeinflussen können. Dies unterstreicht die Bedeutung

eines besseren Verständnisses dieser chronisch infizierenden Viren und ihrer Auswirkungen auf den Stoffwechsel und die Ökologie ihrer Wirtsorganismen in unterschiedlichen Lebensräumen.

Table of Contents

2. Introduction

2.1 Archaea, the third domain of life

The discovery of "Archaebacteria" by Carl Woese 45 years ago, and the recognition of them as a lineage that formed a new domain in the tree of life (Woese and Fox 1977), forever changed our understanding of microbial diversity and the evolutionary trajectories of cellular life on Earth. The evolutionary relationships of this group, commonly known as archaea today, continues to be a topic of heated discussion and pose new question for researchers in many different areas of study. Although the inherent morphological similarities between the bacterial and archaeal cells initially suggested that they were closely related, archaea have proven to be more of a mosaic of bacterial and eukaryotic features.

While archaea and bacteria share many metabolic pathways of their central metabolism and energy conservation, the proteins related to DNA metabolism and information processing are more similar to their eukaryotic counterparts (Allers and Mevarech 2005). Nevertheless, a large proportion of archaeal-encoded genes, are exclusive to this group, and have proven to be key for the discovery of new protein biochemistry and metabolisms (Könneke et al. 2005; Laso-Pérez et al. 2016). Additionally, the chemical composition of archaeal membranes is distinct from bacteria and eukaryotes (Madigan et al. 2013).

Archaea were initially recognized as rare microbes that thrived only in extreme environments such as solar salters and hot springs. However, the advent of next generation sequencing technologies revealed their universal geographical distribution. Subsequently, archaea were recognized to be playing key metabolic roles in every known ecosystem (Spang, Caceres, and Ettema 2017; Woese, Kandler, and Wheelis 1990).

2.2 The explosion of archaeal diversity

Archaea were originally proposed to be divided in two major lineages by Carl Woese in 1990, Crenarchaeota and the Euryarchaeota (Woese, Kandler, and Wheelis 1990). While the Crenarchaeota clade was a relatively physiologically homogeneous group of thermo- and thermoacidophile organisms, and was thought to be the ancestral phenotype of archaea, the Euryarchaeota presented a constellation of more diverse organisms. This clade comprehended the methanogens and their relatives, along with extreme halophiles and thermophiles of the *Themoplasma*, *Thermococcus* and *Pyrococcus* genera (Woese, Kandler, and Wheelis 1990).

The subsequent discovery of novel archaeal lineages, that diverged from the proposed two-archaeal kingdom structure and that displayed different life styles or novel metabolisms, hinted towards a far larger diversity of

the archaeal domain. One initial discovery was the identification of 16rRNA sequences from uncultivated organisms in a hot spring from Yellowstone National Park. These sequences branched deeper in to the archaeal tree than any previously classified archaea, and were proposed to form the kingdom Korarchaeota (Barns et al. 1996). This was followed by the discovery of the nanosized *Nanoarchaea equitans* in a hot submarine vent, which was classified into the new phyla Nanoarchaeota, further expanding the known diversity of the domain, while also providing first evidences of symbiotic lifestyles in archaea (Huber et al. 2002). Furthermore, the isolation of ammonia-oxidizing organisms from the phylum Thaumarchaeota from marine samples, added to the complexity of the archaeal domain and reaffirmed the notion that, rather than being restricted to extreme environments, archaea were playing important roles in global carbon and nitrogen biogeochemical cycles (Könneke et al. 2005; DeLong 1992).

The development of next generation sequencing technologies and their use in environmental surveys in the 2010's has aided to the discovery of multiple novel clades, and reshaped our understanding of the metabolic versatility and lifestyles of archaea. The recovery of metagenomic assembled genomes (MAGs) from the deep branching Aigarchaeota phylum, showed that archaea encode intriguing eukaryotic features, among which were eukaryotic-like ubiquitin-related genes (Nunoura et al. 2011). Similarly, another monophyletic group with versatile metabolisms, the Bathyarchaeota, was discovered to be extremely abundant in marine sediments worldwide (Evans et al. 2015). Members of this clade likely perform degradation of several carbon sources and recalcitrant carbon compounds, as well as exhibit homologues to *mcr* genes, providing evidence that methanogenesis was not strictly restricted to the Euryarchaeota clade as previously thought (Baker et al. 2020; Evans et al. 2015).

The ever-expanding genomics data lead to the discovery of several related lineages known as the DPANN superphylum (i.e. Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota, Woesearchaeota and Pacearchaeota, Altiarchaeota, Huberarchaeota and Micrarchaeota). These groups, shared some features with the previously described Nanoarchaeota, such as small cell size and reduced genomes, lacking several biosynthetic pathways, along with unique introns in the 16s rRNA and tRNA genes, all signatures of a symbiotic lifestyle (Baker et al. 2010; 2020; Youssef et al. 2015).

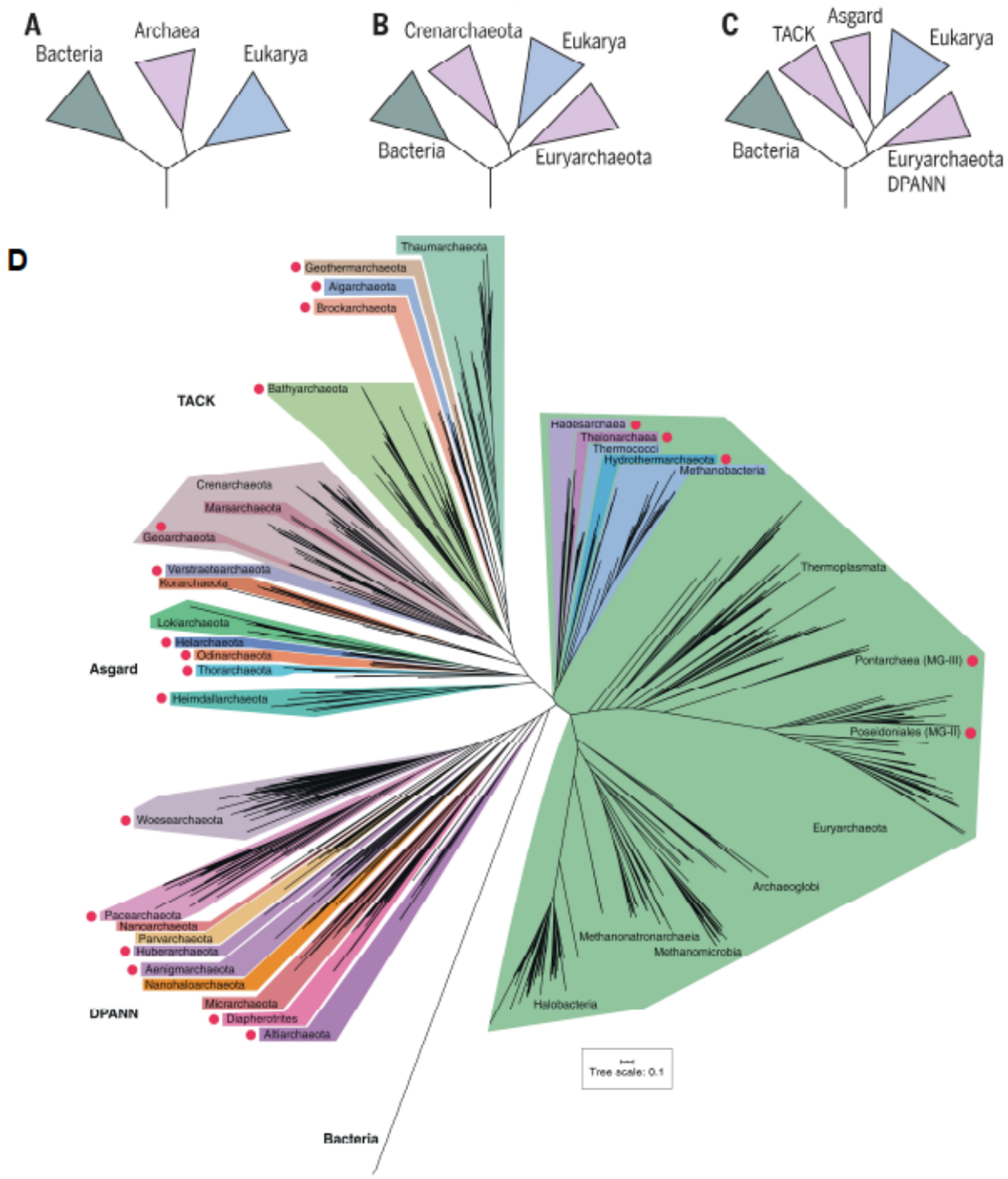


Figure 1: Chronology of archaeal taxonomy and classification. Adapted from (Spang et al., 2017 and Baker et al 2020). **A.** First separation of the archaea as a distinctive domain-ranked clade. **B.** Separation of the archaeal domain in two major phyla Euryarchaeota and Crenarchaeota. **C.** Archaeal phylogenetic tree, accounting the discovery and isolation of new species along with the development of massive sequencing technologies. **D.** Recent version of the archaeal phylogeny, accounting the explosion of archaeal diversity through metagenomic assembled genomes (MAGs).

2.3 The role of archaea in eukaryotic evolution

In recent years, the study of a particular clade of archaea, discovered in marine sediments, has reignited the discussion on the evolutionary origin of eukaryotes. The Asgard archaea clade, composed by Heimdallarchaeota, Thorarchaeota, Odinarchaeota, Helarchaeota and the recently cultivated Lokiarchaeota (Imachi et al. 2020), are abundant in deep-sea sediments. These groups have been shown to encode an unprecedented number of genes homologue to eukaryotic signature proteins (ESPs), such as cytoskeleton-associated actin, the endosomal sorting complex required for transport (ESCRT) which are involved in membrane trafficking and proteasome formation, and ubiquitin modifying systems (Spang et al. 2015). Interestingly, phylogenetic analyses showed that this clade forms a monophyletic group together with the eukaryotes, branching from within the archaea in the tree of life, which would support a two-domains tree of life (Eme et al. 2017; Zaremba-Niedzwiedzka et al. 2017). Detailed studies of the physiology of Lokiarchaeon *Candidatus* Prometheoarchaeum syntrophicum revealed that there is a syntrophic exchange of formate and hydrogen with a co-cultured sulfate-reducing delta-proteobacterium. Moreover, microscopy indicated that *Ca.* P. syntrophicum is able to generate membrane protrusion or appendages that might have provided a mechanisms for the engulfment of an endosymbiont (Imachi et al. 2020). In this evolutionary scenario, an ancestor of the Asgard archaea would have been the host for an endosymbiotic proteobacterium, that later evolved in to the mitochondria, a process that gave rise to the early eukaryotic cells (Baker et al. 2020). However, this hypothesis is far from a consensus in the scientific community and the relationship between archaea and eukaryotes, as well as the role of archaea in the evolution of eukaryotes, remains a highly controversial topic.

The continuous discovery of novel archaeal clades, together with improvements in the bioinformatics field, has left the taxonomy and phylogenetic relationships among archaeal clades in constant flux. The fact that most clades are only known as MAGs, together with the successive addition, splitting and renaming of archaeal clades has led to major inconsistencies in their taxonomic classification. Thus, efforts to standardize the classification of archaeal lineages have become indispensable to aid the advance of the field. As such, a standardized archaeal taxonomy was recently develop (Rinke et al. 2021). Combining the concatenated phylogeny of 122 single-copy marker genes and a relative evolutionary divergence-based rank, the Genome Taxonomy Database (GTDB) reclassified the archaeal taxonomy into 16 phyla. These monophyletic groups comprise the recently discovered groups, as well as reclassify the former Euryarchaeota and DPANN into several independent phyla, as well as unify the formerly known TACK superphylum (Thaumarchaeota–Aigarchaeota–Crenarchaeota–Korarchaeota) into the single phylum Thermoproteota (Fig 2).

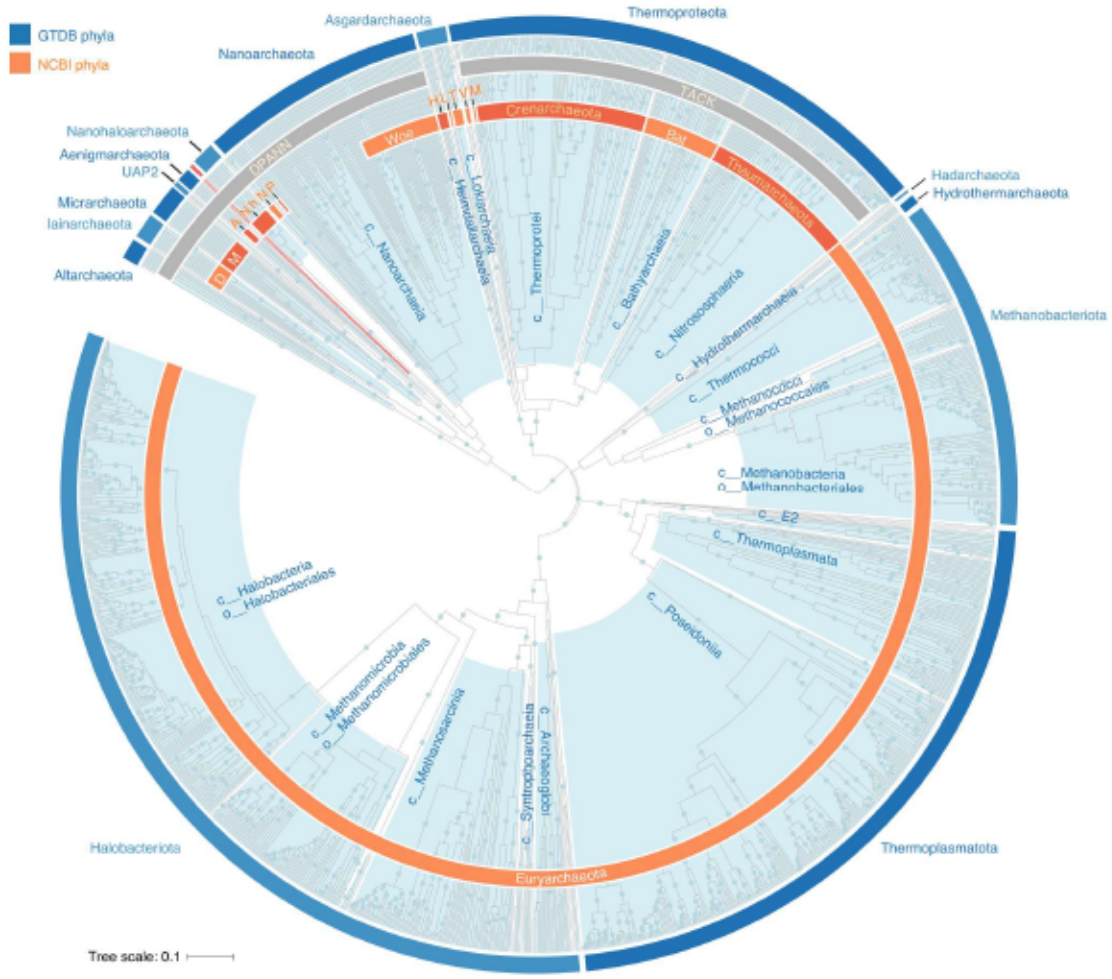


Figure 2: Rank-normalized archaeal GTDB taxonomy Adapted from (Rinke et al., 2021). The blue ring denotes the rank-normalized taxonomy, while inner orange ring indicate the former NCBI phyla. Blue labels indicate phyla (outer labels), class and order (inner labels). Blue dot marks indicate Bootstrap values over 90. Scale bar indicates relative evolutionary divergence (RED).

Nonetheless, only a few of the archaeal clades have actual cultivated representatives. Therefore, their physiology and metabolism remains largely hypothetical, providing the opportunity to uncover new biochemical pathways and ecological relationships in this diverse and enigmatic group.

2.4 Archaeal model systems

Given our limited ability to cultivate the vast majority of archaeal clades, current knowledge of their physiology primarily comes from only a handful of groups such as hyperthermophilic archaea and methanogens, while a large portion of these isolated organisms belonging to the Halobacteriota phylum (former Euryarchaeota). This group includes organisms that exhibit a broad diversity of metabolisms and occur in many different environments. The first members of this clade were characterized even before archaea

was considered a separate domain, as the first methane-producing organisms (Stadtman and Barker 1951). These so-called methanogens, together with organisms isolated from high salt environments from the class Halobacteria, comprise today the largest number and diversity of lineages that have been brought into culture (Baker et al. 2020). Methanogens usually thrive in anoxic environments, where they produce methane as a byproduct of their energy metabolism. However, methanogenesis is not a monophyletic trait, as members of several phyla, i.e. Halobacteriota, Methanobacteriota, and Thermoproteota (former TACK), have the genomic repertoire required to carry out methanogenesis and, in some instances, actively produce methane in culture (Berghuis et al. 2019; Elkins et al. 2008; Spang, Caceres, and Ettema 2017; Laso-Pérez et al. 2016).

Members of the sister class Halobacteria are commonly found in high salt environments, were they usually outnumber bacteria and are the dominant members of the microbial community (Oren 2002), which represents an advantage for their cultivability. This has led to the isolation and characterization of numerous species and the development of multiple genetic tools, making them a model system for the study of archaea cell biology and genetics (Allers et al. 2004; Soppa 2006)

2.5 Hypersaline environments

Hypersaline environments are those were salt concentrations are higher than seawater, and are extremely diverse in terms of their origin, overall ion composition and concentrations (McGenity and Oren 2012). Generally, salt-water bodies can occur inland, in endorheic basins (lakes that do not flow to the sea), or in coastal regions under tidal influence. Also fluids containing concentrations that exceed >5% of salt (known as brines), can be found in deep-sea mud pools, ancient subglacial brines, seldom-flooded salt deposits or in marine-ice brine channels in polar regions (Boetius and Joye 2009). Regardless of their origin, halophilic organisms have successfully colonized hypersaline environments. Halophilic organisms are classified in three major categories: slight halophiles, with an optimal growth at 0.2–0.85M NaCl (1–5%); moderate halophiles, which grow optimally at 0.85–3.4 M NaCl (5–20%); and extreme halophiles, whose optimal growth is at 3.4–5.1 M NaCl (20–30%) (DasSarma and DasSarma 2012).

In order to colonize these environments successfully, organisms present different adaptations to salinity conditions, as they need to balance their internal solute concentration with their environment and achieve osmotic balance. One of the most prominent strategy is the accumulation of compatible solutes, such as glycerol, trehalose or glycine betaine. The organisms synthetize this small, often uncharged compounds to increase the osmolarity and match the salt concentrations of their environment. This provides the organisms with plenty of flexibility to adapt to different and changing salt concentrations, at the expense of a relatively high energetic cost of synthetizing these compounds and is typically observed in halophilic bacteria and eukaryotes. Another strategy, commonly known as salt-in, consists of the accumulation of ionic compounds

(mainly K^+) to match the salt concentrations of their environment; the salt-in strategy is widespread in Halobacteria and obligate halophiles. This strategy is relatively non-expensive when compared to the synthesis of compatible solutes, as it uses transport systems to uptake solutes. However, it requires the adaptation of the entire enzymatic machinery and physiology to these high salt concentrations, and it is likely the product of a long evolutionary process (Oren 2011). These organisms are characterized by a highly acidic proteome and enzymes that are only active under high-salt conditions (DasSarma and DasSarma 2012; Becker et al. 2014).

Hypersaline environments present a relatively low community complexity, however, communities can reach high densities of 10^7–10^8 cells ml^{-1} or higher in brine and sediments (Oren 1994). Eukaryotic macroorganisms are limited to a few species, such as the globally distributed algae *Dunalliela*, the metazoan *Artemia salina* and a few fungal representatives (Oren 2005; Benlloch et al. 2001; Gunde-Cimerman, Ramos, and Plemenitaš 2009). (Oren 2002). Similarly, only a few bacterial groups have been identified in these environments, with members of the *Salinibacter* genus, from the Bacteroidetes clade, as one of the better-studied representatives. Interestingly, unlike other bacteria, *Salinibacter* exhibits a number of distinctive features, such as an acidic proteome and preferring the salt-in strategy. These features were likely acquired through horizontal gene transfer from the Halobacteria, and help to understand their widespread distribution in hypersaline environments (Antón et al. 1999; Mongodin et al. 2005).

Archaea from the Halobacteria class are the most abundant and diverse organisms in hypersaline environments. They dominate microbial communities and their relative abundance positively correlates with salt concentrations, with minimum concentrations of salt required for their growth (10-15%).

Interestingly, bacteriovory by unicellular eukaryotes was shown to be rather less pronounced in hypersaline environments, and correlates negatively with the increase in salinity (Guixa-Boixareu et al. 1996). Thus, the role of viruses is particularly intriguing in these environments, as the abundance of viral particles can often exceed prokaryotic cells in one or two orders of magnitude, and cell lysis by viruses appears far more important than bacteriovory. Furthermore, the abundance of viral-like particles (VLPs) observed in these environments positively correlates with the increase in salinity (Sime-Ngando et al. 2011; Oren, Bratbak, and Heldal 1997; Boujelben et al. 2012). This makes viruses major regulators of microbial community composition in hypersaline environments (Guixa-Boixareu et al. 1996).

2.6 The Halobacteria class

Similarly to methanogens, haloarchaea were discovered before the formal separation of archaea into their own domain, with *Halobacterium salinarum* being isolated more than a century ago from cod fish (Harrison and

Kennedy 1922; Ventosa and Oren 1996). Since then, they have served as a basis for important scientific advances, as it was in *H. salinarum* that the discovery of the light-driven proton pump bacteriorhodopsin, which employs a rhodopsin-retinol system to pump protons for energy conservation was made (Oesterhelt and Stoeckenius 1971). Members of the Halobacteria class (haloarchaea) are mostly chemoorganotrophs and the isolation of additional type strains, like *Haloferax volcanii* (Mullakhanbhai and Larsen 1975), provided further insights into gene transfer mechanisms (Mevarech and Werczberger 1985). Meanwhile, the development of genetic tools to manipulate this organisms have greatly contributed to our understanding of archaeal genetics and physiology (Cline and Ford Doolittle 1987; Cline et al. 1989).

Haloarchaea can be classified into nine families with cultivated representatives (*Haladaptaceae*, *Halalkalicoccaceae*, *Haloarculaceae*, *Halobacteriaceae*, *Haloferacaceae*, *Natrialbaceae*, *Natronoarchaeaceae* and *Saliniarchaeaceae*) and at least seven other clades that remain uncultivated (Rinke et al. 2021). Interestingly, members of the different families of haloarchaea exhibit a wide range of morphologies, none more intriguing than the squared-shape archaeon *Haloquadratum walsbyi* (*Haloferacaceae*). This archaeon was reported for the first time in 1980 by Tony Walsby in brine samples collected in the Sinai Peninsula (Walsby 1980). Subsequently, through 16S rRNA sequences and fluorescence in-situ hybridization (FISH), *H. walsbyi* was recognized as a prominent member of microbial communities worldwide (Antón et al. 1999). While it took more than 20 years to bring it to culture (Burns et al. 2007), this fascinating archaeon and others with similarly surprising morphologies, such as the triangular archaeon *Haloarcula japonica* (Takashina et al. 1990), redefined the boundaries of prokaryotic morphological complexity. Additionally, it posed further questions on how these organisms are so successful and abundant despite their obvious disadvantages in terms of surface-to-volume ratio (Young 2006).

Archaeal cells lack peptidoglycan, and instead, are surrounded by a rigid proteinaceous surface layer (commonly known as S-layer). This surface layer is normally composed of one or to proteins (or glycoproteins), that organize in different lattices (Rodrigues-Oliveira et al. 2017). In haloarchaea, most organisms exhibit a S-layer glycoprotein organized in hexagonal lattice and complemented with pentameric defects in highly curved areas of the cell, further providing continuity to the cell surface, as it has been recently shown in atomic-resolution descriptions of *H. volcanii* native cells (von Kügelgen, Alva, and Bharat 2021)

Haloarchaea exhibit a wide range of metabolisms. The majority are oxygenic heterotrophs or photoheterotrophs (by encoding rhodpsins for energy conservation) that degrade sugars through classic glycolysis and some modified version of the Entner-Douderoff (ED) pathway (Sutter et al. 2016). Many can directly grow on simple carbon sources such as glycerol, which is highly abundant in hypersaline environments due to its use as compatible solute by other organisms (Kauri, Wallace, and Kushner 1990; Zhou et al. 2008). Alternatively, the archaeon *H. volcanii* can grow on acetate as sole carbon source, which under

aerobic conditions is commonly excreted as waste product and can inhibit the growth of some organisms at high concentrations (Kuprat et al. 2020).

Under anaerobic conditions, members of the Halobacteriaceae family can normally grow by fermentation of arginine and using alternative electron acceptors, such as nitrate in the case of the archaeon *H. mediterranei* (Lledó et al. 2004). Additionally, sulfur compound can be used as alternative electron acceptors, such as dimethyl sulfoxide (DMSO) or elemental sulfur (Sorokin et al. 2016; Qi et al. 2016).

One particular organism, *Haloferax volcanii,* has been the subject of extensive research in various areas during the past decades, because *H. volcanii* is easy to cultivate and can grow under aerobic or anaerobic conditions. Additionally, a number of genetic and biochemical tools have been developed to study its physiology and genetics, transforming it into a model organism for haloarchaea (Pohlschroder and Schulze 2019). The study of this archaeon has provided insights into mechanisms of DNA replication, transcriptional regulation and cell division, morphology and surface organization (von Kügelgen, Alva, and Bharat 2021; Breuert et al. 2006).

2.7 The virus concept and taxonomic organization

Life as we know it centers around the replication and transmission of genetic information in some form of nucleic acids to the offspring. Cells contain distinctive units of replication with a degree of autonomy, such as a chromosome or an episomal element like a plasmid (Jacob, Brenner, and Cuzin 1963). However not all genetic elements are under the centralized control of the cellular replication machinery, and these semi-autonomous genetic elements, known as mobile genetic element (or MGEs), exist within the gene pools of almost every organism. These elements often have their own regulation and transmission mechanism, and given that they operate at least partly independent, they are distinct units of evolution and follow their own evolutionary trajectories (Koonin and Starokadomskyy 2016). The diversity of replicators and their genetic content is incredibly high and the attempts to organize and differentiate MGEs from bona fide viruses under strict boundaries has often failed, because in the genetic space of MGEs, no genes are universal to all viruses (Koonin, Senkevich, and Dolja 2006).

Then, how do we differentiate any given MGE from a virus?

In a simplified way, viruses are molecular parasites essentially composed of nucleic acids (single or double stranded DNA or RNA) and an envelope, typically a protein shell called capsid, or a lipid membrane decorated with virion proteins. Strictly, viruses are defined in the International Code of Virus Classification and Nomenclature (ICVCN) as:

"...a type of MGEs that encode at least one protein that is a major component of the virion encasing the nucleic acid of the respective MGE and therefore the gene encoding the major virion protein itself; or MGEs that are clearly demonstrable to be members of a line of evolutionary descent of such major virion protein-encoding entities."

Thus, the defining characteristic of a virus is that it encodes at least one protein that forms the viral particle (virion), which confers the ability to transfer its genetic information; or, that the mobile genetic element in question is a direct descendant of a lineage that we recognize as a true virus.

Viral genomes usually consist of distinct structural and replication modules that frequently recombine with each other and other MGEs. Therefore, it is not rare that we often see different replication strategies within a single group of viruses. This observation, together with the absence of hallmark genes conserved across the full viral spectra due to their different evolutionary origins, have been major roadblocks on the way to generating a consistent and coherent taxonomic classification of viruses.

Despite this, recent efforts to organize viral diversity through the identification and evolutionary analyses of viral hallmark genes has yielded a more robust and coherent classification than the previous nucleic acid-based (Baltimore classification) or operational definition-based classifications. This allowed the identification of at least six superviral hallmark genes (VHGs), which encode structural viral proteins, as well as proteins involved in genome replication (Koonin et al. 2020).

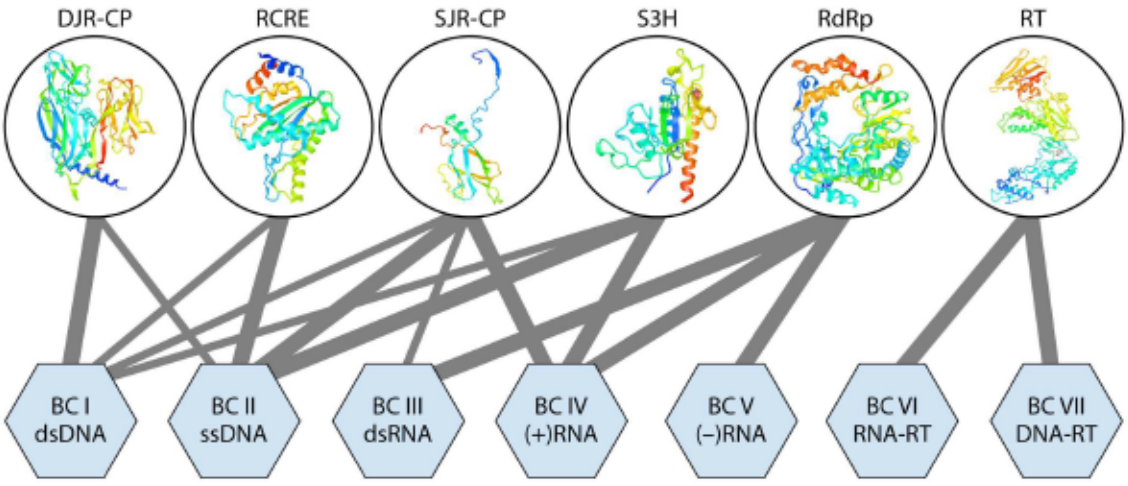


Figure 3: Representation of the 6 "superviral hallmark genes" in virus genomes of the seven Baltimore classes. Adapted from (Koonin et al., 2020). Ribbon diagrams of the representative protein structures show the "superviral hallmark proteins". The lines connect the proteins with the viruses of BCs in which they are present. The thickness of each connecting line roughly reflects the abundance of a given "super hallmark" gene in a given BC. The six hallmark genes encode the following proteins: DJR-CP, double-jelly-roll capsid protein; RCRE, rolling-circle replication (initiation) endonuclease; RdRp, RNA-directed RNA polymerase; RT, reverse transcriptase; S3H, superfamily 3 helicase; SJR-CP, single-jelly-roll capsid protein.

The phylogenetic relationships of these super hallmark genes in combination with gene sharing network analysis has laid the foundations to the establishment of four viral realms: *Riboviria*, *Monodnaviria*, *Varidnaviria*, and *Duplodnaviria* (Koonin et al. 2020). Subsequently, two additional viral realms have been added to account for orphan clades that have different evolutionary histories and form their own lineages. The realm *Ribozyviria* was created to account for hepatitis D virus (HDV), which depends on hepatitis B virus as a helper, has a ribozyme-containing negative-sense single-stranded circular RNA genome and encodes a protein that binds the genomic RNA forming a ribonucleoprotein complex (Szirovicza 2022). The second realm added was the *Adnaviria*, which contains archaeal filamentous viruses that display a linear A-form dsDNA genome, and their capsid proteins that belong to independent lineages, not related to other viruses (Krupovic et al. 2021).

Altogether, the distribution of VHGs and realm organization point once more towards a polyphyletic nature of viruses, and suggests that they likely emerged from multiple independent evolutionary events.

2.8 Virus origin and evolution

As discussed above, viruses do not share a universal hallmark gene, and therefore it is likely that they had different evolutionary origins. Accordingly, at least three major hypothesis have been proposed to explain the origin of viruses (Forterre 2006). The virus first hypothesis (or virus world) proposes that viruses preceded the cellular entities, and that they are somehow intermediaries between the prebiotic and cellular life. The reduction hypothesis, postulates that viruses are actually extreme cases of genome reduction of cellular forms, which became obligated intracellular parasites and lost their ribosomal and other unnecessary features. Finally, the escape hypothesis states that viruses originated in multiple occasions from selfish MGEs, which acquired structural proteins from their hosts and adopted an infectious behavior. Given viral diversity and polyphyly, it is plausible that more than one, if not all or a combination of the proposed scenarios give rise different group of viruses, however, it is unlikely that bona fide viruses were able to exist without a host.

Consequently, hybrid models for the origin of viruses have been proposed, such as a chimeric origin for viruses. In this scenario, the replication machinery of MGEs was likely present before cellular life arose, as many viral or MGE-derived replication proteins do not have homologs in their cellular counterparts. If true, it is within reason to suggest that at least some group of viruses arose from the recruitment and adaptation of cellular proteins by an MGE (Krupovic, Dolja, and Koonin 2019).

Further evidence for this is that some of the most common capsid proteins seem to have evolved from some cellular ancestor. Such it is the case for capsid proteins from the lineage of the single jellyroll (SJR) fold, which are the most common among (+) RNA and ssDNA viruses. Structural comparison of these proteins

revealed that they are similar to DJR domains from at least four major groups of proteins: sugar binding enzymes (glycosyl hydrolases), subtilisin-like proteases, nucleoplasmins (molecular chaperones) and cytokines of the TNF superfamily (Krupovic and Koonin 2017). The double jellyroll fold, which is present in a large variety of dsDNA viruses, has been proposed to have evolved through gene duplication of SJR capsid genes. This, together with the fact that the SJR fold is widespread in a large diversity of cellular enzymes, speaks to an ancestral origin, which places the recruitment of this fold shortly after the emergence of cellular life.

Regardless of the individual evolutionary trajectories of each group of viruses, one fact stands out, as it is unmistakable that the evolutionary history of viruses is heavily intertwined with the one off non-viral mobile genetic elements. With capsidless MGEs not only playing a key role in the onset of viruses, but also interacting, coevolving and actively recombining with viruses today, as they all share similar niches as intracellular genetic parasites.

2.9 Ecological roles of viruses and their lifestyles

Organisms from every branch of the tree of life, regardless of their ecosystem, genomic repertoire or lifestyle, are subject to viral infection, and hypersaline environment are not an exception. With an estimated of $4.8x10^{30}$ viruses on the planet, viral particles outnumber cells in ratios around 0.1 to 30:1 virus-to-microbe in most environments (Cobián Güemes et al. 2016). These high abundancies are a product of viruses actively replicating and releasing into the environment via lysis of their host cells or another alternative mechanism. Thus, viruses strongly influence the community composition and the release of intracellular content upon lysis of the host (Wilhelm and Suttle 1999). However, viruses transmitting their genome in virions comprise only a portion of the virosphere in an environment, as a number of viruses are intracellular (as episomal or integrated genetic elements) and undergoing different life cycles that might not account to virus particle abundance in the environment at a given time.

Viruses are actively participating in horizontal gene transfer, particularly when undergoing a lysogenic cycle, in which the genome of the virus is integrated into the host chromosome and is inherited along with the host genetic material by the offspring. Alternatively to this vertical transmission of the virus, upon induction, the viral genome can be excised and enter the lytic cycle, which often results in the incorporation of host genes that can be transmitted to a new host. Gene transfer mediated by viruses is called transduction, and it can occur in a random or specific manner. In generalized transduction, any given portion of the host genome can be packaged in the viral particle. Conversely, in specialized transduction, only the regions adjacent to the integration site of the viral genome are carried along and enclosed in to the capsid (Soucy, Huang, and Gogarten 2015).

This dichotomy between lysis and lysogeny has been the prevalent framework in the study of prokaryotic viruses, because these mechanisms have been extensively studied for bacteriophages. However, nature itself is rarely dichotomic but rather a complex mixture of traits that, when properly sampled, reveals more of a continuum than discrete units. In this regard, viral life cycles are not the exception and their life strategies are at least as vast and diverse as the ones of the hosts they infect. One example of this diversity is pseudolysogeny, where the virus nucleic acids are neither integrated into the host as a provirus, nor released into the supernatant as particles through a lytic cycle, with the viral genome mostly replicating asynchronously to the host chromosome (Ripp and Miller 1998). This uncoupling to the cell cycle and division leads to an asymmetrical inheritance of the virus by the daughter cells and mixed populations, were only a fraction of the offspring carries the viral genome and are often resistant to new infections (Cenens et al. 2015). Different signals have been suggested to trigger this mechanism, such as nutrient depletion or the lack of response to SOS-triggering inducers, however, none of them are conclusive, and it is still up for debate whether pseudolysogeny is a *de facto* lifestyle or just snapshots of a differently regulated lysogenic cycle (Mäntynen et al. 2021).

The most prominent example of alternative life cycles are chronic infections, where virion release does not result in host cell lysis, but rather viral particles are secreted continuously through the host membrane. This particular lifestyle is very common in multicellular eukaryotes, where for example it has been estimated that every human harbors between 8 and 12 different chronic infecting viruses at any given time even if no symptoms of disease are displayed (Virgin, Wherry, and Ahmed 2009). Viruses that generate chronic infections in eukaryotes are often able to establish relatively long-term relationships with their hosts, as viral and host genes coevolve, and viral genes and host immune system tend to balance each other. It is also well documented that viruses have played a key role in the evolution of eukaryotes as approximately 9% of the human genome consists of endogenous retroviral DNA (Ryan 2016). In eukaryotes, it is often the case that chronic viruses infect a particular cell line within the organism, and often find mechanisms to persist or evade the immune response of the host and replicate at low rates. The relevance of many of these viruses in the medical or agricultural fields have led to a relatively robust body of practical and theoretical body of work on virus-host interactions and the evolution of chronic viruses.

However, it is unmistakable that these types of infective cycles must face extremely different condition in prokaryotic cells, given the inherent unicellular versus multicellular nature of prokaryotes and eukaryotes. Thus, while our knowledge of the molecular mechanisms behind lytic and lysogenic infections is relatively sound, our understanding of chronic infections of viruses in prokaryotes in scarce.

The first identified viruses being released without impairing host growth in prokaryotes were filamentous phages infecting *E. coli* K12 almost 60 years ago (Hoffmann-Berling and Mazé 1964). These viruses possess a non-enveloped filamentous morphology with a circular ssDNA genome and are today classified into the

Inoviridae family (Knezevic and Adriaenssens 2021). These viruses (inoviruses) replicate via the rolling-circle mechanism and either exist as episomal elements or integrate into the host genome. Until recently only a handful of isolates were known, and they were thought to be more of a rarity, however, comprehensive global metagenomic studies have shown that inoviruses are actually distributed across all biomes on Earth (Roux et al. 2019). While the majority infect Gamma and Betaproteobacteria, these viruses count among their hosts' members of at least 22 different bacterial phyla and potentially even members of the archaeal domain; suggesting they could have ecological and evolutionary influence across the entire prokaryotic diversity.

Another example are pleomorphic viruses infecting mycoplasma species, who have been shown to exit the cell without lysis of the host using a budding-like process (Maniloff, Das, and Christensen 1977; Dybvig et al. 1985). Despite these well-studied examples, chronic infecting viruses in bacteria represent a small fraction of the virosphere, with most bacterial viruses exhibiting lytic or temperate infection cycles. Conversely, the archaeal virosphere presents a different landscape with a number of archaea-exclusive viral families the exit the host without lysis of the host and cause chronic infections.

2.10 Viruses of archaea

The study of archaea has been brought to the spotlight in the las few decades, driven in part by the realization that they are widespread in every environment and that they play key roles in biogeochemical cycles. Consequently, our understanding of archaeal viruses has also increased and viruses from both extreme environments (e.g. hyperthermophilic and halophilic), and temperate environments (e.g. soil and the marine systems) have been isolated and characterized. Nevertheless, the number of isolated archaeal viruses (~ 200) still trails behind the number of isolated bacterial viruses by at least one order of magnitude. Thus, the biology and mechanisms of interaction between archaeal viruses and their hosts remain understudied. Strikingly, these viruses display an unprecedented morphological and genetic diversity and a number of described viruses for their own lineages that infect uniquely members of the archaeal branch (Prangishvili et al. 2017).

2.10.1 Icosahedral viruses

The most often isolated archaeal viruses present classic tailed morphology consistent with the other members of the *Caudoviricetes* class. Consequently, their structural module is highlighted by capsids presenting the HK97 fold-type, the genes required for the formation of the neck, tail and tail fibers; along with the DNA-packaging machinery, which is commonly mediated by the terminase enzyme (small and large subunits) (Luk et al. 2014). Tailed viruses have been extensively researched in bacterial hosts and mostly perform some form of the lytic or lysogenic cycles. For years, tailed archaeal viruses were seen as members of the same clades as their bacterial counterparts (a product of morphology- driven classification). However, recent studies have shown that they form separated lineages and instead should be classified in at least 14 new families., Further

evidence of their widespread distribution across every Earth biome, highlights once again their ecological relevance (Y. Liu et al. 2021).

There are at least two other viral lineages of tailless icosahedral viruses that were considered to be shared between bacteria and archaea due to the similar morphology, the *Sphaerolipoviridae* and *Turriviridae* families. Members of the *Sphaerolipoviridae* are characterized by small virions that include a protein-rich internal lipid membrane. They also encode a single-jelly-roll (SJR) capsid protein and infect members of the haloarchaea through lytic cycles (Demina et al. 2023). The members of the *Turriviridae* also present an internal lipid membrane, surrounded by an icosahedral capsid of SJR-like major capsid proteins. Particles are decorated with turrets (hence the name), which are also built from a succession of fused SJR-type proteins (Veesler et al. 2013) and they also perform lytic cycles on their *Sulfolobus* hosts (Brumfield et al. 2009).

Notably, among icosahedral virions, the *Portogloboviridae* occupies a unique place among archaeal (and bacterial) viruses. Members of this groups package the circular genome condensed in the A-form around a nucleoprotein (F. Wang et al. 2019). This complex is surrounded by a lipid membrane and a single SJR capsid protein, and the particles are proposed to be released without lysis of the host (Y. Liu et al. 2017).

2.10.2 Filamentous viruses

The study of archaeal viruses has also contributed with numerous viral groups that do not present the widespread icosahedral morphology (Figure 4). Among these are filamentous viruses from the aforementioned recently recognized *Adnaviria* realm, whose members are globally distributed in thermophilic environments (Baquero, Contursi, et al. 2020). Unlike filamentous phages from the *Inoviridae* family that exhibit ssDNA genomes, these viruses possess dsDNA genomes and some of their members exhibit a lipid envelope (Kasson et al. 2017). The *Rudiviridae*, *Lipothriyviridae* and the *Tristromaviria* families, forms this realm. These viruses present a capsid fold non-related to other viruses, with homo or heterodimers of major capsid proteins interacting directly with the DNA and supporting the A-form of the genome and the virions are often decorated with fibers at both ends (F. Wang et al. 2020; Baquero, Liu, et al. 2020). Interestingly, while members of the *Rudiviridae* and *Tristromaviridae* are lytic viruses, the study of Sulfolobus islandicus filamentous virus (SIFV), an enveloped member of the *Lipothixviridae*, showed that virions are assembled in the cytoplasm and acquire their lipid membrane prior to exiting the cell. Once assemble, particle release from the host occurs via a specialized pore-like pyramidal portals that protrude through the cell surface (Baquero et al. 2021). While this virus-associated pyramids (VAPs) have been observed in other lineages of archaeal viruses, i.e. *Rudiviridae* and *Turriviridae,* in those cases viral release leads to release of intracellular content including along with the mature virion particles. Whereas, in the case of SIFV, the growth is not always severely

impeded but infection with the this virus can lead to cell death in an MOI-dependent manner (Arnold et al. 2000; Häring, Vestergaard, Brügger, et al. 2005, Baquero et al. 2021), although the mechanism behind this phenomena is not fully understood.

Another two unrelated groups of filamentous viruses have also been described to infect members of the archaea. The single representative of the first clade is Aeropyrum coil-shaped virus (ACV) from the *Spiraviridae* family, which replicates in the hyperthermophilic archaeon *Aeropyrum* (Mochizuki et al. 2012). The virions from ACV are formed by coiling of a single circular nucleoprotein filament forming hollow cylindrical particles, that contain the ssDNA genome. Similar to inoviruses, this ssDNA virus also performs a chronic infection, exiting the host cell without lysis. However, ACV is not evolutionary related to the bacterial inoviruses nor to any other archaeal virus and many aspects of its lifecycle and genome replication remain unknown. The second single-species clade of filamentous viruses infecting archaea *Clavaviridae* family, Aeropyrum pernix bacilliform virus 1 (APBV1), presents rigid bacilliform virions constructed by a single α-helical protein (unrelated to the *Adnaviria* fold), with a terminal cap-like structure and terminal fibers on one end (Mochizuki et al. 2010). Meanwhile infection with APBV1 generates neither growth retardation nor leads to host lysis, while still being able to produce viral particles.

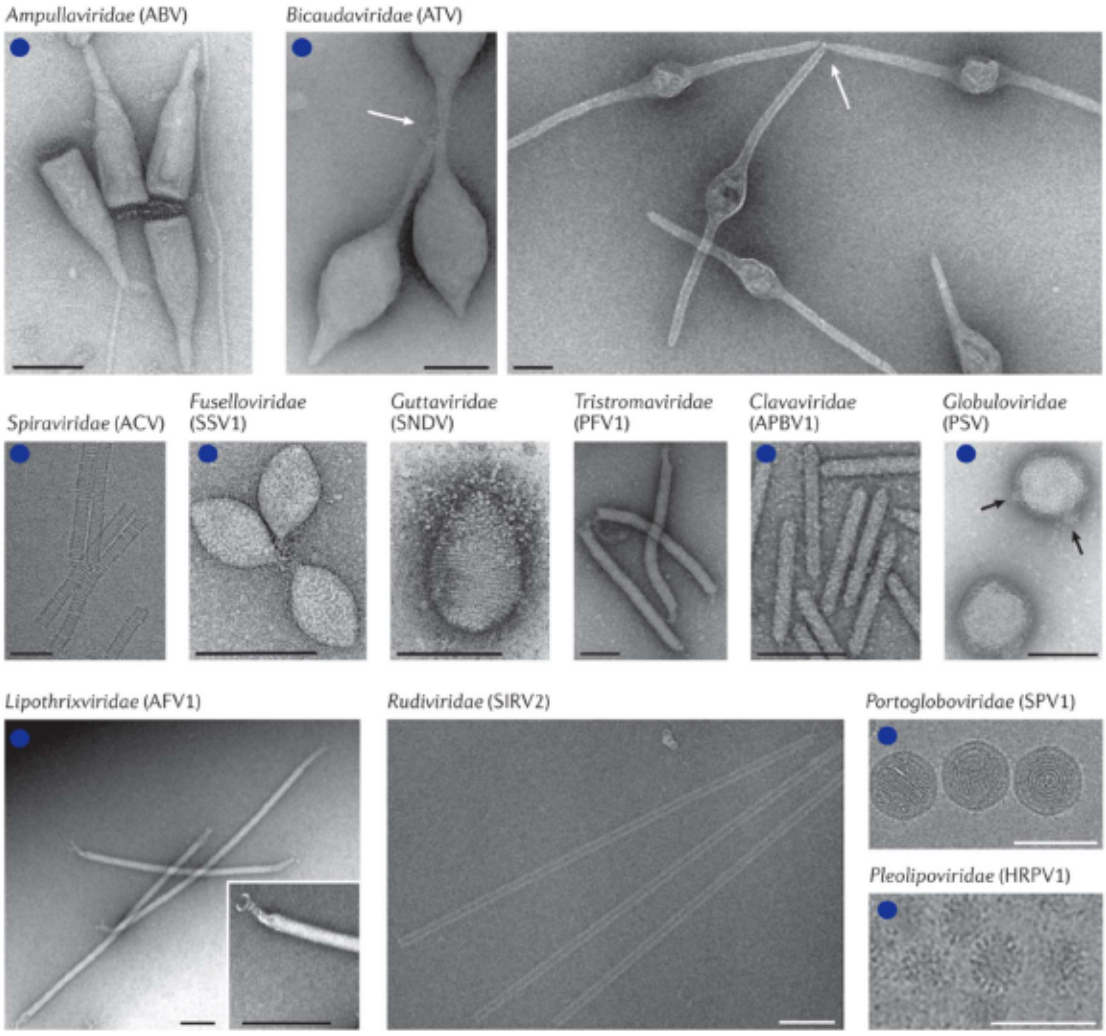


Figure 4: Archaeal exclusive viruses. Adapted from (Prangishvili et al., 2017). Electron micrographs of archaeal viruses. Virus families and species are indicated. Acidianus filamentous virus 1 (AFV1), the inset shows the terminal structure. Acidianus two-tailed virus (ATV), the arrows indicate virion tails (left), which undergo extracellular development (right). Pyrobaculum spherical virus (PSV), the arrows indicate spherical protrusions. Negative stain with uranyl acetate, except for Aeropyrum coil-shaped virus (ACV), Halorubrum pleomorphic virus 1 (HRPV1), Sulfolobus islandicus rod-shaped virus 2 (SIRV2) and Sulfolobus polyhedral virus 1 (SPV1), which are in the vitreous ice. Bars, 100 nm. Groups marked with blue dots have been reported to establish chronic infections on their hosts.

Strikingly, the *Clavaviridae* family has been recently proposed to be evolutionary related to spindle-shaped viruses (F. Wang et al. 2022). Structural comparison of the major structural protein of APBV1 and Sulfolobus monocaudavirus 1 (SMV1), showed remarkable similarities, as both present highly hydrophobic subunits that form a helical-hairpin-like structure, and display similar glycosylation patterns.

2.10.3 Spindle-shaped viruses

Spindle-shaped viruses can be classified according to their morphology and genomic relationships into two main clades: the first one consisting of three families, the *Halspviridae, Thaspviridae* and *Fuselloviridae;* and the second clade formed by the *Bicaudaviridae* family (Krupovic et al. 2014; Bath et al. 2006; Kim et al. 2019; F. Wang et al. 2022).

The *Bicaudaviridae* family contains viruses with relatively large spindle shaped virions with one or two tails, which are not present in the particle upon release, but are rather developed outside the host (Häring, Vestergaard, Rachel, et al. 2005). The tails vary in length, but can reach up to 400nm and are proposed to aid virion attachment to their host. This group of viruses exhibit dsDNA genomes and encode for an AAA ATPase, which is homologous to DnaA a protein responsible for mediating the replication initiation in bacteria. It has been hypothesized that it could play a similar role in the replication of tailed spindle-shaped viruses (Iranzo et al. 2016). These viruses are also believed to exit the cell through a budding mechanism, likely mediated by homologs of the endosomal sorting complexes required for transport (ESCRT) machinery from the host that also participates in the production of extracellular vesicles (EVs) (Quemin et al. 2016; J. Liu et al. 2021)

The second clade, the *Halspviridae* family, is represented by a single-species, the Haloarcula hispanica virus 1 (HIS1) (Bath and Dyall-Smith 1998). This lemon-shaped virus was originally isolated from hypersaline environments and shares little-to-non sequence similarity with other spindle shaped viruses. It encodes a type B DNA polymerase for replication, and while HIS1 was originally proposed to exit the cell through lysis of the host, recent evidence suggests that similar to other spindle-shaped viruses, particles are released rather through a non-lytic mechanism, without disruption of the cellular membrane (Honga et al. 2015; Svirskaitė et al. 2016).

Likewise, the *Thaspviridae* family includes a pair of viruses infecting the ammonia-oxidizing archaeon *Nitrosopumilus* (Kim et al. 2019). The viruses Nitrosopumilus spindle-shaped virus 1 and 2 (NSV1 and NSV2 respectively) present the classic spindle morphology and short tails at one pole, and similarly to HIS1, they encode and presumably replicate through a type B DNA polymerase. These viruses have been observed to replicate in a non-lytic cycle, but infection does lead to growth inhibition of the host and the oxidation of ammonia, the main energy metabolism, is impacted during infection. Interestingly, NSVs represent some of the few isolated archaeal viruses from temperate environments, with evidence suggesting that they are widely distributed not only in the marine environment, but also in sediments and soils (Kim et al. 2019; Gazi et al. 2021).

The *Fuselloviridae* family includes numerous isolates infecting *Sulfolobus* and *Acidianus* species, and are relatively small spindle-shaped particles, whose dsDNA is packaged within the virion as a nucleoprotein complex (Quemin et al. 2015). These viruses are abundant and wide spread in hot spring around the world (Krupovic et al. 2014). Interestingly, it has been shown that the nucleoprotein complex of these viruses are extruded through the host membrane without lysis of the host (Quemin et al. 2016). Transcriptomic analysis of *Sulfolobus solfataricus* infected with the fusellovirus SSV1, revealed that there are only small changes in the transcriptional program of the host upon infection (Frõls et al. 2007). Meanwhile it has been shown that

SSV1 prefers to integrate into the host genome (though UV inducible), resulting in relatively low titers in the supernatants (Schleper, Kubo, and Zillig 1992).

2.10.4 Further unorthodox morphologies

As previously noted for spindle-shaped viruses, the *Portogloboviridae* and some filamentous viruses, several other clades of archaeal viruses exhibit this chronic non-lytic life cycle, including amongst others *Ampullaviridae, Globuloviridae*. Intriguingly, the vast majority of these viruses are currently classified within "orphan" families, as only a single or only a handful of representatives have been isolated, and often they present unique structural proteins and consequently, their position within the virosphere remains up for debate.

As it is the case for several aforementioned archaeal viral clades, the *Ampullaviridae* family is formed by a single-species genus, with Acidianus bottle-shaped virus (ABV) presenting one of the most enigmatic morphologies among archaeal viruses (Peng et al. 2007; Häring, Rachel, Peng, et al. 2005). The virions are composed of a conical core of supercoiled nucleoprotein, which attaches to a ring–like structure at the broader end of the virion, and has around 20 short filaments inserted perpendicular to the base (Figure 4). Additionally virions present an envelope that surround the cone section that has been proposed to participate in host attachment and adsorption. Furthermore, ABV encodes a type B DNA polymerase for genome replication, and it exits the cell in a non-lytic manner. Uncultivated relatives have been identified in metagenomic data from several hot spring systems around the world, suggesting that they are ubiquitous in acid hyperthermophilic environments (Baquero, Contursi, et al. 2020).

The *Globullaviridae* family is formed by two species of enveloped viruses that infect anaerobic hyperthermophilic archaea from the genera *Pyrobaculum* and *Termoproteus* (Häring et al. 2004) . The particles consist of spherical virions with a protein capsid and an external lipid membrane surrounding a nucleoprotein core of dsDNA genome. Unlike the *Ampullaviridae*, these viruses do not encode a replication protein, but they do encode a number of DNA binding proteins that might play a role in genome replication. Globuloviruses cause a chronic infection as particles are released from the host without lysis, taking up the lipid membrane in the budding process.

2.10.5 Pleomorphic viruses

Pleomorphic viruses from the *Pleolipoviridae* family represent a particular example among archaeal viruses. These viruses present a relatively simple structure, with the virions formed by a lipid membrane decorated with a single membrane-anchored protein (spike protein), and one or two different internal membrane proteins (Pietilä et al. 2009). Remarkably, members of this clade can present either ssDNA or dsDNA genomes, as well as hybrid versions of dsDNA with single-stranded interruptions (Senčilo et al. 2012; Pietilä et al. 2012). Pleolipoviruses have been isolated exclusively from members of the haloarchaea, with 15 viral species from

diverse hypersaline environments described and recognized by the ICTV, (Bamford et al. 2017; Mizuno et al. 2019).

Pleolipoviruses establish a chronic infection in their host, and exit the cell through an undescribed budding mechanism, while only causing mild growth retardation. Some pleolipoviruses can also enter lysogenic cycles as the can be found integrated as proviruses in haloarchaea genomes, and some isolates encode integrase genes (J. Wang et al. 2018; Demina et al. 2016).

The atomic structure of the ectodomains of the spike protein from two related pleolipoviruses (HRPV-2 and HRPV-6) has been determined, revealing an unprecedented V-shape fold, which is unrelated to any currently known structure of other viral or cellular protein (El Omari et al. 2019). These proteins presented two subdomains of roughly equal dimensions, and it has been proposed that the N-terminal subdomain unfolds its bundle of alpha helixes upon interaction with the host receptor and specifically the N2-subdomain is responsible for host recognition. Cryo-electron tomography revealed that this protein is in a monomeric state on the surface of the virion (Figure 5). Additional evidence has been presented supporting that the spike protein that decorates the virions participates in in host recognition and is able to trigger membrane fusion (Bignon, Chou, and Roine 2022).

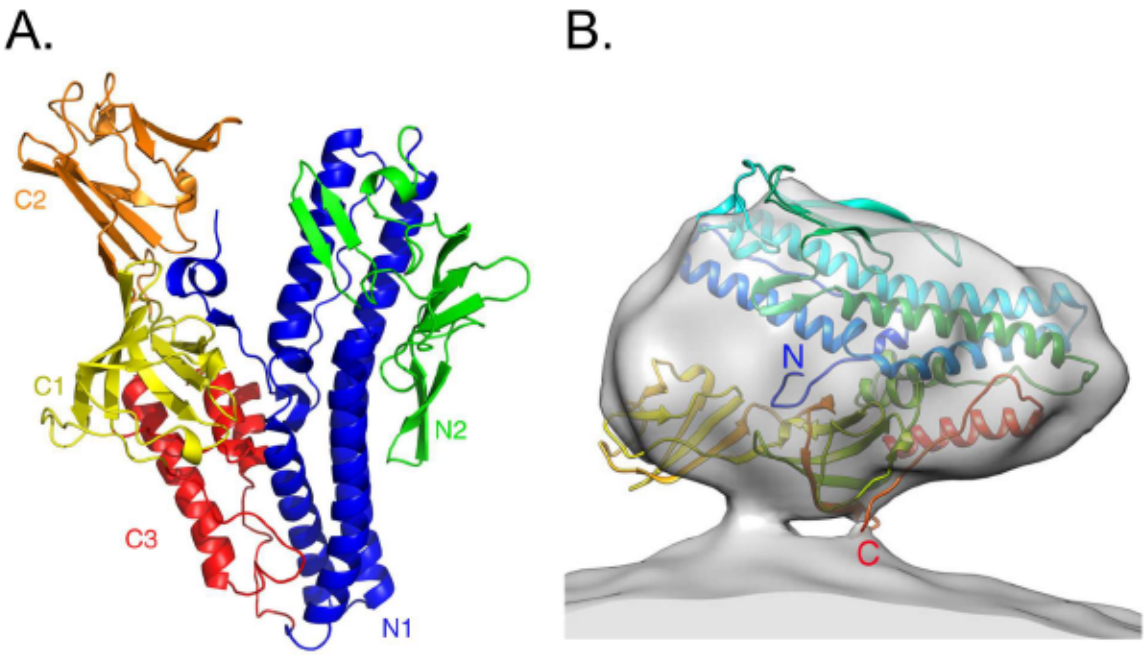


Figure 5: Pleolipovirus spike protein. Adapted from (El Omari et al., 2019). **A.** Cartoon representation of Halorubrum pleomorphic virus 2 (HRPV-2) spike protein (VP5). Colors correspond to C-term and N-term subdomains. **B.**VP5 crystallographic structure fitted into its density solved by cryo-electron tomography and subtomogram averaging.

Interestingly, all described pleolipoviruses previous to this work, were shown to have an extremely narrow host range, being able to infect only a single species or (Atanasova et al. 2012; Demina and Oksanen 2020). However, how these viruses interact with their hosts at the molecular level, or the transcriptional responses behind these infections remains unknown.

The *Pleolipoviridae* family is currently subdivided into three different genera: the *Alpha, Beta,* and *Gammapleolipovirus*. While all pleolipovirus present a syntenic core of conserved genes, the different genera present a clear case of modular organization, where the structural and replication modules seem to have follow different evolutionary trajectories. While the *Gammapleolipovirus* include two viruses infection *Haloarcula hispanica*, HIS2 and Hardyhisp2. These viruses encodes a protein primed type B DNA polymerase, while the other two genera do not encode a recognizable polymerase protein. The isolates from *Alphapleolipovirus* putatively replicate through the rolling-circle, as they encode two non-orthologous version of rolling-circle replication (initiation) endonuclease (RCRE). These preoteins were potentially acquired from two different groups of plasmids, the pGRB1-like and pTP2-like plasmid, which were originally isolated from *Halobacterium* and *Thermococcus*, respectively (Kazlauskas et al. 2019; Gorlas et al. 2013; Ebert, Goebel, and Pfeifer 1984). The rolling-circle cycle is a type of DNA strand displacement replication, where only one strand is replicated at a time. The initiation protein is commonly an endonuclease that generates a nick in the origin of replication on the virus/plasmid genome, and uses the host replication machinery to rapidly generate ssDNA copies of the template. The later are complemented with a second strand for transcription and packaging, or as it is the case for the majority of single-stranded viruses from the *Monodnaviria* realm (e.g. *Inoviridae*, *Circoviridae*, *Geminiviridae*, *Microviridae*, *Nanoviridae*), are directly enclosed into the caspsid (Malathi and Renuka Devi 2019; Te Riele, Michel, and Ehrlich 1986).

Conversely, the *Betapleolipovirus* do not encode any RCRE homolog, and in turn, they present a putative DNA binding protein located in a similar position on their genomes, which shares no similarity with any known replication-related nuclease or cellular protein (Krupovic et al. 2018). While the current genera organization is based on overall genome similarity and the type of replication protein does not constitute official demarcation criteria, it is unmistakable that the three genera have followed different evolutionary trajectories, and that their replication modules have been acquired likely through recombination with different extant and/or extinct lineages of MGEs.

The distinct modularity of this clades is also observed in other groups of prokaryotic and eukaryotic viruses (V., Dolja, and Krupovic 2017; Koonin, Krupovic, and Dolja 2023; Kazlauskas et al. 2019), and can often make taxonomic classification of these groups harder to elucidate. Particularly in the case of the *Pleolipoviridae* family, the presence of RCREs in alphapleolipoviruses, together with the variety of single and double-stranded genomes, as well as the unique spike protein fold, continuously obscures the taxonomic

classification of this clade. Consequently, pleolipovirus are currently classified within the *Monodnaviria* realm, and are single-members of the kingdom *Trapavirae*, phylum *Saleviricota*, class *Huolimaviricetes* and order *Haloruvirales*. This classification is driven by the presence of the RCRE, which is one of the aforementioned super-viral hallmark genes (VHGs). However, as discussed abundantly in the literature and summarized in the viral concept section, the defining trait of a *bona fide* virus is that it encodes a gene for the major virion protein. Therefore, in order to properly understand the internal structure and organization of this family, as well as their evolutionary relationships with other viruses or MGEs, it is key to resolve the diversity and evolution of both the structural and replication modules.

2.11 References

Allers, Thorsten, and Moshe Mevarech. 2005. "Archaeal Genetics - The Third Way." *Nature Reviews Genetics* 6 (1): 58–73. https://doi.org/10.1038/nrg1504.

Allers, Thorsten, Hien Ping Ngo, Moshe Mevarech, and Robert G. Lloyd. 2004. "Development of Additional Selectable Markers for the Halophilic Archaeon *Haloferax Volcanii* Based on the LeuB and TrpA Genes." *Applied and Environmental Microbiology* 70 (2): 943–53. https://doi.org/10.1128/AEM.70.2.943-953.2004.

Antón, J, E Llobet-Brossa, F Rodríguez-Valera, and R Amann. 1999. "Fluorescence in Situ Hybridization Analysis of the Prokaryotic Community Inhabiting Crystallizer Ponds." *Environmental Microbiology* 1: 517–23. http://www.scopus.com/inward/record.url?eid=2-s2.0-0033301308&partnerID=MN8TOARS.

Arnold, Hans Peter, Wolfram Zillig, Ulrike Ziese, Ingelore Holz, Marie Crosby, Terry Utterback, Jan F. Weidmann, et al. 2000. "A Novel Lipothrixvirus, SIFV, of the Extremely Thermophilic Crenarchaeon *Sulfolobus*." *Virology* 267 (2): 252–66. https://doi.org/10.1006/viro.1999.0105.

Atanasova, Nina S., Elina Roine, Aharon Oren, Dennis H. Bamford, and Hanna M. Oksanen. 2012. "Global Network of Specific Virus-Host Interactions in Hypersaline Environments." *Environmental Microbiology* 14 (2): 426–40. https://doi.org/10.1111/j.1462-2920.2011.02603.x.

Baker, Brett J., Valerie De Anda, Kiley W. Seitz, Nina Dombrowski, Alyson E. Santoro, and Karen G. Lloyd. 2020. "Diversity, Ecology and Evolution of Archaea." *Nature Microbiology* 5 (7): 887–900. https://doi.org/10.1038/s41564-020-0715-z.

Baker, Brett J., Luis R. Comolli, Gregory J. Dick, Loren J. Hauser, Doug Hyatt, Brian D. Dill, Miriam L. Land, Nathan C. VerBerkmoes, Robert L. Hettich, and Jillian F. Banfield. 2010. "Enigmatic, Ultrasmall, Uncultivated Archaea." *Proceedings of the National Academy of Sciences of the United States of America* 107 (19): 8806–11. https://doi.org/10.1073/pnas.0914470107.

Bamford, Dennis H., Maija K. Pietilä, Elina Roine, Nina S. Atanasova, Ana Dienstbier, and Hanna M. Oksanen. 2017. "ICTV Virus Taxonomy Profile: *Pleolipoviridae*." *Journal of General Virology* 98 (12): 2916–17. https://doi.org/10.1099/jgv.0.000972.

Baquero, Diana P., Patrizia Contursi, Monica Piochi, Simonetta Bartolucci, Ying Liu, Virginija Cvirkaite-Krupovic, David Prangishvili, and Mart Krupovic. 2020. "New Virus Isolates from Italian Hydrothermal Environments Underscore the Biogeographic Pattern in Archaeal Virus Communities." *ISME Journal* 14 (7): 1821–33. https://doi.org/10.1038/s41396-020-0653-z.

Baquero, Diana P., Anastasia D. Gazi, Martin Sachse, Junfeng Liu, Christine Schmitt, Maryse Moya-Nilges, Stefan Schouten, David Prangishvili, and Mart Krupovic. 2021. "A Filamentous Archaeal Virus Is Enveloped inside the Cell and Released through Pyramidal Portals." *Proceedings of the National Academy of Sciences* 118 (32): e2105540118. https://doi.org/10.1073/pnas.2105540118.

Baquero, Diana P., Ying Liu, Fengbin Wang, Edward H. Egelman, David Prangishvili, and Mart Krupovic. 2020. Structure and Assembly of Archaeal Viruses. Advances in Virus Research. 1st ed. Vol. 108. *Elsevier Inc.* https://doi.org/10.1016/bs.aivir.2020.09.004.

Barns, Susan M., Charles F. Delwiche, Jeffrey D. Palmer, and Norman R. Pace. 1996. "Perspectives on Archaeal Diversity, Thermophily and Monophyly from Environmental RRNA Sequences." *Proceedings of the National Academy of Sciences of the United States of America* 93 (17): 9188–93. https://doi.org/10.1073/pnas.93.17.9188.

Bath, Carolyn, Tania Cukalac, Kate Porter, and Michael L. Dyall-Smith. 2006. "His1 and His2 Are Distantly Related, Spindle-Shaped Haloviruses Belonging to the Novel Virus Group, *Salterprovirus*." *Virology* 350 (1): 228–39. https://doi.org/10.1016/j.virol.2006.02.005.

Bath, Carolyn, and Michael L. Dyall-Smith. 1998. "His1, an Archaeal Virus of TheFuselloviridae Family That Infects *Haloarcula Hispanica*." *Journal of Virology* 72 (11): 9392–95. https://doi.org/10.1128/jvi.72.11.9392-9395.1998.

Becker, Erin A., Phillip M. Seitzer, Andrew Tritt, David Larsen, Megan Krusor, Andrew I. Yao, Dongying Wu, et al. 2014. "Phylogenetically Driven Sequencing of Extremely Halophilic Archaea Reveals Strategies for Static and Dynamic Osmo-Response." *PLoS Genetics* 10 (11). https://doi.org/10.1371/journal.pgen.1004784.

Benlloch, S., S. G. Acinas, J. Antón, López-López, S. P. Luz, and F. Rodríguez-Valera. 2001. "Archaeal Biodiversity in Crystallizer Ponds from a Solar Saltern: Culture versus PCR." *Microbial Ecology* 41 (1): 12–19. https://doi.org/10.1007/s002480000069.

Berghuis, Bojk A., Feiqiao Brian Yu, Frederik Schulz, Paul C. Blainey, Tanja Woyke, and Stephen R. Quake. 2019. "Hydrogenotrophic Methanogenesis in Archaeal Phylum *Verstraetearchaeota* Reveals the Shared Ancestry of All Methanogens." *Proceedings of the National Academy of Sciences of the United States of America* 116 (11): 5037–44. https://doi.org/10.1073/pnas.1815631116.

Bignon, Eduardo A, Kevin R Chou, and Elina Roine. 2022. "Halorubrum Pleomorphic Virus-6 Membrane Fusion Is Triggered by an S-Layer Component of Its Haloarchaeal Host."

Boetius, Antje, and Samantha Joye. 2009. "Thriving in Salt." *Science* 324 (5934): 1523–25. https://doi.org/10.1126/science.1172979.

Boujelben, Ines, Pablo Yarza, Cristina Almansa, Judith Villamor, Sami Maalej, Josefa Antón, and Fernando Santos. 2012. "Virioplankton Community Structure in Tunisian Solar Salterns." *Applied and Environmental Microbiology* 78 (20): 7429–37. https://doi.org/10.1128/AEM.01793-12.

Breuert, Sebastian, Thorsten Allers, Gabi Spohn, and Jörg Soppa. 2006. "Regulated Polyploidy in Halophilic Archaea." *PLoS ONE* 1 (1). https://doi.org/10.1371/journal.pone.0000092.

Brumfield, Susan K., Alice C. Ortmann, Vincent Ruigrok, Peter Suci, Trevor Douglas, and Mark J. Young. 2009. " Particle Assembly and Ultrastructural Features Associated with Replication of the Lytic Archaeal Virus Sulfolobus Turreted Icosahedral Virus ." *Journal of Virology* 83 (12): 5964–70. https://doi.org/10.1128/jvi.02668-08.

Burns, David G., Peter H. Janssen, Takashi Itoh, Masahiro Kamekura, Zhou Li, Grant Jensen, Francisco Rodríguez-Valera, Henk Bolhuis, and Mike L. Dyall-Smith. 2007. "*Haloquadratum Walsbyi* Gen. Nov., Sp. Nov., the Square Haloarchaeon of Walsby, Isolated from Saltern Crystallizers in Australia and Spain." *International Journal of Systematic and Evolutionary Microbiology* 57 (2): 387–92. https://doi.org/10.1099/ijs.0.64690-0.

Cenens, William, Angela Makumi, Sander K. Govers, Rob Lavigne, and Abram Aertsen. 2015. "Viral Transmission Dynamics at Single-Cell Resolution Reveal Transiently Immune Subpopulations Caused by a Carrier State Association." *PLoS Genetics* 11 (12): 1–19. https://doi.org/10.1371/journal.pgen.1005770.

Cline, S. W., and W. Ford Doolittle. 1987. "Efficient Transfection of the Archaebacterium *Halobacterium Halobium*." *Journal of Bacteriology* 169 (3): 1341–44. https://doi.org/10.1128/jb.169.3.1341-1344.1987.

Cline, S. W., W. L. Lam, R. L. Charlebois, L. C. Schalkwyk, and W. F. Doolittle. 1989. "Transformation Methods for Halophilic Archaebacteria." *Canadian Journal of Microbiology* 35 (1): 148–52. https://doi.org/10.1139/m89-022.

Cobián Güemes, Ana Georgina, Merry Youle, Vito Adrian Cantú, Ben Felts, James Nulton, and Forest Rohwer. 2016. "Viruses as Winners in the Game of Life." *Annual Review of Virology* 3 (1): 197–214. https://doi.org/10.1146/annurev-virology-100114-054952.

DasSarma, Shiladitya, and Priya DasSarma. 2012. "Halophiles." *ELS*. https://doi.org/10.1002/9780470015902.a0000394.pub3.

DeLong, E. F. 1992. "Archaea in Coastal Marine Environments." *Proceedings of the National Academy of Sciences of the United States of America* 89 (12): 5685–89. https://doi.org/10.1073/pnas.89.12.5685.

Demina, Tatiana A., Nina S. Atanasova, Maija K. Pietilä, Hanna M. Oksanen, and Dennis H. Bamford. 2016. "Vesicle-like Virion of Haloarcula Hispanica Pleomorphic Virus 3 Preserves High Infectivity in Saturated Salt." *Virology* 499: 40–51. https://doi.org/10.1016/j.virol.2016.09.002.

Demina, Tatiana A., Mike Dyall-Smith, Matti Jalasvuori, Shishen Du, and Hanna M. Oksanen. 2023. "ICTV Virus Taxonomy Profile: *Sphaerolipoviridae* 2023." *Journal of General Virology* 104 (3): 1–2. https://doi.org/10.1099/jgv.0.001830.

Demina, Tatiana A., and Hanna M. Oksanen. 2020. "Pleomorphic Archaeal Viruses: The Family *Pleolipoviridae* Is Expanding by Seven New Species." *Archives of Virology* 165 (11): 2723–31. https://doi.org/10.1007/s00705-020-04689-1.

Dyall-Smith, Mike, Friedhelm Pfeiffer, Pei-Wen Chiang, and Sen-Lin Tang. "Genome sequence of Hardyhisp2, a Gammapleolipovirus infecting haloarcula hispanica." *Microbiology Resource Announcements* 10, no. 19 (2021): 10-1128.

Dybvig, Kevin, Jan A. Nowak, Todd L Sladek, and Jack Maniloff. 1985. "Identification of an Enveloped Phage, Mycoplasma Virus L172, That Contains a 14-Kilobase Single-Stranded DNA Genome." *Journal of Virology* 53 (2): 384–90. https://doi.org/10.1128/microbe.8.242.2.

Ebert, Klaus, Werner Goebel, and Felicitas Pfeifer. 1984. "Homologies between Heterogeneous Extrachromosomal DNA Populations of *Halobacterium Halobium* and Four New Halobacterial Isolates Klaus." *Molecular Genetics and Genomics* 194: 91–97.

Elkins, James G., Mircea Podar, David E. Graham, Kira S. Makarova, Yuri Wolf, Lennart Randau, Brian P. Hedlund, et al. 2008. "A Korarchaeal Genome Reveals Insights into the Evolution of the Archaea." *Proceedings of the National Academy of Sciences of the United States of America* 105 (23): 8102–7. https://doi.org/10.1073/pnas.0801980105.

Eme, Laura, Anja Spang, Jonathan Lombard, Courtney W. Stairs, and Thijs J.G. Ettema. 2017. "Archaea and the Origin of Eukaryotes." *Nature Reviews Microbiology* 15 (12): 711–23. https://doi.org/10.1038/nrmicro.2017.133.

Evans, Paul N., Donovan H. Parks, Grayson L. Chadwick, Steven J. Robbins, Victoria J. Orphan, Suzanne D. Golding, and Gene W. Tyson. 2015. "Methane Metabolism in the Archaeal Phylum *Bathyarchaeota* Revealed by Genome-Centric Metagenomics." *Science* 350 (6259): 434–38.

https://doi.org/10.1126/science.aac7745.

Forterre, Patrick. 2006. "The Origin of Viruses and Their Possible Roles in Major Evolutionary Transitions." *Virus Research* 117 (1): 5–16. https://doi.org/10.1016/j.virusres.2006.01.010.

Fröls, Sabrina, Paul M.K. Gordon, Mayi Arcellana Panlilio, Christa Schleper, and Christoph W. Sensen. 2007. "Elucidating the Transcription Cycle of the UV-Inducible Hyperthermophilic Archaeal Virus SSV1 by DNA Microarrays." *Virology* 365 (1): 48–59. https://doi.org/10.1016/j.virol.2007.03.033.

Gazi, Khaled S., Jong Geol Kim, Olorunleke Y. Olufowora, Ui Ju Lee, and Sung Keun Rhee. 2021. "Prevalence of Nitrosopumilus Spindle-Shaped Viruses in Different Ecogeographical Locations." *Korean Journal of Microbiology* 57 (2): 99–105. https://doi.org/10.7845/kjm.2021.1020.

Gorlas, Aurore, Mart Krupovic, Patrick Forterre, and Claire Geslin. 2013. "Living Side by Side with a Virus: Characterization of Two Novel Plasmids from *Thermococcus Prieurii*, a Host for the Spindle-Shaped Virus TPV1." *Applied and Environmental Microbiology* 79 (12): 3822–28. https://doi.org/10.1128/AEM.00525-13.

Guixa-Boixareu, Núria, Juan I. Calderón-Paz, Mikal Heldal, Gunnar Bratbak, and Carlos Pedrós-Alió. 1996. "Viral Lysis and Bacterivory as Prokaryotic Loss Factors along a Salinity Gradient." *Aquatic Microbial Ecology* 11 (3): 215–27. https://doi.org/10.3354/ame011215.

Gunde-Cimerman, Nina, Jose Ramos, and Ana Plemenitaš. 2009. "Halotolerant and Halophilic Fungi." *Mycological Research* 113 (11): 1231–41. https://doi.org/10.1016/j.mycres.2009.09.002.

Häring, Monika, Xu Peng, Kim Brügger, Reinhard Rachel, Karl O. Stetter, Roger A. Garrett, and David Prangishvili. 2004. "Morphology and Genome Organization of the Virus PSV of the Hyperthermophilic Archaeal Genera *Pyrobaculum* and *Thermoproteus*: A Novel Virus Family, the *Globuloviridae*." *Virology* 323 (2): 233–42. https://doi.org/10.1016/j.virol.2004.03.002.

Häring, Monika, Reinhard Rachel, Xu Peng, Roger A. Garrett, and David Prangishvili. 2005. " Viral Diversity in Hot Springs of Pozzuoli, Italy, and Characterization of a Unique Archaeal Virus, Acidianus Bottle-Shaped Virus, from a New Family, the *Ampullaviridae* ." *Journal of Virology* 79 (15): 9904–11. https://doi.org/10.1128/jvi.79.15.9904-9911.2005.

Häring, Monika, Gisle Vestergaard, Kim Brügger, Reinhard Rachel, Roger A. Garrett, and David Prangishvili. 2005. "Structure and Genome Organization of AFV2, a Novel Archaeal Lipothrixvirus with Unusual Terminal and Core Structures." *Journal of Bacteriology* 187 (11): 3855–58. https://doi.org/10.1128/JB.187.11.3855-3858.2005.

Häring, Monika, Gisle Vestergaard, Reinhard Rachel, Lanming Chen, Roger A. Garrett, and David Prangishvili. 2005. "Independent Virus Development Outside a Host." *Nature* 436 (25): 1102. https://doi.org/10.1038/nature4361101a.

Harrison, FC., and ME. Kennedy. 1922. "The Discoloration of Cured Codfish." *Trans Royal Society Canada Sect III* 16 (101–152).

Hoffmann-Berling, Hartmut, and René Mazé. 1964. "Release of Male-Specific Bacteriophages From Surviving Host Bacteria." *Strain* 313: 305–13.

Honga, Chuan, Maija K. Pietilä, Caroline J. Fu, Michael F. Schmid, Dennis H. Bamford, and Wah Chiu. 2015. "Lemon-Shaped Halo Archaeal Virus His1 with Uniform Tail but Variable Capsid Structure." *Proceedings of the National Academy of Sciences of the United States of America* 112 (8): 2449–54. https://doi.org/10.1073/pnas.1425008112.

Huber, Harald, Michael J. Hohn, Reinhard Rachel, Tanja Fuchs, Verena C. Wimmer, and Karl O. Stetter. 2002. "A New Phylum of Archaea Represented by a Nanosized Hyperthermophilic Symbiont." *Nature* 417 (6884): 63–67. https://doi.org/10.1038/417063a.

Imachi, Hiroyuki, Masaru K. Nobu, Nozomi Nakahara, Yuki Morono, Miyuki Ogawara, Yoshihiro Takaki, Yoshinori Takano, et al. 2020. "Isolation of an Archaeon at the Prokaryote–Eukaryote Interface." *Nature* 577 (7791): 519–25. https://doi.org/10.1038/s41586-019-1916-6.

Iranzo, Jaime, Eugene V. Koonin, David Prangishvili, and Mart Krupovic. 2016. "Bipartite Network Analysis of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements." *Journal of Virology* 90 (24): 11043–55. https://doi.org/10.1128/jvi.01622-16.

Jacob, F., S. Brenner, and F. Cuzin. 1963. "On the Regulation of DNA Replication in Bacteria." *Cold Spring Harbor Symposia on Quantitative Biology* 28 (0): 329–48. https://doi.org/10.1101/sqb.1963.028.01.048.

Kasson, Peter, Frank Dimaio, Xiong Yu, Soizick Lucas-Staat, Mart Krupovic, Stefan Schouten, David Prangishvili, and Edward H. Egelman. 2017. "Model for a Novel Membrane Envelope in a Filamentous Hyperthermophilic Virus." *ELife* 6: 1–17. https://doi.org/10.7554/eLife.26268.

Kauri, Tiiu, Rebecca Wallace, and Donn J. Kushner. 1990. "Nutrition of the Halophilic Archaebacterium, *Haloferax Volcanii*." *Systematic and Applied Microbiology* 13 (1): 14–18. https://doi.org/10.1016/S0723-2020(11)80174-8.

Kazlauskas, Darius, Arvind Varsani, Eugene V. Koonin, and Mart Krupovic. 2019. "Multiple Origins of Prokaryotic and Eukaryotic Single-Stranded DNA Viruses from Bacterial and Archaeal Plasmids."

Nature Communications 10 (1): 1–12. https://doi.org/10.1038/s41467-019-11433-0.

Kim, Jong Geol, So Jeong Kim, Virginija Cvirkaite-Krupovic, Woon Jong Yu, Joo Han Gwak, Mario López-Pérez, Francisco Rodriguez-Valera, Mart Krupovic, Jang Cheon Cho, and Sung Keun Rhee. 2019. "Spindle-Shaped Viruses Infect Marine Ammoniaoxidizing Thaumarchaea." *Proceedings of the National Academy of Sciences of the United States of America* 116 (31): 15645–50. https://doi.org/10.1073/pnas.1905682116.

Knezevic, Petar, and Evelien M. Adriaenssens. 2021. "ICTV Virus Taxonomy Profile: *Inoviridae*." *Journal of General Virology* 102 (7): 10–11. https://doi.org/10.1099/jgv.0.001614.

Könneke, Martin, Anne E. Bernhard, José R. De La Torre, Christopher B. Walker, John B. Waterbury, and David A. Stahl. 2005. "Isolation of an Autotrophic Ammonia-Oxidizing Marine Archaeon." *Nature* 437 (7058): 543–46. https://doi.org/10.1038/nature03911.

Koonin, Eugene, Valerian V. Dolja, Mart Krupovic, Arvind Varsani, Yuri I Wolf, Natalya Yutin, F. Murilo Zerbini, and Jens H. Kuhn. 2020. "Global Organization and Proposed Megataxonomy of the Virus World." *Microbiology and Molecular Biology Reviews* 84 (2). https://doi.org/10.1128/MMBR.00061-19CE.

Koonin, Eugene V., Mart Krupovic, and Valerian V. Dolja. 2023. "The Global Virome: How Much Diversity and How Many Independent Origins?" *Environmental Microbiology* 25 (1): 40–44. https://doi.org/10.1111/1462-2920.16207.

Koonin, Eugene V., Tatiana G. Senkevich, and Valerian V. Dolja. 2006. "The Ancient Virus World and Evolution of Cells." *Biology Direct* 1 (September). https://doi.org/10.1186/1745-6150-1-29.

Koonin, Eugene V., and Petro Starokadomskyy. 2016. "Are Viruses Alive? The Replicator Paradigm Sheds Decisive Light on an Old but Misguided Question." *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences* 59: 125–34. https://doi.org/10.1016/j.shpsc.2016.02.016.

Krupovic, Mart, Virginija Cvirkaite-Krupovic, Jaime Iranzo, David Prangishvili, and Eugene V. Koonin. 2018. "Viruses of Archaea: Structural, Functional, Environmental and Evolutionary Genomics." *Virus Research* 244 (November 2017): 181–93. https://doi.org/10.1016/j.virusres.2017.11.025.

Krupovic, Mart, Valerian V. Dolja, and Eugene V. Koonin. 2019. "Origin of Viruses: Primordial Replicators Recruiting Capsids from Hosts." *Nature Reviews Microbiology* 17 (7): 449–58. https://doi.org/10.1038/s41579-019-0205-6.

Krupovic, Mart, and Eugene V. Koonin. 2017. "Multiple Origins of Viral Capsid Proteins from Cellular Ancestors." *Proceedings of the National Academy of Sciences* 114 (12): E2401–10. https://doi.org/10.1073/pnas.1621061114.

Krupovic, Mart, Jens H. Kuhn, Fengbin Wang, Diana P. Baquero, Valerian V. Dolja, Edward H. Egelman, David Prangishvili, and Eugene V. Koonin. 2021. "*Adnaviria*: A New Realm for Archaeal Filamentous Viruses with Linear A-Form Double-Stranded DNA Genomes Mart." *Journal of Virology* 95 (15).

Krupovic, Mart, Emmanuelle R. J. Quemin, Dennis H. Bamford, Patrick Forterre, and David Prangishvili. 2014. "Unification of the Globally Distributed Spindle-Shaped Viruses of the Archaea." *Journal of Virology* 88 (4): 2354–58. https://doi.org/10.1128/jvi.02941-13.

Kügelgen, Andriko von, Vikram Alva, and Tanmay A.M. Bharat. 2021. "Complete Atomic Structure of a Native Archaeal Cell Surface." *Cell Reports* 37 (8). https://doi.org/10.1016/j.celrep.2021.110052.

Kuprat, Tom, Ulrike Johnsen, Marius Ortjohann, and Peter Schönheit. 2020. "Acetate Metabolism in Archaea: Characterization of an Acetate Transporter and of Enzymes Involved in Acetate Activation and Gluconeogenesis in *Haloferax Volcanii*." *Frontiers in Microbiology* 11 (December): 1–16. https://doi.org/10.3389/fmicb.2020.604926.

Laso-Pérez, Rafael, Gunter Wegener, Katrin Knittel, Friedrich Widdel, Katie J. Harding, Viola Krukenberg, Dimitri V. Meier, et al. 2016. "Thermophilic Archaea Activate Butane via Alkyl-Coenzyme M Formation." *Nature* 539 (7629): 396–401. https://doi.org/10.1038/nature20152.

Liu, Junfeng, Virginija Cvirkaite-Krupovic, Pierre Henri Commere, Yunfeng Yang, Fan Zhou, Patrick Forterre, Yulong Shen, and Mart Krupovic. 2021. "Archaeal Extracellular Vesicles Are Produced in an ESCRT-Dependent Manner and Promote Gene Transfer and Nutrient Cycling in Extreme Environments." *ISME Journal* 15 (10): 2892–2905. https://doi.org/10.1038/s41396-021-00984-0.

Liu, Ying, Tatiana A. Demina, Simon Roux, Pakorn Aiewsakun, Darius Kazlauskas, Peter Simmonds, David Prangishvili, Hanna M. Oksanen, and Mart Krupovic. 2021. "Diversity, Taxonomy, and Evolution of Archaeal Viruses of the Class *Caudoviricetes*." *PLoS Biology* 19 (11): 1–26. https://doi.org/10.1371/journal.pbio.3001442.

Liu, Ying, Sonoko Ishino, Yoshizumi Ishino, Gérard Pehau-Arnaudet, Mart Krupovic, and David Prangishvili. 2017. "A Novel Type of Polyhedral Viruses Infecting Hyperthermophilic Archaea." *Journal of Virology* 91 (13): 1–14. https://doi.org/10.1128/jvi.00589-17.

Lledó, B., R. M. Martínez-Espinosa, F. C. Marhuenda-Egea, and M. J. Bonete. 2004. "Respiratory Nitrate

Reductase from Haloarchaeon *Haloferax Mediterranei*: Biochemical and Genetic Analysis." *Biochimica et Biophysica Acta - General Subjects* 1674 (1): 50–59. https://doi.org/10.1016/j.bbagen.2004.05.007.

Luk, Alison W.S., Timothy J. Williams, Susanne Erdmann, R. Thane Papke, and Ricardo Cavicchioli. 2014. "Viruses of Haloarchaea." *Life*. MDPI AG. https://doi.org/10.3390/life4040681.

Madigan, Michael T., John M. Martinko, David A. Stahl, and David P. Clark. 2013. *Brock Biology of Microorganisms (13th Ed.)*. Vol. 12. https://doi.org/10.1007/s13398-014-0173-7.2.

Malathi, V. G., and P. Renuka Devi. 2019. "SsDNA Viruses: Key Players in Global Virome." *VirusDisease* 30 (1): 3–12. https://doi.org/10.1007/s13337-019-00519-4.

Maniloff, Jack, Jyotirmoy Das, and J. R. Christensen. 1977. "Viruses of Mycoplasmas and Spiroplasmas." *Adv. Virus Res.*, no. 21: 343–380. http://dx.doi.org/10.1016/s0065-3527(08)60765-4.

Mäntynen, Sari, Elina Laanto, Hanna M Oksanen, Minna M Poranen, and Samuel L Díaz-Muñoz. 2021. "Black Box of Phage-Bacterium Interactions: Exploring Alternative Phage Infection Strategies." *Open Biology* 11 (9): 210188. https://doi.org/10.1098/rsob.210188.

McGenity, Terry J., and Aharon Oren. 2012. "Hypersaline Environments." In *Life at Extremes: Environments, Organisms and Strategies for Survival. Wallingford UK: Cabi*, 402–37. https://doi.org/10.1007/978-1-4020-9212-1_115.

Mevarech, M., and R. Werczberger. 1985. "Genetic Transfer in *Halobacterium Volcanii*." *Journal of Bacteriology* 162 (1): 461–62. https://doi.org/10.1128/jb.162.1.461-462.1985.

Mizuno, Carolina M., Bina Prajapati, Soizick Lucas-Staat, Telesphore Sime-Ngando, Patrick Forterre, Dennis H. Bamford, David Prangishvili, Mart Krupovic, and Hanna M. Oksanen. 2019. "Novel Haloarchaeal Viruses from Lake Retba Infecting *Haloferax* and *Halorubrum* Species." *Environmental Microbiology* 21 (6): 2129–47. https://doi.org/10.1111/1462-2920.14604.

Mochizuki, Tomohiro, Mart Krupovic, Gérard Pehau-Arnaudet, Yoshihiko Sako, Patrick Forterre, and David Prangishvili. 2012. "Archaeal Virus with Exceptional Virion Architecture and the Largest Single-Stranded DNA Genome." *Proceedings of the National Academy of Sciences of the United States of America* 109 (33): 13386–91. https://doi.org/10.1073/pnas.1203668109.

Mochizuki, Tomohiro, Takashi Yoshida, Reiji Tanaka, Patrick Forterre, Yoshihiko Sako, and David Prangishvili. 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 (2): 347–54. https://doi.org/10.1016/j.virol.2010.03.046.

Mongodin, E. F., K. E. Nelson, S. Daugherty, R. T. DeBoy, J. Wister, H. Khouri, J. Weidman, et al. 2005. "The Genome of *Salinibacter Ruber*: Convergence and Gene Exchange among Hyperhalophilic Bacteria and Archaea." *Proceedings of the National Academy of Sciences of the United States of America* 102 (50): 18147–52. https://doi.org/10.1073/pnas.0509073102.

Mullakhanbhai, Moiz F., and Helge Larsen. 1975. "*Halobacterium Volcanii* Spec. Nov., a Dead Sea Halobacterium with a Moderate Salt Requirement." *Archives of Microbiology* 104 (1): 207–14. https://doi.org/10.1007/BF00447326.

Nunoura, Takuro, Yoshihiro Takaki, Jungo Kakuta, Shinro Nishi, Junichi Sugahara, Hiromi Kazama, Gab Joo Chee, et al. 2011. "Insights into the Evolution of Archaea and Eukaryotic Protein Modifier Systems Revealed by the Genome of a Novel Archaeal Group." *Nucleic Acids Research* 39 (8): 3204–23. https://doi.org/10.1093/nar/gkq1228.

Oesterhelt, Dieter, and Walther Stoeckenius. 1971. "Rhodopsin-like Protein from the Purple Membrane of *Halobacterium Halobium*." *Nature New Biology* 233 (39): 149–52.

Omari, Kamel El, Sai Li, Abhay Kotecha, Thomas S. Walter, Eduardo A. Bignon, Karl Harlos, Pentti Somerharju, et al. 2019. "The Structure of a Prokaryotic Viral Envelope Protein Expands the Landscape of Membrane Fusion Proteins." *Nature Communications* 10 (1). https://doi.org/10.1038/s41467-019-08728-7.

Oren, Aharon. 1994. "The Ecology of the Extremely Halophilic Archaea." *FEMS Microbiology Reviews* 13 (4): 415–39. https://doi.org/10.1016/0168-6445(94)90063-9.

———. 2002. "Molecular Ecology of Extremely Halophilic Archaea and Bacteria." *FEMS Microbiology Ecology* 39 (1): 1–7. https://doi.org/10.1016/S0168-6496(01)00200-8.

———. 2005. "A Hundred Years of *Dunaliella* Research: 1905–2005." *Saline Systems* 1 (1): 1–14. https://doi.org/10.1186/1746-1448-1-2.

———. 2011. "Thermodynamic Limits to Microbial Life at High Salt Concentrations." *Environmental Microbiology* 13 (8): 1908–23. https://doi.org/10.1111/j.1462-2920.2010.02365.x.

Oren, Aharon, Gunnar Bratbak, and Mikal Heldal. 1997. "Occurrence of Virus-like Particles in the Dead Sea." *Extremophiles* 1 (3): 143–49. https://doi.org/10.1007/s007920050027.

Peng, Xu, Tamara Basta, Monika Häring, Roger A. Garrett, and David Prangishvili. 2007. "Genome of the Acidianus Bottle-Shaped Virus and Insights into the Replication and Packaging Mechanisms." *Virology* 364 (1): 237–43. https://doi.org/10.1016/j.virol.2007.03.005.

Pietilä, Maija K., Nina S. Atanasova, Violeta Manole, Lassi Liljeroos, Sarah J. Butcher, Hanna M. Oksanen, and Dennis H. Bamford. 2012. "Virion Architecture Unifies Globally Distributed Pleolipoviruses Infecting Halophilic Archaea." *Journal of Virology* 86 (9): 5067–79. https://doi.org/10.1128/jvi.06915-11.

Pietilä, Maija K., Elina Roine, Lars Paulin, Nisse Kalkkinen, and Dennis H. Bamford. 2009. "An SsDNA Virus Infecting Archaea: A New Lineage of Viruses with a Membrane Envelope." *Molecular Microbiology* 72 (2): 307–19. https://doi.org/10.1111/j.1365-2958.2009.06642.x.

Prangishvili, David, Dennis H. Bamford, Patrick Forterre, Jaime Iranzo, Eugene V. Koonin, and Mart Krupovic. 2017. "The Enigmatic Archaeal Virosphere." *Nature Reviews Microbiology* 15 (12): 724–39. https://doi.org/10.1038/nrmicro.2017.125.

Qi, Qiuzi, Yoshiyasu Ito, Katsuhiko Yoshimatsu, and Taketomo Fujiwara. 2016. "Transcriptional Regulation of Dimethyl Sulfoxide Respiration in a Haloarchaeon, *Haloferax Volcanii*." *Extremophiles* 20 (1): 27–36. https://doi.org/10.1007/s00792-015-0794-6.

Quemin, Emmanuelle R.J., Petr Chlanda, Martin Sachse, Patrick Forterre, David Prangishvili, and Mart Krupovic. 2016. "Eukaryotic-like Virus Budding in Archaea." *MBio* 7 (5): 3–7. https://doi.org/10.1128/mBio.01439-16.

Quemin, Emmanuelle R J, Maija K Pietilä, Hanna M Oksanen, Patrick Forterre, W Irene C Rijpstra, and Stefan Schouten. 2015. "Spindle-Shaped Virus 1 Contains Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids" 89 (22): 11681–91. https://doi.org/10.1128/JVI.02270-15.Editor.

Riele, H. Te, B. Michel, and S. D. Ehrlich. 1986. "Single-Stranded Plasmid DNA in *Bacillus Subtilis* and *Staphylococcus Aureus*." *Proceedings of the National Academy of Sciences of the United States of America* 83 (8): 2541–45. https://doi.org/10.1073/pnas.83.8.2541.

Rinke, Christian, Maria Chuvochina, Aaron J. Mussig, Pierre Alain Chaumeil, Adrián A. Davín, David W. Waite, William B. Whitman, Donovan H. Parks, and Philip Hugenholtz. 2021. "A Standardized Archaeal Taxonomy for the Genome Taxonomy Database." *Nature Microbiology* 6 (7): 946–59. https://doi.org/10.1038/s41564-021-00918-8.

Ripp, Steven, and Robert V. Miller. 1998. "Dynamics of the Pseudolysogenic Response in Slowly Growing Cells of *Pseudomonas Aeruginosa*." *Microbiology* 144 (8): 2225–32. https://doi.org/10.1099/00221287-144-8-2225.

Rodrigues-Oliveira, Thiago, Aline Belmok, Deborah Vasconcellos, Bernhard Schuster, and Cynthia M. Kyaw.

2017. "Archaeal S-Layers: Overview and Current State of the Art." *Frontiers in Microbiology* 8 (DEC): 1–17. https://doi.org/10.3389/fmicb.2017.02597.

Roux, Simon, Mart Krupovic, Rebecca A. Daly, Adair L. Borges, Stephen Nayfach, Frederik Schulz, Allison Sharrar, et al. 2019. "Cryptic Inoviruses Revealed as Pervasive in Bacteria and Archaea across Earth's Biomes." *Nature Microbiology* 4 (11): 1895–1906. https://doi.org/10.1038/s41564-019-0510-x.

Ryan, Francis Patrick. 2016. "Viral Symbiosis and the Holobiontic Nature of the Human Genome." *Apmis*. https://doi.org/10.1111/apm.12488.

Schleper, C., K. Kubo, and W. Zillig. 1992. "The Particle SSV1 from the Extremely Thermophilic Archaeon *Sulfolobus* Is a Virus: Demonstration of Infectivity and of Transfection with Viral DNA." *Proceedings of the National Academy of Sciences of the United States of America* 89 (16): 7645–49. https://doi.org/10.1073/pnas.89.16.7645.

Senčilo, Ana, Lars Paulin, Stefanie Kellner, Mark Helm, and Elina Roine. 2012. "Related Haloarchaeal Pleomorphic Viruses Contain Different Genome Types." *Nucleic Acids Research* 40 (12): 5523–34. https://doi.org/10.1093/nar/gks215.

Sime-Ngando, Télesphore, Soizick Lucas, Agnès Robin, Kimberly Pause Tucker, Jonathan Colombet, Yvan Bettarel, Elie Desmond, et al. 2011. "Diversity of Virus-Host Systems in Hypersaline Lake Retba, Senegal." *Environmental Microbiology* 13 (8): 1956–72. https://doi.org/10.1111/j.1462-2920.2010.02323.x.

Soppa, Jörg. 2006. "From Genomes to Function: Haloarchaea as Model Organisms." *Microbiology* 152 (3): 585–90. https://doi.org/10.1099/mic.0.28504-0.

Sorokin, Dimitry Y., Ilya V. Kublanov, Sergei N. Gavrilov, David Rojo, Pawel Roman, Peter N. Golyshin, Vladlen Z. Slepak, et al. 2016. "Elemental Sulfur and Acetate Can Support Life of a Novel Strictly Anaerobic Haloarchaeon." *ISME Journal* 10 (1): 240–52. https://doi.org/10.1038/ismej.2015.79.

Soucy, Shannon M., Jinling Huang, and Johann Peter Gogarten. 2015. "Horizontal Gene Transfer: Building the Web of Life." *Nature Reviews Genetics* 16 (8): 472–82. https://doi.org/10.1038/nrg3962.

Spang, Anja, Eva F. Caceres, and Thijs J.G. Ettema. 2017. "Genomic Exploration of the Diversity, Ecology, and Evolution of the Archaeal Domain of Life." *Science* 357 (6351). https://doi.org/10.1126/science.aaf3883.

Spang, Anja, Jimmy H. Saw, Steffen L. Jørgensen, Katarzyna Zaremba-Niedzwiedzka, Joran Martijn, Anders E. Lind, Roel Van Eijk, Christa Schleper, Lionel Guy, and Thijs J.G. Ettema. 2015. "Complex Archaea

That Bridge the Gap between Prokaryotes and Eukaryotes." *Nature* 521 (7551): 173–79. https://doi.org/10.1038/nature14447.

Stadtman, Thressa C., and H. A. Barker. 1951. "STUDIES ON THE METHANE FERMENTATION IX: The Origin of Methane in the Acetate and Methanol Fermentations by *Methanosarcina*." *Journal of Bacteriology* 61 (1): 81–86.

Sutter, Jan Moritz, Julia Beate Tästensen, Ulrike Johnsen, Jörg Soppa, and Peter Schönheit. 2016. "Key Enzymes of the Semiphosphorylative Entner-Doudoroff Pathway in the Haloarchaeon *Haloferax Volcanii*: Characterization of Glucose Dehydrogenase, Gluconate Dehydratase, and 2-Keto-3-Deoxy-6-Phosphogluconate Aldolase." *Journal of Bacteriology* 198 (16): 2251–62. https://doi.org/10.1128/JB.00286-16.

Svirskaitė, Julija, Hanna M. Oksanen, Rimantas Daugelavičius, and Dennis H. Bamford. 2016. "Monitoring Physiological Changes in Haloarchaeal Cell during Virus Release." *Viruses* 8 (3). https://doi.org/10.3390/v8030059.

Szirovicza, Leonóra. 2022. "Characterization and Molecular Biology of Daletvirus Boae - a Founding Member of the Realm Ribozyviria."

Takashina, Tomonori, Tetsuo Hamamoto, Kiyotaka Otozai, William D. Grant, and Koki Horikoshi. 1990. "*Haloarcula Japonica* Sp. Nov., a New Triangular Halophilic Archaebacterium." *Systematic and Applied Microbiology* 13 (2): 177–81. https://doi.org/10.1016/S0723-2020(11)80165-7.

V., Eugene Koonin, Valerian V. Dolja, and Mart Krupovic. 2017. "Origins and Evolution of Viruses of Eukaryotes: The Ultimate Modularity." *Physiology & Behavior* 176 (12): 139–48. https://doi.org/10.1016/j.physbeh.2017.03.040.

Veesler, David, Thiam Seng Ng, Anoop K. Sendamarai, Brian J. Eilers, C. Martin Lawrence, Shee Mei Lok, Mark J. Young, John E. Johnson, and Chi Yu Fu. 2013. "Atomic Structure of the 75 MDa Extremophile Sulfolobus Turreted Icosahedral Virus Determined by CryoEM and X-Ray Crystallography." *Proceedings of the National Academy of Sciences of the United States of America* 110 (14): 5504–9. https://doi.org/10.1073/pnas.1300601110.

Ventosa, A., and A. Oren. 1996. "*Halobacterium Salinarum* Nom. Corrig., a Name to Replace *Halobacterium Salinarium* (Elazari-Volcani) and to Include *Halobacterium Halobium* and *Halobacterium Cutirubrum*." *International Journal of Systematic Bacteriology* 46 (1): 347. https://doi.org/10.1099/00207713-46-1-347.

Virgin, Herbert W., E. John Wherry, and Rafi Ahmed. 2009. "Redefining Chronic Viral Infection." *Cell* 138 (1): 30–50. https://doi.org/10.1016/j.cell.2009.06.036.

Walsby, A. E. 1980. "A Square Bacterium." *Nature* 283 (5742): 69–71. https://doi.org/10.1038/283069a0.

Wang, Fengbin, Diana P Baquero, Leticia C Beltran, Zhangli Su, Tomasz Osinski, Weili Zheng, David Prangishvili, Mart Krupovic, and Edward H Egelman. 2020. "Structures of Filamentous Viruses Infecting Hyperthermophilic Archaea Explain DNA Stabilization in Extreme Environments." *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.2011125117.

Wang, Fengbin, Virginija Cvirkaite-Krupovic, Matthijn Vos, Leticia C. Beltran, Mark A.B. Kreutzberger, Jean Marie Winter, Zhangli Su, et al. 2022. "Spindle-Shaped Archaeal Viruses Evolved from Rod-Shaped Ancestors to Package a Larger Genome Graphical." *Cell* 185 (8): 1297-1307.e11. https://doi.org/10.1016/j.cell.2022.02.019.

Wang, Fengbin, Ying Liu, Zhangli Su, Tomasz Osinski, Guilherme A.P. de Oliveira, James F. Conway, Stefan Schouten, Mart Krupovic, David Prangishvili, and Edward H. Egelman. 2019. "A Packing for A-Form DNA in an Icosahedral Virus." *Proceedings of the National Academy of Sciences of the United States of America* 116 (45): 22591–97. https://doi.org/10.1073/pnas.1908242116.

Wang, Jiao, Yingchun Liu, Ying Liu, Kaixin Du, Shuqi Xu, Yuchen Wang, Mart Krupovic, and Xiangdong Chen. 2018. "A Novel Family of Tyrosine Integrases Encoded by the Temperate Pleolipovirus SNJ2." *Nucleic Acids Research* 46 (5): 2521–36. https://doi.org/10.1093/nar/gky005.

Wilhelm, Steven W, and Curtis a Suttle. 1999. "Viruses and Nutrient Cycles in the Sea Aquatic Food Webs." *BioScience* 49 (October): 781–88. https://doi.org/10.2307/1313569.

Woese, C. R., and G. E. Fox. 1977. "Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms." *Proceedings of the National Academy of Sciences of the United States of America* 74 (11): 5088–90. https://doi.org/10.1073/pnas.74.11.5088.

Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. "Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya." *Proceedings of the National Academy of Sciences of the United States of America* 87 (12): 4576–79. https://doi.org/10.1073/pnas.87.12.4576.

Young, Kevin D. 2006. "The Selective Value of Bacterial Shape." *Microbiology and Molecular Biology Reviews* 70 (3): 660–703. https://doi.org/10.1128/mmbr.00001-06.

Youssef, Noha H., Christian Rinke, Ramunas Stepanauskas, Ibrahim Farag, Tanja Woyke, and Mostafa S. Elshahed. 2015. "Insights into the Metabolism, Lifestyle and Putative Evolutionary History of the Novel

Archaeal Phylum '*Diapherotrites*.'" *ISME Journal* 9 (2): 447–60. https://doi.org/10.1038/ismej.2014.141.

Zaremba-Niedzwiedzka, Katarzyna, Eva F. Caceres, Jimmy H. Saw, DIsa Bäckström, Lina Juzokaite, Emmelien Vancaester, Kiley W. Seitz, et al. 2017. "Asgard Archaea Illuminate the Origin of Eukaryotic Cellular Complexity." *Nature* 541 (7637): 353–58. https://doi.org/10.1038/nature21031.

Zhou, Guangyin, David Kowalczyk, Matthew A. Humbard, Sunil Rohatgi, and Julie A. Maupin-Furlow. 2008. "Proteasomal Components Required for Cell Growth and Stress Responses in the Haloarchaeon *Haloferax Volcanii*." *Journal of Bacteriology* 190 (24): 8096–8105. https://doi.org/10.1128/JB.01180-08.

3. Aim of this dissertation

Viruses that establish a chronic infection are widespread among archaeal groups. However, while non-lytic infections are of great importance in eukaryotic systems, their impact on host metabolism, as well as many aspect of virus-host interaction remain poorly understood in prokaryotic systems. As these viruses can have a major impact on cell behavior, chronically infected cells can have different biochemical and ecological roles within microbial communities.

Pleolipoviruses represent an intriguing group, because they exhibit many unusual and puzzling characteristics. They present a rare diversity of nucleic acid composition, as they have evolved single and double-stranded, as well as hybrid versions of DNA genomes. They exhibit a relatively simple virion structure, which resembles the one from multiple eukaryotic viruses, which could aid the understanding of structures involved in prokaryotic membrane fusion processes. They have an unknown and likely complex evolutionary origin, as they present a unique protein fold of their major capsid protein, and an entangled and diverse replication machinery. In addition, they are able to establish chronic infections, with little to no impact on host physiology. This lifestyle likely requires intricate and fine-tuned interaction with host metabolism. Furthermore, their hosts of isolation include the culture-amenable haloarchaea, and include model organisms with genetic tools available. This is a key variable to elucidate the mechanisms that drive virus-host interaction under controlled and manipulable conditions.

During the course of this study, a novel species of pleolipovirus was isolated and comprehensively characterized. However, this novel virus challenged many established concepts about pleolipovirus physiology and evolution, and therefore the aim of this study was to understand this virus-host interaction and its driving factors, as well as to investigate the evolutionary history of pleolipoviruses. The specific aims of this study were

I. To establish a reliable and consistent virus-host system in a suitable haloarchaeal host as a model for study of chronic virus infections.
II. To investigate the life cycle of pleolipoviruses and the molecular mechanisms behind the establishment of chronic infections.
III. To explore the diversity of pleolipoviruses, their phylogenetic relationships and their evolutionary trajectories

4. Chapter 1.

Isolation of the first virus causing a chronic infection in the archaeal model organism *Haloferax volcanii* reveals antiviral activities of a provirus

Tomas Alarcón-Schumacher, Adit Naor, Uri Gophna and Susanne Erdmann

Published in Proceedings of the National Academy of Sciences.

2022 Aug 30;119(35):e2205037119.

PNAS RESEARCH ARTICLE | MICROBIOLOGY

Isolation of a virus causing a chronic infection in the archaeal model organism *Haloferax volcanii* reveals antiviral activities of a provirus

Tomas Alarcón-Schumacher[a], Adit Naor[b,1], Uri Gophna[b], and Susanne Erdmann[a,2]

Edited by Gustavo Palacios, Icahn School of Medicine at Mount Sinai, New York, NY; received March 23, 2022; accepted June 28, 2022 by Editorial Board Member Adolfo Garcia-Sastre

Viruses are important ecological, biogeochemical, and evolutionary drivers in every environment. Upon infection, they often cause the lysis of the host cell. However, some viruses exhibit alternative life cycles, such as chronic infections without cell lysis. The nature and the impact of chronic infections in prokaryotic host organisms remains largely unknown. Here, we characterize a novel haloarchaeal virus, Haloferax volcanii pleomorphic virus 1 (HFPV-1), which is currently the only virus infecting the model haloarchaeon *Haloferax volcanii* DS2, and demonstrate that HFPV-1 and *H. volcanii* are a great model system to study virus–host interactions in archaea. HFPV-1 is a pleomorphic virus that causes a chronic infection with continuous release of virus particles, but host and virus coexist without cell lysis or the appearance of resistant cells. Despite an only minor impact of the infection on host growth, we uncovered an extensive remodeling of the transcriptional program of the host (up to 1,049 differentially expressed genes). These changes are highlighted by a down-regulation of two endogenous provirus regions in the host genome, and we show that HFPV-1 infection is strongly influenced by a cross-talk between HFPV-1 and one of the proviruses mediated by a superinfection-like exclusion mechanism. Furthermore, HFPV-1 has a surprisingly wide host range among haloarchaea, and purified virus DNA can cause an infection after transformation into the host, making HFPV-1 a candidate for being developed into a genetic tool for a range of so far inaccessible haloarchaea.

archaea | virus | chronic infection | CRISPR

Significance

Viruses use a wide range of strategies to take over host cells for virus replication. While nonlytic infections are common in eukaryotic systems, little is known about those in prokaryotes. We report a virus causing a persistent infection in the model organism *Haloferax volcanii*. The infection cycle is characterized by high viral titers and stability over multiple generations but low impact on cell viability. The virus has a major impact on cell behavior, mainly by down-regulating host genes in unprecedented quantities, while interacting with a provirus region in the host genome, revealing an undescribed superinfection exclusion mechanism. This unique virus has an unusually wide host range and the potential to be developed into a versatile genetic tool in archaea.

Prokaryotic viruses are the most abundant and genetically diverse biological entities, and they have the potential to modulate microbial abundances, community structure, and the evolutionary trajectory of their hosts (1–3). The impact of viruses on their particular host and within the microbial community is largely determined by their lifestyle. Traditionally, viruses have been mainly considered as killing machines that do not have their own metabolism; instead, they hijack the host metabolic machinery and redirect it toward virus production that subsequently leads to the lysis of the host cell. Some viruses can enter into a lysogenic state, in which they integrate their genome into the host chromosome, are replicated with the chromosome, and, subsequently, are inherited by the cell progeny without host cell lysis. Nevertheless, environmental changes can induce the excision of the provirus to enter into a lytic life cycle.

When using traditional methods for virus isolation, such as plaque assay, that are based on host cell lysis or severe growth delay, viruses that cause chronic infections are often overlooked. During a so-called "productive chronic infection," virus particles are consistently released without lysis of the host cell, typically, by budding through the host cell membrane and incorporating host lipids (4–7). Chronic infections are common among eukaryotic viruses. However, only a few chronic viruses have been described for prokaryotes (8, 9), and the majority of them infect archaea (10–14).

Archaea share cellular characteristics with both bacteria and eukaryotes while exhibiting some unique characteristics, and so do their viruses (15). Accordingly, archaea and bacteria share many physiological and morphological traits, and a large proportion of the lytic viruses infecting archaea isolated to date have the classic head-tailed morphology of bacteriophages (16, 17). Meanwhile, recent characterization of novel archaeal clades has tightened the phylogenetic relationships between archaea and eukaryotes, and revealed that several intracellular processes, such as intracellular trafficking and vesicle formation and export, are carried out by homologous groups of proteins in archaea and eukaryotes (18, 19). These processes are crucial for the success of nonlytic viral infections, and the isolation and characterization of novel archaeal viruses has shown that virion morphogenesis and egress mechanisms in archaea largely resemble the

Authors contributions: T.A.-S. and S.E. conceived and led the study and performed the primary writing of the manuscript; A.N. and U.G. generated the provirus mutant strain; and T.A.-S., A.N., U.G., and S.E. participated in the analysis and interpretation of the data and contributed to the writing of the manuscript.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.P. is a guest editor invited by the Editorial Board.

[1]Present address: Specific Diagnostics, San Jose, CA 951343.

[2]To whom correspondence may be addressed. Email: serdmann@mpi-bremen.de.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2205037119/-/DCSupplemental.

Published August 22, 2022.

budding of enveloped viruses in eukaryotes (20–23). Therefore, the study of chronic virus in archaea has the potential to provide insights into the molecular evolution of virus–host interactions in eukaryotes.

Given the relatively small number of characterized chronic infections in prokaryotes, virus–host interaction in archaea and the impact on microbial community dynamics and evolution of this particular lifestyle remains largely unexplored. Moreover, recent studies indicate that chronic cycles are more widespread in nature than previously thought (6, 9). The development of a stable yet controllable chronic infection in archaea will, therefore, significantly advance our understanding of the molecular basis and ecological impact of chronic viruses.

In most environments, viruses outnumber their host by a factor of 10, which means that not only must viruses evolve to overcome host defense mechanisms but also, due to a limited number of host cells, there is the need to outcompete other related viruses. Cells harbor a wide spectrum of mobile genetic elements from viral and nonviral origins such as plasmids, transposons, casposons, or proviruses (24). Particularly, proviruses are commonly found in archaeal and bacterial genomes, and ~60% of the genomes contain, at least, intact functional forms or, defective in varying degrees, virus-like sequences (25, 26). At some point in their evolution, some of these proviruses have been immobilized by their hosts, and, through selection, the cells sometimes conserve viral sequences encoding for advantageous traits (27). For example, viral genes encoding for superinfection exclusion mechanisms, whose aim is to block further infections by competing viruses (28), can result in the emergence of a new defense system against viral infection (29, 30). Superinfection exclusion has been largely studied for bacterial viruses, and the nature and molecular basis of these mechanisms comprise an extremely wide variety of targets, such as blocking the adsorption of viruses through modifications to the cell envelope, inhibiting the genome replication of a competing virus, or inhibiting the activity of lysozymes when lytic cycles are underway (30–32). However, the possibility of superinfection has, so far, only been observed for an isolated archaeal rod-shaped virus from the family *Rudiviridae*, although the mechanism remains unknown (33). Superinfection exclusion has also been proposed to occur in icosahedral, nonlytic viruses of the family *Portogloboviridae* through virus-encoded CRISPR arrays (34).

Here we characterize a virus causing a chronic productive life cycle, isolated for the model organisms *Haloferax volcanii* DS2 (and DS70 and its descendants). *H. volcanii* is one of the very few archaea for which genetic tools are available, and it is commonly used to study archaeal cell biology (35–37). To the best of our knowledge, there is currently no other virus available that infects this species (38). This new model virus–host system allows analysis of the influence of a chronic life cycle on basic cellular processes such as genome replication and cell division (37, 39) in a prokaryotic organism in detail. First insights reveal a surprisingly high virus release and virus-to-host ratio (VHR), while having little impact on the overall fitness of the host. Moreover, upon infection, Haloferax volcanii pleomorphic virus 1 (HFPV-1) induces specific changes in the transcriptional program of *H. volcanii*, particularly downregulating transcription of genes in endogenous provirus regions of the host. The latter suggests that virus–virus interactions play a key role in the infective cycle of HFPV-1, which is supported by experiments showing that the presence or absence of one of these proviruses heavily modulates the outcome of infection.

Furthermore, to date, HFPV-1 is the only pleolipovirus that is capable of infecting a wide range of hosts across diverse haloarchaeal clades, thereby exhibiting promising traits for development as a genetic tool for diverse archaea, including those that cannot be genetically manipulated yet.

Experimental Procedures

Sampling Sites and Culture Conditions. Sediment samples were collected from Lake Tyrrell (35°20′38″S, 142°50′00″E) on December 23, 2018 (Department of Environment, Land, Water and Planning, Victoria, Australia, permit number 10008945) and kept at 4 °C until further use. For the isolation of viruses, enrichment cultures were established. Approximately 500 mg of sediment were completely dissolved in 100 mL of rich medium (Hv-YPC media) (40), supplemented with 1 mL of trace element solution and 3 mL of vitamin solution [21], referred to as Hv-YPC+. Dissolved samples were incubated aerobically at 45 °C with constant agitation (120 rpm) for 14 d. Ampicillin and kanamycin were used to prevent bacterial growth (final concentrations of 0.1 and 0.03 mg/mL, respectively).

Isolation and Purification of Viruses. Enrichment cultures were centrifuged at 4,500 × *g* for 45 min to pellet the cells. The supernatant was recovered, and viruses were subsequently precipitated with polyethylene glycol (PEG) 6000 (10% wt/vol final concentration) and incubation at 4 °C overnight. Then, viral preparations were collected by centrifugation (13,000 × *g*, 45 min, 4 °C). Pellets were resuspended in 18% buffered salt water (BSW) (180 g of NaCl, 25 g of MgCl2, 29 g of $MgSO_4$, and 5.8 KCl per L), sterile filtered (pore size 0.2 µm), and used for plaque assay and, after another filtration (pore size 0.2 µm), in liquid culture infection assays. Viral preparations were screened for viruses infecting *H. volcanii* DS2 (German Collection of Microorganisms and Cell Cultures GmbH). Briefly, 5×10^9 cells of *H. volcanii* were grown in Hv-YPC+ media at 45 °C, harvested from an exponential phase growing culture (optical density at 600 nm of one), mixed with 20 µL of viral suspension from each enrichment, and incubated for 2 h at room temperature to allow viral adsorption. Then, treated cells were either 1) used for plaque assay by mixing with 10 mL of top layer agar (base agar 10g/L and top layer 4 gr/L), plating, and incubating for a week at 28 °C or 45 °C or 2) used to inoculate a 100-mL liquid culture that was monitored by measuring optical density changes. No plaque formation was observed, but one liquid culture with a slight growth retardation was chosen for further analysis. Virus particles were isolated from culture supernatants as described above for enrichment cultures. For infection assays, the virus solution was filtered two times (pore size 0.2 µm), and cell contamination was excluded by inoculating the virus solution into media and incubation for 2 wk to exclude growth. For genomic DNA extraction, the virus solution was treated with 20 µL of DNase I and 10 µL of RNase to reduce host genomic DNA contamination, and was further purified on a CsCl (0.45 g CsCl/mL 15% BSW) density gradient (38,000 rpm for 22 h, 4 °C, SW 41 Ti Swinging-Bucket rotor, Beckman & Coulter). Bands containing virus particles were extracted with a syringe, diluted in one volume of 18% BSW, and reprecipitated with PEG (final concentration 10%, 4 °C, overnight), and the resulting pellet was washed two times with 18% BSW before DNA extraction. DNA was extracted using genomic DNA extraction kit (Bioline) according to the manufacturer's instructions. Virus particles were observed by transmission electron microscopy (TEM). Virus-containing solution was adsorbed for 5 min to carbon-coated copper grids (company) and stained for 1 min with 2% uranyl acetate (wt/vol in water). Electron micrographs were generated

using an FEI Tecnai F20 S-TWIN or JEM2100 Plus at 200-kV acceleration voltage.

Analytic Digest of Genomic DNA, Genome Sequencing, and Protein Content Analysis. For analytic digestion of genomic DNA, 1 μg of DNA was digested with restriction enzyme *Hind*III (37 °C, 1 h). To determine methylation patterns, genomic DNA was digested with *Dpn*I and *Bsp*143I (37 °C, 1 h). To search for single-stranded interruptions, the genome HFPV-1 DNA was digested with *Bal*31 and Exonuclease III (30 °C for 10 min). All products were separated on 1% agarose gels and visualized by staining with SYBR Safe (Invitrogen).

For DNA sequencing, libraries were prepared from purified DNA using FS DNA Library (NEBNext Ultra), and samples were sequenced with Illumina HiSeq2500 (Max Planck-Genome-Centre), with paired end 2 × 250 bp (paired end) read length. The sequenced reads were quality trimmed using the software Cutadapt (41), with a minimum quality of 30 and a minimum length of 50 (-q 30, -m 50). Subsequent assembly of high-quality reads was performed with assembler SPAdes v3.13.1 from ref. 42. Contigs coverage was calculated using BBmap v38.06 (43) with a minimum identity of 99% (minid = 0.99). Protein prediction was performed using Prodigal (44). Functional annotation of predicted proteins was done with the software package HH-suite3 (45) against the PDB70 database (release pdb70 200108) and the Big Fantastic database (BFD) (46). The genome is available in the National Center for Biotechnology Information (NCBI) database under accession number OM621814. Protein content analyses of in-solution sample preparation were performed using mass spectrometry (MS) as described in ref. 47. For gel-separated samples, the protein profiles were analyzed by using a Tris-Glycine sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE) gel electrophoresis (12% acrylamide in resolving gel). Gels were stained with fast Coomassie staining (50% ethanol, 10% acetic acid, 0.5% wt/vol Coomassie Blue). The protein content of selected regions or the entire gel lane line separated into four to six slices were in-gel reduced with Dithiothreitol (DTT), alkylated with iodoacetic acid (JAA), and digested overnight with trypsin (Promega). Resulting peptide mixtures were extracted twice by exchange of 5% formic acid (FA) and acetonitrile and dried down. Peptides were resuspended in 25 μL of 5% FA, and a 5-μL aliquot was analyzed by liquid chromatography–MS/mass spectroscopy (LC-MS/MS) on a nano-ultra-high performance LC (nano-UPLC) system interfaced to a Q Exactive HF Orbitrap or an LTQ Orbitrap Velos mass spectrometer (Thermo Fischer Scientific). The nano-UPLC was equipped with an Acclaim PepMap100 C18 75-μm internal diameter (i.d.) × 20-mm trap column and a 75 μm × 15 cm analytical column (3 μm/100 A, Thermo Fisher Scientific). Peptides were separated using an 80- or 180-min linear gradient for gel bands and unfractionated samples, respectively; solvent A was 0.1% aqueous FA, and solvent B was 0.1% FA in neat acetonitrile. Spectra were acquired using the Data-dependent acquisition (DDA) method and Top 20 approach; lock mass set on m/z = 445.1200. Three blank runs were performed after each sample analysis to minimize carryover.

All MS/MS samples were analyzed using Mascot (Matrix Science; version 2.2.04). Mascot was set up to search against the *H. volcanii* DS2 and HFPV-1 predicted proteins, assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.025 Da and a parent ion tolerance of 5.0 PPM; label: $^{13}C_6{}^{15}N_2$ of lysine; label: $^{13}C_6{}^{15}N_4$ of arginine, oxidation of methionine, acetyl of the n-terminus, carbamidomethyl of cysteine, and propionamide of cysteine were specified in Mascot as variable modifications. Scaffold (version Scaffold_4.10.0, Proteome Software Inc.) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 91.0% probability by the Peptide Prophet algorithm (48, 49) with Scaffold delta-mass correction. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Only proteins with more than two unique peptides were considered in downstream analyses as quality control, while relative abundances were calculated by averaging the c values of biological triplicates.

Virus Infectivity and Kinetics. To study the life cycle, cultures of *H. volcanii* were synchronized using an adaptation of the "stationary phase method" (50). For this, a single colony was picked in liquid culture and grown in Hv-YPC+ media up to an optical density at 600 nm (OD_{600}) ≈ 1. Then a 20-fold dilution step in fresh media was performed (final OD_{600} = 0.05), and cultures were then regrown up to an OD_{600} of one. Iterative dilution and growth of the culture were repeated three times before considering a culture synchronized. For infection with HFPV-1, 1×10^9 cells ($OD_{600} \approx 1$) from synchronized cultures were collected by centrifugation, resuspended in 500 μL of fresh media, and infected with HFPV-1 virus with a multiplicity of infection (MOI) of 10. After incubation (2 h, room temperature), cells were transferred into liquid cultures, and growth was monitored by optical density (OD_{600}) every 6 h. Viral titer was quantified for free and intracellular virus by qPCR. Briefly, samples of 1 mL in biological replicates were collected and pelleted(11,000 × *g*, 10 min, room temperature). Supernatants were recovered, and PEG 6000 was added to a 10% wt/vol final concentration and stored at 4 °C, and virus particles were precipitated as described above. Cell pellets were washed two times with 1 mL of fresh Hv-YPC+ media, flash frozen with liquid N_2, and stored at −20 °C upon DNA or RNA extraction. DNA was extracted using a genomic DNA extraction kit (Bioline) according to the manufacturer's instructions. Long-term cultures were established by diluting infected cultures every 7 d. Infection was assessed by PCR with specific primers targeting the genome sequence of HFPV-1 (TyrVUF 5′-acgaacgagaacaccgacc-3′ and TyrVUR 5′-tgatgacgaatccaacgagcag-3′). Each PCR was performed with the Q5 High-Fidelity DNA Polymerase (New England Biolabs) and contained 0.02 U/μL polymerase, primer concentration of 0.1 μM for both forward and reverse, 1× of Q5 Reaction Buffer, and 1× Q5 High GC Enhancer. The following program was used: 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 68 °C for annealing, and 30 s at 72 °C for elongation. Results were visualized in 1% agarose gels stained with SYBR Safe (Invitrogen).

Quantification of *H. volcanii* and HFPV-1 genome copy number (gcn) were carried out using a CFX96 Touch Real-Time PCR (Bio-Rad Laboratories, Inc.) and the software CFX Manager Software. Host DNA polymerase II small subunit gene (HVO_0003) was amplified with a specific primer set (FpolB 5′-cccgaatcaggacgaagaac-3′ and RpolB 5′-atttgaggtgctcggagaac-3′). Each reaction (10 μL) contained 1× SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and 300 nM of each primer. For viral quantification, a primer set targeting the internal structural protein 3 was designed (FVP3 5′-ttgcgtacgcggtatctgtc-3′ and RVP3 5′-agcttctccgcatcgtcttt-3′). The following amplification thermal cycling program was used for both primer sets: 5 min at 95 °C, followed by 40 cycles of 30 s at 95 °C and 30 s at 68 °C, with readings taken between each cycle. Efficiencies of the assays

were 95 to 100%, with R^2 values ≥ 0.99 for all assays. The specificity of the qPCR was confirmed by unique signals in melting curves and gel electrophoresis of PCR products.

Transcriptomic Analyses. RNA extraction of frozen cell pellets in three biological replicates was performed with the Zymo Direct-zol RNA miniprep Kit (R2051). RNA concentration and integrity were assessed using nanodrop DS-11 Spectrophotometer (DeNovix) according to the manufacturer instructions. Ribosomal RNA (rRNA) was depleted prior to sequencing using the rRNA depletion Kit riboPOOL, for *H. volcanii*, siTOOLs Biotech. Libraries were prepared with library kit NEBNext Ultra II RNA Library Prep Kit for Illumina, and sequencing was performed on an Illumina HiSeq3000 sequencer, using sequencing kit, following a 1 × 150 run. Sequencing depth of the samples ranged from 4.3 million to 7.2 million reads (mean = 6.5 million, SD = 0.7 million reads). Raw reads were quality trimmed with Cutadapt (41), with a minimum quality of 30 and a minimum length of 50 (-q 30, -m 30), and all reads containing "N" bases were removed. Filtered reads were mapped to *H. volcanii* DS2 strain and HFPV-1 genome using BBmap v38.06 with an identity threshold of 99% (minid = 0.99). Differential expression analyses were performed with R package DESeq2 (51) and plotted using ggplots2 (52). Briefly, raw counts were normalized by sequencing depth and geometric mean (51), and differential expression was calculated comparing the differences in read counts between infected and uninfected cultures by each time point (with three biological replicates for the infected and two biological replicates for the controls). Genes with P values < 0.01, false discovery rates < 0.05 (p adjusted < 0.05), and a fold change of at least two times (log2FC ≥ 1 or ≤ −1) were considered differentially expressed (DE). Raw data were submitted to European Nucleotide Archive (ENA) under Project Number PRJEB50750.

Global Distribution and Alternative Hosts of HFPV-1. The viral genome was searched against the Integrated Microbial Genomes/virus (IMG/VR) database (53) using Basic Local Alignment Search Tool (BLASTn) (54), e value < 10^{-5}, to detect uncultivated relatives from metagenomes or previous isolates. Original sequencing read files (metagenome or host isolate sequencing projects) for each target were then retrieved from relevant BLAST hits (identity ≥ 70% and query coverage ≥ 50%.). HFPV-1 presence in these samples was determined by read mapping using BBmap (minid = 0.95), and HFPV-1 was considered present when the genome had at least ≥75% of genome length with ≥1× coverage (55). Furthermore, to establish the geographical distribution of the virus, an additional BLAST against the Virus CRISPR Spacer Database from IMG/VR was performed. This database contains short spacer sequences retrieved from CRISPR-Cas arrays from isolated and metagenomic genomes deposited in the databases. Only BLAST hits with an e value < 10^{-5} and identity ≥ 90% were considered as true positives.

Host Range Assessment of HFPV-1. Eight different strains of haloarchaea were tested to determine the host range of HFPV-1 (*SI Appendix*, Table S1). Exponential phase cultures were infected with HFPV-1 with an MOI of ∼10, following the same procedure aforementioned, and incubated at 28 °C with constant agitation (120 rpm). Growth kinetics were monitored through optical density, while cells were collected by centrifugation (10,000 × *g*, 10 min), washed two times with fresh media (to avoid false positives arising from free virus in the supernatant), and screened for HFPV-1 infection by PCR (see *Virus Infectivity and Kinetics*). PCR-positive strains were selected, viral particles were isolated from culture supernatants by PEG precipitation (as described above), and HFPV-1 particle production was assessed by qPCR (see *Virus Infectivity and Kinetics*). In order to confirm infectivity, viral particles isolated from the supernatant of alternative hosts were used to infect virus-free *H. volcanii*. Infection was confirmed by PCR as previously described.

Transformation of *H. volcanii* with HFPV-1 Genomic DNA. Transformation of *H. volcanii* was carried out by using PEG 600 as described previously in ref. 56. Briefly, cultures of *H. volcanii* were grown to an OD_{600} of ∼0.8 and then harvested by centrifugation (30 min 4,500 × *g*). The pellet was gently washed one time in 2 mL of spheroplasting solution (1 M NaCl, 27 mM KCl, 50 mM Tris·HCl, sucrose 15%) and then gently resuspended in 1 mL of spheroplasting solution. The solution was chelated with 0.5 M (ethylenedinitrilo)tetraacetic acid (pH 8). Then four different concentrations of HFPV-1 genomic DNA were used for transformation (100 pg, 1 ng, 10 ng, and 100 ng DNA, respectively). Additionally, two genomic DNA samples of HFPV-1 that were treated with DNase and RNase (DNase I and RNase H, respectively; New England Biolabs) were used as controls. After addition of DNA, 250 µL of 60% PEG_{600} was added and mixed gently, and incubated at room temperature for 1 h. Then, cells were washed two times and transferred into a regeneration solution (Hv-YPC+ media with 15% sucrose) for 3 h at 28 °C before transfer into liquid culture. Successful infection was assessed by PCR and isolation of virus particles from culture supernatants as described above (host range assessment).

Knockout Mutant Strains Generation and Determination of HFPV-1 Infection. Transformation and construction of deletion mutants in *H. volcanii* H133 (Δ*pyrE2* Δ*trpA*, Δ*leuB*, and Δ*hdrB*) based on selection with uracil were performed as described previously (57). Plasmid construction was done by classical restriction enzyme-based molecular cloning. Inserts were amplified from wild-type (DS70) genomic DNA (using genomic DNA extraction kit: Bioline, according to the manufacturer's instructions) via PCR using Phusion polymerase (NEB). All restriction endonucleases used were purchased from NEB. Restriction enzyme digestions and PCR were performed according to the manufacturer's protocols. Provirus Halfvol 3 (ΔHVO_1422 to HVO_1434) was generated by first generating a deletion mutant of HVO_1434. Two fragments were generated using the following primer and restriction enzymes: A1434up5 (*Not*I) 5′-aacgcggccgcagtcgttgatggcgttgccc-‘3, A1434up3 (*Xba*I) 5′-aactctagactccgagagaccgcacgcg-‘3 and A1434down5 (*Eco*RI) 5′-aacgaattcaggaacctaccctgcctggt-‘3, A1434down3 (*Hind*III) 5′-aacaagcttagcgaactcgcgcatcagag-‘3. Fragments were cloned into pTA131 (57) and transformed into *H. volcanii* after isolation from *dam*$^-$/*dcm*$^-$ competent *E. coli* cells (NEB). The remaining part of provirus Halfvol 3 was deleted in ΔHVO_1434 after cloning the fragment generated before with A1434up5 and A1434up3 and a fragment generated using primers A1422down5 (*Eco*RI) 5′-aacgaattccgacatatccgaacgcgagag-‘3 and A1422down3 (*Hind*III) 5′-aacaagcttgtgttccgatggccgtgac-‘3 into pTA131. The construct was demethylated by isolation from *dam*$^-$/*dcm*$^-$ competent *E. coli* and transformed into ΔHVO_1434, and, after selection with uracil, pop-out was performed as described. The resulting ΔHVO_1422 to HVO_1434 deletion strain was confirmed by PCR (Halfvol3F 5′-ccgatgtactgacccttgac-‘3, Halfvol3R 5′-agacatacccgtactgttc-‘3) and Sanger sequencing of the resulting PCR product. Virus infectivity on the provirus mutant strain was determined as described above for wild-type *H. volcanii* (*Virus Infectivity and Kinetics*). Experiments were performed in three biological replicates, and the

parental strain was used as a wild-type control. Hv-YPC+ was supplemented with uracil (50 μg/mL), tryptophan (50 μg/mL), leucine (50 μg/mL), and hypoxanthin (40 μg/mL) to support growth.

Results and Discussion

Isolation of a Virus Infecting the Model Haloarchaeon *H. volcanii*. The halophilic archaeon *H. volcanii* is widely used as a model organism for the study of archaeal cell biology and biochemistry and is one of few archaea with an extensive genetic toolset available (36, 58). A number of viruses infecting halophilic archaea have been isolated (17, 59); however, no virus was available for a model haloarchaeon that would allow a detailed study of virus–host interactions in halophilic archaea. To isolate a virus specifically targeting *H. volcanii*, enrichment cultures in media specified for *H. volcanii* at optimal growth temperature of *H. volcanii* were established, using samples taken at hypersaline Lake Tyrrell, Australia (60). Virus-containing supernatants were generated from enrichment cultures and tested on *H. volcanii*. Growth retardation was observed in one of the treated *H. volcanii* cultures, and DNA containing virus-like particles (VLPs) were retrieved from the supernatant. TEM of the samples revealed the presence of nonsymmetrical pleomorphic particles, with sizes ranging from 50 nm to 80 nm (Fig. 1*A*), and represents the only virus isolated for this species; thus we named it Haloferax volcanii pleomorphic virus 1 (HFPV-1).

Genomic Characteristics of HFPV-1 and Taxonomic Classification. Analytic digests of HFPV-1 genomic DNA revealed a double-stranded DNA genome with single-stranded interruptions (*SI Appendix*, Fig. S1). The presence of a conserved DNA motif that precedes single-stranded discontinuities in the genome of Halorubrum pleomorphic virus 3 (HRPV-3) (61) was in silico identified in HPFV-1 genome (GCCCA motif $n = 3$); however, experimental evidence is further needed to confirm that it is a common trait in HFPV-1 as well. Meanwhile, no dam methylation was detected on the genome of HFPV-1 (Fig. 1*B*). Illumina sequencing and genome assembly yielded a contig of 7,924 base pairs (bp) with overlapping ends, which circularizes to a single molecule of 7,869 bp and encodes for 11 open reading frames (ORFs) (*SI Appendix*, Fig. S2). The average GC content of HFPV-1 is 59.3%, which is significantly lower than the GC content of the main chromosome of *H. volcanii* (66.6%), and more similar to *H. volcanii* plasmids pHV1, pHV2, and pHV4 (56%, 55.5%, and 61.7% respectively) (62).

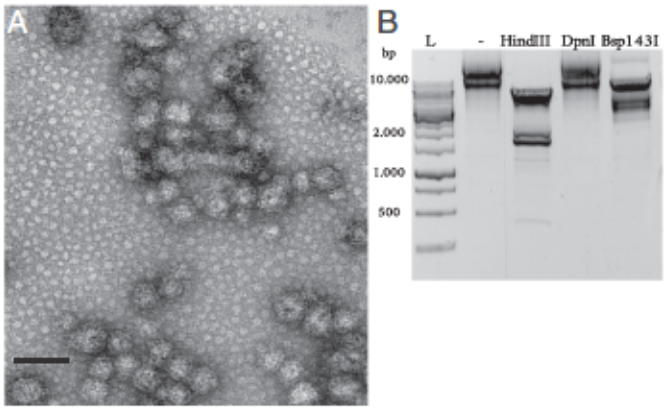


Fig. 1. (*A*) Transmission electron micrographs of HFPV-1. Viral concentrates were sterile filtered (0.2 μm) and purified through CsCl density gradients. Particles were negatively stained with uranyl acetate and imaged at 200 kV on an FEI Tecnai TF20. (Scale bar, 100 nm.) (*B*) Analytic restriction digest of HFPV-1 genomic DNA with different restriction enzymes. Molecular weight size marker (L) is shown (gene ruler 1 kb; DNA ladder, Thermo Fisher Scientific). Untreated (-) and restriction digestion with enzymes *Hind*III, *Dpn*I, and *Bsp*143I. DNA was separated on 1% agarose gels and stained with SYBR Safe (Invitrogen) at a final concentration of 1×. Data represent one sample of three biological replicates. The restriction digest patterns were observed in four repetitions.

The genome of HFPV-1 displayed no significant similarity at the nucleotide level with any previous viral isolate, while low identity values at the protein level were observed when compared with members of the *Pleolipoviridae* family (Fig. 2). Further comparison against multiple viral databases (see *Experimental Procedures* for details) revealed the presence of two putative structural proteins (ORF2 and ORF4) homologous to the spike proteins of *Halogeometricum sp.* pleomorphic virus 1 HGPV-1 (63), a member of the *Betapleolipovirus* genus, although the similarity was rather low (identity = 45%). ORF3 has little similarity to any known protein and only has a homolog in HGPV-1 among related viruses, but its function remains unknown. Three other proteins (ORF5, ORF6, and ORF7) presented a significant homology to the conserved core cluster of proteins within the *Pleolipoviridae* family (Fig. 2). ORF7 is proposed to have NTPase activity, and both ORF6 and ORF7 are predicted to have transmembrane domains, but their function and role in the infective cycle of the virus has not been characterized. Further assessment of the hypothetical proteins encoded by HFPV-1 using hidden Markov models showed that two predicted proteins, ORF1 and ORF9, were similar to transcriptional regulators (probabilities of $P = 99.44$ and 84.96, respectively).

No replication protein could be identified in the genome of HFPV-1. However, it possesses a predicted DNA binding protein (ORF10, $P = 91.77$) at the position where related pleolipoviruses encode their predicted replication protein (Fig. 2). This is of particular interest because, thus far, the taxonomic affiliation of the members of the *Pleolipoviridae* normally correlated with their proposed replication mechanisms. While the alphapleolipoviruses are thought to replicate via a rolling circle replication endonuclease (61), and the gammapleolipoviruses possess a protein-primed family B DNA polymerase (pPolB polymerase) (59), the betapleolipoviruses, whose members are the ones with highest homology to HFPV-1, present a conserved homolog of an uncharacterized replication protein (Rep protein; Fig. 2) (64). The lack of the conserved replication protein present in all related viruses and the lower identity values at the protein level suggest that HFPV-1 is a rather divergent member of the *Betapleolipovirus* genus.

Protein Composition of HFPV-1 Particles Differs from Closely Related Viruses, and Particles Are Stable at a Wide Range of Salt Concentrations. MS of virus particles confirmed the presence of ORF4, the predicted spike protein, as the most abundant structural protein present in the virus particles (average peptide count = 1,700; Dataset S1). Interestingly, no signal was detected for the predicted internal membrane protein ORF2, neither by individual SDS/PAGE band sequencing (*SI Appendix*, Fig. S3) nor by MS of whole purified viral particles, despite presenting trypsin digestion sites. This suggests that, unlike for its closest relative HGPV-1, this protein is not within virions, or another method such as N-terminal sequencing would be required to detect it. Additionally, four other proteins, ORF3, ORF5, ORF6, and ORF7, were detected in virions (Dataset S1). While no function could be predicted for ORF3 and ORF5, we detected transmembrane domains in ORF6 and ORF7. However, whether these proteins are involved in virion assembly or the infection process remains to

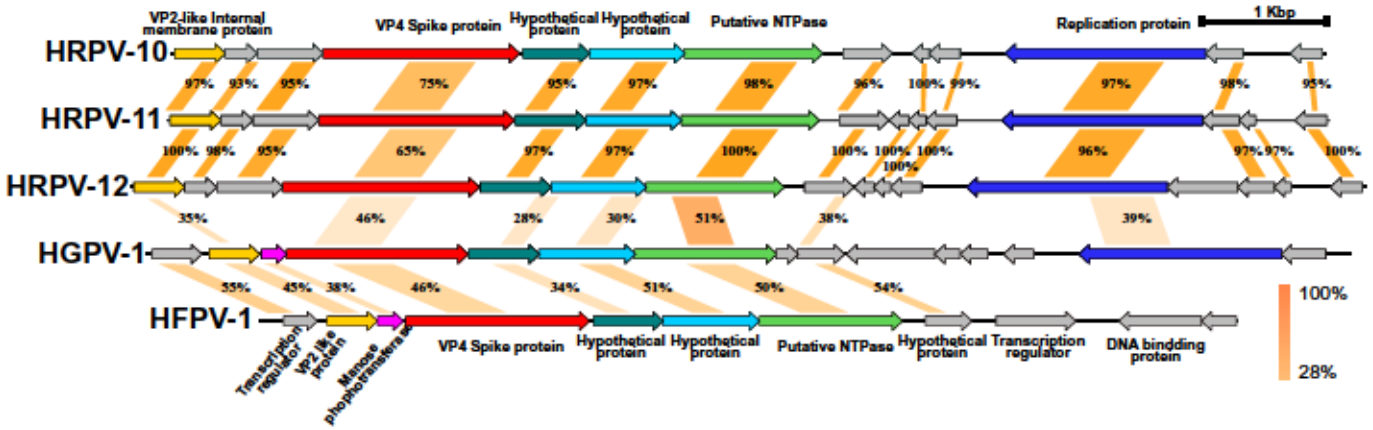


Fig. 2. Genome comparison of HFPV-1 with closest related pleolipoviruses. Homolog genes are indicated with the same colors. Vertical colored bars represent the percentage of identity between protein pairs. HRPV-10, HRPV-11, and HRPV-12: Halorubrum pleomorphic viruses; HGPV-1, Halogeometricum pleomorphic virus 1.

be elucidated. Noteworthy is that a number of host-encoded proteins were also detected in the viral preparations in significant amounts, such as the S-layer protein and several transporters (Dataset S1). However, the detection of these proteins in the preparations can be attributed to membrane vesicles present in the sample. These vesicles are naturally produced by *H. volcanii* independently of viral infection, and they copurify with HFPV-1, due to their similar physical properties.

Transformation of uninfected *H. volcanii* cells with purified HFPV-1 DNA generates a successful infection, indistinguishable from the infection with virus particles (*SI Appendix*, Fig. S4); we therefore conclude that none of the viral proteins is crucial for initiation of the virus life cycle once the virus DNA has entered the host cell.

Sensitivity to different conditions was assessed through a spot test (65), revealing that viral particles remain stable in 18% BSW at 4 °C for over a year. Infectivity was maintained upon freezing with liquid nitrogen after 2 mo; however, samples stored at −20 °C lost their infectivity within a day. Assessment of thermal stability through spot tests revealed that particles remain infective after 1-h incubation at 50 °C but are inactivated at 60 °C and above. Particles also exhibited remarkable versatility under osmotic stress, remaining infective in salt concentrations ranging from 5 to 25%. Furthermore, the ability to generate an infection by transformation of purified virus DNA provides an advantage for long-term storage and highlights the potential of HFPV-1 to be developed into a genetic tool.

HFPV-1 Causes Only a Minor Impact on Host Growth Despite High gcn and High Viral Titers. Infection of *H. volcanii* cells in an exponential phase with HFPV-1 with an MOI of 10 caused only a slight growth retardation of the host in liquid culture (Fig. 3). No decrease in the turbidity of the media was observed, well into stationary phase, supporting the nonlytic nature of HFPV-1 infection. Long-term experiments revealed that the virus and the host form a stable relationship in liquid cultures, where viral infection and release is detectable by PCR for at least 3 mo (weekly dilution), without the emergence of resistant cells nor extinction (complete lysis) of the host (*SI Appendix*, Fig. S5).

Consequently, during plaque assays with purified viral particles, no zones of cell lysis were observed, regardless of temperature (28 °C to 45 °C) and salt concentration (15 to 22%). Therefore, the viral titer was determined by measuring viral gcn in the supernatant of liquid cultures by qPCR. Virus particles could be first detected in culture supernatants at 24 h post infection (h p.i.). Subsequently, an exponential increase in the concentration of free virus in the supernatant was observed, reaching up to 1×10^{10} gcn per mL^{-1} 84 h p.i. (Fig. 3). Assessment of the intracellular viral titer, and subsequent calculation of the VHR, unveiled a high number of viral genomes within each cell, with VHR values up to $\sim 10^3$ at 72 h p.i. (average VHR = 977.5; *SI Appendix*, Fig. S6). However, *H. volcanii* is a polyploid organism, and can have up to 20 copies of the genome during exponential phase (66); therefore, the total number of viral genomes within each cell could be severely underestimated.

Transcriptional Profile of the Viral Genome during Infection. So far, most studies that investigated changes in gene expression during viral infection in prokaryotes have focused on the impact of the infection by lytic viruses (67). However, little is known about how the transcriptional program changes upon a long-term chronic infection in a prokaryotic organism. HFPV-1 produces a highly stable and reproducible infection cycle in *H. volcanii*, and is therefore a good model for analyses of infected populations.

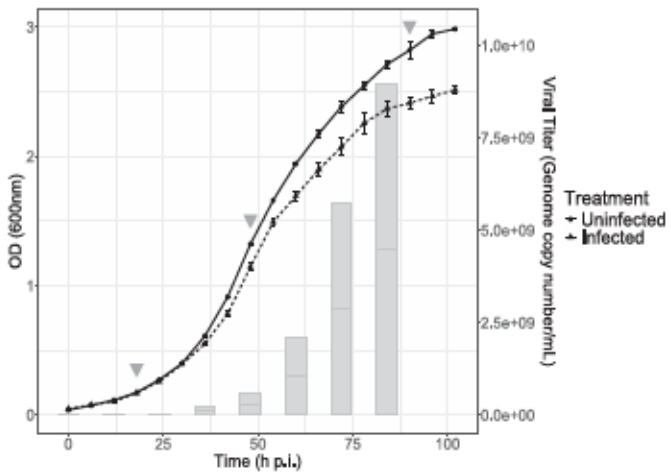


Fig. 3. Virus life cycle: growth curve of uninfected (black circles) and infected (black triangles) *H. volcanii* DS2 cultures. Error bars represent the SD for biological triplicates on both treatments. Gray bars represent the number of free virus particles in the supernatant of liquid cultures, assessed through qPCR using specific probes targeting the HFPV-1 genome. Inverted gray triangles indicate time points chosen for transcriptomic analyses.

In order to shed light on the nonlytic life cycle of HFPV-1, we took samples from synchronized infected and control cultures at three different stages of the viral life cycle (Fig. 3): 1) lag phase, from 0 h to 24 h p.i., with little to no virus release and by growth rate undistinguishable to the uninfected cultures; 2) exponential phase, from 24 h to 72 h p.i., characterized by growth retardation, continuous release of virus particles, and the highest numbers of intracellular virus genomes; and 3) stationary phase, from 72 h to 102 h p.i., characterized by media exhaustion and limited growth, while reaching the highest viral titers in the supernatant. The respective representative samples were taken in three biological replicates at 18, 48, and 90 h p.i. for lag, exponential, and stationary phases, respectively.

Transcriptomic analyses revealed that all predicted viral genes are expressed across the three different growth phases, comprising increasing portions of the total transcriptome as the infection spreads, from an average of ~2% in lag phase to ~7% and 16% during exponential and stationary phases, respectively (Dataset S2). ORF4, which encodes the structural spike protein, along with the ORF10, a DNA binding protein, are the most highly expressed virus genes (average transcripts ≈ 150,000). Interestingly, the ORF10 location in the genome of HFPV-1 matches the genomic location of the proposed replication protein in other *Betapleolipoviruses* but shares no homology with these replication proteins or any other known protein. The position in the genome, together with the high expression levels, and the lack of any other recognizable replication-associated gene, suggests that ORF10 could be a representative of an undescribed family of replication proteins. Additionally, the second predicted structural protein, ORF2, is the third most highly expressed gene. However, the protein product was not detected in purified virus particles, despite the fact that it has four predicted transmembrane domains, and its role in particle formation and infection remains unknown.

Global Pattern of Host Metabolic Takeover in *H. volcanii* Is Phase Dependent. Global analysis of the transcriptional program of the host showed that the three defined phases of infection have remarkably different profiles, with 1,049 genes being DE at least at one of the three phases. Given that the transcriptional profile of each of the growth stages was extremely different, further comparisons to establish the DE genes were always performed between infected and uninfected controls for each time point independently.

The largest number of DE genes was observed during the exponential phase (887 DE genes), with most of the genes being down-regulated (597 genes), meanwhile, the lag phase exhibits the least changes, with only 171 DE genes. During this phase of infection, only relatively few changes in the transcriptomic profile of *H. volcanii* were observed, with 159 genes being up-regulated and only 12 genes being down-regulated (Dataset S2). The two top-ranked overexpressed genes are proteins that encode a DoxX domain (HVO_0763 and HVO_A0548), and, while their specific function in *H. volcanii* remains unknown, characterized members of this protein family are associated with oxidoreductase complexes, specifically linking radical detoxification with thiol homeostasis in *Mycobacterium tuberculosis* (68). Notably, a number of universal stress proteins containing the UspA domain are consistently overexpressed. These small cytoplasmic proteins are generally overexpressed when the cell is exposed to stress agents in an unspecific manner (69). In addition, several oxidoreductase genes, electron carriers, and halocyanines (e.g., HVO_2150, HVO_1119), as well as the RpaB2 gene (HVO_0291), were also up-regulated. Interestingly, the single-stranded DNA-binding protein HVO_0291 has been shown to be one of the most highly up-regulated genes in *H. volcanii* upon exposure to H_2O_2 (70), further suggesting that the cells are undergoing oxidative stress during viral infection. Conversely, no significant up-regulation of any of the known specific immune response pathways was identified (71), with only a distant homolog to an RNase protein being up-regulated (HVO_1114), which could potentially target viral transcripts.

Interestingly, two groups of different transcriptional regulators modulating carbon metabolism appear to be up-regulated, with two genes belonging to the IclR family (HVO_B0040 and HVO_A0527), which regulate different carbon–related pathways (72), and one to the DeoR family (HVO_1501), which is commonly a repressor of sugar metabolism (73). The increase in the expression of these general carbon repressor-like regulators could be the reason for the subsequent growth retardation and may help hijack resources for virus production.

All together, the data largely suggest that, while the cells are enduring some degree of oxidative stress, which arises from the metabolic imbalance generated by the viral infection, the viral genome manages to remain hardly noticeable, thus avoiding triggering any major host defense mechanism at this stage of infection. The latter could be related to the absence of dam methylation on the viral genome, which allows it to escape primary host defense mechanisms against exogenous DNA.

HFPV-1 Induces a Massive Down-Regulation of the Host Transcriptome during Exponential Phase. The exponential phase presents a radically different landscape compared to the earlier stage of infection. This is highlighted by a massive transcriptional down-regulation of 597 genes, more than doubling the up-regulated ones (290 genes). This general down-regulation is, in itself, unusual when compared to previously reported studies on viral takeover of host metabolism in prokaryotes. Research in lytic virus-induced transcriptional changes carried out in *Sulfolobus* showed a trend of little to no down-regulation (74, 75). An exception is the case of the lytic virus SIRV2, infecting *Sulfolobus islandicus*, which was reported to have between 30% and 50% of the host genome differentially expressed during infection (76). However, the number of down-regulated genes is approximately the same as the number of up-regulated genes, and the magnitude of increase in transcription of the up-regulated genes was much higher than the decrease in the down-regulated genes (76). Intriguingly, during the infection of *Sulfolobus solfataricus* with the nonlytic virus SSV2, only around 5% of the genome was reported to be differentially expressed, with around 80 genes down-regulated (77, 78), which is considerably lower compared to the impact of HFPV-1. Similarly, microarray analyses of the nonlytic filamentous phage M13 that infects *E. coli*, one of the few examples of persistent infections in bacteria, also showed that only a small proportion of the genes are differentially expressed upon infection (79). In conclusion, to our knowledge, HFPV-1 generates one of the largest overall transcriptional down-regulations induced by a virus infecting archaea, and poses intriguing questions on how this relatively simple virus, encoding only 11 genes, is capable of establishing a successful and long-term infection.

In terms of cluster of orthologous genes (COG) functional categories, the most DE genes are found in the amino acids metabolism and transport (E), with 122 DE genes (112 down-regulated and only 10 up-regulated). Within this category, the most significant changes are observed for an ABC-type transport system for branched chain amino acids and dipeptides, which show up to a 34.2- to 48.5-fold decrease (HVO_0899,

HVO_0900, HVO_2801, and HVO_0901). Other categories such as energy production and conversion (C), with 82 DE genes (65 down-regulated and only 17 up-regulated), and inorganic transport and metabolism (P), with 69 DE genes (54 down-regulated and only 15 up-regulated), are strongly down-regulated (Fig. 4). In the case of energy metabolism, several NADH dehydrogenase complex components (e.g., HVO_0980, HVO_0981, and HVO_0982), as well as subunits of the A-type ATP synthase (HVO_0313, HVO_0315, and HVO_0316), are also down-regulated in a range of 3.7- to 4.6-fold change. The latter is likely to result in a disruption of the recycling of reductive power and ATP generation, which translates into less available energy for the cellular metabolism and could explain the observed growth retardation. Interestingly, genes involved in cell division exhibited a more moderate down-regulation, with only one of the essential proteins for cell division in *H. volcanii*, FtsZ2, that primarily participates in the constriction mechanism (37, 80), showing lower levels of expression (3.2-fold change), while other essential proteins like FtsZ1 or SepF are not differentially expressed. This is in contrast to STSV2, a nonlytic virus infecting *S. islandicus*, that leads to a strong down-regulation of cell division genes and a massive increase in cell size and DNA content (81).

Regarding cell motility, HFPV-1 generates an overall down-regulation of the archaellum machinery, with the archaellin A1 and A2 (HVO_1210 and HVO_1211) and the flagella-related proteins ArlCE, ArlD, ArlF, and ArlG (HVO_1213, HVO_1203, HVO_1214, and HVO_1215). These genes are proposed to be homologous to the FlaC/D/E complex, which is involved in transducing intracellular/extracellular signals that influence the activity of the flagellum. The FlaF gene, on the other hand, encodes for a b-sandwich protein that anchors the archaellum in the archaeal cell envelope by binding the S-layer protein, while the FlaG gene function is still unknown (82). However, despite those structural and signaling genes of the archaellum being down-regulated, analyses of electron micrographs of infected cells revealed that they possess an archaellum and are optically indistinguishable from uninfected cells (*SI Appendix*, Fig. S7). However, the actual mobility of the cells needs to be further investigated by motility assays.

The CRISPR-Cas System of *H. volcanii* Is Ineffective against HFPV-1 Infection. The CRISPR-Cas system is one of the most sophisticated and effective defense mechanisms known in archaea and bacteria (83), and it has been shown, previously, that it can be effective against other pleolipoviruses (84). *H. volcanii* encodes a single and fully functional CRISPR-Cas system of type I-B (71); however, it does not possess spacers targeting the HFPV-1 genome, and it does not prevent the development of a successful infection by HFPV-1. Upon infection with HFPV-1, several Cas proteins are down-regulated, with Cas8 (HVO_A0206), which is essential for the interference reaction, and Cas7, that mediates the sequence specificity of the defense system, enabling recognition by base-pairing with the invader DNA (71), being the genes subjected to the highest decrease (5.6-fold change). Meanwhile Cas6, which is essential for CRISPR RNA maturation but otherwise not required for the defense reaction, remains unaffected by the infection. Interestingly, the endonuclease Cas1 (HVO_A0211), which participates in new spacer insertion and acquisition, is down-regulated during exponential phase (2.0-fold change). The latter could explain why the host is not able to defend itself against infection with HFPV-1, given that the CRISPR arrays of *H. volcanii* do not have spacers against the genome of the virus, and the down-regulation of Cas1 could inhibit the process of acquisition of new spacers. Analysis of the transcriptomic data showed that there is no evidence for acquisition of new spacers, as there are no transcripts related to the CRISPR array that contain both viral sequences and repeats from one of the arrays, which supports this hypothesis.

Sustained Viral Production during Stationary Phase Despite Cellular Arrest. The transcriptional changes during the stationary phase followed a similar pattern to the one observed during exponential growth, highlighted by an exacerbated trend toward the down-regulation of host genes (216 down-regulated) (Dataset S2), although no down-regulation of any defense mechanism, including the CRISPR system, was observed. Interestingly, this massive down-regulation is concurrent with the largest proportion of viral mRNA (~16% of total mRNA) and consecutive to the highest values of intracellular viral load (VHR $\approx 10^3$ at 72 h p.i.).

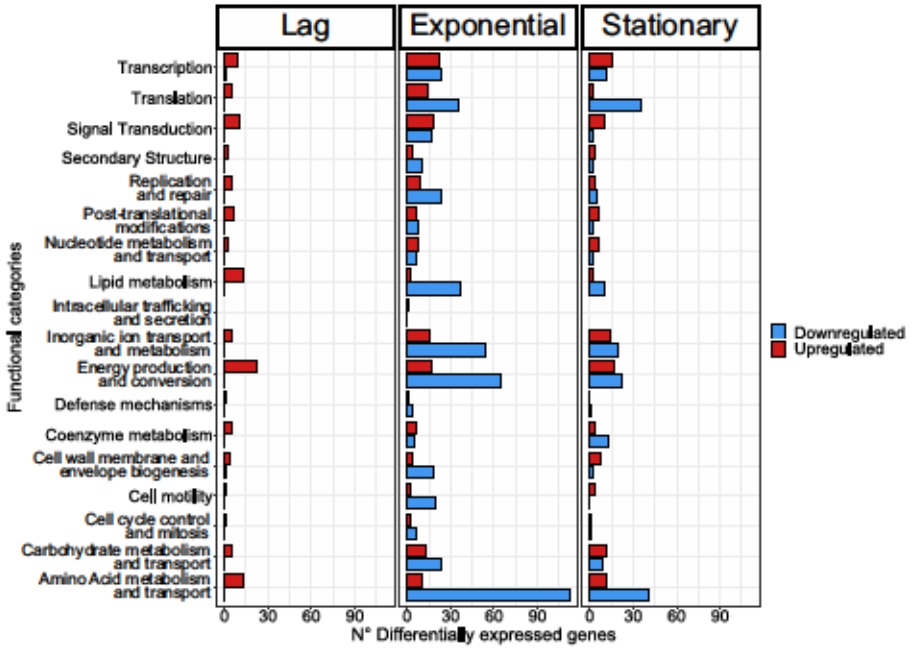


Fig. 4. Functional profile of differentially expressed genes. Bars represent the number of differentially expressed genes assigned to each particular functional category at a given growth stage, that is, lag, exponential, and stationary phases. Functional classification of *H. volcanii* genome was performed using the COG database. The colors of the bars indicate whether genes are up-regulated (red) or down-regulated (blue). Genes that were not assigned to any functional category are not displayed on the plot (for detailed information, see Dataset S2).

The latter suggests that, although cells enter division arrest, they remain highly productive viral factories, with the viral replication not being negatively affected by the lower nutrient availability and the CRISPR system remaining ineffective, despite the increasingly high amount of viral genetic material in the cells.

On the other hand, several genes that showed significant up-regulation, with more than a 30-fold increase in expression, are related to energy production and conversion (i.e., HVO_B0291, HVO_2411, HVO_1083, and HVO_2251) and two universal stress proteins (i.e., HVO_2337 and HVO_0401) (Dataset S2). Also, an acetate-CoA ligase (HVO_1000), which is a key enzyme that catalyzes the recycling of acetate to acetyl-CoA, to compensate starvation during stationary phase in *H. volcanii* (85), is threefold up-regulated. The data suggest that, upon infection with HFPV-1, the infected cells are forced to fulfill at least the energy requirements of the machinery dedicated to virus production.

The shutdown of numerous cellular metabolic functions as cells become quiescent could redirect most of the available resources to viral production, resulting in the highest proportions of viral RNA at 90 h p.i. Subsequently, as these alternative reservoirs are depleted, viral production is also halted, which would explain the decrease in the VHR values after 96 h p.i.

HFPV-1 Specifically Targets Provirus-Like Sequences in the Genome of *H. volcanii*. The genome of *H. volcanii* encodes at least two prophage regions that are likely defective, or temperate viruses from previous infections (Fig. 5). These proviruses are thought to have been maintained by the host because they encode an unknown evolutionary advantage. In this scenario, one of the reasons for the difficulties of isolating viruses infecting *H. volcanii* could be that the defective proviruses encode an unknown virus–virus exclusion mechanism. Evidence for the latter arose during in silico analyses of single-cell sequenced bacterial genomes, where cells with prophage sequences correlated with reduced coinfection events by other prophages (86). In *H. volcanii*, two of the identified proviruses can be related to members of the *Pleolipoviridae* family (87).

Remarkably, both prophage regions related to pleolipoviruses showed down-regulated genes during exponential and stationary phase (Fig. 5), with provirus region Halfvol1 having several genes among the highest ranked down-regulated genes. Particularly, HVO_0269 exhibited a fold change of more than 150× with respect to the uninfected cultures, the highest among any differentially expressed gene. The genes HVO_0268, HVO_0270, and HVO_0271 showed a very similar pattern, with decreases ranging from 40- to 60-fold change. Interestingly, there is sequence similarity, at the nucleotide level, with HFPV-1 in the downstream end of genes (HVO_0272 to HVO_0274, with a 77% similarity) which are also down-regulated (~10-fold change). However, there is no sequence similarity between the viral genome and the highest down-regulated genes (HVO_0269 to HVO_0271) or any of the other intergenic regions. Surprisingly, a single gene within this region exhibited the opposite trend (HVO_0262), being up-regulated in exponential phase at a 2.3-fold change. This suggests that the regulation of this genomic region rather depends on multiple promotors and/or mechanisms, viral, cellular, or both, that modulate, simultaneously, the gene expression.

The second prophage region (HVO_1422 to HVO_1434) seems to be a defective provirus related to alphapleolipoviruses, that also contains several genes with high homology to the

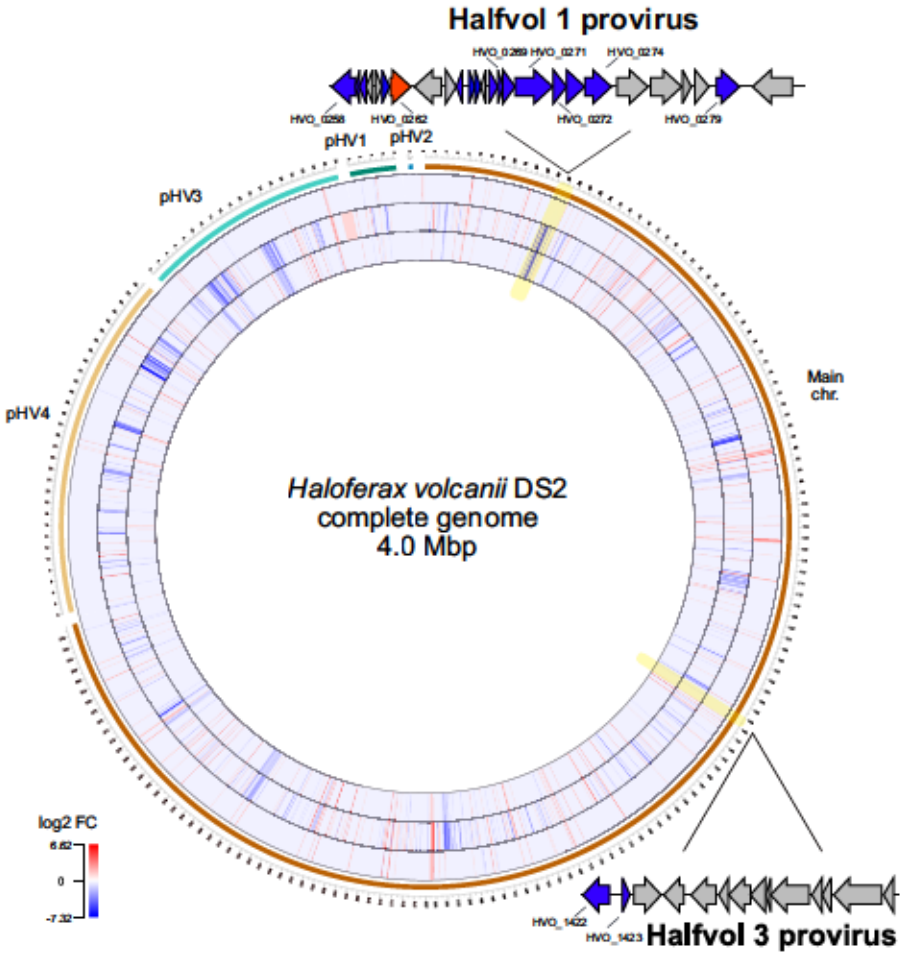


Fig. 5. Specific differential expression of prophage regions in the *H. volcanii* genome: heat map representation of gene expression profile of the host *H. volcanii*. The magnitudes of expression level changes are displayed as the $\log_2$ fold change (log2 FC) for up-regulated genes (red) and down-regulated genes (blue). Each heat map ring corresponds to one of the growth phases studied, that is: lag phase, outer ring; exponential phase, middle ring; and stationary phase, inner ring. Prophage regions phylogenetically related to pleolipoviruses are zoomed in, and genes that are differentially expressed during at least one growth phase are highlighted in blue (for details, see Dataset S2).

pHK2 plasmid from *Haloferax* sp. Aa2.2 (88). Two genes are down-regulated during the exponential and stationary phases, including an integrase (HVO_1422) and a transcriptional regulator (HVO_1423). Interestingly, HVO_1423 shows homology to the phage PhiH1 repressor protein, which can provide *Halobacterium salinarum* cells with immunity to infection by phiH1 virus (89, 90), and, potentially, could be playing a similar role in *H. volcanii*, by acting as a viral defense mechanism.

The third prophage region, which is proposed to be a degenerate provirus related to tailed viruses, exhibits a more patched pattern. Here, an integrase gene and unknown protein (HVO_2259 and HVO_2266) are moderately down-regulated (an approximately twofold to fourfold change), while the genes HVO_2254 and HVO_2255 are being overexpressed during the exponential phase. However, most genes remain unchanged upon viral infection.

Provirus Region Modulates the Outcome of HFPV-1 Infection. Interactions between coinfecting viruses or between a virus and mobile genetic elements has been shown to have significant effects on viral fitness and life cycles (91, 92). Given the fact that genes from both proviral regions are down-regulated during infection with HFPV-1, we explored putative virus–virus interactions. To this purpose, a mutant strain lacking the provirus region Halfvol 3 was generated and subjected to infection with HFPV-1. Growth assays showed a much stronger growth retardation in the mutant strains lacking the provirus region (Fig. 6*A*). This suggests that, in the absence of this genome segment, HFPV-1 generates a more acute infection, which, in turn, suggests that the provirus region encodes one or more genes that interact with the primary infecting virus, and that mitigate the overall effects of the infection and decrease the capacity of HFPV-1 virus to generate additional copies of its genome. This is strongly supported by the differences in the observed VHR, where the infected mutant strain reaches up to 10 times higher VHR values than the parental strain (Fig. 6*B*). Interestingly, only two genes in this region are being down-regulated (HVO_1422 and HVO_1423, as described above). HVO_1423 as a potential repressor represents the best candidate to be responsible for inhibition of HFPV-1 infection, while, in return, the down-regulation of HVO_1422, the integrase, by HFPV-1 possibly prevents the excision of the provirus.

HFPV-1 Is a Promiscuous and Globally Distributed Virus. Previous studies have suggested that head-tail viruses from hypersaline environments have an unusually broad host range, as several viruses are able to infect multiple genera or secondary hosts isolated in distant locations (63, 93). However, the opposite trend has been observed for the members of the *Pleolipoviridae* family, as all isolates described, so far, were shown to be extremely species specific, infecting only a single host, commonly isolated from the same sampling site (11, 63).

Nonetheless, given that HFPV-1 was originally isolated from salt crust from Lake Tyrrell, Australia, and *H. volcanii* was isolated in the Dead Sea, we performed in silico analyses to identify potential additional hosts for HFPV-1. Particularly, when searching against the isolate and metagenome-derived CRISPR spacers database in the IMG/VR database, v.2.0 (53), we identified 10 spacers that were significant matches against the HPFV-1 genome, with identities ranging from 97 to 100% and an average length of 36 bp (Dataset S3). All of the spacers derived from sequenced genomes of previously isolated *Haloferax* species (i.e., *H. lucentense, H. denitrificans, H. alexandrinus, and H. masiliensis*), while no match was detected in the metagenomics-derived sequences. Furthermore, assessment of 59 genomes of previously sequenced *Halobacteria* isolates in Becker et al. (94) revealed that the genome of HFPV-1 could be detected in 9 of them, using previously described thresholds for viral detection in genomic data (read mapping identity $\geq$ 95%, genome coverage $\geq$ 75%) (55). This suggests that the isolates were likely infected with HFPV-1 or a closely related variant at the time of isolation, but the infection went unnoticed. These isolates belong to multiple genera (i.e., *Haloferax, Natronococcus, Halococcus, and Haloterrigena*; *SI Appendix*, Table S2), and indicate a wide host range of HFPV-1.

Hence, we assessed the host range of HFPV-1 under laboratory conditions using a collection of previously isolated haloarchaea. This yielded several species that exhibited susceptibility to infection by HFPV-1, while also being able to produce consistently viral progeny and generating a chronic infection similar to the one observed in *H. volcanii*. Surprisingly, the susceptible hosts included related genera within the *Haloferacaceae* family (i.e., *Haloquadratum* and *Halorubrum*), and also crossed higher taxonomic rank, infecting a member from a different family (i.e., *Haloarculaeace*) (*SI Appendix*, Table S1 and Fig. S8). Quantification of HFPV-1 gcn in culture supernatants from these alternative hosts reveled that *H. volcanii* produced the highest number of virus particles among the assessed organisms(2.542×10^{11} copies per mL^{-1} per $OD_{600}{}^{-1}$; *SI Appendix*, Table S3). Interestingly, another member of the same family the square-shaped *Haloquadratum walsbyi*, for which no virus has been yet isolated, displayed the second highest efficiency of viral production (2.808×10^{10} copies per mL^{-1} per $OD_{600}{}^{-1}$). Meanwhile, the most distantly related organism, *Haloarcula japonica* (*Haloarculaeace*), exhibited the lowest efficiency (2.730×10^{8} copies per mL^{-1} per $OD_{600}{}^{-1}$). Nevertheless, infected cultures of these

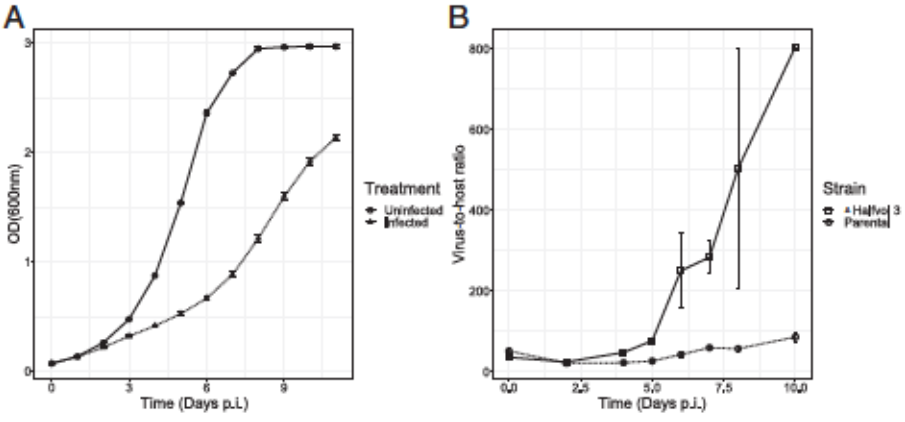


Fig. 6. Virus life cycle in provirus knockout strain. (*A*) Growth curve of uninfected (circle) and infected (black triangles) ΔHalfvol 3 mutant strains cultures. (*B*) VHR calculated by qPCR of virus and host gcn within cells for the parental (open circles) and the mutant strain (open squares). Error bars represent the SD for biological triplicates on both control and treatments.

alternative hosts do not display any significant growth retardation (*SI Appendix*, Fig. S9), which suggests that the virus hijacks host resources and that metabolism is not as pronounced as observed in *H. volcanii*.

Notably, this work represents a report of a promiscuous pleolipovirus, and, while haloarchaeal myoviruses have a very broad host range that crosses the genus level (63, 95), HFPV-1 displays an even broader host range, successfully infecting members of distant lineages at the family level. This versatility of HFPV-1, added to the extreme metabolic remodeling of the host during infection, while possessing a very small genetic repertoire, suggests the existence of global transcriptional regulators on the virus genome and a very conserved receptor on the cell surface across different host lineages. This makes HFPV-1, including its promotors and transcriptional regulators, a great candidate to be developed into a genetic tool, expanding the genetic toolset of model haloarchaea and possibly allowing genetic manipulation of so far inaccessible haloarchaeal strains.

Conclusions

Here we isolated a pleolipovirus infecting the model organism *H. volcanii*. HFPV-1 exhibits a chronic life cycle, being released from host cells without cell lysis and causing only a minor growth retardation despite high viral titers. HFPV-1 does not produce plaques either on *H. volcanii* or on other host organisms, demonstrating that plaque assays are often not suitable to isolate viruses causing a chronic infection and to determine their host range. Furthermore, transcriptional analyses of infected cultures revealed that HFPV-1 infection leads to a radical alteration of the host transcriptional program, with a great proportion of the differentially expressed genes being significantly down-regulated. In particular, we observed that provirus genes in the genome of *H. volcanii* are strongly down-regulated upon HFPV-1 infection, while a knockout strain lacking one of the proviruses experienced a more severe growth retardation than the wild type upon infection. Therefore, we conclude that the provirus encodes a defense mechanism that is likely based on a repressor gene, and that these virus–virus interactions drive the success and temporal stability of HFPV-1 infection. Furthermore, upon HFPV-1 infection, an inhibition of the host CRISPR-Cas immune system was observed, and none of the other known host defense mechanisms seem to be active for so far unknown reasons.

The establishment of this stable virus–host system in *H. volcanii* opens up the possibility of investigating the influence of a chronic infection on all cellular processes in an archaeal model organism. Additionally, the chronic nature of HFPV-1 infection, together with its wide host range and the high copy number of its genome, makes HFPV-1 a great candidate to be developed as a genetic tool, similar to chronic bacteriophages (6). The small genome could potentially be manipulated within *H. volcanii* and then used for high-yield protein expression in *H. volcanii* or to enable genetic modification of so far inaccessible haloarchaea.

Data, Materials, and Software Availability. HFPV-1 virus genome was submited to the National Center for Biotechnology Information GenBank database (accession number OM621814) (96) and raw reads from RNA sequencing data have been deposited in the European Nucleotide Archive (project number PRJEB50750) (97).

ACKNOWLEDGMENTS. We thank Karsten Thiel and the Fraunhofer Institute for Manufacturing Technology and Advanced Materials, Bremen, Germany, for helping out with TEM. We thank Dr. Anna Shevchenko and the Max-Planck-Institute of Molecular Cell Biology and Genetic, Dresden, Germany, for MS analysis of virus proteins. We thank Daniela Thies and Ingrid Kunze (MPI for Marine Microbiology, Bremen, Germany) for assistance with some of the experiments. We acknowledge Anita Marchfelder for helpful discussions. Finally, we thank the MPI for Marine Microbiology and the Max-Planck-Society for continuous support.

Author affiliations: [a]Max-Planck-Institute for Marine Microbiology, Archaeal Virology, Celsiusstrasse 1, Bremen, 28359 Germany; and [b]The Shmunis School of Biomedicine and Cancer Research, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

1. A. G. Cobián Güemes *et al.*, Viruses as winners in the game of life. *Annu. Rev. Virol.* **3**, 197–214 (2016).
2. S. W. Wilhelm, C. A. Suttle, Viruses and nutrient cycles in the sea aquatic food webs. *Bioscience* **49**, 781–788 (1999).
3. A. E. Zimmerman *et al.*, Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* **18**, 21–34 (2019).
4. C. Howard-Varona, K. R. Hargreaves, S. T. Abedon, M. B. Sullivan, Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME J.* **11**, 1511–1520 (2017).
5. J. Maniloff, J. Das, J. R. Christensen, Viruses of mycoplasmas and spiroplasmas. *Adv. Virus Res.* **21**, 343–380 (1977).
6. S. Mäntynen, E. Laanto, H. M. Oksanen, M. M. Poranen, S. L. Díaz-Muñoz, Black box of phage-bacterium interactions: Exploring alternative phage infection strategies. *Open Biol.* **11**, 210188 (2021).
7. M. K. Pietilä, E. Roine, L. Paulin, N. Kalkkinen, D. H. Bamford, An ssDNA virus infecting archaea: A new lineage of viruses with a membrane envelope. *Mol. Microbiol.* **72**, 307–319 (2009).
8. M. Krupovic; ICTV Report Consortium, ICTV virus taxonomy profile. *J. Gen. Virol.* **99**, 617–618 (2018).
9. S. Roux *et al.*, Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* **4**, 1895–1906 (2019).
10. K. S. Gazi, J. G. Kim, O. Y. Olufowora, U. J. Lee, S. K. Rhee, Prevalence of Nitrosopumilus spindle-shaped viruses in different ecogeographical locations. *Korean J. Microbiol.* **57**, 99–105 (2021).
11. M. K. Pietilä, E. Roine, A. Sencilo, D. H. Bamford, H. M. Oksanen, *Pleolipoviridae*, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. *Arch. Virol.* **161**, 249–256 (2016).
12. Y. Liu *et al.*, A novel type of polyhedral viruses infecting hyperthermophilic archaea. *J. Virol.* **91**, e00589-17 (2017).
13. R. Hartman *et al.*, Discovery and characterization of thermoproteus spherical pilliferous virus 1: A spherical archaeal virus decorated with unusual filaments. *J. Virol.* **94**, e00036-20 (2020).
14. D. Prangishvili, M. Krupovic; ICTV Report Consortium, ICTV virus taxonomy profile. *J. Gen. Virol.* **99**, 1357–1358 (2018).
15. A. Nasir, P. Forterre, K. M. Kim, G. Caetano-Anollés, The distribution and impact of viral lineages in domains of life. *Front. Microbiol.* **5**, 194 (2014).
16. M. K. Pietilä, T. A. Demina, N. S. Atanasova, H. M. Oksanen, D. H. Bamford, Archaeal viruses and bacteriophages: Comparisons and contrasts. *Trends Microbiol.* **22**, 334–344 (2014).
17. A. W. S. Luk, T. J. Williams, S. Erdmann, R. T. Papke, R. Cavicchioli, Viruses of haloarchaea. *Life (Basel)* **4**, 681–715 (2014).
18. A. C. Lindås, E. A. Karlsson, M. T. Lindgren, T. J. G. Ettema, R. Bernander, A unique cell division machinery in the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18942–18946 (2008).
19. J. Liu *et al.*, Archaeal extracellular vesicles are produced in an ESCRT-dependent manner and promote gene transfer and nutrient cycling in extreme environments. *ISME J.* **15**, 2892–2905 (2021).
20. J. Liu *et al.*, Functional assignment of multiple ESCRT-III homologs in cell division and budding in *Sulfolobus islandicus*. *Mol. Microbiol.* **105**, 540–553 (2017).
21. E. R. J. Quemin *et al.*, Eukaryotic-like virus budding in archaea. *MBio* **7**, 3–7 (2016).
22. D. P. Baquero *et al.*, New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *ISME J.* **14**, 1821–1833 (2020).
23. T. A. Demina, H. M. Oksanen, Pleomorphic archaeal viruses: The family *Pleolipoviridae* is expanding by seven new species. *Arch. Virol.* **165**, 2723–2731 (2020).
24. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: Guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2019).
25. S. Casjens, Prophages and bacterial genomics: What have we learned so far? *Mol. Microbiol.* **49**, 277–300 (2003).
26. S. Roux, S. J. Hallam, T. Woyke, M. B. Sullivan, Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* **4**, e08490 (2015).
27. S. Benler, E. V. Koonin, Phage lysis-lysogeny switches and programmed cell death: Danse macabre. *BioEssays* **42**, e2000114 (2020).
28. N. Cumby, A. M. Edwards, A. R. Davidson, K. L. Maxwell, The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J. Bacteriol.* **194**, 5012–5019 (2012).
29. J. Bondy-Denomy, A. R. Davidson, When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. *J. Microbiol.* **52**, 235–242 (2014).
30. J. Bondy-Denomy *et al.*, Prophages mediate defense against phage infection through diverse mechanisms. *ISME J.* **10**, 2854–2866 (2016).
31. A. A. Vostrov, O. A. Vostrukhina, A. N. Svarchevsky, V. N. Rybchin, Proteins responsible for lysogenic conversion caused by coliphages N15 and φ80 are highly homologous. *J. Bacteriol.* **178**, 1484–1486 (1996).
32. M. J. Lu, U. Henning, Superinfection exclusion by T-even-type coliphages. *Trends Microbiol.* **2**, 137–139 (1994).
33. E. R. J. Quemin *et al.*, First insights into the entry process of hyperthermophilic archaeal viruses. *J. Virol.* **87**, 13379–13385 (2013).
34. S. Medvedeva *et al.*, Virus-borne mini-CRISPR arrays are involved in interviral conflicts. *Nat. Commun.* **10**, 5204 (2019).

35. J. Soppa, From genomes to function: Haloarchaea as model organisms. *Microbiology (Reading)* **152**, 585–590 (2006).
36. D. T. Nieuwlandt, C. J. Daniels, An expression vector for the archaebacterium *Halobacterium volcanii*. *J. Bacteriol.* **172**, 7104–7110 (1990).
37. Y. Liao, S. Ithurbide, C. Evenhuis, J. Löwe, I. G. Duggin, Cell division in the archaeon *Haloferax volcanii* relies on two FtsZ proteins with distinct functions in division ring assembly and constriction. *Nat. Microbiol.* **6**, 594–605 (2021).
38. S. D. Nuttall, M. L. Dyall-Smith, HF1 and HF2: Novel bacteriophages of halophilic archaea. *Virology* **197**, 678–684 (1993).
39. D. Auslannikava *et al.*, Evolution of genome architecture in archaea: Spontaneous generation of a new chromosome in *Haloferax volcanii*. *Mol. Biol. Evol.* **35**, 1855–1868 (2018).
40. M. Dyall-Smith, Halohandbook Version 7.2 (2009). https://haloarchaea.com/halohandbook/.
41. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads kenkyuhi hojokin gan rinsho kenkyu jigyo. *EMBnet. J.* **17**, 10–12 (2013).
42. A. Bankevich *et al.*, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
43. B. Ushnell, BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), Version 38.06 (2014).
44. D. Hyatt, *et al.*, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
45. L. Zimmermann *et al.*, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
46. M. Steinegger, M. Mirdita, J. Söding, Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nat. Methods* **16**, 603–606 (2019).
47. O. Knittelfelder *et al.*, Shotgun lipidomics combined with laser capture microdissection: A tool to analyze histological zones in cryosections of tissues. *Anal. Chem.* **90**, 9868–9878 (2018).
48. A. I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
49. A. Keller, A. I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
50. R. G. Cutler, J. E. Evans, Synchronization of bacteria by a stationary-phase method. *J. Bacteriol.* **91**, 469–476 (1966).
51. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2009).
53. D. Paez-Espino *et al.*, IMG/VR: A database of cultured and uncultured DNA viruses and retroviruses. *Nucleic Acids Res.* **45**, D457–D465 (2017).
54. C. Camacho *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
55. S. Roux, J. B. Emerson, E. A. Eloe-Fadrosh, M. B. Sullivan, Benchmarking viromics: An *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**, e3817 (2017).
56. S. W. Cline, W. L. Lam, R. L. Charlebois, L. C. Schalkwyk, W. F. Doolittle, Transformation methods for halophilic archaebacteria. *Can. J. Microbiol.* **35**, 148–152 (1989).
57. T. Allers, H. P. Ngo, M. Mevarech, R. G. Lloyd, Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* **70**, 943–953 (2004).
58. M. Pohlschroder, S. Schulze, *Haloferax volcanii*. *Trends Microbiol.* **27**, 86–87 (2019).
59. C. Bath, T. Cukalac, K. Porter, M. L. Dyall-Smith, His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology* **350**, 228–239 (2006).
60. J. B. Emerson, B. C. Thomas, K. Andrade, K. B. Heidelberg, J. F. Banfield, New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. *Appl. Environ. Microbiol.* **79**, 6755–6764 (2013).
61. A. Senčilo, L. Paulin, S. Kellner, M. Helm, E. Roine, Related haloarchaeal pleomorphic viruses contain different genome types. *Nucleic Acids Res.* **40**, 5523–5534 (2012).
62. A. L. Hartman *et al.*, The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* **5**, e9605 (2010).
63. N. S. Atanasova, E. Roine, A. Oren, D. H. Bamford, H. M. Oksanen, Global network of specific virus-host interactions in hypersaline environments. *Environ. Microbiol.* **14**, 426–440 (2012).
64. M. Krupovic, V. Cvirkaite-Krupovic, J. Iranzo, D. Prangishvili, E. V. Koonin, Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res.* **244**, 181–193 (2018).
65. A. Gorlas, C. Geslin, A simple procedure to determine the infectivity and host range of viruses infecting anaerobic and hyperthermophilic microorganisms. *Extremophiles* **17**, 349–355 (2013).
66. S. Breuert, T. Allers, G. Spohn, J. Soppa, Regulated polyploidy in halophilic archaea. *PLoS One* **1**, e92 (2006).
67. C. Howard-Varona *et al.*, Phage-specific metabolic reprogramming of virocells. *ISME J.* **14**, 881–895 (2020).
68. S. Nambi *et al.*, The oxidative stress network of *Mycobacterium tuberculosis* reveals coordination between radical detoxification systems Subhalaxmi. *Cell Host Microbe.* **17**, 829–837 (2015).
69. T. Nyström, F. C. Neidhardt, Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Mol. Microbiol.* **11**, 537–544 (1994).
70. D. R. Gelsinger, J. DiRuggiero, Transcriptional landscape and regulatory roles of small noncoding RNAs in the oxidative stress response of the haloarchaeon *Haloferax volcanii*. *J. Bacteriol.* **200**, e00779-17 (2018).
71. L. K. Maier, M. Dyall-Smith, A. Marchfelder, The adaptive immune system of *Haloferax volcanii*. *Life (Basel)* **5**, 521–537 (2015).
72. A. J. Molina-Henares, T. Krell, M. E. Guazzaroni, A. Segura, J. L. Ramos, Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. *FEMS Microbiol. Rev.* **30**, 157–186 (2005).
73. V. Anantharaman, L. Aravind, Diversification of catalytic activities and ligand interactions in the protein fold shared by the sugar isomerases, eIF2B, DeoR transcription factors, acyl-CoA transferases and methenyltetrahydrofolate synthetase. *J. Mol. Biol.* **356**, 823–842 (2006).
74. E. Okutan *et al.*, Novel insights into gene regulation of the rudivirus SIRV2 infecting *Sulfolobus* cells. *RNA Biol.* **10**, 875–885 (2013).
75. A. C. Ortmann *et al.*, Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with *Sulfolobus* turreted icosahedral virus. *J. Virol.* **82**, 4874–4883 (2008).
76. T. E. F. Quax *et al.*, Massive activation of archaeal defense genes during viral infection. *J. Virol.* **87**, 8419–8428 (2013).
77. S. Fusco *et al.*, Transcriptome analysis of *Sulfolobus solfataricus* infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system. *Biochimie* **118**, 322–332 (2015).
78. Y. Ren, Q. She, L. Huang, Transcriptomic analysis of the SSV2 infection of *Sulfolobus solfataricus* with and without the integrative plasmid pSSVI. *Virology* **441**, 126–134 (2013).
79. F. Karlsson, A. C. Malmborg-Hager, A. S. Albrekt, C. A. K. Borrebaeck, Genome-wide comparison of phage M13-infected vs. uninfected *Escherichia coli*. *Can. J. Microbiol.* **51**, 29–35 (2005).
80. P. Nußbaum, M. Gerstner, M. Dingethal, C. Erb, S. V. Albers, The archaeal protein SepF is essential for cell division in *Haloferax volcanii*. *Nat. Commun.* **12**, 3469 (2021).
81. J. Liu *et al.*, Virus-induced cell gigantism and asymmetric cell division in archaea. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022578118 (2021).
82. S. V. Albers, K. F. Jarrell, The archaellum: An update on the unique archaeal motility structure. *Trends Microbiol.* **26**, 351–362 (2018).
83. S. Bradde, A. Nourmohammad, S. Goyal, V. Balasubramanian, The size of the immune repertoire of bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5144–5151 (2020).
84. M. Li, R. Wang, D. Zhao, H. Xiang, Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* **42**, 2483–2492 (2014).
85. T. Kuprat, M. Ortjohann, U. Johnsen, P. Schönheit, Glucose metabolism and acetate switch in archaea: Analysis of the enzymes involved in *Haloferax volcanii*. *J. Bacteriol.*, *203* (2021). 10.1128/JB.00690-20.
86. S. L. Díaz-Muñoz, Viral coinfection is shaped by host ecology and virus-virus interactions across diverse microbial taxa and environments. *Virus Evol.* **3**, vex011 (2017).
87. M. Dyall-Smith, F. Pfeiffer, P.-W. Chiang, S.-L. Tang, The novel halovirus Hardycor1, and the presence of active (induced) proviruses in four haloarchaea. *Genes (Basel)* **12**, 149 (2021).
88. M. L. Holmes, F. Pfeifer, M. L. Dyall-Smith, Analysis of the halobacterial plasmid pHK2 minimal replicon. *Gene* **153**, 117–121 (1995).
89. M. Dyall-Smith, F. Pfeifer, A. Witte, D. Oesterhelt, F. Pfeiffer, Complete genome sequence of the model halovirus phiH1 (ΦH1). *Genes (Basel)* **9**, 493 (2018).
90. R. Ken, N. R. Hackett, Halobacterium halobium strains lysogenic for phage phi H contain a protein resembling coliphage repressors. *J. Bacteriol.* **173**, 955–960 (1991).
91. S. Erdmann, S. Le Moine Bauer, R. A. Garrett, Inter-viral conflicts that exploit host CRISPR immune systems of *Sulfolobus*. *Mol. Microbiol.* **91**, 900–917 (2014).
92. A. Luque, C. B. Silveira, Quantification of lysogeny caused by phage coinfections in microbial communities from biophysical principles. *mSystems* **5**, e00353-20 (2020).
93. C. M. Mizuno *et al.*, Novel haloarchaeal viruses from Lake Retba infecting *Haloferax* and *Halorubrum* species. *Environ. Microbiol.* **21**, 2129–2147 (2019).
94. E. A. Becker *et al.*, Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet.* **10**, e1004784 (2014).
95. N. S. Atanasova, T. A. Demina, A. Buivydas, D. H. Bamford, H. M. Oksanen, Archaeal viruses multiply: Temporal screening in a solar saltern. *Viruses* **7**, 1902–1926 (2015).
96. T. Alarcon-Schumacher, A. Naor, U. Gophna, S. Erdmann. Haloferax pleomorphic virus 1 (HFPV-1). National Center for Biotechnology Information GenBank database. https://www.ncbi.nlm.nih.gov/nuccore/OM621814. Deposited 8 February 2022.
97. T. Alarcon-Schumacher, A. Naor, U. Gophna, S. Erdmann. Haloferax volcanii infected with HFPV-1 virus. European Nucleotide Archive. https://www.ebi.ac.uk/ena/browser/view/PRJEB50750. Deposited 8 February 2022.

Supplementary material

Isolation of the first virus causing a chronic infection in the archaeal model organism *Haloferax volcanii* reveals antiviral activities of a provirus

Tomas Alarcón-Schumacher, Adit Naor, Uri Gophna and Susanne Erdmann

Supplementary Information

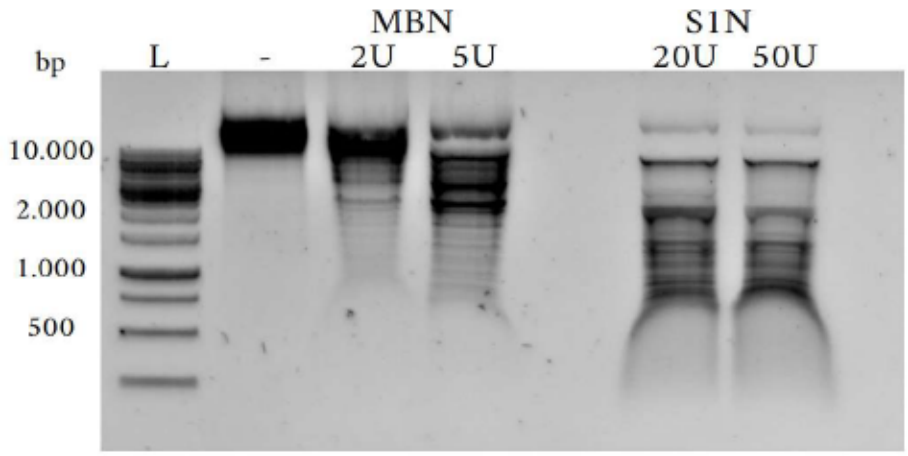
Supplementary Figures

Supplementary Tables

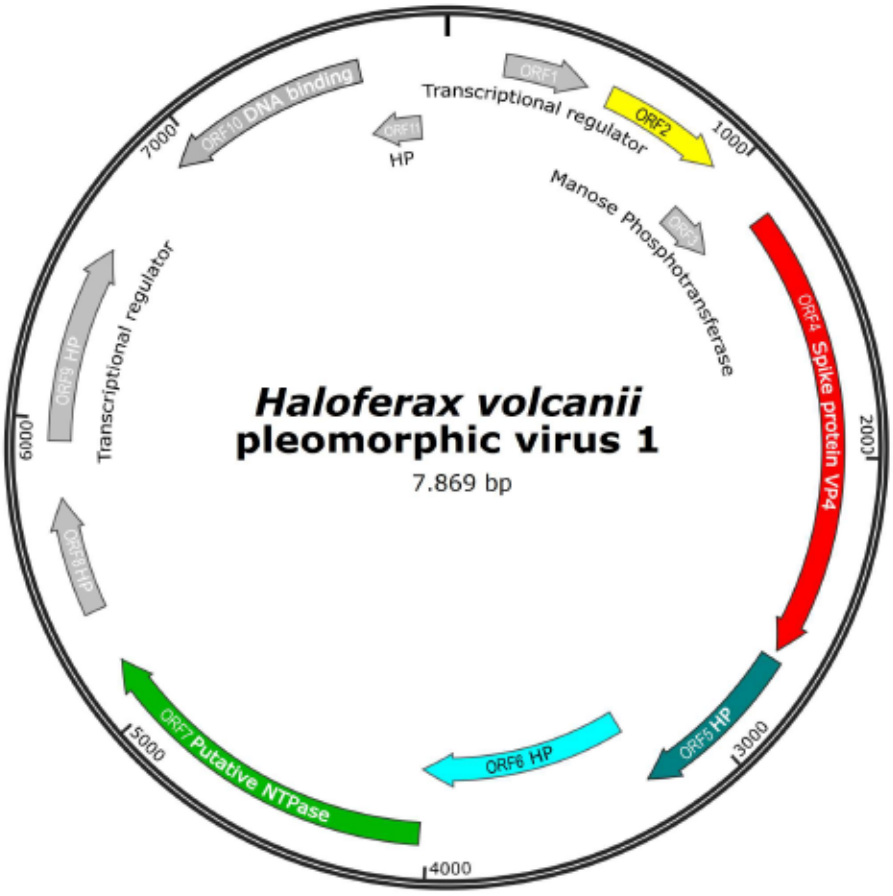
Datasets

Supplementary References

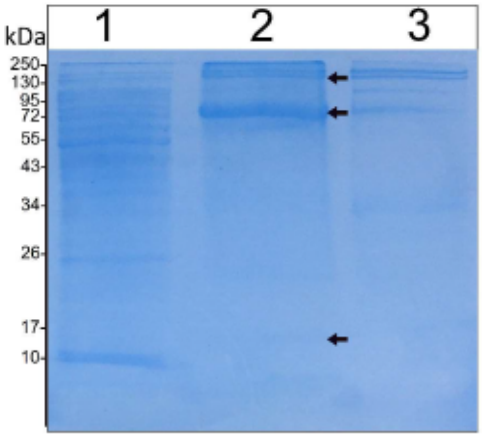
Supplementary Figures



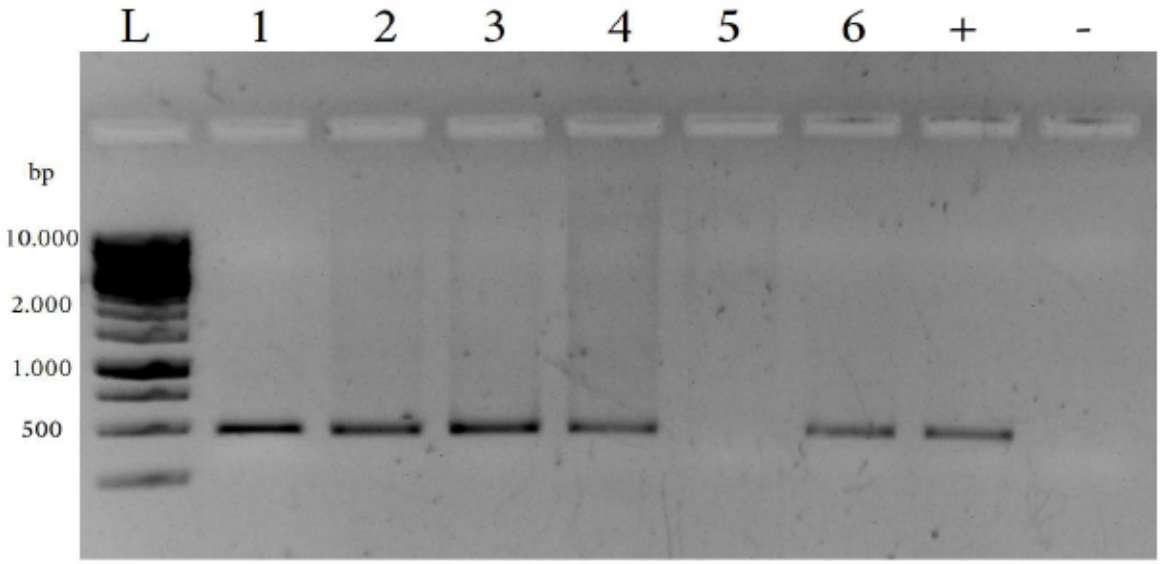
Supplementary figure 1: HFPV-1 genome contains ssDNA regions. Analytic digestion of native purified virus DNA with Mung bean nuclease (MBU) and endonuclease S1 (S1N) to confirm the presence of single stranded regions. Molecular weight size marker (L) is shown (Gene ruler 1kb DNA ladder Thermo Fisher Scientific). Untreated (-), 2-3. Treatment with nuclease MBU at two different concentrations (2U and 5U respectively), 5, 6. Treatment with nuclease S1 at two different concentrations (20U and 50U respectively). DNA was separated on 1% agarose gels and stained with SYBR™ Safe (Invitrogen™) at a final concentration 1x. Data represent one sample of two biological replicates. The restriction digest patterns were observed in two repetitions.



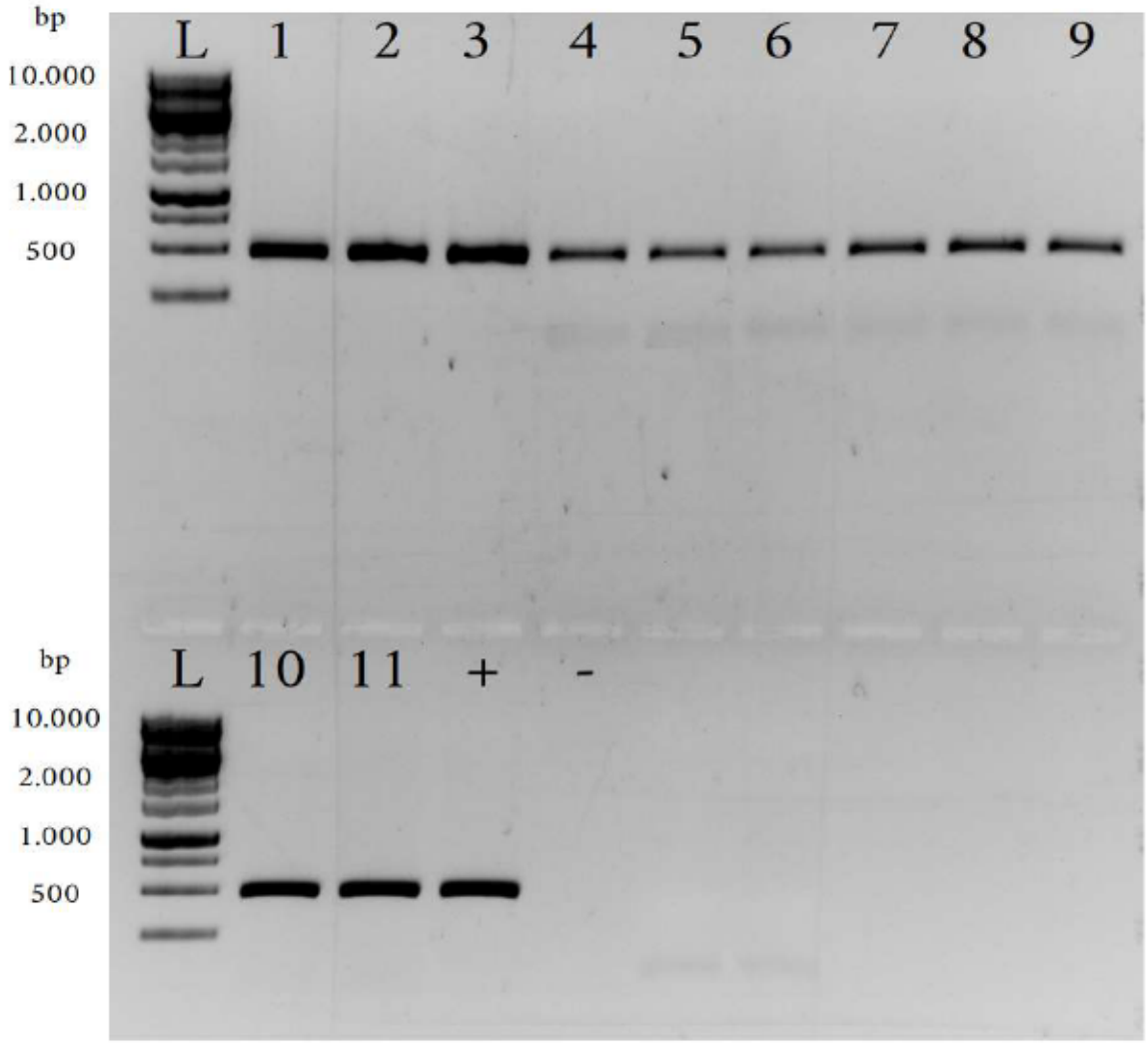
Supplementary figure 2: **Genome map of HFPV-1**. Proteins were predicted using Prodigal (2). In colors are highlighted proteins that had matches against other isolated pleolipoviruses proteins. HP, hypothetical proteins; VP2, structural internal virion protein 2-like.



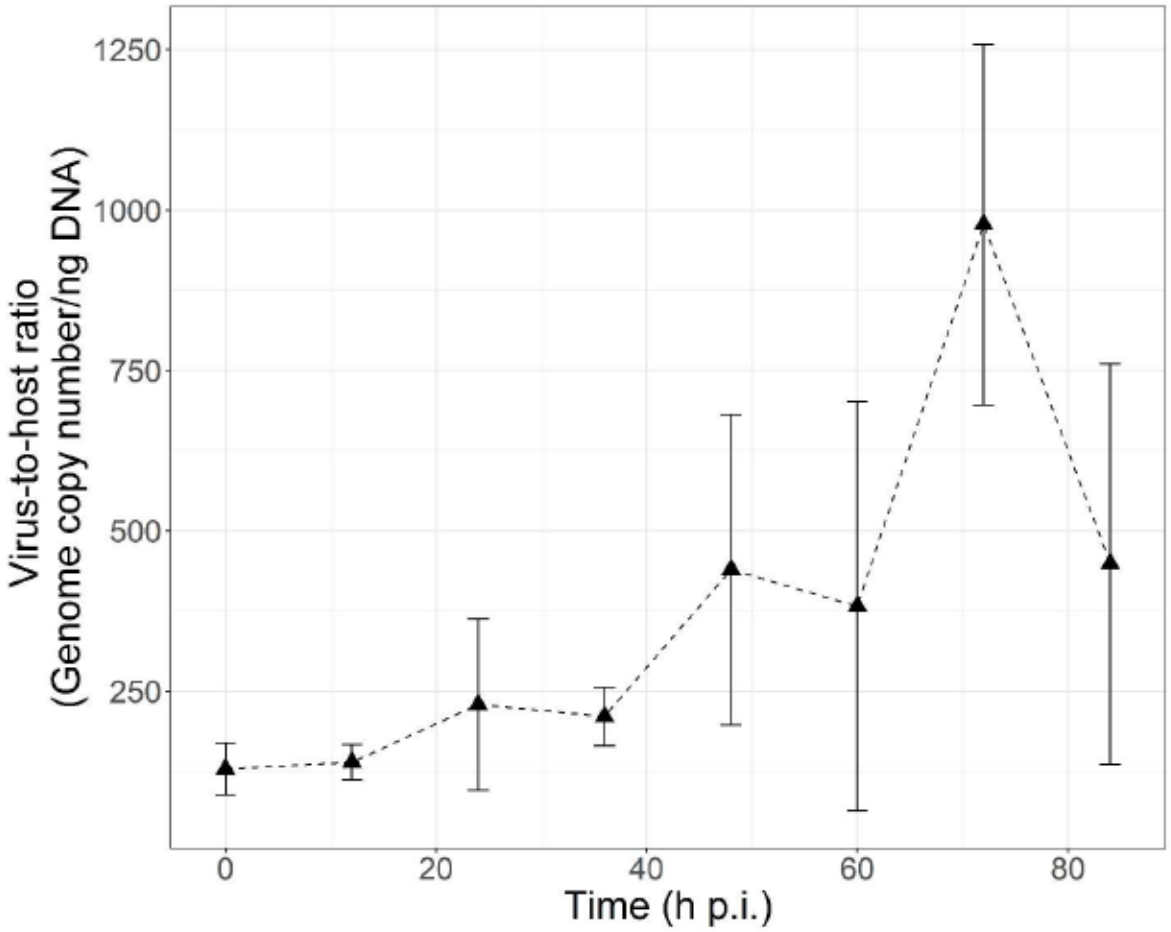
Supplementary figure 3. Protein composition of HFPV-1 virus particles. SDS PAGE gel stained with fast Coomassie blue staining, (1) *H. volcanii* DS2 cell extract, (2) HFPV-1 purified virus particles, (3) membrane vesicles produced by uninfected *H. volcanii*. Arrows indicate additional faint bands observed after the staining.



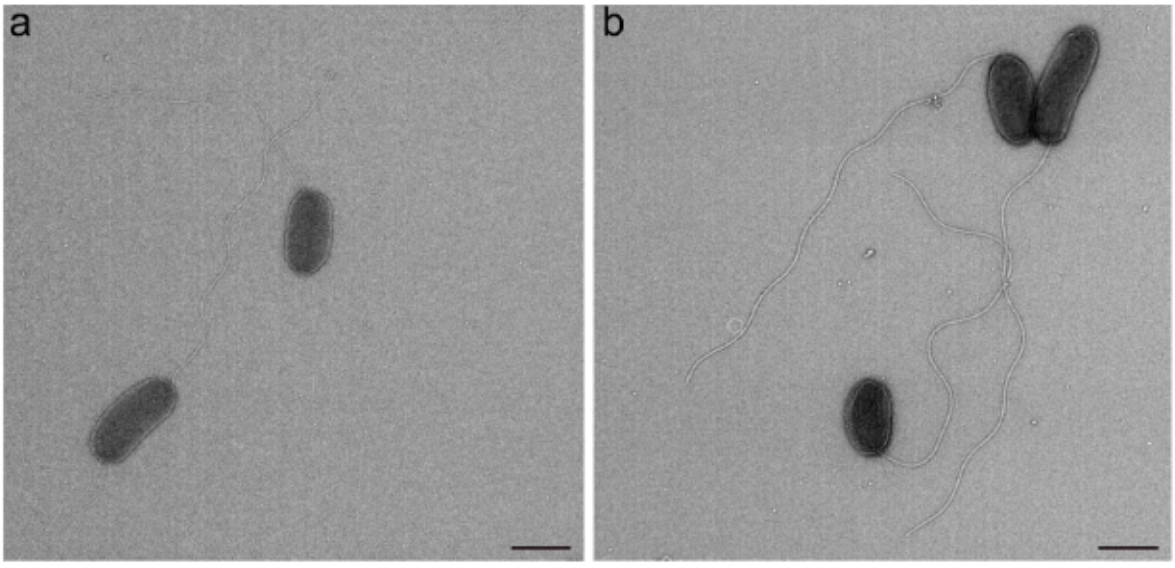
Supplementary figure 4: Infection of *H. volcanii* with HFPV-1 after transformation with purified genomic DNA. Infection was confirmed by PCR assay (see *Experimental procedures* for details). Molecular weight size marker (L) is shown (Gene ruler 1kb DNA ladder Thermo Fisher Scientific). 1-4. PCR on host cells after transformation with different concentrations of purified HFPV-1 DNA (100pg, 1ng, 10ng, 100ng DNA, respectively), 5. PCR on host cells after transformation with DNase treated virus genomic DNA (DNase I, RNase free, New England Biolabs). 6. PCR on host cells after transformation with RNase treated virus genomic DNA (RNase H, DNase free, New England Biolabs) prior to transformation. (+) positive control, (-) negative control. DNA was separated on 1% agarose gels and stained with SYBR™ Safe (Invitrogen™) at a final concentration 1x.



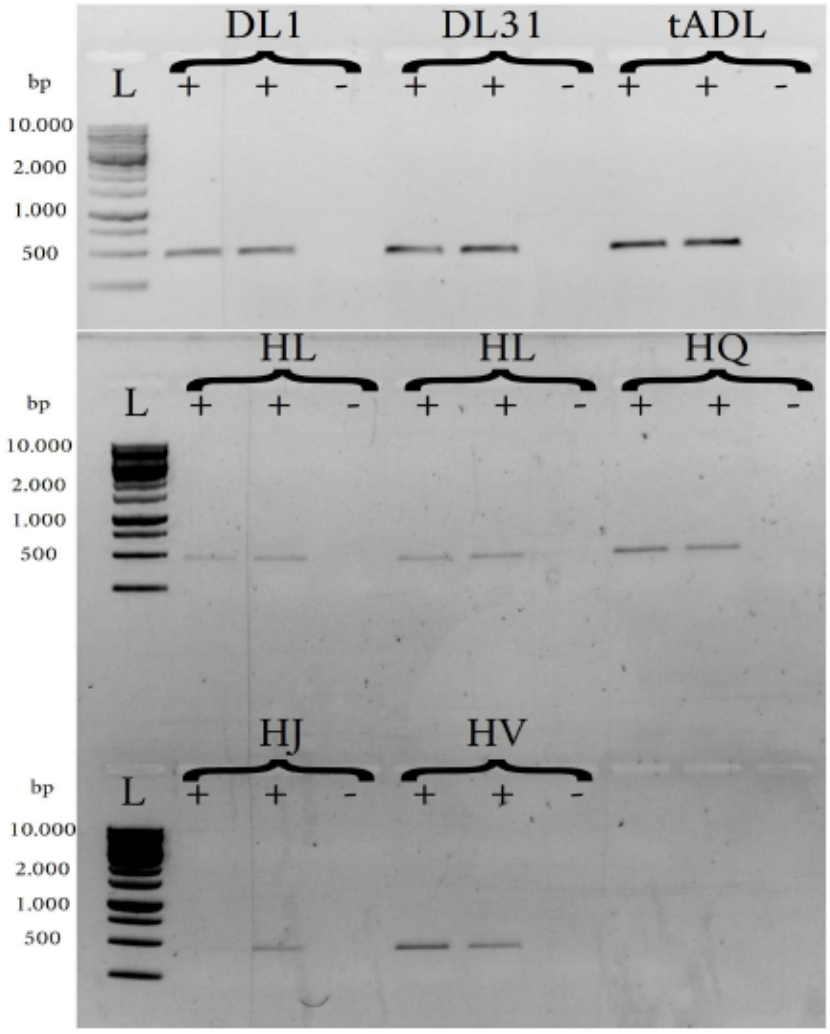
Supplementary figure 5: Long-term infection of *H. volcanii* with HFPV-1. PCR assay to confirm ongoing infections of HFPV-1 virus in *H. volcanii* cultures (see *Experimental procedures* for details). Molecular weight size marker (L) is shown (Gene ruler 1kb DNA ladder Thermo Fisher Scientific). 1-11. Weekly sample of infected cultures (weeks one to eleven), with cultures being diluted to a 0.05 initial OD. DNA was separated on 1% agarose gels and stained with SYBR™ Safe (Invitrogen™) at a final concentration 1x.



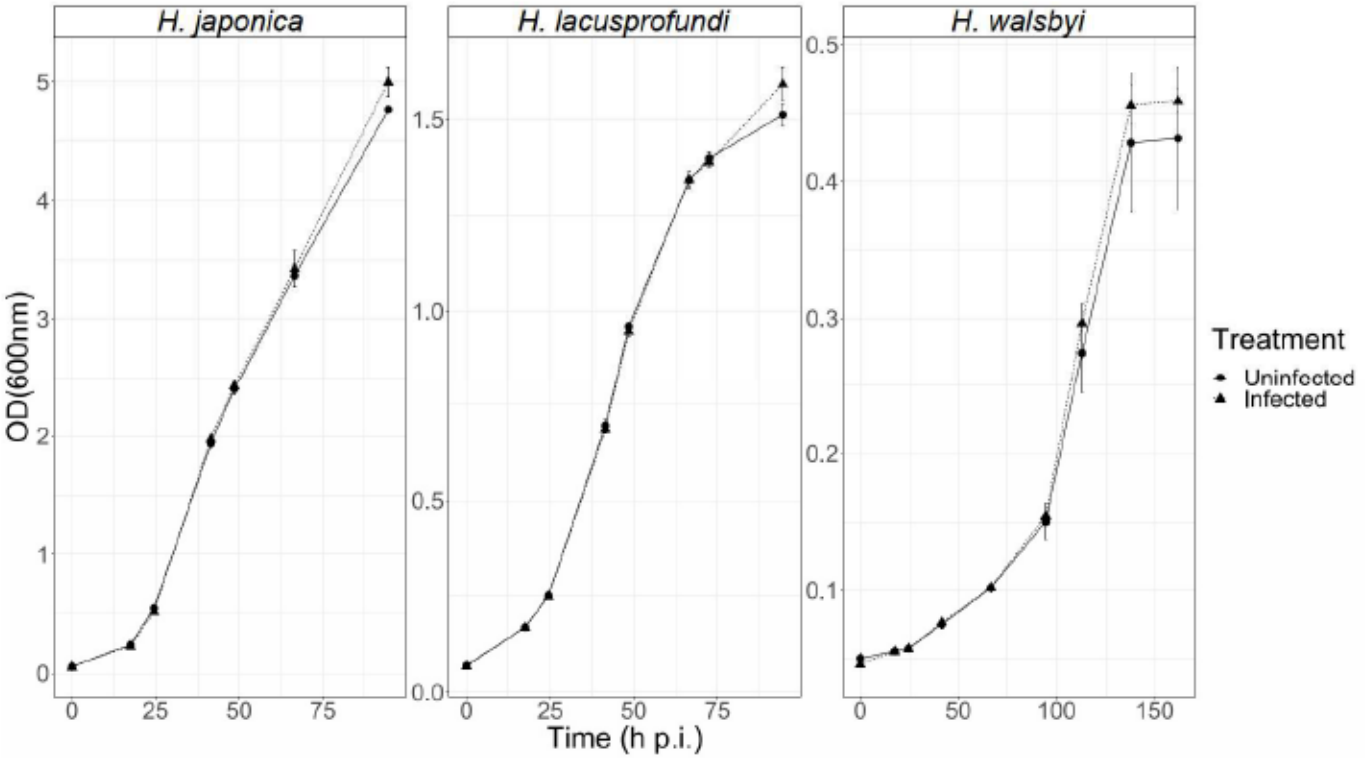
Supplementary figure 6: Virus-to-host ratio (VHR) kinetics. VHR was calculated as the ratio between HFPV-1 genome copy numbers and host genome copy number assessed by qPCR assays. Error bars, represent the standard deviation within biological triplicates. h p.i.: hours post infection.



Supplementary figure 7: Electron micrograph of infected cultures of *H. volcanii*. **a.** Uninfected, **b.** Infected. Samples were negatively stained with uranyl acetate (2%) and imaged at 200kV on a JEM-2100 Plus (JEOL). Scale bar1µm



Supplementary figure 8: Host range assessment of HFPV-1. PCR assay to confirm ongoing infections of HFPV-1 virus in different haloarchaeal strains. Molecular weight size marker (L) is shown (Gene ruler 1kb DNA ladder Thermo Fisher Scientific). Infected (+) and uninfected (-) cultures of seven different species were tested. DL1, DL31 and tADL (DeMaere et al. 2013); HL, *Halorubrum lacusprofundi* (Franzmann et al. 1989); HQ, *Haloquadratum walsbyi* (Burns et al. 2007); HJ, *Haloarcula japonica (*Takashina et al. 1991); HV, *Haloferax volcanii* DS2 (Mullakhanbhai and Larsen 1975). DNA was separated on 1% agarose gels and stained with SYBR™ Safe (Invitrogen™) at a final concentration 1x.



Supplementary figure 9: Life cycle of HFPV-1 on alternative hosts. Growth curve of uninfected (black circles) and infected (black triangles) *H. lacusprofundi*, *H. walsbyi* and *H. japonica* cultures. Error bars, represent the standard deviation for biological triplicates on both treatments. h p.i.: hours post infection.

Supplementary Tables

Supplementary Table 1: List of strains for host range assessment

Strain ID	Reference
Halorubrum lacusprofundi	Franzmann et al. 1989 (3)
Haloquadratum walsbyi	Burns et al. 2007 (4)
Haloarcula japonica	Takashina et al. 1991 (5)
Haloferax volcanii DS2	Mullakhanbhai and Larsen 1975 (6)
Halobacterium sp. DL1	DeMaere et al. 2013 (7)
Halophilic archaeon DL31	DeMaere et al. 2013 (7)
Halohasta litchfieldiae tADL	DeMaere et al. 2013 (7)

Supplementary Table 2: Isolates genomes with reads mapping to the HFPV-1 genome.

Isolate name	Coverage (%)
Natronococcus jeotgali DSM 18795	99.93
Haloferax alexandrinus JCM 10717	93.02
Halorubrum terrestre JCM 10247	100
Natrialba taiwanensis DSM 12281	100
Natrialba hulunbeirensis JCM 10989	80.32
Haloterrigena limicola JCM 13563	95.19
Natrialba chahannaoensis JCM	89.5
Halorubrum californiensis DSM	92.14
Haloferax prahovense DSM	79.18
Haloarcula argentinensis DSM	54.83
Halococcus thailandensis JCM 13552	50.64

Supplementary Table 3: HFPV-1 genome copy number production in alternative hosts

Viral absolute abundance (viral titer) was assessed by qPCR at 94h p.i.. Values show the average of three biological replicates as well as the standard deviation (SD). Efficiency of viral production was calculated by normalizing the viral titer with the optical density (OD).

Strain	Viral titer (copy number x mL^{-1})	SD	Efficiency (copy number x mL^{-1} x OD^{-1})	SD
H. lacusprofundi	$4.386x10^9$	$4.87x10^8$	$3.265x10^9$	$3.61x10^8$
H. walsbyi	$2.864x10^9$	$1.92x10^9$	$2.808x10^{10}$	$1.90x10^{10}$
H. japonica	$9.323x10^8$	$9.70x10^7$	$2.730x10^8$	$3.44x10^7$
H. volcanii	$5.108x10^{11}$	$1.42x10^{11}$	$2.542x10^{11}$	$7.67x10^{10}$

Datasets

Dataset 1: Protein composition of HFPV-1 particles assessed through mass spectrometry **(Excel file)**
https://www.pnas.org/doi/suppl/10.1073/pnas.2205037119/suppl_file/pnas.2205037119.sd01.xlsx

Dataset 2: Differentially expressed genes of *H. volcanii* upon infection with HFPV-1 virus **(Excel file)**
https://www.pnas.org/doi/suppl/10.1073/pnas.2205037119/suppl_file/pnas.2205037119.sd02.xlsx

Dataset 3: Potential additional host of HFPV-1 search via Blast search against IMGVR Isolate Spacers database (1) **(Excel file)**
https://www.pnas.org/doi/suppl/10.1073/pnas.2205037119/suppl_file/pnas.2205037119.sd03.xlsx

Supplementary References

1. D. Paez-Espino, *et al.*, IMG/VR: A database of cultured and uncultured DNA viruses and retroviruses. *Nucleic Acids Res.* **45**, D457–D465 (2017).

2. D. Hyatt, *et al.*, Prodigal : prokaryotic gene recognition and translation initiation site identification (2010).

3. P. D. Franzmann, *et al.*, *Halobacterium lacusprofundi* sp. nov., a Halophilic Bacterium Isolated from Deep Lake, Antarctica. *Syst. Appl. Microbiol.* **11**, 20–27 (1988).

4. D. G. Burns, *et al.*, *Haloquadratum walsbyi* gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. *Int. J. Syst. Evol. Microbiol.* **57**, 387–392 (2007).

5. T. Takashina, T. Hamamoto, K. Otozai, W. D. Grant, K. Horikoshi, *Haloarcula japonica* sp. nov., a New Triangular Halophilic Archaebacterium. *Syst. Appl. Microbiol.* **13**, 177–181 (1990).

6. M. F. Mullakhanbhai, H. Larsen, *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* **104**, 207–214 (1975).

7. M. Z. Demaere, *et al.*, High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16939–16944 (2013).

5. Chapter 2.

Understanding the impacts of chronic viral infections in host fitness and ecology

Tomas Alarcón-Schumacher[1], Fabian Roeloffs[2,3], Julian Schmitz[2,3], Alexander Grünberger[2,3] and Susanne Erdmann[1#]

[1] Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany

[2] Multiscale Bioengineering, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

[3] Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany

[#]**Corresponding author**: Susanne Erdmann, Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany

Email: serdmann@mpi-bremen.de Phone: +49 421 2028 7340

Authors Contributions

T.A-S. and S.E. conceived and led the study and performed the primary writing of the manuscript. T. A-S., F.R, J.S. and A.G. collected and analyzed the experimental data. All authors participated in the analysis and interpretation of the data and contributed to the writing of the manuscript.

Competing Interest Statement:

The authors declare no competing interests.

Keywords: Archaea, virus, chronic infection.

Abstract

Interactions between viruses and their hosts depend largely on viral life style and infection strategy. While lytic viruses heavily influence host abundance and community composition, lysogenic and chronic viruses have more moderate, but long-lasting effects on cell behavior, metabolism and evolutionary trajectories. Nevertheless, how chronic infecting viruses interact with their hosts remains severely understudied in prokaryotes. Chronic infecting viruses have so far rarely been found in bacteria, however, numerous archaeal viruses establish long-term chronic infections with their hosts. Here we characterize the main physicochemical factors driving the interaction between the archaeon *H. volcanii* and the chronic infecting virus HFPV-1. We demonstrate that viral titers are negatively correlated with temperature, and are severely reduced at 45 °C. Our data shows that protein folding and degradation are strongly upregulated under viral infection at 45 °C, likely due to misfolding of viral proteins. Furthermore, we show that host motility is compromised upon viral infection. Furthermore, we showed that post-translational addition of glycans plays a key role in viral fitness, but does not affect particle adsorption nor host range of the virus, suggesting that the main role of glycosylation is related to virus particle stability at changing environmental conditions. Finally, we showed that non-N-glycosylated particles maintain the unique wide host range of HFPV-1, suggesting that its receptor could be a relatively highly conserved protein in a range of haloarchaea. In conclusion, we uncovered the most relevant drivers mediating the chronic infection and affecting virus and host fitness, improving our understanding of chronic infections in prokaryotic halophilic organisms.

Introduction

Viruses are the most abundant biological entities on earth, and are major ecological, evolutionary and biogeochemical drivers in every environment (Cobián Güemes et al., 2016). They are capable of infecting organisms across the tree of life and possess the ability to highjack and reshape hos metabolism in various ways, that range from regulating expression levels of host genes, to introducing metabolic innovation through viral encoded auxiliary metabolic genes (Rosenwasser et al., 2016). Upon viral replication, infected cells seemingly turn into viral factories, which metabolic program can greatly differ from their uninfected relatives within the population (Howard-Varona et al., 2020). The resultant biological entity, is in a transient metabolic state derived from the interaction between virus and host gene expression and is known as "virocell" (Forterre, 2011). However, viruses use different strategies to overcome host defenses and to highjack hos machinery, and therefore, the degree of metabolic reprogramming of the host can greatly differ from virus to virus. Virus-host interactions can range from purely parasitic relationships, like the ones performed by lytic viruses, to different degrees of commensalism or even mutualism, as it can be the case of chronic and lysogenic

infections. Viruses can for example transfer genes, expanding the metabolic repertoire of the host, or can negatively impact competing uninfected strains (DeWerff et al., 2022; Obeng et al., 2016).

Interestingly, viruses establishing long-term chronic infections are in a unique position, given that they are transmitted both horizontally, via virus particle production, as well as vertically along with cell division (Weitz et al., 2019). Thus, as these viruses evolve, they must balance selective pressures for traits that enhance both virus and host fitness, intrinsically driving virus-host interaction toward mutualism. A mutualistic relationship has been shown for the chronic infecting archaeal virus SSV9 and its host *Sacharolobus islandicus*. Infected cells produce a virus-encoded toxin that induce growth inhibition in uninfected cells, while the viral replication occurs with minimal effects on host fitness (Dewerff et al., 2020; DeWerff et al., 2022). Nevertheless, if at all present, mutualistic interactions are not always readily identifiable and the vast majority of chronic infecting viruses are thought to be purely parasitizing their respective hosts.

Recently, the virus Haloferax pleomorphic virus 1 (HFPV-1), causing a chronic productive life cycle infecting the archaeon *Haloferax volcanii* was isolated and characterized (Alarcón-Schumacher et al., 2022). The study of this virus-host system revealed that, alike other prokaryotic viruses infecting halophilic archaea described to the date, HFPV-1 was able to sustain unprecedentedtly high copy numbers inside the cell and high virus particle titers, while having little impact on the overall fitness of the host. Furthermore, *H. volcanii* infected cells with HFPV-1 display a drastically different transcriptional program, highlighted by specific downregulation of over a third of the host-encoded genes and a specific repression of proviruses encoded in *H. volcanii* genome through an undescribed virus exclusion mechanism (Alarcón-Schumacher et al., 2022). Remarkably, HFPV-1 is the only described pleolipovirus that is capable of infecting a wide range of hosts across diverse haloarchaeal clades, thereby exhibiting promising traits for the development as a genetic tool for diverse archaea including those that cannot be genetically manipulated yet. However, the factors modulating virus and host fitness and interactions in this virus host-system, remain understudied.

Moreover, it is common that structural proteins of archaeal viruses are glycosylated or that some viruses even carry their own glycosyltransferases enzymes (Larson et al., 2006; Quemin et al., 2015). Glycan modifications have been proposed to enhance protein stability under harsh environmental conditions (Jayaprakash & Surolia, 2017; F. Wang et al., 2019). Similarly, pleolipoviruses also can present glycan modifications on the structural spike protein (Kandiba et al., 2012; Pietilä et al., 2009). It has been hypothesized that glycosylation of the spike protein plays key roles on host recognition, as often these interactions are mediated by sugar-sugar interactions (Kandiba et al., 2012). We have recently shown that glycosylation plays an important role for particle stability for the HFPV-1 virus during its infection cycle in *Halorubrum lacusprofundi* (Gebhard et al., 2023). However, the role of glycosylation on host recognition and adsorption remains understudied. In *H. volcanii* N-glycosylation, which is the covalent linkage of glycan moieties selected asparagine residues has

been extensively described, and it has been shown that among other proteins, key structural features such as the S-layer glycoprotein or archaellins are glycosylated (Eichler & Maupin-Furlow, 2013). N-glycosylation involves a series of glycosyltransferases (genes AglJ, AglG, AglI and AglE) that add monosaccharides to a lipid carrier (dolichyl phosphate). Then a flipase translocates the carrier to the cell surface and the oligosaccharyltransferase (AglB) transfer the sugar to asparagine residues of target proteins. For the addition of the final sugar, another glycosyltransferase (AglD) adds a mannose to an additional lipid carrier, which is then translocated to the extracellular by the flipase AglR, and added to the tetrasacharide by the glycosyltransferase AglS. Nevertheless, the effect of N-glycosylation on HFPV-1 host range and the particular selective advantage that glycosylation might confer remain mostly enigmatic.

In this study, we aimed to understand how different environmental factors affects virus-host interactions, and observed that; increase in temperature has a severe impact on HFPV-1 horizontal transmission. Furthermore, we showed that the transcriptional program of the host was drastically modified upon infection and that there was a general upregulation of protein degradation and repair machinery, likely due to poor adaptation of viral proteins to higher temperatures. Finally, we showed that glycosylation modifications is a key component for virion stability, but does not play major role for adsorption to host cells nor in the host range of HFPV-1.

Experimental procedures

Viral stocks production and purification

Viral stocks of HFPV-1 were produced on different *H. volcanii* strains (Supplementary table 1) following the protocol established in (Alarcón-Schumacher et al., 2022). Briefly, ~10^9 cells of *H. volcanii* were harvested by centrifugation at 10,000 x g for 10 minutes, at room temperature (RT). Cells were then resuspended in 1mL of fresh Hv-YPC media (Dyall-Smith, 2009), and HFPV-1 was added at a multiplicity of infection (MOI) of 10. Cultures were left 2 hours at RT for adsorption and subsequently upscaled to 500 mL and incubated for 5 days at 28 °C, 120 rpm. Cells were then removed by centrifugation (9,000 x g, 45 minutes). Virus particles were precipitated from the supernatant with polyethylene glycol (PEG) 6,000 (10% w/v final concentration) and incubation at 4 °C overnight. Viral particles were recovered by centrifugation (13,000 x g, 45 min, 4 °C, fixed-angle rotor JA14 Beckmann and Coulter). Pellets were resuspended in 18% buffered sea water (BSW) (180g NaCl, 25g MgCl2, 29g MgSO4 and 5.8 KCl per liter), cell debris were removed by centrifugation (10 min., 10,000 x g, RT), and the viral suspension was sterile filtered two times with polycarbonate syringe filters (pore size 0.2 µm, Sartorius, Göttingen, Germany). Viral stocks were stored at 4 °C and cell contamination was assessed by inoculating the virus solution into 15 mL of fresh HvYPC media and incubation for 2 weeks at 28 °C, at 120 RPM, and growth was monitored via OD600 nm measurements and plating.

Virus infectivity and kinetics

For studying the effect of abiotic factors on viral and host fitness, we performed full infection cycles with HFPV-1 under different conditions (Supplementary Table 2). Standard conditions were HvYPC media, 28 °C and 18% salinity (NaCl), and only one factor was modified in each particular treatment. To study the infective cycle, *H. volcanii* cultures (DS2 wild type strain) were synchronized as described in Alarcon-Schumacher et al., 2022, using the adaptation of the "Stationary phase method" (Cutler & Evans, 1966), with three dilution steps in each case. For infection, $1x10^9$ *H. volcanii* cells (OD_{600} ~ 1) from synchronized cultures were collected by centrifugation (4,500 x g, 45 min), resuspended in 500 µL of fresh media and infected with HFPV-1 virus with an MOI of 10and incubated for 2h at room temperature for adsorption. Subsequently, non-adsorbed viral particles were removed as following: cell-virus mixtures were separated by centrifugation (10,000 x g, 10 min), and the supernatant containing non-adsorbed virus was discarded. Then, cell were washed by resuspending in 1 mL of fresh media, followed by subsequent centrifugation (10,000 x g, 10 min) and removal of the supernatant. To avoid carryover of unadsorbed particles, the washing step was repeated twice, and then, cultures were transferred into 150 mL of respective media.

Growth was monitored by optical density (OD_{600}) twice a day and viral titer of free and intracellular virus genome copy numbers, as well as host copy numbers, were assessed by qPCR as described in (Alarcón-Schumacher et al., 2022). Briefly, DNA was extracted with the genomic DNA extraction kit (Bioline, London, UK). Quantification of *H. volcanii* and HFPV-1 genome copy number were carried out using a CFX96 Touch Real-Time PCR (Bio-Rad Laboratories, Inc., Hercules, CA, USA). For qPCR assessment, each reaction eas 10µL, and contains contained 1X SsoAdvanced Universal SYBR™ Green Supermix (Bio-Rad) and 300 nM of each primer. qPCR set up was configures to 5 min at 95 °C, followed by 35 cycles of 30s at 95 °C and 30s 68 °C, and readings taken between each cycle. Product specificity was assessed through a melting curve (0.5 °C increase). Efficiencies of the assays were 95–100%, with R^2 values ≥0.99 for all assays.

Transcriptomic analyses

RNA extraction of flash-frozen cell pellets was carried out with the Zymo Direct-zol RNA miniprep Kit (R2051). RNA integrity and concentration was measured by spectrophotometry, using a nanodrop DS-11 (DeNovix). To enhance the range of detected genes, ribosomal RNA transcripts were depleted prior to sequencing using the rRNA depletion Kit riboPOOL™, specific for *Haloferax volcanii* (siTOOLs Biotech®). Libraries were prepared with library kit NEBNext® Ultra™ II RNA Library Prep Kit for Illumina and sequencing was performed on an Illumina HiSeq3000 sequencer, following paired-end 2 x 150 run. Bioinformatic analyses were carried out as described in (Alarcón-Schumacher et al., 2022). Briefly, quality trimming of raw reads was performed with software Cutadapt (Martin, 2013). A minimum quality of 30 and a minimum length of 50 thresholds were established (-q 30, -m 30). Filtered reads were mapped to *Haloferax*

volcanii DS2 genome (NC_013967.1 main chromosome; and 4 plasmid, NC_013964.1, NC_013965.1 NC_013966.1 NC_013968.1) and HFPV-1 (OM621814.1) with BBmap v38.06. An identity threshold of 99% (minid = 0.99) was established. Differential expression analyses were performed with R package DESeq2 (Love et al., 2014) and plotted using ggplots2 (Wickham, 2009). Genes with p-values < 0.01, false discovery rates < 0.05 (p-adjusted < 0.05) and a fold change of at least two times (log2FC ≥ 1 or ≤ -1) were considered differentially expressed (DE).

Single-nucleotide polymorphisms and strain variation assessment

In order to investigate if temperature has an impact on the fidelity of viral genome replication we investigated the presence of single-nucleotide polymorphisms (SNPs). For this purpose, we used the software inStrain (Olm et al., 2021), to profile intra-population genetic diversity. SNPs were assessed in transcriptomic datasets of infected cultures at 45 °C (this study), and 28 °C (from Alarcon-Schumacher et. al., 2022) were compared. For this, reads mapping to HFPV-1 genome (see "*Transcriptomic analysis*") were extracted. Subsequently bamfiles were generated, sorted, indexed, and provided to inStrain for SNPs. Pipelines "profile" and "compare" were sequentially run with default parameters.

Viral persistency assessment

In order to assess th persistence of the infection and the proportion of cell infected within the population, serial dilution of infected late-exponential cultures of *H. volcanii* grown at 28 °C and 45 °C were plated in Hv-YPC solid agar plates (10 g/L agar). Plates were incubated at 28 °C for 5 days and dilutions that produced single colonies were selected for further analysis. Subsequently, colony PCR was performed on the obtained colonies. For this, each individual colony was picked and transferred into 10 µL of H_2O, and then diluted in water on a 1:10 ratio. PCR targeting HFPV-1 genome (primers TyrVUF 5'- acgaacgagaacaccgacc-3' and TyrVUR 5'-tgatgacgaatccaacgagcag-3'), with 0.02 U/µL of Q5® High-Fidelity DNA Polymerase (New England Biolabs), 0.1µM of each forward and reverse primers, 1X of Q5 Reaction Buffer and 1X Q5 High GC Enhancer. The following program was used: 5 min 95 °C, followed by 35 cycles of 30s at 95 °C, 30s 68 °C for annealing and 30s at 72 °C for elongation. Results were visualized in 1% agarose gels stained with SYBR™ Safe (InvitrogenTM).

KO mutant strains generation and determination of HFPV-1 infection

Transformation and construction of deletion mutants in *H. volcanii* H53 (Δ*pyrE2 and* Δ*trpA)* and H71 (Δ*pyrE2,* Δ*trpA and* Δ*CetZ1*) based on selection with uracil and tryptophan were performed as described previously (Allers et al., 2004). Briefly, plasmid construction was done by classical restriction enzyme-based molecular cloning. Inserts were amplified from wild type (DS70) genomic DNA (using genomic DNA

extraction kit: Bioline, London, UK, according to the manufacturer's instructions) via polymerase chain reaction (PCR) using PHUSION® polymerase (New England Biolabs, Frankfurt, Germany). Restriction endonucleases used here were acquired from New England Biolabs, Frankfurt, Germany. Virus infectivity on the provirus mutant strain was determined as described above for wild-type *H. volcanii ('Virus infectivity and kinetics')*. Experiments were performed in three biological replicates and the parental strain was used as wild-type control. Hv-YPC+ was supplemented with uracil (50ug/ml) and tryptophan (50ug/ml).

Host range assessment of HFPV-1 of glycosylation defective HFPV-1 particles.

In order to confirm infectivity of glycosylation defective particles, viral particles were produced and isolated (see *"Viral stocks production and purification"*) from the supernatant of two different deletion strains for genes involved in the glycosylation pathway: ΔAglB and ΔAglD (Abu-Qarn & Eichler, 2006). Wild type particles were produced in H26 strain (Δ*pyrE2*). Subsequently, particles were tested for infectivity as follows. Exponential phase cultures were infected with HFPV-1 with a MOI of ~10, following the same procedure aforementioned, and incubated at 28 °C with constant agitation (120 rpm). Cells were collected by centrifugation (10,000 g, 10 min), washed two times with fresh media and screened for HFPV-1 infection by PCR (see "*Viral infectivity and kinetics*" section).

Viral adsorption kinetics

To assess the effect of glycosylation on the interaction between virus and host-cell surface an adsorption assay was performed. Since HFPV-1 is a chronic virus and does not produce distinguishable plaques on soft agar, traditional two-layer approaches were ineffective. Therefore, adsorption was measured in liquid cultures as following: 10^9 host cells in mid-exponential phase (OD ~0.8) were harvested by centrifugation (10,000 x g, 10 min., RT), cells were then resuspended in 1 mL of fresh media and HFPV-1 virus was added with and MOI of 0.5 and incubated at RT. Then biological triplicates were harvested at different times (0, 10, 15, 30, 60 and 120 minutes). For harvesting, cells were removed by centrifugation (10,000 x g, 10 min., RT) and viral particles were precipitated with PEG 6,000 (10% w/v final concentration) and incubated at 4 °C overnight. Virus were then retrieved by centrifugation (13,000 x g, 45 min, 4°C) and DNA extraction was performed with genomic DNA extraction kit (Bioline, London, UK). Then, viral genome copey number in the supernatant was measured by qPCR as described in "*Virus infectivity and kinetics*". Percentage of unadsorbed virus was then calculated in relation to the quantity of virus originally added and plotted using ggplots2 (Wickham, 2009)

Results and discussion

Salinity has no significant effects on virus fitness

The establishment of chronic infections that persists in microbial communities has effects that go beyond the classic population control of lytic phages. These virocell-type entities can have radically different ecological roles and different capabilities to respond to changes in their environment than uninfected relatives within the

microbial communities (Rosenwasser et al., 2016). In order to understand the changes on the ecology of infected cells and the adaptability of the chronic virus-host system of HFPV-1 and *H. volcanii*, we aimed to establish the impact of environmental conditions on the fitness of both virus and host.

It has been previously suggested for filamentous viruses causing chronic infections in *Vibrio alginolyticus* that changes in salinity significantly decreases the fitness of the host upon infection (Goehlich et al., 2019). Hypersaline lakes or solar salterns are subjected to seasonal rains and draughts or tidal floods that significantly change the salt concentrations of the environment. Therefore, the ability to produce an adequate response to osmotic stress is crucial for haloarchaeal organisms. Thus, we assessed the response of infected and uninfected cultures of *H. volcanii* to a range of different salinities (Supplementary table 2). Interestingly, viral infection was successful at all salinities, and all infected cultures showed the characteristic slight growth retardation as observed previously (Alarcón-Schumacher et al., 2022) (Supplementary figure 1). Despite slight slower growth rates at 25% salinity, the overall impact of viral infection on host fitness was not significantly different ($p > 0.05$ Wilcoxon test). Furthermore, viral titers in the supernatant showed a slight decrease at 25% salinity, however this was not significant after normalizing for growth rates. Is has been shown that *H. volcanii* can grow in a wide range of salinities, with relative impact on growth rates (Mullakhanbhai & Larsen, 1975). Furthermore, a study of the response mechanism upon osmotic stress has shown that the 3-hydroxy-3-methylglutaryl Coenzyme A reductase (hmgA, HVO_2583), which is involved in the synthesis of isoprenoids and other lipds, was more abundant in cultures growth at higher salinities -23% -NaCl- (Bidle et al., 2007). Additionally, proteomic analysis have shown that among others, the stress response transcriptional regulator PspA (HVO_2636), which is part of a general stress response mechanism, is upregulated at higher salinities (Bidle et al., 2008). Nevertheless, in previous transcriptomic analysis of infected cultures of *H. volcanii* with HFPV-1, the transcription of the genes previously associated to osmotic stress response remained unchanged, despite a large proportion of other genes being affected (Alarcón-Schumacher et al., 2022). Thus, it is likely that interference of the virus infection with these mechanisms has been negatively selected during virus-host evolution, due to the relevance of osmotic balance control in high-salt conditions.

Different carbon sources limit host growth rates but not viral fitness

Among other described factor regulating host fitness upon infection, it has been reported that lysogenic phages can indirectly have strong detrimental effects on host growth rate depending on the carbon source available (Chen et al., 2005). The archaeon *H. volcanii* is able to grow on a minimal media, with only ammonia as nitrogen source and variety of simple single carbon sources, such as pyruvate or acetate (Kauri et al., 1990; Kuprat et al., 2020). Interestingly, upon infection with HFPV-1, several key genes of gluconeogenesis pathway were previously reported to be among the highest downregulated genes, i.e. fructose-1,6-bisphosphatase (-32 fold change) (Alarcón-Schumacher et al., 2022). Thus, we explored the effect of growth with different carbon

sources on virus and host fitness. However, similarly to salinity, HFPV-1 was successfully able to infect host cells during growth on single carbon sources. Meanwhile, host growth rates were minimally impacted when comparing infected and uninfected cultures, as infected cultures grown on pyruvate and simple mixture of carbon sources did not present the characteristic growth retardation (Supplementary Figure 1). Furthermore, the different carbon sources used, did not significantly changed the ability to produce large amounts of particles, as the titers were not significantly different from those reached under optimal conditions for HFPV-1 replication (Figure 1).

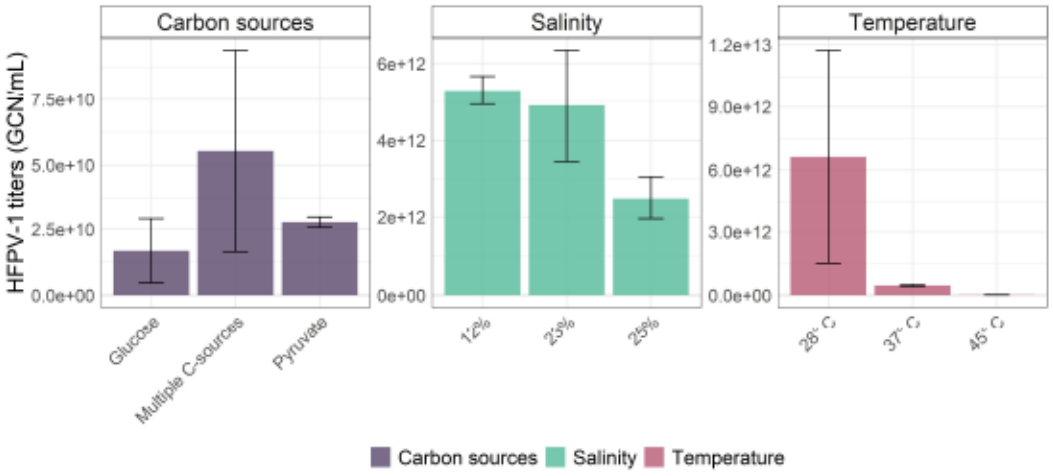


Figure 1 Viral titers under different growth conditions. Colored bars represent the viral titer of *H. volcanii* cultures infected with HFPV-1 under different contitions. Titers were measured by qPCR and normalized by the maximum growth rate. Error bars represent the SD for biological triplicates for each treatment. Salinity percentages represent the concentration in weight/volume of NaCl.

Increased temperature has a negative impact on virus fitness

Temperature is one of the environmental factors that has been shown to impact virus-host interactions on multiple levels. Environmental studies have shown that increases in temperature determines the switching from lytic to lysogenic cycles in planktonic microbial communities (Abdulrahman Ashy & Agustí, 2020; Brum et al., 2016). Similarly, global studies have shown that temperature is among the top–ranked the factors that modulate viral population structures (Brum et al., 2015; Sunagawa et al., 2015).

Interestingly, when testing the effect of temperature on HFPV-1, we discovered that increase in temperature negatively correlates with extracellular viral titers (Figure 1). Significant differences in maximum titers were observed between growth at 45 °C and 28 °C (Kruskal–Wallis p-value < 0.001, Dunn's p-adjusted < 0.05), with up to three orders of magnitude difference at the studied temperatures (10^{13} and 10^{10} gcn/mL at 28 ° and 45 °C respectively. Interestingly, there were not significant differences in the intracellular copy numbers of HFPV-1 (Figure 2) which suggests that the virus is replicating at similar rates and maintaining high copy numbers within the host cells (see section "*Viral induced transcriptional reprogramming negatively impacts host motility*") .

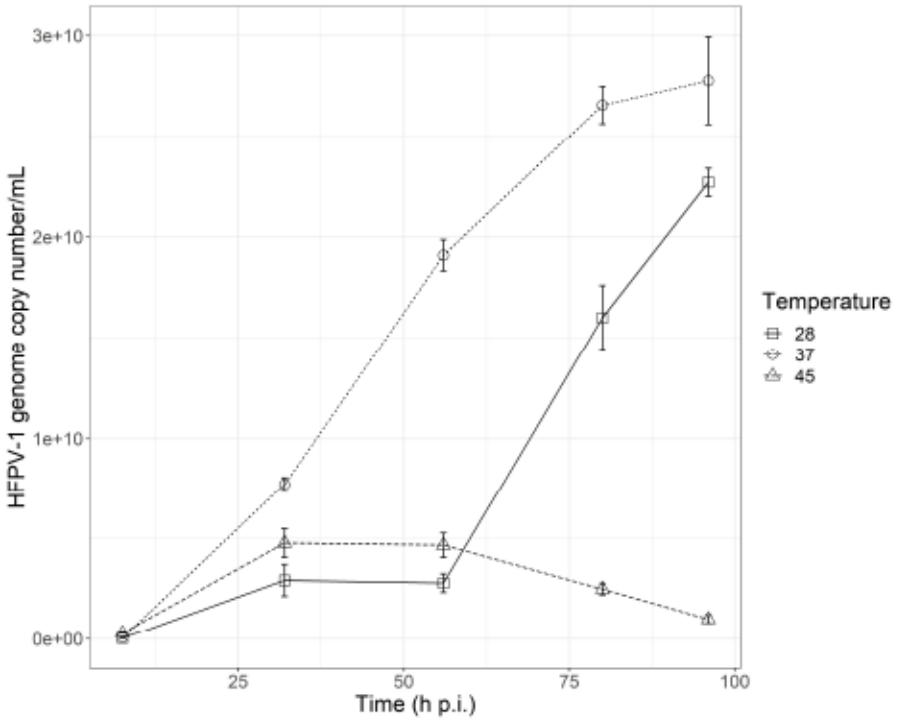


Figure 2 Intracellular genome copy numbers of HFPV-1. Genome copy numbers were meassured by qPCR targetting HFPV-1 genome. Lines depict the temporal changes in intracellular concentrations at 28 °C (squares), 37 °C (circles) and 45 °C (triangles). Error bars represent the SD for biological triplicates for each treatments.

Interestingly, the opposite effect has been observed previously in the bacterium *Listeria Monocytogenes*, with a temperature-dependent resistance to broad-host-range phages that results in phage infection being blocked at lower temperature (30 °C v/s 37 °C) (Kim & Kathariou, 2009). Conversely, in the unicellular eukaryote *Micromonas pusilla*, temperature increase shortened the latent periods (up to 50%), and increased the burst size up to 40% (Maat et al., 2017). Furthermore, moderate increases in temperature form 30° to 37°C have been reported to induce proviruses and modulate the lysis-lysogeny decision of phages of *Burkholderia* (Chu et al., 2011; Shan et al., 2014). Given the negative correlation and significant decrease in viral titers at higher temperatures, we further investigate the effect of temperature variations on the phenotype of these chronically infected populations at the transcriptional level.

Infection with HFPV-1 inhibits cell shape transition in H. volcanii at high temperatures

Upon infection, viruses can have different impacts on host physiology. During lytic infections, these impacts are usually temporary, and in lysogenic cycles, the gene expression of viral genes is at rather low levels with minor impacts on host physiology. However, chronic infecting viruses have a great and long lasting effect on the host metabolic program and cell cycle. It has previously been observed that the spindle-shaped SSV9 virus can arrest the cell division machinery and induce dormancy and severe growth retardation (Bautista et al., 2015). In *Sacharolobus islandicus*, the establishment of an infection by the chronic spindle-shaped virus

STSV2 results in cell gigantism and asymmetric cell division, mainly by transcriptional repression of the cell division machinery (Liu et al., 2021).

We have previously reported that *H. volcanii* cultures infected with HFPV-1 are morphologically undistinguishable from uninfected ones under optimal temperature for viral production at 28 °C (Alarcón-Schumacher et al., 2022). No significant differences were observed regarding the motility of host cells or structural features such as cell shape and size, or presence and number of archaealla. However, microscopic analysis of single cells grown in microfluidic chambers at 45 °C showed significant differences to cells grown at 28 °C upon infection with HFPV-1 (Figure 3, Supplementary figure 2). HFPV-1 infected cultures growing at 28 °C showed no difference to uninfected ones in morphotypes abundance (rod, cocci/disc and triangular-shaped), with rod-shaped cells being the dominant morphotypes during late exponential phase (up to 70% of total cell, at 110 hours). Conversely, at 45 °C little to none rod-shaped *H. volcanii* cells were detected in HFPV-1 infected cultures (up to 3% of total cells, at 24h), while uninfected cultures exhibited up 20% of the cells in the rod-motile shape during the same growth stage (Figure 3).

In haloarchaea, cell division and the morphology of cells was shown to be regulated by tubulin-like proteins, with *H. volcanii* encoding eight different families, 6 of which are CetZ-like and two FtsZ-like proteins (Hartman et al., 2010). In *H. volcanii,* the FtsZ proteins (FtsZ1 and 2) are regulating cell division, by forming the constriction ring and driving the separation of daughter cells (Liao et al., 2021). On the other hand, CetZ-like proteins are not essential for division, however, particularly the CetZ1gene (HVO_2204) was shown to play a crucial role in controlling cell shape, being required for rod cell shape development (Duggin et al., 2015). In the absence of CetZ1, none-rod morphologies have been shown to occur, with cocci shaped cells dominating the population, and seldom triangular or non-defined morphologies have been reported.

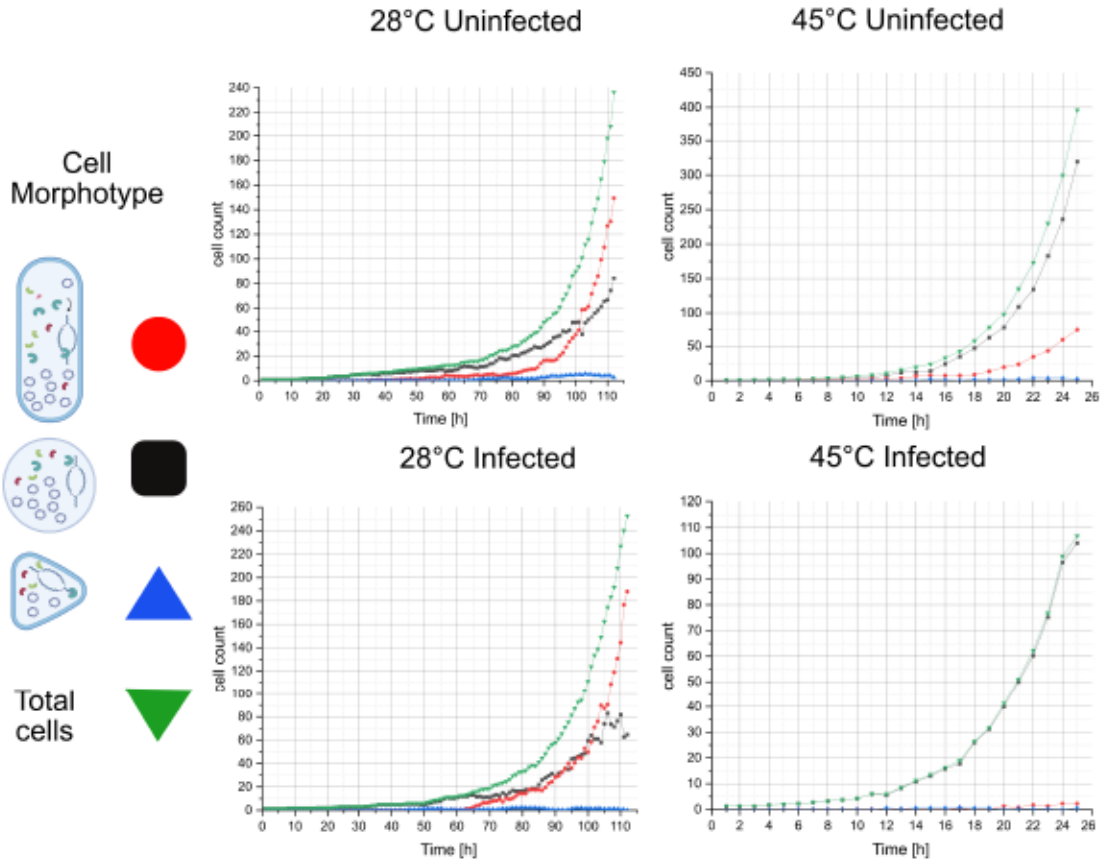


Figure 3 Cell morphotypes enumeration. Growth single cell cultures of *H. volcanii* in microfluidic chambers. Different observed morphotypes as highlighted in the left panel are rod-shaped (red circles), cocci/disc-shaped (black squares), triangular-shaped (blue triangles) and total cell counts (green inverted triangles). Points represent the average count for each cell morphotypes in biological triplicates for both treatments.

Based on the absence of rod-shaped phenotypes at 45 °C, we hypothesized, that at higher temperatures, the host reprogramming induced by the viral infection could result in transcriptional repression of the CetZ1 gene that regulates shape-transition.

CetZ1 was shown to be located at the host membrane, and cell shape has a great influence on the amount of available host cell surface for budding. Therefore, we assessed whether a CetZ1 knockout would have an impact on virus particle production. Surprisingly, the deletion of CetZ1, and thereby the loss of the ability to transition to the motile rod-shape, did not result in significant differences neither in viral titers in the supernatant, nor in the intracellular copy number of HFPV-1 when compared to the parental strain (Supplementary figure 3). This suggests that the budding efficeincy of HFPV-1 virions is not cell shape dependent, and that the lower titers observed at higher temperatures are caused by a different molecular or physical mechanism.

Viral induced transcriptional reprogramming negatively impacts host motility

The analysis of the transcriptional data of HFPV-1 infected cultures of *H. volcanii* revealed that, unlike the observed lower titers, transcription of viral genes occurrs at high rates at 45 °C (Supplementary table 3). Notably, the proportion of the total mRNA pool occupied by HFPV-1 transcripts was larger than at the optimal temperature for viral particle production (28 °C). During exponential phase, at 18 and 48 hours post-infection, HFPV-1 transcripts represent ~ 15% and 19% of total mRNA respectively (SD = 0.287 and 0.884), while at 28 °C this fraction reaches up to ~2% and 7% at the same timepoints. Even more, the largest proportion of transcripts reported at 28 °C occurs during stationary phase, and comprises up to ~16% of the total mRNA, which is at least comparable, if not lower than the values observed at 45 °C. The high proportion of viral mRNA suggests that at least oexpression of HFPV-1 genes is not limited at 45 °C, which goes in line with previous evidence of high copy numbers of the viral genome in the cellular fraction.

Interestingly, host transcriptional changes under HFPV-1 infection at 45 °C revealed a large number of differentially expressed (DE) genes, with up to 1465 genes beeing impacted in a growth-phase dependent manner (Supplementary table 4). Thus, we compared independently the transcriptional changes at each stage of infection between infected and uninfected controls, namely 18 hours, 48 hours and 90 hours post infection -h p.i.-, corresponding to early exponential, late exponential and stationary, respectively. During early exponential growth (18 h p.i.) we observed that up to 91 genes were downregulated (Supplementary table 4). Interestingly, the majority of highly downregulated genes were related to cell motility and signal transduction, with two archaellins genes (HVO_1210 and HVO_1211 genes) being the most repressed genes (-13.5 and -12 fold change, respectively). These genes encode for the protein subunits which polymerizes to form the filaments of archaeal flagella (Jarrell & Albers, 2012). Meanwhile, several genes that connect the chemotaxis systems and the archaellum were also strongly downregulated (-4.9, -4.3, -3.2, and -2.8 fold change, for ArlCE, ArlD, ArlF and ArlG respectively). Furthermore, two cluster of chemotaxis related genes were also repressed in infected cultures. The first cluster include the central regulator of chemotaxis CheA (HVO_1223), adaptor protein CheW (HVO_1225), methylesterase CheB (HVO_1224) and a CheR-like methyltransferase (HVO_1222); and the second one with genes CheY response regulator (HVO_1207), phosphatase CheC (HVO_1206) and deamidase CheD (HVO_1205) (Hartman et al., 2010). While a low levels of repression of archaellum-related genes was also observed at 28 °C (Alarcón-Schumacher et al., 2022), it did not affected the development of flagella. However, as observed during microscopy analysis at 45 °C (Figure 3), *H. volcanii* infected cells do not carry out the shape transition to rod-motile morphology, which correlates with the observed downregulation patterns. Meanwhile, CetZ1 expression levels exhibited a puzzling pattern. CetZ1 was shown to be downregulated during early exponential phase (-2 fold change), correlating with the absence of rod-shaped cells. Conversely, CetZ1 was overexpressed during late exponential phase in infected cultures (3.1-fold change). The accumulation of CetZ1 protein on the host membrane could reduce the surface available

for viral proteins or disrupt their organization, indirectly interfering with the budding process. This could provide an explanation for the lower viral titer in the supernatant, however further experimental evidence on this is required.

Interestingly, the filamentous phage SW1 that establishes a chronic infection in *Shewanella piezotolerans* has also been shown to regulate the lateral flagella movement in a temperature-dependent fashion, with viral infection inducing an upregulation of the flagella apparatus at lower temperatures (Jian et al., 2013). The fact that *H. volcanii* infected cells are arrested in non-motile states has vast implications in terms of nutrient acquisition and environmental response, and thus, it might significantly influence host fitness and ecological interactions within microbial communities.

Intriguingly, a gene encoding for a CopG domain protein (HVO_0381) is highly upregulated in infected cells at 45 °C (8.6 fold change) but not at 28 °C. CopG related proteins are involved in copy number regulation of plasmids (Robinson & Bell, 2007; H. Wang et al., 2015), with mutations of these genes leading to uncontrolled increase in copy numbers numbers of rolling circle-replicating plasmids by several fold (Rasooly & Rasooly, 1997). Related members of the *Alphapleolipovirus* genus replicate using a homologue of the rolling-circle replication endonuclease (RCRE), but HFPV-1 does not encode orthologues to any known RCRE or other replication protein, and the mechanism for genome replication has not yet been elucidated (Alarcón-Schumacher et al., 2022). Thus, the overexpression of this gene could be an unspecific response to high copy numbers of episomal DNA. Overexpression of this protein could possibly regulate the replication of the viral genome, being responsible for the reduction of viral titers at 45 °C when compared to lower temperatures. Furthermore, several proteins of unknown function, as well as, proteins with a predicted helix-turn-helix (HTH) domain that likely bind DNA or act as transcriptional regulators, were observe to be highly upregulated in infected cultures at 45 °C (i.e. HVO_2701, HVO_2065, HVO_A0394, HVO_D0005) but not differentially expressed at 28 °C. However, their involvement on gene regulation or virus takeover remains enigmatic.

Temperature-induced protein misfolding and aggregation of viral proteins

The late exponential phase displayed the highest numbers of significantly DE genes, with 391 genes being downregulated and 853 being overexpressed. Interestingly, the expression pattern was dramatically different to the changes previously observed during the exponential phase at the optimal temperature for viral replication (28 °C), were downregulation of host genes was primarily observed (Alarcón-Schumacher et al., 2022). While the CRISPR system of *H. volcanii* (Maier et al., 2015), as well as the cluster of genes encoding for the broad-anti-phage DISARM defense system (Ofir et al., 2018) were strongly downregulated under infection at 28 °C (Alarcón-Schumacher et al., 2022), no differential expression of any defense system was observed under infecton at 45 °C. Furthermore, no evidence of acquisition of spacers against the viral genome was detected, suggesting that these specific and broad spectrum defense sytems remain largely innefective

against infection with HFPV-1. Conversely, HFPV-1 induces a similar pattern of downregulation of the provirus region Halfvol1 (HVO_0258 to HVO_0280) (Dyall-Smith et al., 2021) to the one observed at 28 °C (Supplementary table 4). However, while at 28 °C these genes are among the top-ranked dowreglated genes with up to -60 fold change , the impact is less severe at 45 °C, with fold changes from -2 to -2.5 lower in the infected cultures, This implicates that the mechanism encoded by the provirus that possibly interfers with other incoming viruses such as HFPV-1 is less downregulated by HFVP-1 at higher temperatures, and therefore, could also be responsible for reduced copy numbers of HFPV-1 at 45 °C.

Intriguingly, one gene of this proviral region (HVO_266) that shows homology to the HicA component of a bacterial toxin-antitoxin (TA) system was upregulated at 45 °C 48 h p.i. (4-fold change) but not at 28 °C. HicA is the toxin component of the TA system and purified HicA was shown to be able to degrade ribosomal RNA in vitro (Thomet et al., 2019). Toxin-antitoxin systems are often present in bacterial and archaeal genomes and they can be used as a defense mechanism. Degradation or downregulation of the antitoxin component upon viral infection can trigger an abortive infection leading to cell dormancy or cell death (Hampton et al., 2020). Considering the fact that there is a highly specific downregulation of this provirus region upon HFPV-1 infection at both 28 °C and 45 °C, we hypothesized that this particular toxin upregulation could be a viral induced effect, and that toxin production by infected cells could play a role in the interaction between infected and uninfected cells and in a boader ecological context.

While the vast majority of DE genes upregulated upon viral infection at 45 °C, we could not identify a specific virus-associated response. The highest levels of upregulation was observed for a gene encoding for a protein of unknown function (HVO_A0172, 28-fold change), and universal stress proteins (HVO_0401 and HVO_2337, with 26 and 13-fold change respectively). Additionally, among the the highest overexpressed gens were multiple proteins associated with the functional categories of translation, transcription and post-transcriptional modifications (Figure 5). Upregulation of universal stress proteins has been observed earlier in virus infected cells (Mercier et al., 2022), and could be a general repsonse to virus infection. Upregulation of translation, transcription and energy conversion metabolisms is likely driven by the virus to ensure production of virus proteins and particles.

Notably, genes related to the proteolityc and protein-folding pathway were highly overexpressed, including homologues for the three subunits of the thermosome (HVO_0133, HVO_0455 and HVO_0778). In *H. volcanii*, the three chaperonines homologues that form the thermosome complex are functionally redundant, because each gene is individually dispensable an participate in the refolding of proteins (Kapatai et al., 2006).

Furthermore, several AAA-type ATPases related to the sructure of the proteasome (CDC48 subfamily homologues HVO_1327, HVO_2380 and HVO_2700) as well as the alpha and beta subunits of the

proteasome (HVO_1091 and HVO_1562) are highly upregulated. These genes are infvolved in degradation of proteins in archaeal cells in an ATP-dependent manner (Kaczowka & Maupin-Furlow, 2003; Reuter & Maupin-Furlow, 2004). In the proteasome of *Thermoplasmata acidophilum,* one of the best studied systems, the 20S core peptidase, a barrel-like structure composed of four heptameric rings of alpha and beta subunits combines with the CDC48 AAA-type ATPase, forming a complex that can bind non-native proteins and facilitate their degradation (Barthelme et al., 2014; Nitsch et al., 1997).

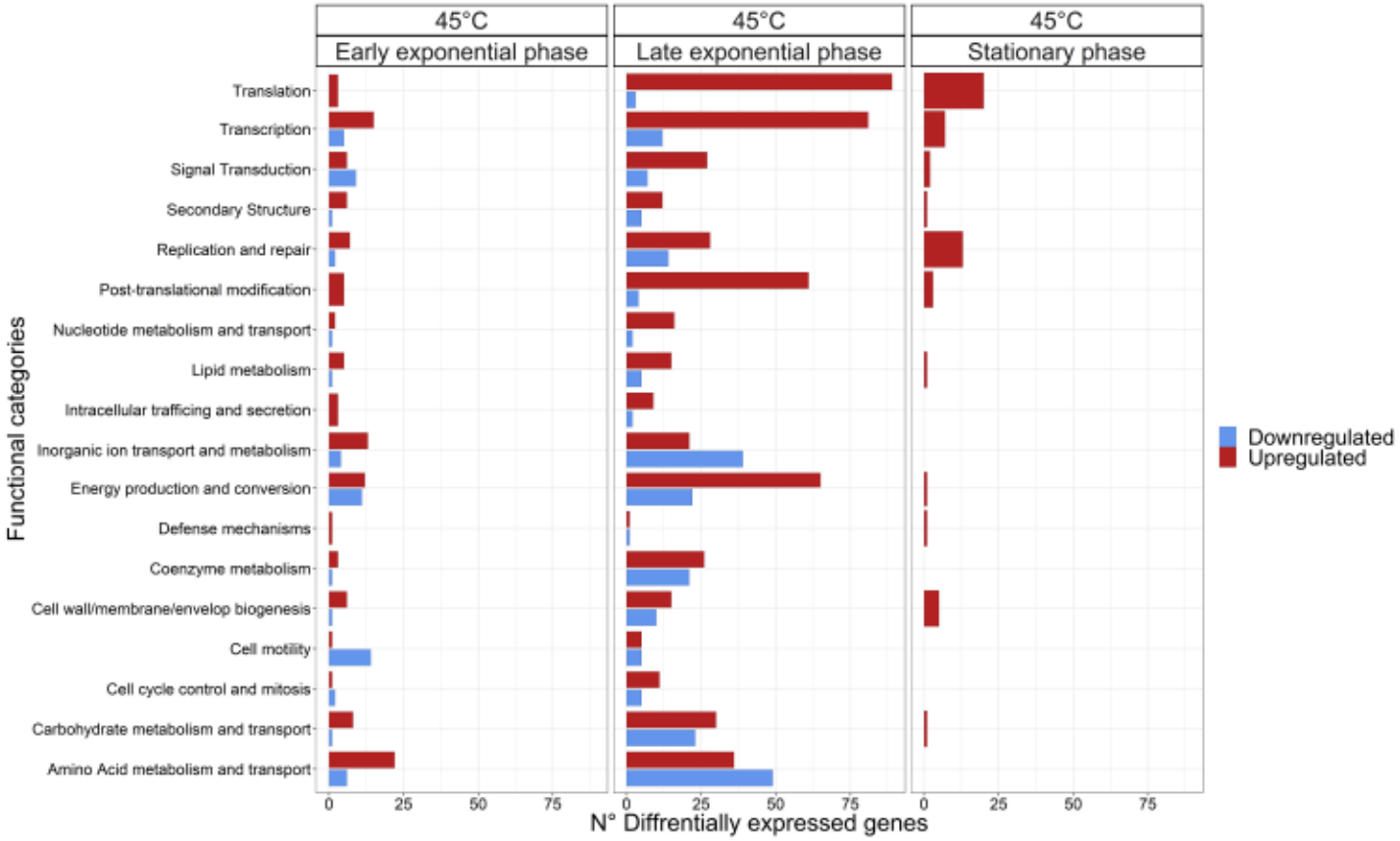


Figure 4 Functional profile of differentially expressed genes. Bars represent the number of differentially expressed genes assigned to each particular functional category at a given growth stage, namely, early exponential, late exponential, and stationary phases. Functional classification of *H. volcanii* genome was performed using the COG database. The colors of the bars depic genes up- (red) or downregulated (blue).

Additionally, homologues of the stress-induced DnaJ/DnaK and GrpE system are highly upregulated upon infection with HFPV-1 at 45 °C. These genes participate in the reponse to osmotic and temperature-related stress, and are chaperonines capable of repairind stress-induced damage or misfolding (Schroder et al., 1993), further suggesting that considerable protein misfolding is ocurring in infected cultures. Since most of the protein degradation and folding machineary are highly overexpressed only upon infection at 45 °C and viral genes are among the top-ranked in overall expression levels, we hypothesized that a considerable fraction of translated viral proteins are less stable at 45 °C, or more prone to misfolding. Missfolded virus proteins could trigger this response of the host proteolytic machinery. Nevertheless we have previously shown that viral

particles remain infective after incubation at temperatures up to 50 °C (Alarcón-Schumacher et al., 2022), suggesting that at least the structural proteins are stable at 45 °C, and that misfolded viral proteins are likely cytosolic and accumulate in the intracellular.

HFPV-1 stability depends on post-translational addition of glycans

In order to elucidate the role of glycosylation in HFPV-1 fitness, we measured viral production in two different mutants of the glycosylation pathway in ΔAglB, the oligosaccharyltransferase responsible for addition of the tetrasaccharide to the S-layer; and ΔAglD, the glycosyltransferase involved in the addition of the last sugar (mannose). Particles produced in these mutants were considered as lacking N-glycosylation (ΔAglB) or with defective glycosylation (ΔAglD). Interestingly, both glycosylation deletion mutants showed reduced titers of HFPV-1 in the supernatant when compared to the parental strain, however only the reduction in the ΔAglB deletion strain was statistically significant (Kruskal–Wallis p-value < 0.001, Dunn's p-adjusted < 0.05) (Figure 5). Given the differences in titer, we further performed an adsorption assay with particles produced either in the ΔAglB deletion strain or in the parental strain. Interestingly, the assay showed high variability of particle adsorption at different time points with some being statistically different between non-glycosylated and native virions. However, overall percentage of adsorbed viruses reached similar values for mutant and parental produced virions, two hours post infection. Thus, we concluded that given the variance in general there is no significant effect of the glycosylation on viral adsorption.

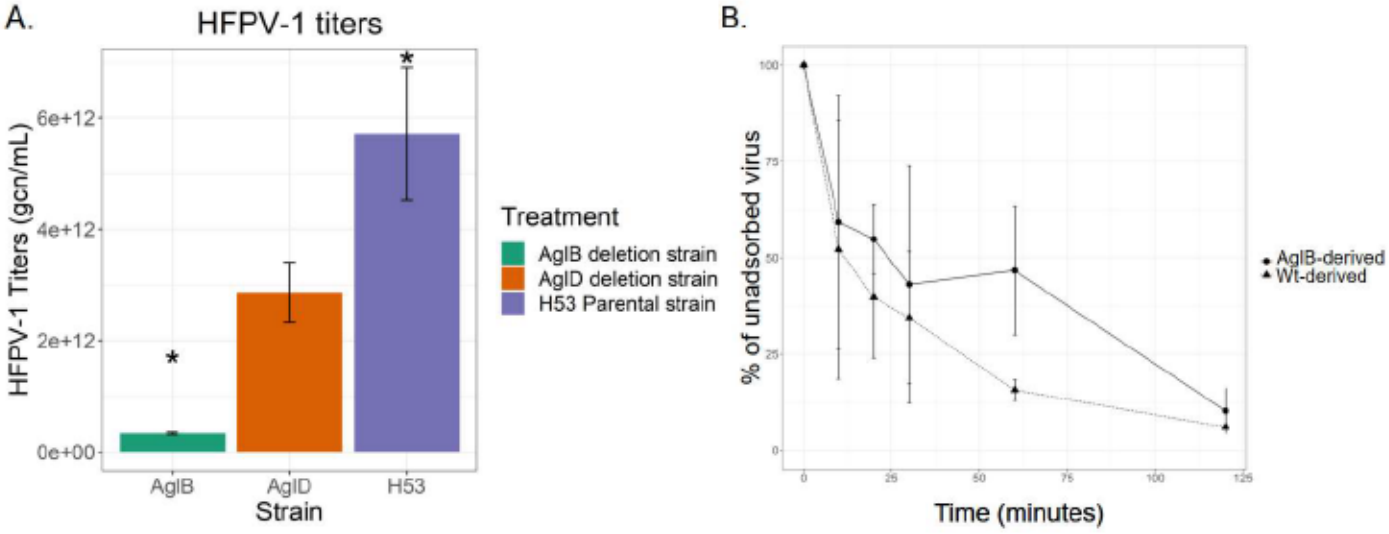


Figure 5 Effect of glycosilation on virus-host interactions. **A.** Viral titers on different glycsylation mutants. Bars represent HFPV-1 titers on *H. volcanii* cultures 90 hours post infection. Titers were measured by qPCR and normalized by the maximum growth rate. Error bars represent the SD for biological triplicates for each treatment. Stars indicate significant differencess between treatments. **B.** Adsorption kinetics of HFPV-1. Viral particles produce in either the parental strain (H53) or the AglB deletion mutant were adsorben in wild type *H. volcanii* cells. Percentage of unadsorbed virus were calculated based on qPCR meassurements of free virus normalized by initial the amount of added virus. Adsorption was performed with independent biological triplicates for each timepoint and error bars represent the SD of such replicates at each timepoint and preparation.

This was further confirmed when testing wether virion particles produced in the ΔAglB and ΔAglD strains are capable of generating a detectable infection on *Haloarcula japonica* (family *Haloarculaceae*), the most distant relative of *H. volcanii* (family *Haloferacaceae*) known to be susceptible to HFPV-1 infection. Particles produced in both strains were able to produce a successful infection in *H. japonica* (Figure supplementary 5), suggesting that HFPV-1 maintained its extremely broad host range despite defective or absence of glycosylation, highlighting that N-glycosylation is does not seem to play a key role in host recognition for HFPV-1. This could prove crucial for future applications of HFPV-1 as a genetic tool, as particles produced in organisms with different glycosylation patters would not lose infectivity or exhibit lower adsorption efficiencies.

Conclusions

In this study, we explored the main variables influencing host and virus fitness during the establishment of a chronic infection of the virus Haloferax pleomorphic virus 1 on the haloarchaeal host *Haloferax volcanii*. HFPV-1 is able to perform successful infections and to propagate in a diverse range of conditions. We showed that, neither salinity (low salt conditions, or saturated media) nor different carbon sources had a significant impact on viral fitness, while infection only had minimal impact on host fitness. Conversely, we uncovered that an increase in temperature to 45 °C negatively correlated with viral fitness a temperature at which the maximum growth rates have been observed for the host. The differences in virus titers at this temperature could not be associated with activation of defense mechanisms of the host, whose expression remains largely unchanged. Nevertheless, infection with HFPV-1 had significant impact in host cell shape at 45 °C, inhibiting the transition to rod-shaped cell during exponential phase and downregulating archaellum and chemotaxis related genes, which likely results in severely hampered motility for the host. The inability of infected cells to transition into the motile rod-shape and to respond to chemotactic signal molecules could represent a major disadvantage for the host when resources become scarce, or upon facing competition in a broader community context. Additionally, we observed that the proteolytic and protein folding and repair machinery is highly upregulated upon infection at 45 °C, and therefore, we suggested that viral proteins are likely not well adapted to higher temperatures.

Lastly, we showed that glycosylation of the spike protein plays a key role for the production of HFPV-1 particles in *H. volcanii*, which is likely due to reduced particle stability as suggested earlier for HFPV-1 in *Halorubrum lacusprofundi* (Gebhard et al., 2023). However, similar to results from *H. lacusprofundi*, there are no significant difference in adsorption rates between N-glycosylated and non-N-glycosylated particles in *H. volcanii*, further suggesting that sugar interaction are not essential for receptor recognition. Most interestingly, non-N-glycosylated particles maintain the unique wide host range of HFPV-1, indicating that the receptor for HFPV-1 must be a protein with high sequence conservation amongst a wide range of

haloarchaea or depend on different forces such as hydrophobic interactions. The versatility of HFPV-1 under different conditions, as well as it independence on post-translational modification such as glycosylation, are increasingly promising traits for the development of HFPV-1 as a genetic tool for haloarchaea without an available genetic system.

Data availability

Raw reads from metagenomic data generated in this study were submitted to ENA-EMBL under project number PRJEB63689.

Acknowledgments

We thank Daniela Thies, Ingrid Kunze, Kathryn Lambert-Slosarska and Margret Menke (MPI for Marine Microbiology, Bremen, Germany) for assistance with some of the experiments. Finally, we want to thank the Max-Planck-Institute for Marine Microbiology and the Max-Planck-Society for continuous support.

References

Abdulrahman Ashy, R., & Agustí, S. (2020). Low Host Abundance and High Temperature Determine Switching from Lytic to Lysogenic Cycles in Planktonic Microbial Communities in a Tropical Sea (Red Sea). *Viruses*, *12*(7). https://doi.org/10.3390/v12070761

Abu-Qarn, M., & Eichler, J. (2006). Protein N-glycosylation in Archaea: Defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. *Molecular Microbiology*, *61*(2), 511–525. https://doi.org/10.1111/j.1365-2958.2006.05252.x

Alarcón-Schumacher, T., Naor, A., Gophna, U., & Erdmann, S. (2022). Isolation of a virus causing a chronic infection in the archaeal model organism *Haloferax volcanii* reveals antiviral activities of a provirus. *Proceedings of the National Academy of Sciences*, *119*(35), 1–12. https://doi.org/https://doi.org/10.1073/pnas.2205037119

Allers, T., Ngo, H., Mevarech, M., & Lloyd, R. G. (2004). Development of Additional Selectable Markers for the Halophilic Archaeon *Haloferax volcanii* Based on the leuB and trpA Genes Thorsten. *Applied and Environmental Microbiology*, *70*(2). https://doi.org/10.1128/AEM.70.2.943

Barthelme, D., Chen, J. Z., Grabenstatter, J., Baker, T. A., & Sauer, R. T. (2014). Architecture and assembly of the archaeal Cdc48·20S proteasome. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17). https://doi.org/10.1073/pnas.1404823111

Bautista, M. A., Zhang, C., & Whitaker, R. J. (2015). Virus-induced dormancy in the archaeon *Sulfolobus islandicus*. *MBio*, *6*(2), 1–8. https://doi.org/10.1128/mBio.02565-14

Bidle, K. A., Hanson, T. E., Howell, K., & Nannen, J. (2007). HMG-CoA reductase is regulated by salinity at the level of transcription in *Haloferax volcanii*. *Extremophiles*, *11*(1), 49–55. https://doi.org/10.1007/s00792-006-0008-3

Bidle, K. A., Kirkland, P. A., Nannen, J. L., & Maupin-Furlow, J. A. (2008). Proteomic analysis of *Haloferax volcanii* reveals salinity- mediated regulation of the stress response protein PspA Kelly. *Microbiology*, *154*(1), 1436–1443. https://doi.org/10.1099/mic.0.2007/015586-0.Proteomic

Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W., & Sullivan, M. B. (2016). Seasonal time bombs: Dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME Journal*, *10*(2), 437–449. https://doi.org/10.1038/ismej.2015.125

Brum, J. R., Ignacio-espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., & Chaffron, S. (2015). Patterns and ecological drivers of ocean viral communities. *Science*, *348*(6237), 1261498-1–11. https://doi.org/10.1126/science.1261498

Chen, Y., Golding, I., Sawai, S., Guo, L., & Cox, E. C. (2005). Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS Biology*, *3*(7), 1276–1282. https://doi.org/10.1371/journal.pbio.0030229

Chu, T. C., Murray, S. R., Hsu, S. F., Vega, Q., & Lee, L. H. (2011). Temperature-induced activation of freshwater Cyanophage AS-1 prophage. *Acta Histochemica*, *113*(3), 294–299. https://doi.org/10.1016/j.acthis.2009.11.003

Cobián Güemes, A. G., Youle, M., Cantú, V. A., Felts, B., Nulton, J., & Rohwer, F. (2016). Viruses as Winners in the Game of Life. *Annual Review of Virology*, *3*(1), 197–214. https://doi.org/10.1146/annurev-virology-100114-054952

Cutler, R. G., & Evans, J. E. (1966). Synchronization of bacteria by a stationary-phase method. *Journal of Bacteriology*, *91*(2), 469–476. https://doi.org/10.1128/jb.91.2.469-476.1966

Dewerff, S. J., Bautista, M. A., Pauly, M., Zhang, C., & Whitaker, R. J. (2020). Killer archaea: Virus-mediated antagonism to CRISPR-immune populations results in emergent virus-host mutualism. *MBio*, *11*(2). https://doi.org/10.1128/MBIO.00404-20

DeWerff, S. J., Zhang, C., Schneider, J., & Whitaker, R. J. (2022). Intraspecific antagonism through viral toxin encoded by chronic Sulfolobus spindle-shaped virus . *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1842). https://doi.org/10.1098/rstb.2020.0476

Duggin, I. G., Aylett, C. H. S., Walsh, J. C., Michie, K. A., Wang, Q., Turnbull, L., Dawson, E. M., Harry, E.

J., Whitchurch, C. B., Amos, L. A., & Löwe, J. (2015). CetZ tubulin-like proteins control archaeal cell shape. *Nature*, *519*(7543), 362–365. https://doi.org/10.1038/nature13983

Dyall-Smith, M. (2009). *Halohandbook*. https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf

Dyall-Smith, M., Pfeiffer, F., Chiang, P.-W., & Tang, S.-L. (2021). The Novel Halovirus Hardycor1, and the Presence of Active (Induced) Proviruses in Four Haloarchaea. https://doi.org/10.3390/genes12020149

Eichler, J., & Maupin-Furlow, J. (2013). Post-translation modification in Archaea: Lessons from *Haloferax volcanii* and other haloarchaea. *FEMS Microbiology Reviews*, *37*(4), 583–606. https://doi.org/10.1111/1574-6976.12012

Forterre, P. (2011). Manipulation of cellular syntheses and the nature of viruses: The virocell concept. *Comptes Rendus Chimie*, *14*(4), 392–399. https://doi.org/10.1016/j.crci.2010.06.007

Gebhard, L. J., Vershinin, Z., Alarcón-Schumacher, T., Eichler, J., & Erdmann, S. (2023). Influence of N-Glycosylation on Virus–Host Interactions in *Halorubrum lacusprofundi* L. *Viruses*, *15*(1496). https://doi.org/https://doi.org/10.3390/v15071469

Goehlich, H., Roth, O., & Wendling, C. C. (2019). Filamentous phages reduce bacterial growth in low salinities. *Royal Society Open Science*, *6*(12), 2–9. https://doi.org/10.1098/rsos.191669

Hampton, H. G., Watson, B. N. J., & Fineran, P. C. (2020). The arms race between bacteria and their phage foes. *Nature*, *577*(7790), 327–336. https://doi.org/10.1038/s41586-019-1894-8

Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T. M., Maupin-Furlow, J., Pohlschroder, M., Daniels, C., Pfeiffer, F., Allers, T., & Eisen, J. A. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE*, *5*(3). https://doi.org/10.1371/journal.pone.0009605

Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H. Bin, Andreopoulos, B., Brewer, H. M., Glavina del Rio, T., Adkins, J. N., Paul, S., Sullivan, M. B., & Duhaime, M. B. (2020). Phage-specific metabolic reprogramming of virocells. *ISME Journal*. https://doi.org/10.1038/s41396-019-0580-z

Jarrell, K. F., & Albers, S. V. (2012). The archaellum: An old motility structure with a new name. *Trends in Microbiology*, *20*(7), 307–312. https://doi.org/10.1016/j.tim.2012.04.007

Jayaprakash, N. G., & Surolia, A. (2017). Role of glycosylation in nucleating protein folding and stability.

Biochemical Journal, *474*(14), 2333–2347. https://doi.org/10.1042/BCJ20170111

Jian, H., Xiao, X., & Wanga, F. (2013). Role of filamentous phage SW1 in regulating the lateral flagella of *Shewanella piezotolerans* strain WP3 at low temperatures. *Applied and Environmental Microbiology*, *79*(22), 7101–7109. https://doi.org/10.1128/AEM.01675-13

Kaczowka, S. J., & Maupin-Furlow, J. A. (2003). Subunit topology of two 20S proteasomes from *Haloferax volcanii*. *Journal of Bacteriology*, *185*(1), 165–174. https://doi.org/10.1128/JB.185.1.165-174.2003

Kandiba, L., Aitio, O., Helin, J., Guan, Z., Permi, P., Bamford, D. H., Eichler, J., & Roine, E. (2012). Diversity in prokaryotic glycosylation: An archaeal-derived N-linked glycan contains legionaminic acid. *Molecular Microbiology*, *84*(3), 578–593. https://doi.org/10.1111/j.1365-2958.2012.08045.x

Kapatai, G., Large, A., Benesch, J. L. P., Robinson, C. V., Carrascosa, J. L., Valpuesta, J. M., Gowrinathan, P., & Lund, P. A. (2006). All three chaperonin genes in the archaeon *Haloferax volcanii* are individually dispensable. *Molecular Microbiology*, *61*(6), 1583–1597. https://doi.org/10.1111/j.1365-2958.2006.05324.x

Kauri, T., Wallace, R., & Kushner, D. J. (1990). Nutrition of the Halophilic Archaebacterium, *Haloferax volcanii*. *Systematic and Applied Microbiology*, *13*(1), 14–18. https://doi.org/10.1016/S0723-2020(11)80174-8

Kim, J. W., & Kathariou, S. (2009). Temperature-dependent phage resistance of Listeria monocytogenes epidemic clone II. *Applied and Environmental Microbiology*, *75*(8), 2433–2438. https://doi.org/10.1128/AEM.02480-08

Kuprat, T., Johnsen, U., Ortjohann, M., & Schönheit, P. (2020). Acetate Metabolism in Archaea: Characterization of an Acetate Transporter and of Enzymes Involved in Acetate Activation and Gluconeogenesis in *Haloferax volcanii*. *Frontiers in Microbiology*, *11*(December), 1–16. https://doi.org/10.3389/fmicb.2020.604926

Larson, E. T., Reiter, D., Young, M., & Lawrence, C. M. (2006). Structure of A197 from Sulfolobus Turreted Icosahedral Virus: a Crenarchaeal Viral Glycosyltransferase Exhibiting the GT-A Fold . *Journal of Virology*, *80*(15), 7636–7644. https://doi.org/10.1128/jvi.00567-06

Liao, Y., Ithurbide, S., Evenhuis, C., Löwe, J., & Duggin, I. G. (2021). Cell division in the archaeon *Haloferax volcanii* relies on two FtsZ proteins with distinct functions in division ring assembly and constriction. *Nature Microbiology*, *6*(5), 594–605. https://doi.org/10.1038/s41564-021-00894-z

Liu, J., Cvirkaite-Krupovic, V., Baquero, D. P., Yang, Y., Zhang, Q., Shen, Y., & Krupovic, M. (2021). Virus-

induced cell gigantism and asymmetric cell division in archaea. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(15), 1–10. https://doi.org/10.1073/pnas.2022578118

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Maat, D. S., Biggs, T., Evans, C., van Bleijswijk, J. D. L., van Der Wel, N. N., Dutilh, B. E., & Brussaard, C. P. D. (2017). Characterization and temperature dependence of arctic micromonas polaris viruses. *Viruses*, *9*(6), 6–9. https://doi.org/10.3390/v9060134

Maier, L. K., Dyall-Smith, M., & Marchfelder, A. (2015). The adaptive immune system of Haloferax volcanii. *Life*, *5*(1), 521–537. https://doi.org/10.3390/life5010521

Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads kenkyuhi hojokin gan rinsho kenkyu jigyo. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Mercier, C., Thies, D., Zhong, L., Raftery, M. J., Cavicchioli, R., & Erdmann, S. (2022). In depth characterization of an archaeal virus-host system reveals numerous virus exclusion mechanisms. *BioRxiv*, 2022.10.18.512658. https://www.biorxiv.org/content/10.1101/2022.10.18.512658v2%0Ahttps://www.biorxiv.org/content/10.1101/2022.10.18.512658v2.abstract

Mullakhanbhai, M. F., & Larsen, H. (1975). Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Archives of Microbiology*, *104*(1), 207–214. https://doi.org/10.1007/BF00447326

Nitsch, M., Klumpp, M., Lupas, A., & Baumeister, W. (1997). The thermosome: Alternating α and β-subunits within the chaperonin of the archaeon *Thermoplasma acidophilum*. *Journal of Molecular Biology*, *267*(1), 142–149. https://doi.org/10.1006/jmbi.1996.0849

Obeng, N., Pratama, A. A., & Elsas, J. D. van. (2016). The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. *Trends in Microbiology*, *24*(6), 440–449. https://doi.org/10.1016/j.tim.2015.12.009

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., & Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature Microbiology*, *3*(1), 90–98. https://doi.org/10.1038/s41564-017-0051-0

Olm, M. R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B. A., Morowitz, M. J., & Banfield, J. F. (2021). inStrain profiles population microdiversity from metagenomic data and sensitively detects shared

microbial strains. *Nature Biotechnology*. https://doi.org/10.1038/s41587-020-00797-0

Pietilä, M. K., Roine, E., Paulin, L., Kalkkinen, N., & Bamford, D. H. (2009). An ssDNA virus infecting archaea: A new lineage of viruses with a membrane envelope. *Molecular Microbiology*, *72*(2), 307–319. https://doi.org/10.1111/j.1365-2958.2009.06642.x

Quemin, E. R. J., Pietilä, M. K., Oksanen, H. M., Forterre, P., Rijpstra, W. I. C., Schouten, S., Bamford, D. H., Prangishvili, D., & Krupovic, M. (2015). Sulfolobus Spindle-Shaped Virus 1 Contains Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids. *Journal of Virology*, *89*(22), 11681–11691. https://doi.org/10.1128/jvi.02270-15

Rasooly, A., & Rasooly, R. S. (1997). How rolling circle plasmids control their copy number. *Trends in Microbiology*, *5*(11), 440–446. https://doi.org/10.1016/S0966-842X(97)01143

Reuter, C. J., & Maupin-Furlow, J. A. (2004). Analysis of proteasome-dependent proteolysis in *Haloferax volcanii* cells, using short-lived green fluorescent proteins. *Applied and Environmental Microbiology*, *70*(12), 7530–7538. https://doi.org/10.1128/AEM.70.12.7530-7538.2004

Robinson, N. P., & Bell, S. D. (2007). Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(14), 5806–5811. https://doi.org/10.1073/pnas.0700206104

Rosenwasser, S., Ziv, C., Creveld, S. G. van, & Vardi, A. (2016). Virocell Metabolism: Metabolic Innovations During Host–Virus Interactions in the Ocean. *Trends in Microbiology*, *24*(10), 821–832. https://doi.org/10.1016/j.tim.2016.06.006

Schroder, H., Langer, T., Hartl, F., & Bukaul, B. (1993). DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *EMBO Journal*, *12*(1), 4137–4144.

Shan, J., Korbsrisate, S., Withatanung, P., Adler, N. L., Clokie, M. R. J., & Galyov, E. E. (2014). Temperature dependent bacteriophages of a tropical bacterial pathogen. *Frontiers in Microbiology*, *5*(NOV), 1–7. https://doi.org/10.3389/fmicb.2014.00599

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., D'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., … Bork, P. (2015). Structure and function of the global ocean microbiome. *Science (New York, N.Y.)*, *348*(6237), 1261359. https://doi.org/10.1126/science.1261359

Thomet, M., Trautwetter, A., Ermel, G., & Blanco, C. (2019). Characterization of HicAB toxin-antitoxin

module of *Sinorhizobium meliloti*. *BMC Microbiology*, *19*(1), 1–12. https://doi.org/10.1186/s12866-018-1382-6

Wang, F., Liu, Y., Su, Z., Osinski, T., de Oliveira, G. A. P., Conway, J. F., Schouten, S., Krupovic, M., Prangishvili, D., & Egelman, E. H. (2019). A packing for A-form DNA in an icosahedral virus. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(45), 22591–22597. https://doi.org/10.1073/pnas.1908242116

Wang, H., Peng, N., Shah, S. A., Huang, L., & She, Q. (2015). Archaeal Extrachromosomal Genetic Elements. *Microbiology and Molecular Biology Reviews*, *79*(1), 117–152. https://doi.org/10.1128/mmbr.00042-14

Weitz, J. S., Li, G., Gulbudak, H., Cortez, M. H., & Whitaker, R. J. (2019). Viral invasion fitness across a continuum from lysis to latency†. *Virus Evolution*, *5*(1), 1–9. https://doi.org/10.1093/ve/vez006

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

Understanding the impacts of chronic viral infections in host fitness and ecology

Tomas Alarcón-Schumacher[1], Fabian Roeloffs[2,3], Julian Schmitz[2,3], Alexander Grünberger[2,3] and Susanne Erdmann[1]

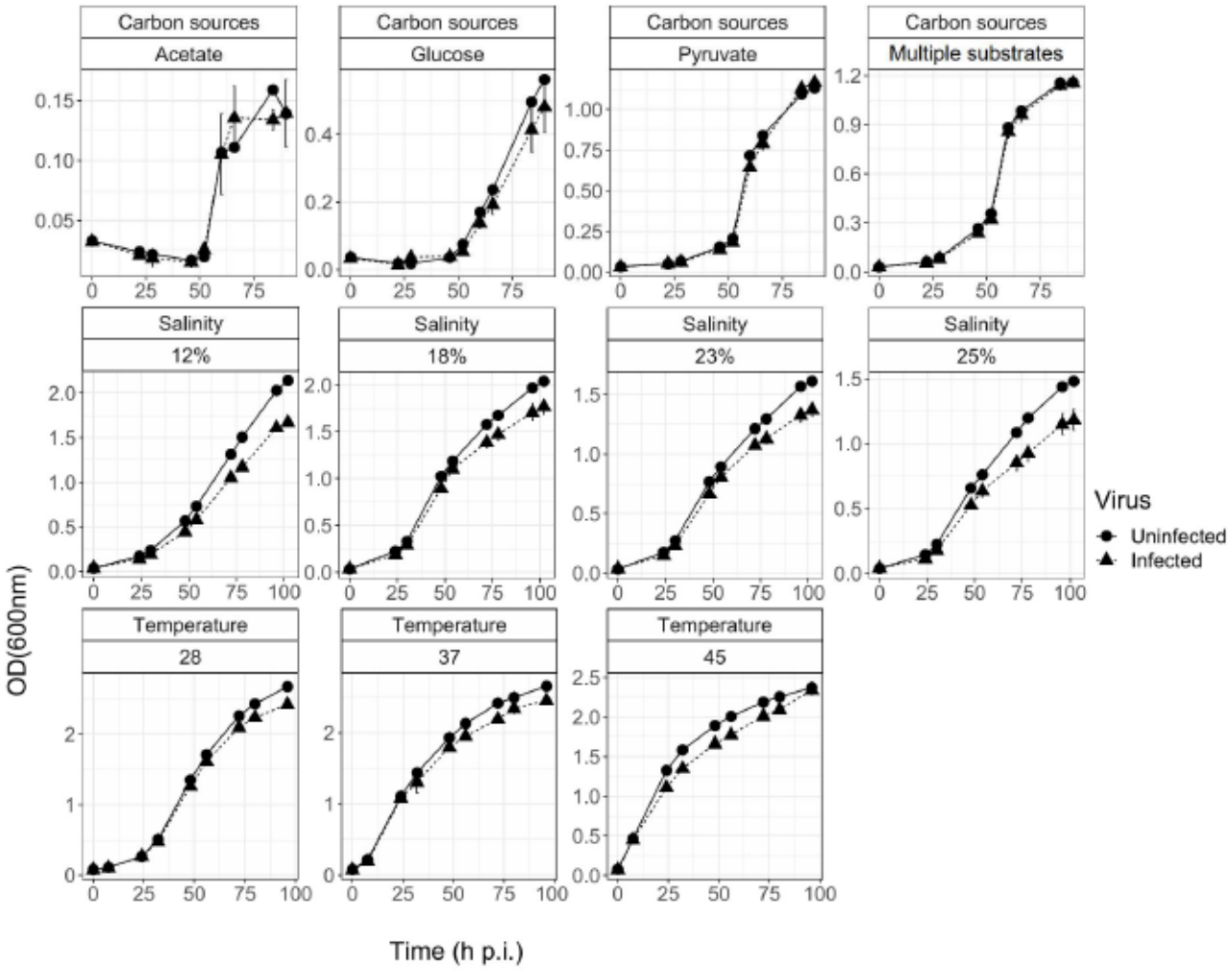
Supplementary Information

Supplementary Figures

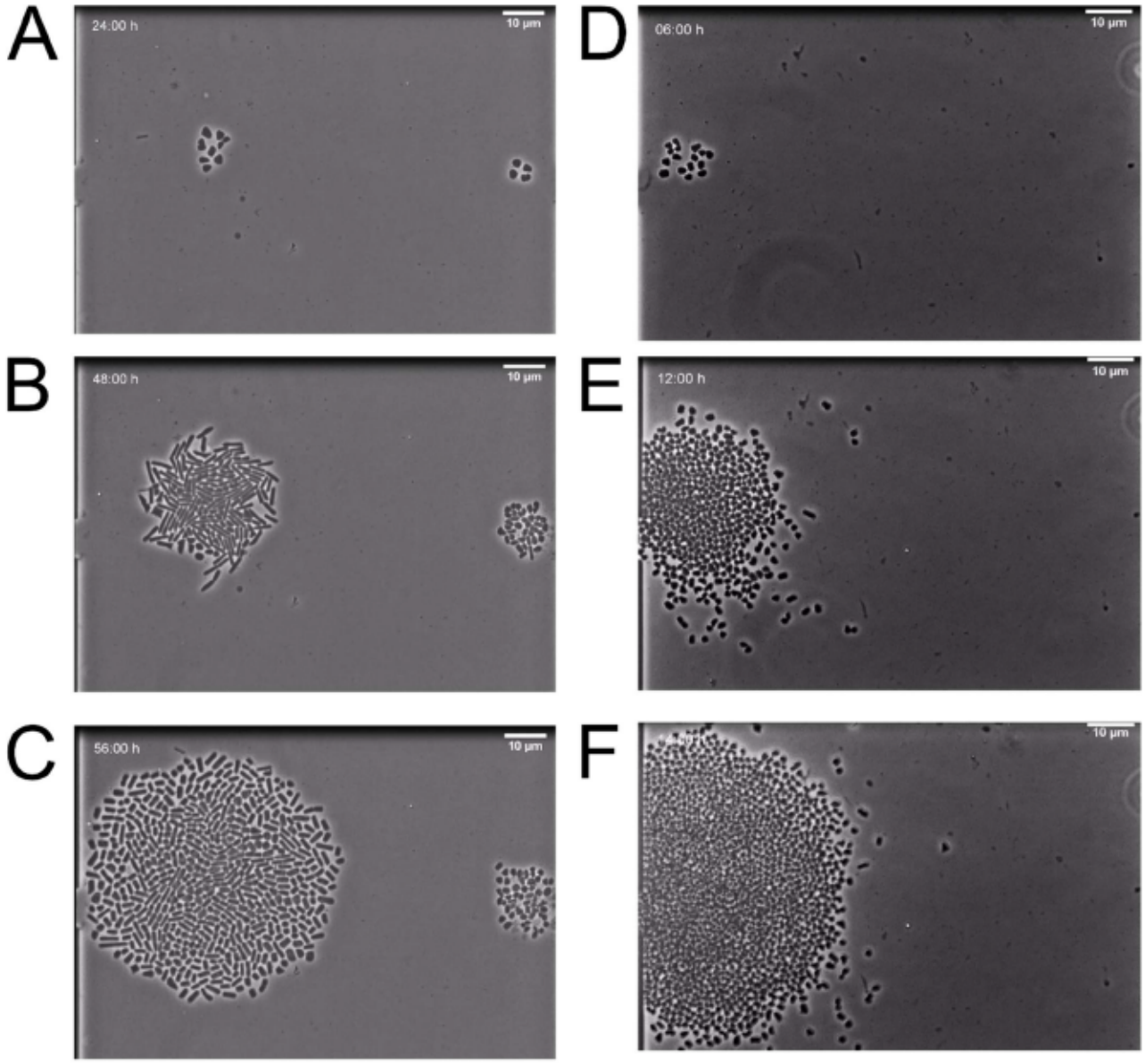
Supplementary Tables

Supplementary References

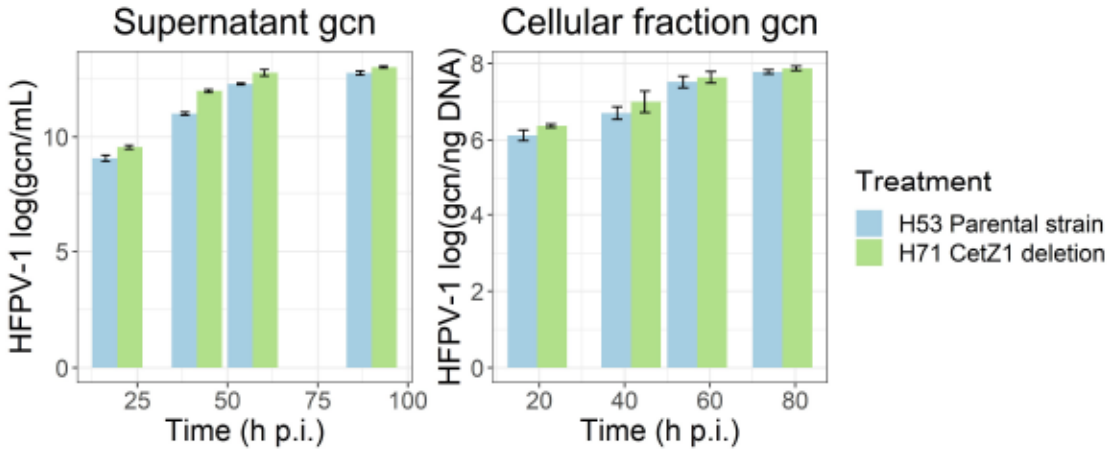
Supplementary Figures



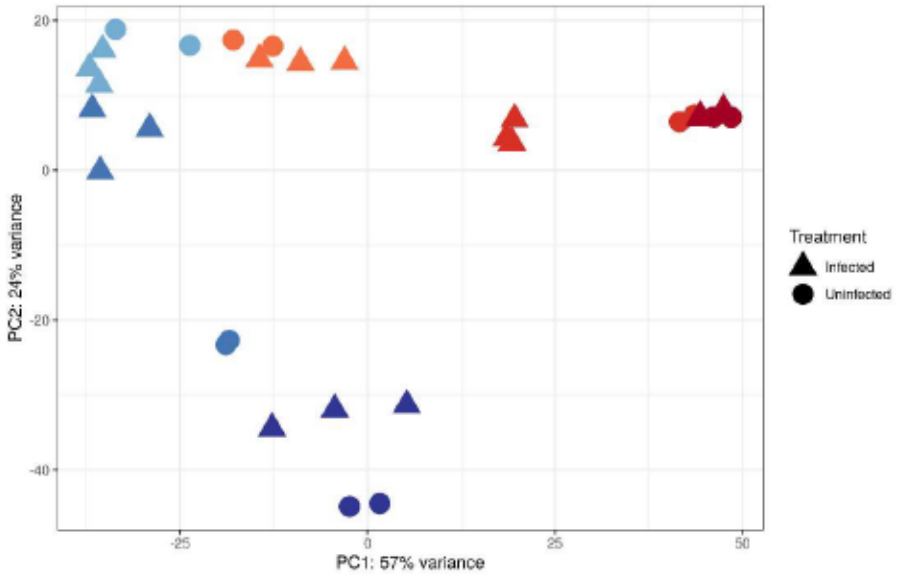
Supplementary figure 1: Life cycle of HFPV-1 under different growth conditions. Growth curve of uninfected (black circles) and infected (black triangles) cultures of *H. volcanii*. Error bars represent the standard deviation for biological triplicates on both treatments. h p.i.: hours post infection.



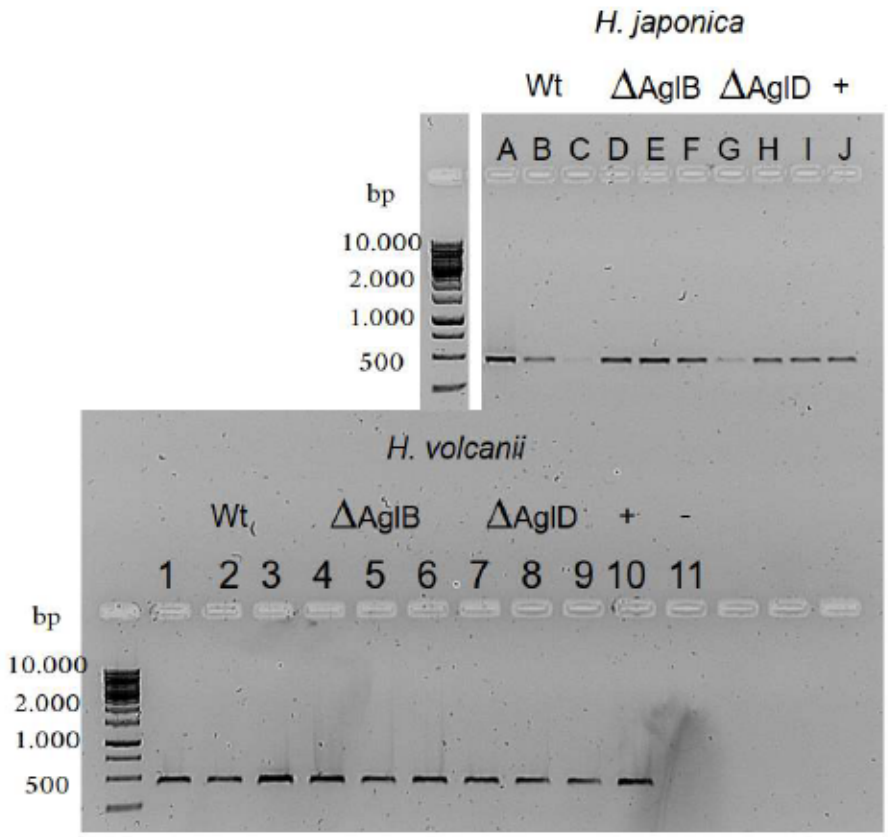
Supplementary figure 2: Cell shape transition upon infection with HFPV-1. Light microscopy of HFPV-1 infeted cultures of *H. volcanii* grown at different temperatures, 28 °C (A - C) and 45 °C (D - F). Time scale (left corner) signals different timepoints registered for each phase of cell growth: lag phase (A, 24h; D, 6h), exponential phase (B, 48h and E, 12h) and stationary phase (C, 56h; F, 14h). Scale bars represent 10µm (right corner)



Supplementary figure 3 Effect of Cetz1 deletion on HFPV-1 fitness. HFPV-1 Genome copy numbers were meassured by qPCR targetting (see methodsBars depict the genome copy numbers (gcn) detected in the supernatant (left panel) and in the cellular fraction (right panel). Error bars represent the SD for biological triplicates for each treatments.



Supplementary figure 4 Overall comparison of transcriptional programs. Principal component analysis of the transcriptional profiles of infected (triangles) an uninfected (circles) cultures of *H. volcanii).*Different shades of each color represent a different growth stage at a different temperatures. In blue, (from light to dark respectively); and in red, early exponentil, late exponential and stationary phase (from orange to dark red). Axis depict the percentage of the variance explained by each of the princial components.



Supplementary figure 5 Host range assessment of glycosylation defective HFPV-1 virions. PCR assay to confirm ongoing infections of HFPV-1 virus in *H. japonica* (A-I) and *H. volcanii* (1-9) cultures (see *Experimental procedures* for details). Molecular weight size marker (L) is shown (Gene ruler 1kb DNA ladder Thermo Fisher Scientific). Infections was assessed for three biological replicates. H26 producet particles (WT), glycosylation deficient particles (ΔAglB), glycosilation deffective particles (ΔAglD), (+) purfied HFPV-1 DNA (-) water PCR negative control. DNA was separated on 1% agarose gels and stained with SYBR™ Safe (Invitrogen™) at a final concentration 1x.

Supplementary Tables

Supplementary Table 1: *H. volcanii* strains. Supplementary Table 1 summarizes the different strains of *H. volcanii* used in this study for cell-shape studies as well as, glycosylation deficient strains.

Strain	*Deletions*
H. volcanii DS2	None
H26	Δ*PyrE2* (Allers et al. 2004)
H53	Δ*PyrE2,* Δ*trpA* (Allers et al. 2004)
H71	Δ*pyrE2,* Δ*trpA and* Δ*CetZ1* (Duggin et al. 2015)
AglB	Δ*pyrE2,* Δ*trpA and aglB* (Abu-Qarn and Eichler 2006)
AglD	Δ*pyrE2,* Δ*trpA and aglD* (Abu-Qarn and Eichler 2006)

Supplementary Table 2: Experimental growth conditions for *H. volcanii* and HFPV-1 virus. Supplementary Table 1 summarizes the different modification to media and culture condition used for infected and uninfected cultures of *H. volcanii*. Salinity modifications to the media were performed only modifying the NaCl concentration, while other ions were not modified. Optimal conditions for HFPV-1 replication are indicated between brackets under "standard conditions"

Factor	*Treatments*
Temperature	28 °C (standard conditions)
	37 °C
	45 °C
Salinity	12% NaCl
	18% NaCl (standard conditions)
	23% NaCl
	25% NaCl
Carbon sources	Acetate (40 mM)
	Glucose (15mM)
	Pyruvate (0.5%)
	Multiple substrates (Glycerol, succinate, pyruvate)
	Hv-YPC (standard conditions)

Supplementary Table 3: Viral reads in HFPV-1 infected cultures. Supplementary Table 3 summarizes the proportion of total mRNA transcripts mapping to HFPV-1 genome from transcriptomic data of infected *H. volcanii* at 28 °C and 45 °C. Time points represent the different growth stages analyzed in this study (45 °C) and previous data from (Alarcón-Schumacher et al. 2022).

Temperature	*Time*	*% of viral reads*	*SD*
28°C	18 h p.i.	2.187	0.026
	48 h p.i.	7.087	0.736
	90 h p.i.	16.330	2.865
45°C	18 h p.i.	15.136	0.287
	48 h p.i.	19.373	0.884
	90 h p.i.	10.198	0.369

Supplementary Table 4: Differentially expressed genes of H. volcanii upon infection with HFPV-1 at 45 °C. Supplementary Table 4 summarizes the top-ranked differentially expressed genes of *H. volcanii* during the different growth stages analyzed in this study. LFC values are equivalemn to Log_2 fold change values.

Gen_ID	*LFC 18h p.i.*	*LFC 48h p.i.*	*LFC 90h p.i.*	*Product*	*COG_category*
HVO_0105	1.3	2.6	0.0	FAD-dependent oxidoreductase	Function Unknown
HVO_0133	2.4	2.9	0.0	thermosome subunit 1	Post-translational modification
HVO_0363	0.0	-1.8	0.0	uncharacterized protein	Function Unknown
HVO_0401	1.9	4.7	1.4	UspA domain protein	Signal Transduction
HVO_0455	2.7	2.9	0.0	thermosome subunit 2	Post-translational modification
HVO_0471	-2.3	0.0	0.0	uncharacterized protein	Replication and repair
HVO_0481	0.0	4.3	1.5	glyceraldehyde-3-phosphate dehydrogenase (NAD)	Carbohydrate metabolism and transport
HVO_0555	-1.9	0.0	0.0	transducer protein Htr15	Signal Transduction
HVO_0618	-1.2	-1.0	0.0	probable secreted glycoprotein	Cell wall/membrane/envelop biogenesis
HVO_0778	1.8	2.2	0.0	thermosome subunit 3	Post-translational modification
HVO_1052	2.1	1.6	0.0	transcription initiation factor TFB	Transcription
HVO_1150	1.4	2.7	1.5	IS200-type transposase HfIRS11	Replication and repair
HVO_1151	0.0	2.8	1.3	IS1341-type transposase HfIRS11	Function Unknown
HVO_1162	1.1	3.2	0.0	Sec-independent translocase protein	Intracellular trafficking and secretion
HVO_1202	-1.7	0.0	0.0	uncharacterized protein	Function Unknown
HVO_1203	-2.1	0.0	0.0	arl cluster protein ArlD	Cell motility
HVO_1205	-2.6	0.0	0.0	taxis cluster protein CheD	Cell motility
HVO_1206	-2.7	0.0	0.0	taxis cluster protein CheC	Cell motility
HVO_1207	-2.7	0.0	0.0	response regulator CheY	Signal Transduction
HVO_1210	-3.8	0.0	0.0	archaellin A1	Cell motility
HVO_1211	-3.6	0.0	0.0	archaellin A2	Cell motility
HVO_1213	-2.3	0.0	0.0	arl cluster protein ArlCE	Cell motility
HVO_1214	-1.7	0.0	0.0	arl cluster protein ArlF	Cell motility
HVO_1215	-1.5	0.0	0.0	arl cluster protein ArlG	Cell motility
HVO_1222	-1.3	-1.0	0.0	protein-glutamate O-methyltransferase CheR	Cell motility
HVO_1223	-1.7	0.0	0.0	taxis sensor histidine kinase CheA	Signal Transduction
HVO_1224	-2.0	0.0	0.0	protein-glutamate methylesterase /	Cell motility
HVO_1225	-2.6	0.0	0.0	purine-binding taxis protein CheW	Cell motility
HVO_1252	2.7	2.2	0.0	uncharacterized protein	Function Unknown
HVO_1359	1.3	3.2	0.0	small CPxCG-related zinc finger protein	Function Unknown
HVO_1488	0.0	3.7	0.0	D-gluconate dehydratase	Cell wall/membrane/envelop biogenesis
HVO_1652	0.0	3.8	0.0	uncharacterized protein	Function Unknown
HVO_1676	2.3	2.6	1.0	transcription initiation factor TFB	Transcription
HVO_1766	1.7	2.4	0.0	ArsR family transcription regulator	Transcription
HVO_1780	3.0	2.3	1.0	HTH domain protein	Translation
HVO_1902	-3.1	0.0	0.0	uncharacterized protein	Function Unknown
HVO_1992	2.0	2.5	0.0	cold shock protein	Transcription
HVO_1999	-2.7	0.0	0.0	transducer protein Htr7	Cell motility
HVO_2000	-3.1	0.0	0.0	uncharacterized protein	Function Unknown
HVO_2006	2.8	1.3	0.0	probable secreted glycoprotein	Function Unknown
HVO_2031	-1.5	0.0	0.0	ABC-type transport system periplasmic	Function Unknown
HVO_2032	-1.7	0.0	0.0	ABC-type transport, ATP-binding protein	Amino Acid metabolism and transport

HVO_2033	-1.5	0.0	0.0	ABC-type transport, permease protein	Function Unknown
HVO_2034	-1.6	0.0	0.0	ABC-type transport, permease protein	Function Unknown
HVO_2037	1.0	2.6	3.6	DUF2078 family protein	Energy production and conversion
HVO_2065	3.8	1.7	0.0	uncharacterized protein	Function Unknown
HVO_2220	-2.2	-1.0	0.0	transducer protein Htr38	Carbohydrate metabolism and transport
HVO_2337	0.0	3.7	0.0	UspA domain protein	Signal Transduction
HVO_2375	2.0	2.3	0.0	ABC-type transport system periplasmic	Inorganic ion transport and metabolism
HVO_2565	0.0	2.8	1.1	DUF171 family protein	Function Unknown
HVO_2606	1.3	2.3	3.5	PQQ repeat protein	Function Unknown
HVO_2676	0.0	3.1	1.2	HesB/IscA family iron-sulfur cluster assembly	Post-translational modification
HVO_2700	3.1	3.6	1.5	AAA-type ATPase (CDC48 subfamily)	Post-translational modification
HVO_2701	4.1	4.2	1.7	uncharacterized protein	Function Unknown
HVO_2753	0.0	3.4	1.1	small CPxCG-related zinc finger protein	Translation
HVO_2756	0.0	2.4	1.4	50S ribosomal protein L10	Translation
HVO_2757	0.0	2.5	1.9	50S ribosomal protein L1	Translation
HVO_2786	1.6	3.1	1.1	thioredoxin	Post-translational modification
HVO_A0022	0.0	-1.5	0.0	uncharacterized protein	Function Unknown
HVO_A0029	0.0	-1.6	0.0	uncharacterized protein	Function Unknown
HVO_A0030	0.0	-1.5	0.0	uncharacterized protein	Function Unknown
HVO_A0052	0.0	-1.7	0.0	uncharacterized protein	Function Unknown
HVO_A0055	0.0	-1.5	0.0	uncharacterized protein	Function Unknown
HVO_A0067	0.0	-1.6	0.0	uncharacterized protein	Function Unknown
HVO_A0068	0.0	-1.5	0.0	uncharacterized protein	Function Unknown
HVO_A0105	0.0	-1.6	0.0	uncharacterized protein	Function Unknown
HVO_A0129	3.3	2.3	0.0	uncharacterized protein	Function Unknown
HVO_A0172	0.0	4.8	0.0	uncharacterized protein	Function Unknown
HVO_A0195	0.0	-1.5	0.0	DUF1616 family protein	Function Unknown
HVO_A0218	0.0	-1.6	0.0	GFO family oxidoreductase	Function Unknown
HVO_A0219	0.0	-1.6	0.0	probable nucleotide sugar dehydrogenase	Cell wall/membrane/envelop biogenesis
HVO_A0350	1.6	2.6	0.0	uncharacterized protein	Function Unknown
HVO_A0351	1.4	2.5	0.0	uncharacterized protein	Function Unknown
HVO_A0394	3.4	1.5	0.0	HTH domain protein	Transcription
HVO_A0399	0.0	-1.6	0.0	uncharacterized protein	Function Unknown
HVO_A0409	0.0	-1.5	0.0	uncharacterized protein	Replication and repair
HVO_A0466	3.4	2.8	0.0	uncharacterized protein	Cell wall/membrane/envelop biogenesis
HVO_B0068	0.0	-1.5	0.0	uncharacterized protein	Function Unknown
HVO_B0070	-1.8	0.0	0.0	pyridoxal phosphate-dependent aminotransferase	Amino Acid metabolism and transport
HVO_B0193	2.3	2.1	0.0	ArsR family transcription regulator	Transcription
HVO_B0248	0.0	2.2	2.3	probable oxidoreductase (short-chain	Lipid metabolism
HVO_B0276	3.2	1.3	0.0	DMT superfamily transport protein	Carbohydrate metabolism and transport
HVO_B0284	1.3	1.2	2.5	translation initiation factor aIF-1 (SUI1	Translation
HVO_B0285	1.3	2.7	1.4	transcription initiation factor TFB	Transcription
HVO_B0305	-1.7	-1.1	0.0	uncharacterized protein	Function Unknown
HVO_B0362	-1.6	-1.6	0.0	uncharacterized protein	Energy production and conversion
HVO_B0363	-1.9	-1.6	0.0	dimethylsulfoxide reductase subunit A	Energy production and conversion
HVO_B0364	-1.9	-1.3	0.0	dimethylsulfoxide reductase subunit B	Energy production and conversion
HVO_B0365	-2.1	-1.4	0.0	dimethylsulfoxide reductase subunit C	Energy production and conversion
HVO_B0366	-2.2	-1.6	0.0	Tat proofreading chaperone DmsD	Function Unknown

HVO_B0367	-1.2	-1.2	0.0	molybdopterin-binding domain protein	Function Unknown
HVO_B0368	-1.2	-1.2	0.0	molybdenum cofactor guanylyltransferase	Coenzyme metabolism
HVO_B0369	-1.5	-1.2	0.0	ABC-type transport system periplasmic	Inorganic ion transport and metabolism
HVO_B0371	0.0	2.5	1.6	aldehyde dehydrogenase	Energy production and conversion
HVO_D0001	2.8	2.5	1.4	NA	Function Unknown
HVO_D0002	2.7	2.5	0.0	uncharacterized protein	Function Unknown
HVO_D0003	3.1	2.4	0.0	homolog to plasmid replication protein RepH	Cell motility
HVO_D0004	2.6	2.4	1.5	uncharacterized protein	Function Unknown
HVO_D0004A	3.5	2.4	1.7	small CPxCG-related zinc finger protein	Function Unknown
HVO_D0005	3.2	2.5	1.4	HTH domain protein	Transcription
HVO_D0006	3.1	2.4	1.4	uncharacterized protein	Function Unknown

Supplementary References

Abu-Qarn, Mehtap, and Jerry Eichler. 2006. "Protein N-Glycosylation in Archaea: Defining Haloferax Volcanii Genes Involved in S-Layer Glycoprotein Glycosylation." *Molecular Microbiology* 61 (2): 511–25. https://doi.org/10.1111/j.1365-2958.2006.05252.x.

Alarcón-Schumacher, Tomas, Adit Naor, Uri Gophna, and Susanne Erdmann. 2022. "Isolation of a Virus Causing a Chronic Infection in the Archaeal Model Organism Haloferax Volcanii Reveals Antiviral Activities of a Provirus." *Proceedings of the National Academy of Sciences* 119 (35): 1–12. https://doi.org/https://doi.org/10.1073/pnas.2205037119.

Allers, Thorsten, Hien-ping Ngo, Moshe Mevarech, and Robert G Lloyd. 2004. "Development of Additional Selectable Markers for the Halophilic Archaeon." *Applied and Environmental Microbiology* 70 (2). https://doi.org/10.1128/AEM.70.2.943.

Duggin, Iain G., Christopher H.S. Aylett, James C. Walsh, Katharine A. Michie, Qing Wang, Lynne Turnbull, Emma M. Dawson, et al. 2015. "CetZ Tubulin-like Proteins Control Archaeal Cell Shape." *Nature* 519 (7543): 362–65. https://doi.org/10.1038/nature13983.

6. Chapter 3.

Revisiting evolutionary trajectories and the organization of the *Pleolipoviridae* family

Tomas Alarcón-Schumacher[1*], Dominik Lücking[1] and Susanne Erdmann[1*]

[1] Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany

* To whom correspondence should be addressed: Susanne Erdmann, Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany, Email: serdmann@mpi-bremen.de Phone: +49 421 2028 7340

Tomas Alarcón-Schumacher, Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany, Email: t.alarcon.sch@gmail.com

Authors Contributions

T.A-S. and S.E. conceived and led the study and performed the writing of the manuscript. D.L. generated the metagenomics data from Australian salt lakes. All authors participated in the analysis and interpretation of the data and contributed to the writing of the manuscript.

Competing Interest Statement:

The authors declare no competing interests.

Classification:

Biological Sciences, Microbiology

Keywords: Archaea, virus, Pleolipoviridae, phylogeny

- ❑ *PLOS genetics*, Under review

Abstract

Viruses infecting archaea are remarkably diverse, while displaying unique morphotypes and genomic features that have not been observed for eukaryotic or bacterial viruses. Particularly, archaeal pleomorphic viruses belonging to the Pleolipoviridae family represent an enigmatic group. They infect halophilic archaea from the Halobacteriota phylum performing productive chronic infections, without lysis of their host, and interestingly, they are thought to have evolved through recombination with different archaeal plasmids. However, most of our understanding of the diversity and evolutionary trajectories of this clade comes from a handful of representatives. Here we present 164 new genomes of pleolipoviruses obtained from metagenomic data of Australian hypersaline lakes, and publicly available metagenomic data, to perform a comprehensive analysis on their diversity and evolutionary relationships. We propose to classify the viruses into five genera within the *Pleolipoviridae* family, with one new genera represented only by virus genomes retrieved during this study. Interestingly, we observed that the genome size of pleolipoviruses appears to be dependent on their ability to integrate into the host. We confirm that they likely reshaped their genomes through recombining with multiple different groups of plasmids; in particular, we show that the proposed genus Epsilonpleolipovirus has evolutionary ties to pRN1-like plasmids, suggesting that this group could be infecting other archaeal phyla. Analyses of the host range showed that all but one virus exhibit an extremely narrow host range, and the analysis of virion spike proteins reveals that the evolution of the spike protein is driven by the interaction with the cellular surface of the hosts. Our study demonstrates that pleolipoviruses are far more diverse than previously thought, and highlight the importance of understanding these chronic-infecting viruses and their impact on host metabolism and ecology in different environments.

Introduction

Archaeal viruses represent one of the most fascinating part of the virosphere. Despite the low number of representatives, when compared to bacterial and eukaryotic viruses, they represent an unparalleled genomic structure, gene content and morphological diversity [1]. While some icosahedral viruses have been shown to be evolutionary related to bacterial and eukaryotic viruses [2,3], the vast majority of archaeal virus groups are unique to archaea and no relatives have been identified in the other domains of life [4]. Moreover, archaea-specific groups of viruses display distinct evolutionary origins, highlighting that viruses likely evolved on numerous multiple independent occasions [5–7]. Virus genomes typically exhibit structural and replication modules, which frequently recombine with each other and with a variety of mobile genetic elements (MGEs) such as transposons and plasmids [8,9].

One of the most intriguing groups of viruses demonstrating evolutionary mixing with different viruses and MGEs is the *Pleolipoviridae* family [3]. Viruses belonging to the *Pleolipoviridae* family, known as pleolipoviruses, are unique in many aspects, with members of this family shown to have different genome types, going from single or double stranded DNA genomes, to even hybrid double stranded genomes with single-stranded interruptions [10,11]. Unlike most prokaryotic viruses, which have a protein shell (capsid), the pleolipovirus virions have a host-derived lipid membrane enclosing their genome. Biochemical analyses have shown the presence of at least two major structural proteins in virions: a spike protein anchored to the membrane and a membrane associated protein facing the particle interior, commonly known as VP4 and VP3-like proteins, due to their position in the genomes of the first isolated representative Halorubrum pleomorphic virus 1 (HRPV-1) [10,12,13]. Virions are pleomorphic particles that range from 50 to 100 nm, which bud off from host cells without causing cell lysis. During this, so called persistent or productive chronic infections, virions are released in large amounts and have the potential to strongly impact host metabolism [10,14]. The majority of currently isolated pleolipoviruses were shown to be host specific, with the sole exception of recently characterized Haloferax pleomorphic virus 1 (HFPV-1), which displays an unusual broad host range, being able to infect members of both the *Haloferacaceae* and *Haloarculaceae* families [14].

Currently, pleolipoviruses are classified into three genera: the *Alpha*, *Beta* and *Gammapleolipovirus*. While the three genera share a stable core of genes consisting of two structural proteins, (i.e. VP3 and VP4 spike protein), plus two hypothetical proteins flanked by the spike protein and a conserved ATPase (ORFs 6 and 7 in HRPV-1), they exhibit a remarkably modular mosaicism in their genomes. This is particularly evident with their replication mechanisms. Viruses from the *Alphapleolipovirus* genus encode two non-orthologous families of rolling circle replication initiation endonuclease (RCRE), which were likely acquired by an ancestral pleolipovirus-like entity in independent events from two different groups of plasmids (pGRB1-like pTP2-like plasmids) [15]. Meanwhile, the betapleolipoviruses encode for a putative replication protein (Rep

protein), that shows no homology to any other known replication protein, and its evolutionary origin remains a mystery [4]. On the other hand, His2 virus, the only isolated member of the *Gammapleolipovirus* genus, encodes for a protein-primed DNA polymerase related to the spindle-shaped virus His-1, with both viruses infecting the same host, *Haloarcula hispanica* [16].

Recent metagenomic studies, as well as the isolation and characterization of divergent members of the *Pleolipoviridae* family have hinted that the diversity of this clade might be larger than previously thought [14,17,18]. Particularly, HFPV-1 exhibits low levels of sequence similarity with other isolated pleolipoviruses and encodes no homolog for any of the replication proteins, raising further questions on the diversity and organization of this clade. Thus, we aimed to expand our knowledge on the diversity of pleolipoviruses and study in depth the phylogenetic relationships between pleolipoviruses and other viruses and MGEs.

Experimental procedures and data analyses

Sampling sites

Sediment salt crust were collected from a total of 11 hypersaline lakes in December 2018 and January 2019 (Figure 1) under the permission from the Department for Environment and Water, South Australia (Permission number: U26817-1) and the Department of Environment, Land, Water Planning, Victoria (Permission number:1008945). DNA was extracted from approximately 1g of sediment with FastDNA™ SPIN Kit for Soils. DNA sequencing libraries (FS DNA Library (NEBNext® Ultra™)) and sequencing (Illumina HiSeq2500 - Rapid Mode) was performed at the Max Planck-Genome-Centre Cologne (Germany), with run condition 2 x 250 bp (paired end reads). Reads were trimmed with Cutadapt [19], allowing a minimum quality of 30 and discarding short and unpaired reads (-q 30, -m 30). Quality-trimmed reads were assembled with metaSPAdes v3.13.1 [20] and protein prediction was performed using Prodigal [21] in metagenomic mode (-p meta). Contig coverage was calculated using BBmap v38.06 [22] with a minimum identity of 90% (minid=0.9). Assembled contigs were then binned with MetaBAT 2 [23] allowing a minimum contig length of 1500 nucleotides (-m 1500), a maximum number of edges of 1000 to increase sensitivity (--maxEdges 1000) and a minimum score of 95 to increase specificity (--minS 95), with the --noAdd flag to diminish contamination issues. For exact reproducibility a seed was also indicated (--seed 1). Quality assessment of generated bins was performed with CheckM [24]. High and Medium-quality MAGs were determined according to the standards developed by the Genomic Standards Consortium [25] and selected for further analyses. Taxonomic classification of selected MAGs was performed with the GTDB-Tk toolkit v1.4.0 [26].

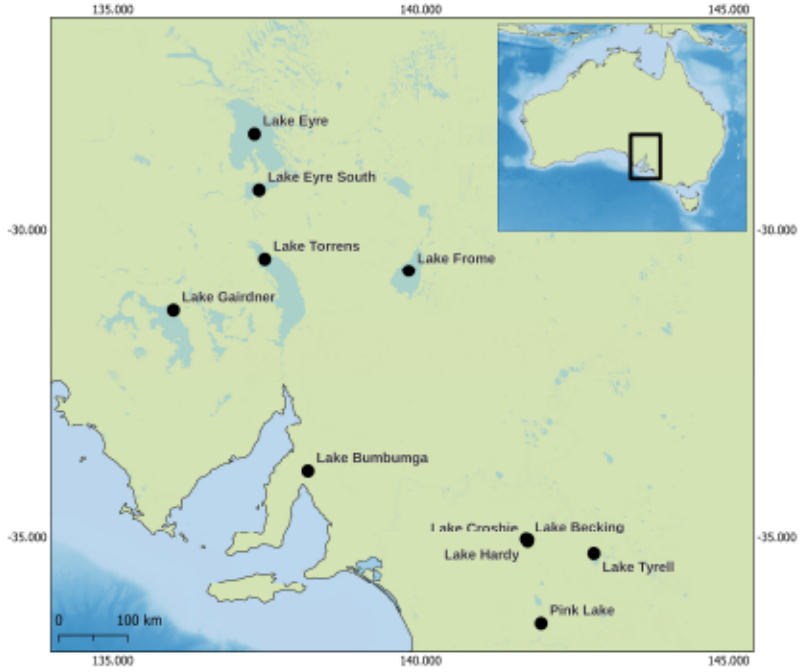


Figure 1: Sampling sites. Hypersaline lakes from south east Australia (South Australia and Victoria states) sampled for this study. Samples were collected during the austral summer of 2018/2019. Map data was obtained from public available databases Natural Earth, ©OpenStreetMap and the General Bathymetric Chart of the Oceans (GEBCO).

Generation of Pleolipovirus database

First, we generated a trusted database for each of the proteins conserved in all the available sequences of 16 previously isolated pleolipoviruses (ORFS 4,6, 7 and 8 in HRPV-1) [14,17]. Then, we used the sequences corresponding to each gene to query the IMG/VR database version 5.1 [27,28] for related viruses using blastp implemented in Diamond [29]. Hits with an e-value $< 10^{-5}$ and a score > 50 were considered significant and added to the database of each core gene. This process was recursively iterated adding the new significant hits from each iteration to the respective database until no new hits were obtained. Then, Hidden Markov model (HMM) were generated for each gene, for which alignments were performed with MAFFT v7.407 [30] and the --localpair and --reorder flags. Subsequently, HMM profiles were generated with HMMER v. 3.2.1 [31]. HMM models for each gene were then used to further identify distant homologs for each gene using the same recursive approach described above. In order to reduce false positive results and to enhance specificity, HMM models were trained against a dataset containing a combination of protein sequences from identified pleolipoviruses (true positives), and an array of proteins of known function in the PFAM database v.35.0 [32] (true negatives). Hits with a minimum e-value of 10^{-5} and above the score threshold (score >= to the lowest by a true positive and 10 units higher than the highest score observed for a true negative in the training dataset), were considered as true positives hits for each model.

Identification of genomes in metagenomic data

Resultant models were then used to screen binned and non-binned contigs, generated in this study, to obtained potential virus genomes belonging to the *Pleolipoviridae* family. Contigs of at least 4.000 bp and with significant hits against at least four of the previously generated models of conserved genes among pleolipoviruses were considered for further analyses. Quality control to identify and exclude non-viral regions was performed with CheckV [33], database version 1.4 (Aug 27, 2022). Due to the limited representation of pleolipoviruses in CheckV database, and in order to properly asses the quality of potential pleolipoviruses-like genomes, the HMM models from conserved proteins previously generated in this work were added to the database and accounted as viral proteins. Additional identification of integrated proviral sequences and the presence of direct and inverted terminal repeats (DTRs and ITRs respectively) were performed with Virsorter2 and geNomad [34,35]. Genomes were then considered complete if they presented terminal repeats (TRs), were identified as complete integrated proviruses or if they displayed CheckV completeness value = 100. Quality-trimmed sequences were clustered at 95% nucleotide identity with the software NUCmer (NUCleotide MUMmer) version 3.1 [36], which corresponds to the species level [37,38].

Genomic and phylogenetic analysis of pleolipoviruses genomes

For genomic characterization, viral contigs retrieved from metagenomic data were combined with genomes of previously isolated pleolipoviruses and those retrieved from the IMG/VR, and redundancy was revised with NUCmer as described above. Average nucleotide distance (ANI) was calculated using PyANI v.0.2.12 [39] and the VIRIDIC web server [40]. Functional annotation of predicted viral proteins was performed InterProScan v5 [41], with DRAM [42], the package HH-suite3 [43] against the PDB70 database (release pdb70 200108) and the BFD database [44]. Functional categories were assigned using eggnog-mapper [45,46].

Genome based phylogeny and classification was performed with VICTOR [47] using the predicted protein sequences and the D6 formula. Gene-shared network analyses were performed with vConTACT [48,49] using the Viral RefSeq-archaea database v.211, blastp as the relations mode (--rel-mode BLASTP) with a minimum e-value of 10^{-3}, a significance threshold in the contig and protein cluster similarity network of 0.5 (--sig 0.5, --mod-sig 0.5) and with the minimum number of contigs a protein cluster must appear of 2 (--mod-shared-min 2). Network results were displayed using R packages ggplot2 and Network [50,51].

Single gene phylogeny

Protein sequences from selected conserved genes were aligned using MAFFT v7.407 [30] with local pair strategy (--localpair). Non-trimmed alignments were used to infer phylogenetic relationships using IQ-TREE 2 [52] with 10.000 ultrafast bootstrap [53] and 10.000 replicates of SH-aLRT branch test and automatic model selection [54]. Results were then visualized using iTOL [55].

Protein structural prediction

Protein structures were predicted using Alphalfold 2 [56] or retrieved from protein structure databases when available [57,58]. Protein structures were visualized with open source PyMOL(TM) Molecular Graphics System, Version 2.2.0 [59]. Resulting structural models were manually inspected and low-confidence (pLDDT < 50) N and C-terminal regions were removed. Subsequently, selected models were used for structural similarity comparison with DALI [60].

Virus-host inference

Host prediction was performed with the integrated phage host predictor Iphop [61], which combines multiple approaches for virus-host prediction, i.e. blast to reference host genomes and CRISPR spacers database; k-mer composition algorithms implemented in tools WIsH, PHP and VHM - s2* [62–64]; and protein content-based prediction implemented in RaFAH [65]. In case of multiple hosts with significant scores (>=90), hosts containing CRISPR matches against the virus were considered as true positives. If no CRISPR spacer was identified, the highest integrated Iphop score value was considered as the most likely host.

Results and Discussion

Generation and curation of pleolipovirus dataset yields 129 pleolipovirus-like elements

To expand our understanding of the evolutionary relationships of the *Pleolipovirirdae* family, and improve our ability to detect potential novel members, we assembled a dataset consisting of the 16 previously isolated pleolipoviruses and 219 uncultivated viral genomes (UViGs) related to this clade from the IMG/VR database (v5.1, 11th August, 2022) [28]. Selected genomes were then screened to identify the conserved cluster of four genes shared by all pleolipoviruses (ORFs 4, 5, 6, and 7 of Haloferax pleomorphic virus 1 (HFPV-1) [14] using homology annotation and HMM profile generation, profile validation, iterative search, and synteny verification. First, for each gene from the protein cluster we used the protein sequences derived from sequenced isolated pleolipoviruses to do a homology search using Diamond [29]. Significant hits for each gene were then aligned and an HMM profile was generated. In the second step, the generated profiles were used to screen the PFAM database (v.35.0) and score threshold to avoid false positives were established (-T values of 30, 36, 35 and 163 for ORFs, 4, 5, 6 and 7 of HFPV-1 respectively). Thirdly, profiles were used to screen the IMG/VR dataset in an iterative way. After each run, significant hits were added to the alignment and an updated profile was generated until exhaustion (three iterations), yielding a total of ~150 significant hits for each gene. Finally, profiles were again queried against the PFAM database and the aforementioned thresholds validated. The gene arrangement of resulting contigs were further assessed for those containing the four conserved genes no more than 4.5kb apart and in the same order as observed in known isolates. This yielded 129 high-confidence UVIGs that were selected for further genome comparisons and phylogenetic analyses.

Novel pleolipovirus-like elements from Australian salt lakes reveal the preference for a productive life cycle

Given that the divergent pleolipovirus HFPV-1 was isolated from enrichment cultures of salt crust from Lake Tyrrell [14], we searched the metagenome of Lake Tyrrell as well as 10 other Australian salt lakes (Figure 1), that were sampled during Australian summer 2018/2019, to further search for novel pleolipovirus genomes. To identify pleolipovirus-like sequences in the metagenomes, we screened individually binned and non-binned assembled data for the presence of the hallmark genes described above. This resulted in 261 contigs containing at least one of the conserved genes (minimum length 2kb). Notably, all but one of the retrieved sequences were relatively short (> 20kb) and part of the non-binned fraction of each metagenome. Contamination assessment with CheckV based on the presence of host genes, showed that only 27 contigs (~ 10 % of the total) corresponded to viruses that were likely integrated as proviruses at the time of sampling (Supplementary Table 1). However, these sequences could not be assigned to any MAG, likely due to their relative short length (average 2.5kb) and the presence of the viral sequences, which often display different k-mer frequencies and coverage values than the ones from their host genomes and distort the binning process. Additionally, functional prediction showed that only a small fraction of the sequences (6 out of 261 contigs) encoded integrase-like genes, indicating the capacity to integrate into the host genome. Altogether, this suggest that the majority of the pleolipoviruses reported here, were likely replicating as extra chromosomal elements, rather than exhibiting an integrated lysogenic life cycle at the time of sampling. This has profound ecological implications, because it has been shown that pleolipoviruses can dramatically remodel the transcriptomic program of their hosts while actively performing an infective cycle [14]. If a large proportion of the reported pleolipoviruses exists as episomal elements, and, given that pleolipoviruses perform productive persistent infections (genome replication and constant particle release without lysis the host cell), these viruses could be greatly impacting host fitness and modulating interactions within a microbial community.

The only pleolipovirus-like element from the binned fraction (Gairdner_Node_15) was found integrated into the metagenome-assembled genome of a new species belonging to the *Halobaculum* genus (Supplementary Table 2), for which no pleolipovirus has been assigned to the date. The pleolipovirus-like element of 10.852 bp is flanked by terminal repeats (TRs), with an integration site likely determined by the tryptophan tRNA adjacent to the integrase gene (Supplementary Figure 1), indicating that it is a complete genome. Functional prediction showed that aside from the core pleolipovirus genes, this genome exhibits a transcriptional regulator (arCOG08095) and a rolling-circle-like replication protein, which likely ties it to the alphapleolipoviruses [66]. However, as it was the case for the vast majority of the reported viruses, most predicted proteins were recalcitrant to annotation and their function and evolutionary origin remains unknown.

Genome wide phylogenomic organization of pleolipoviruses reveals unexpected high diversity

In order to further deepen our understanding the taxonomic composition and the phylogenetic relationships of the mined pleolipoviruses, contigs were combined with the dataset of isolated and UVIG pleolipovirus, clustered at 95% similarity (equivalent to the species level) and only those containing the four conserved gene cluster and a minimum length of 4.5kb, were considered for phylogenomic analyses [37]. This resulted in 184 non-redundant pleolipovirus genomes with an overall average completion of ~81%. Out of these, 100 genomes (~ 54%) were considered complete (see methods), with an average length of 13.5kb (Supplementary Table 3), which is considerably larger than the average genome length of current isolates (10.4kb). However, completeness and quality values estimated with checkV are likely underestimated, as several complete genomes containing terminal repeat or with genome length around the average were labeled incomplete and only medium quality given their extremely divergent gene content. Interestingly, all genomes, including the ones from this study and UVIGs from the IMGVR database, derive from hypersaline environments, indicating that pleolipovirus-like elements are restricted to high salt environments. However, our inability to detect them in other environments could be caused by the low sequence similarity values (Supplementary figure 2) and the dramatic variation in gene content among pleolipoviruses. Additionally, recent work has suggested that some pleolipo-like viruses could be part of the human gut virome of methanogenic archaea [67], further expanding their habitat distribution and suggesting that the relevance of this virus group might be underestimated.

Given the extremely low similarity at nucleotide level between the isolated pleolipoviruses (Supplementary Figure 2), phylogenetic relationships between members of this clade have been previously studied using genome-wide based approaches [68,69]. Thus, we used the predicted proteins sequences and constructed a genome-based phylogeny with VICTOR [47] (Figure 2A). The majority of the retrieved viruses are related to the alpha and betapleolipovirus genera, with the alphapleolipoviruses forming a monophyletic group (although no significant support was obtained for the clade; score = 41). Most genomes within this group encoding for a homolog of a rolling circle replication endonuclease (RCRE), which is consistent with the current classification as the RCRE is one of the hallmark genes of the alphapleolipovirus genus (ORF1 in HRPV-1) [66] (Figure 2A). Conversely, despite the apparent coherence with the topology previously proposed for the *Betapleolipovirus* genus, the vast majority of clades observed in the VICTOR tree present extremely low support values (< 95) indicating that the evolutionary relationships within this clade remain poorly resolved (Figure 2A). Despite sharing several characteristic proteins that serve as discrimination criteria for this genus (i.e. Halorubrum pleomorphic virus 3 ORFs 6 and 9), the betapleolipoviruses seem to form a polyphyletic group. They cluster into two major clades separated by several branches which consistently do not share these conserved features (ORF6, ORF9), suggesting that they could be part of previously undescribed genera. Similarly, several genomes sharing no characteristic feature with neither alpha nor betapleolipoviruses appear

to be branching close to the gammapleolipoviruses. However, none of them exhibit a type B DNA polymerase, while sharing only the core genes and sparsely hypothetical proteins with the other members of the gammapleolipovirus genus. Additionally, gene-sharing network analyses implemented in vConTACT2, which generates cluster that are equivalent to genera level classification [48,49], also revealed that not all genomes fit in the current trichotomy. The analysis disclosed 16 cluster with at least two members. However, almost one third of the genomes (62 genomes) failed to fit into any of these genera-like groups (Supplementary figure 3, Supplementary Table 3), because these genomes shared a large proportion of their proteins with multiple clusters. Similar to the result observed with VICTOR, the network analysis confirms the existence of three main clusters that do not correspond with the three currently described genera (Alpha, Beta and Gammapleolipovirus). The alphapleolipoviruses cluster as a single group, however, the betapleolipoviruses form at least two different cluster. One cluster contains the small-genome isolates HFPV-1, HGPV-1 and HRPV-10 to 12 (approx. 8-9kb) and the other contains amongst others the larger-genome isolates HHPV-3 and HRPV-3 and the integrase encoding isolates HHPV-4 and SNJ2 (Figure 2B). Altogether, this indicates that the diversity within the *Pleolipoviridae* family could be greater than previously thought and further challenges the current classification.

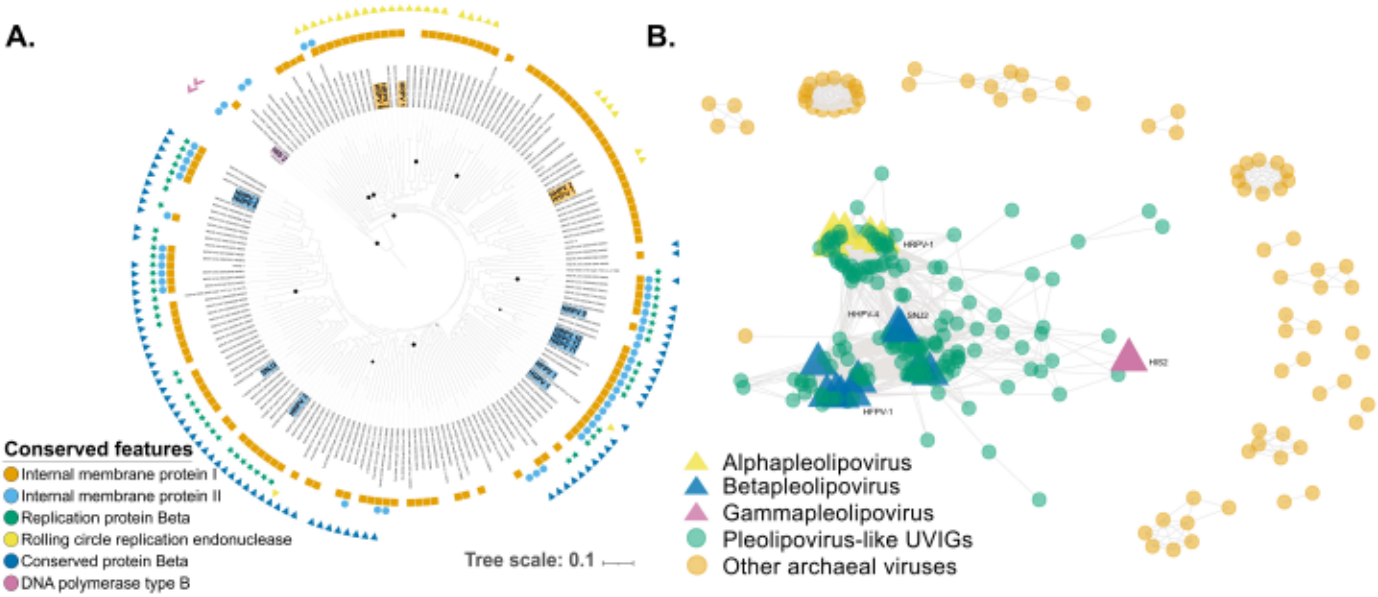


Figure 2: Genome-wide phylogenetic analysis. A. Phylogeny of pleolipoviruses generated with the Genome BLAST Distance Phylogeny (GBDP) algorithm implemented in VICTOR using amino acid sequences. The branches are scaled and represent the GBDP interproteomic distance inferred using formula D6. Supported branches (>95) are demarcated with black circles. Colored labels represent isolated representatives from each genus: *Alphapleolipovirus* (yellow), *Betapleolipovirus* (blue), *Gammapleolipovirus* (magenta). **B.** Gene-shared network of pleolipoviruses generated with vConTACT2. Nodes (circles) represent genomes and edges (lines) indicate shared protein content. Highlighted nodes are isolated representatives from the different genera: *Alphapleolipoviruses* (yellow), *Betapleolipoviruses* (blue) and *Gammapleolipoviruses* (magenta).

Pleolipoviruses also display little to no connections with other archaeal viruses, with the only exception *Natrinema* SNJ1 virus, a member of the *Betasphaerolipovirus* genus, whose members infect halophilic and

thermophilic archaea, and that shares the host with pleolipovirus SNJ2 [70]. The absence of shared features with other viruses has also been observed for some viruses infecting hyperthermophilic archaea, and it has been suggested that this is likely a product of distinct evolutionary origins [6]. Altogether, the presented evidence suggests that the vast majority of proteins encoded by members of the *Pleolipoviridae* family are not shared with other archaeal viruses, and must have been obtained rather early in their evolutionary history or exchanged through recombination with other MGEs that infect the same host.

Network analysis reveals a correlation between genome length and integration

Furthermore, the integrase encoding HHPV-4 and SNJ2 pleolipoviruses are among the nodes with the highest degree of connectivity (node degree of 70 and 69 respectively, Figure 2B), and interestingly, overall comparisons showed that genomes encoding integrases, have in general significantly higher degrees of connectivity than those which do not (Wilcoxon test, p-value = 0.00158). Further analysis revealed that, irrespective of their clustering patterns, pleolipoviruses carrying integrases also have significantly larger genomes (Wilcoxon test, p-value $=2.2 \times 10^{-16}$). On the other hand, no significant correlation was observed between node degree and genome length (correlation coefficient = 0.1 and 0.12 for Pearson and Spearman correlation tests respectively). While pleolipoviruses with the ability to integrate into their host genomes seem to harbor a more extensive gene repertoire, the connectivity between nodes is likely a result of their phylogenetic relationships and not random evolutionary convergence as a byproduct of larger genome sizes. Even though we cannot rule out that some viruses without integrases can integrate into their host genomes, as it has been shown for ssDNA viruses that employ alternative mechanisms for integration [71], the data suggests that pleolipoviruses rely on a self-encoded integrase to expand their genomic repertoire and to reliably to fix the acquired genes in the population.

Single gene phylogeny allows to resolve the evolutionary history of pleolipoviruses

To better resolve the evolutionary relationships between pleolipoviruses, we considered a concatenated phylogenetic approach using the four described core genes. They are conserved across all genera, indicating that they were already present at the origin of this virus group. To ensure that they display a congruent evolutionary history, we first performed single-gene phylogenetic reconstruction for each marker and compared the resulting topologies. The phylogenetic reconstructions of ORF4 and ORF5 (in HFPV-1 genome), the structural spike protein and a hypothetical protein, displayed markedly different topologies (Figure 3), with little similarity to both the current taxonomic classification and the clades described here (Figure 2). The three genera (Alpha-, beta- gamma) exhibit a very polyphyletic nature, indicating that selective

pressures that do not necessarily reflects the phylogenetic relationships between the different clades might be driving the clustering pattern of each of these proteins (see more in Host driven evolution section).

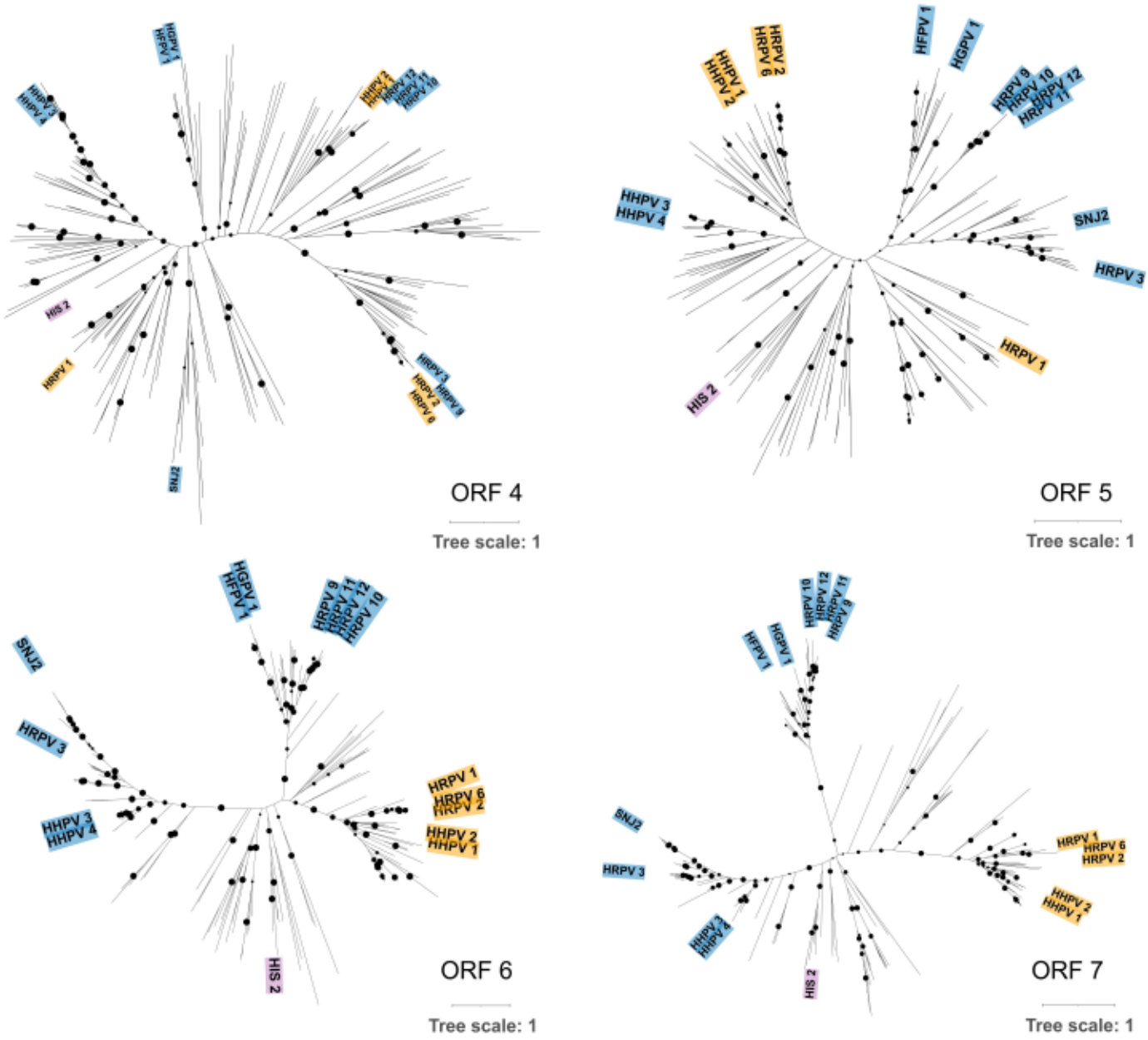


Figure 3: Single gene phylogenies. Phylogenetic tree reconstruction of conserved pleolipoviruses proteins. Sequences were obtained from the 184 pleolipovirus-like genomes dataset generated in this study. ORFs 4 to 7 correspond to the spike protein, two proteins of unknown function and an ATPase respectively, with ORFs numbers designated according to the annotation of Haloferax pleomorphic virus 1 (HFPV-1). Tree was constructed with iqtree with 10.000 ultrafast bootstrap. Supported branches (SH-aLRT >= 80 and ultrafast bootstrap >= 95) are demarcated with black circles. Colored labels represent isolated representatives from each genus: *Alphapleolipovirus* (yellow), *Betapleolipovirus* (blue), *Gammapleolipovirus* (magenta). Scale bar represents the number of substitutions every 100 amino acids.

In contrast, ORF6 and ORF7 displayed a congruent topology (Figure 3), which is consistent with the higher sequence similarity of these two genes across pleolipoviruses genomes. The overall topology showed that the alpha and gammapleolipoviruses form monophyletic clades with high support values (SH-aLRT >= 80% and UFboot >= 95%). Meanwhile, beta-like pleolipoviruses organize into two separated well supported monophyletic clades, consistent with the results obtained with VICTOR and the network analysis with vConTACT2, suggesting that they should be reclassified into separate lineages. Furthermore, using ORF6 and ORF7 as concatenated phylogenetic marker genes, resolves the relationships between previously poorly positioned branches (Figure 4). Additionally, the majority of genomes that do not share the major hallmark genes with the established genera (i.e RCRE, polB polymerase, Rep protein, ORF8-like protein), form a well-supported separate lineage (aLRT >= 100% and UFboot >= 96.1%) (Figure 4). Meanwhile, a small number of genomes allocated close to the gammapleolipoviruses remained recalcitrant persistent to classification, with many of them sharing little to none of the genes that currently serve as genera confining criteria [66]. Altogether, both genome-based, and single gene phylogeny, as well as the network analysis, indicate that they might be divergent members of the gammapleolipovirus genus, or that they form separate clades, further demonstrating that the diversity of the *Pleolipoviridae* family is likely much greater than previously thought.

Proposed reclassification of the Pleolipoviridae with two new genera

Based on the presented evidence, we propose a new organization of the *Pleolipoviridae* family with the addition of two genera: *Deltapleolipovirus* and *Epsilonpleolipovirus* (Figure 4). This reclassification includes the separation of the former betapleolipoviruses into two different lineages, the betapleolipoviruses and the deltapleolipoviruses, with the deltapleolipoviruses including some members that were formerly classified as betapleolipoviruses (i.e. HGPV-1 and HRPV-9 to 12). However, the deltapleolipoviruses are a monophyletic clade that, despite sharing many features with the betapleolipoviruses, diverged relatively early in the evolutionary history of pleolipoviruses (Figure 4).

The proposed new genus *Epsilonpleolipovirus* comprises only uncultivated members, with the complete genome (UVIG) Pink_NODE_1698 chosen as the representative for the clade, as it is the only one flanked by ITRs and shares the main protein features identified in this clade (Figure 5). Consistent with the phylogenetic analysis, the most conserved genes are the genes equivalent to ORF6 and ORF7 in HFPV-1 with up to 60% similarity. Interestingly, the epsilonpleolipoviruses possess a very flexible genome organization, as they display marked differences in gene content with each other and even some of their core genes share little to none protein sequence similarity (Spike protein and ORF-5-like proteins) (Figure 5).

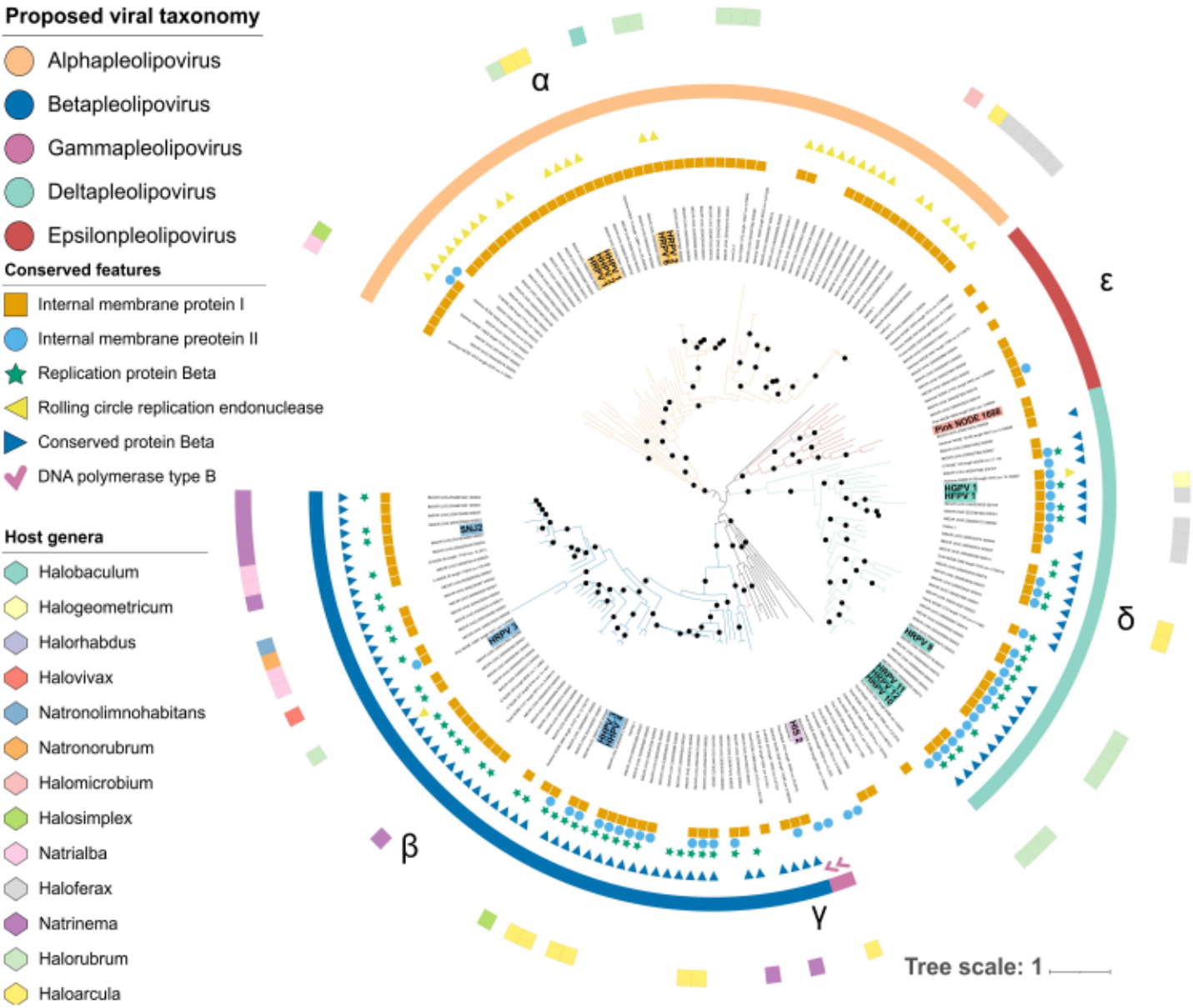


Figure 4: Proposed taxonomy. Phylogenetic inference from concatenation of ORF6 and ORF7-like genes using IQ-Tree. Supported branches (SH-aLRT >= 80 and ultrafast bootstrap >= 95) are demarcated with black circles. Greek characters represent different proposed genus within the Pleolipoviridae family: *Alphapleolipovirus* (α), *Betapleolipovirus* (β), *Gammapleolipovirus* (γ), *Deltapleolipovirus* (δ) and *Epsilonpleolipovirus* (ε). Scale bar represents the number of substitutions every 100 amino acids. From inside out the rings represent: presence of type I internal membrane protein (VP3-like), type II internal membrane protein, Betapleolipovirus-like Replication protein (Rep protein), Rolling circle replication endonuclease (RCRE), ORF8-like protein, Type B DNA polymerase, Proposed Taxonomy, Predicted Host genus.

The spike protein and ORF5 exhibit abnormally large sizes when compared to other pleolipoviruses, however, structural prediction of the spike protein, using the representative genome, reveals a similar v-shaped fold as reported for pleolipoviruses HRPV-2 and HRPV-6 [72] (Figure 6). A host could not be identified for any member of the Epsilonpleolipoviruses, indicating that they might be infecting different taxa than previously described pleolipoviruses, which could be the cause for the observed changes to the spike protein. Through

additional structural prediction of neighbor proteins, we also identified a very divergent variant of the VP3-like internal membrane protein (ORF6 in Pink_NODE_1698, Figure 5), which despite the lack of sequence homology exhibits a very similar predicted fold to the one predicted for the member of the sister clade Deltapleolipovirus HGPV-1 (Supplementary figure 4).

The replication strategy of epsilonviruses seems rather diverse. They do not share any of the hallmark replication genes with the other genera (RCRE, Rep protein nor type B polymerase). However, some (3 out of 14) encode for proteins with a DNA primase/polymerase domain (IPR015330, Supplementary table 4), which has been described in archaeal plasmids, such as pRN1 from *Sulfolobus islandicus* [73], and highlights again the relevance of plasmid-virus recombination in the evolutionary trajectory of this clade. Another member (IMGVR_UViG_3300005273_000010) encodes for a non-related phage P4-like DNA primase (IPR006500, Supplementary table 4). However, the remaining epsilonpleolipovirus members, including three complete genomes, do not exhibit a homolog for any known replication protein. This suggests that they might replicate via a combination of virus-mediated activity and host polymerases, such as the head-tailed myohalovirus PhiH1, that encodes a proliferating-cell nuclear antigen (PCNA), or the pleolipovirus HIS-2, that contain proteins covalently-bound to the termini of their linear genomes [16,74–76]. Additionally, 9 out of 14 epsilonpleolipovirus genomes encode small proteins homologues to HNH-like nucleases (IPR003615, Figure 5). These nucleases are commonly found in archaeal and bacterial tailed viruses, and have been shown to be involved in terminase-mediated genome packaging [76,77]. However, their role in the replication cycle of epsilonpleolipoviruses remains to be elucidated.

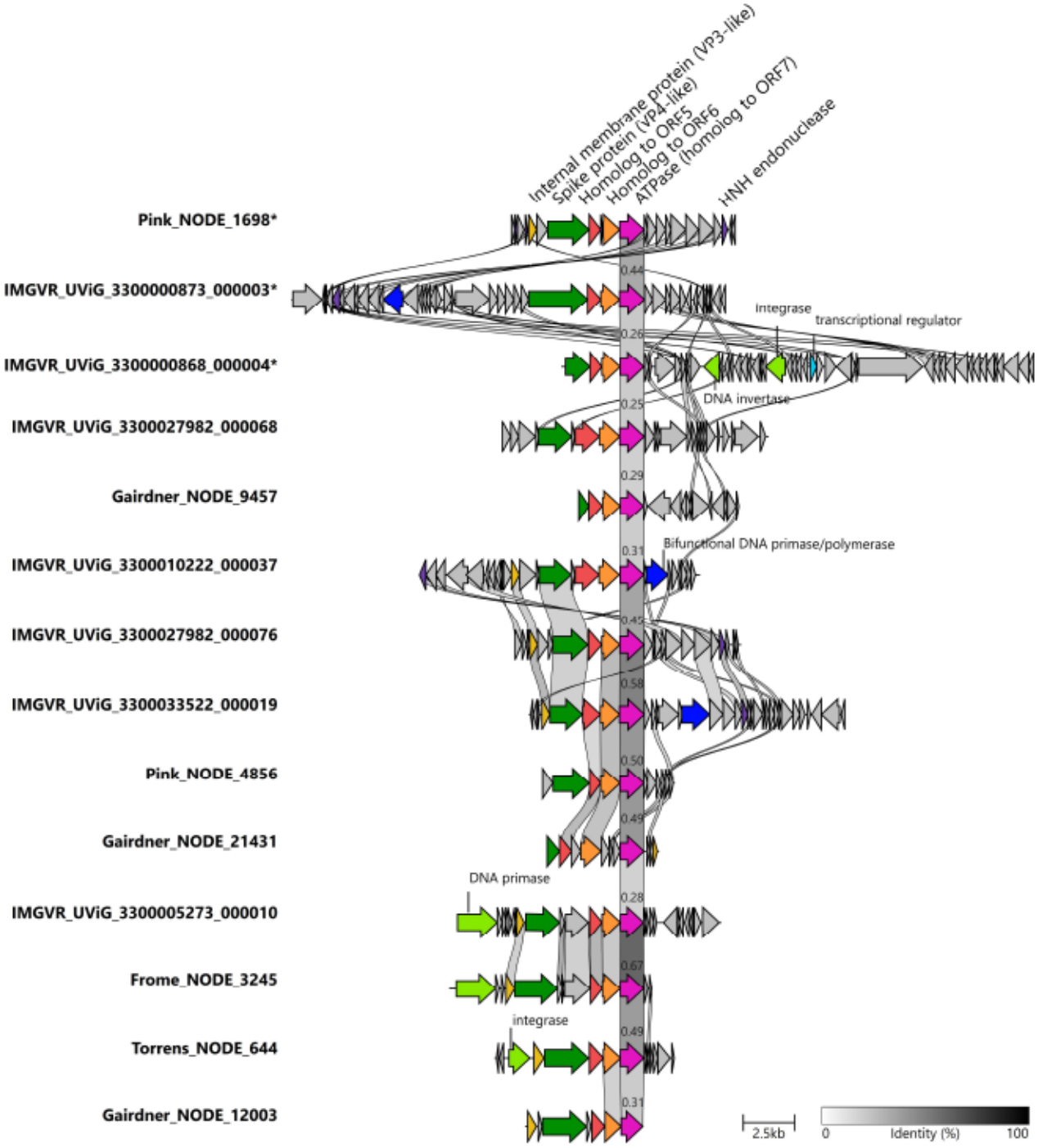


Figure 5: The *Epsilonpleolipovirus* genus. Genomic alignment of members of the proposed *Epsilonpleolipovirus* genus. Similarity values (blastp) are indicated by grayscale shading. Homologues of conserved genes are colored the same as follows: VP3-like protein (light brown), Spike protein (dark green), ORF5-like (red), ORF6-like (orange), ATPase (magenta), HNH-like endonuclease (purple), bifunctional DNA polymerase/primase (blue), transcriptional regulator (light blue) and other DNA-related enzymes, i.e. integrase, primase and invertase (light green). Complete genomes are highlighted with (*) symbol.

Frequent recombination and gene loss and gain of replication-related proteins, shape the structure of the Pleolipoviridae family

Interestingly, the concatenated phylogenetic approach also provided some insights into the evolutionary history and acquisition of the replication machinery. The phylogenetic relationships uncovered in this work are consistent with the proposed scenario for the evolution of pleolipoviruses, with the last common ancestor of the *Pleolipoviridae* family likely acquiring the two variants of RCRE in independent events from different groups of plasmids (pGRB1-like pTP2-like plasmids, or a related ancestor). This event separated the members of this clade from the rest of the lineages giving birth to the *Alphapleolipovirus* genus [3,78]. Later in evolutionary history, the last common ancestor acquired the proposed replication protein of the betapleolipoviruses (Rep protein, corresponding to Halorubrum pleomorphic virus 3 ORF 9) in an independent recombination event from a different group of plasmids or similar MGEs.

However, the phylogenetic reconstruction of the proposed Rep B protein in beta- and deltapleolipoviruses displays a very polyphyletic topology (Supplementary Figure 5), suggesting a more intricate evolutionary trajectory for these groups. Interestingly, the phylogenetic reconstruction of another conserved gene between the beta- and deltapleolipoviruses groups (homolog proteins to ORF8 in HFPV-1), exhibit a clear separation between both clades (Supplementary Figure 6). This monophyletic characteristic shows that, despite being acquired at a similar time than the Rep B protein, it has remained relatively stable across the evolutionary history of both clades. The later suggests, that the Rep B replication protein was likely gained and/or lost multiple times in the evolutionary history. The later highlights once more that recombination events with other MGEs and between different pleolipovirus genera appear to be frequent in this groups of viruses. Furthermore it also suggests that the divergence of the beta and deltapleolipoviruses likely occurred during earlier stages, at a similar time to the acquisition of the ORF8-like protein by the last common ancestor of the pleolipoviruses.

Altogether, while the replication apparatus in the *Alphapleolipovirus* genus remained relatively stable, the Rep protein was lost or replaced on multiple occasions, resulting in different lineages. We hypothesize that in the case of the gammapleolipoviruses the Rep protein was replaced by a PolB type polymerase. Subsequently, different independent gene losses gave rise to the epsilonpleolipoviruses, which lack a canonical replication protein, or the proposed deltapleolipoviruses, where some members do not exhibit any recognizable replication protein. The Deltapleolipovirus HFPV-1 for example, lacks a homolog to the conserved replication proteins and instead encodes a hypothetical protein with a Helix-turn-helix DNA binding domain, which has no homolog in any database [14]. Interestingly, while the alphapleolipoviruses have a conserved replication strategy and a relatively conserved genome structure, the other lineages that went through a number of loss and gain events of the replication proteins have also in general a less stable core genome and are more prone to gene reshuffling and recombination with other MGEs and/or their hosts. The later suggests that the RCRE

might be a more reliable enzyme for genome replication compared to the other uncharacterized proteins, which results in the observed higher genome stability. Under this scenario, given that the alternative mechanisms are more prone to make mistakes, the other clades may often rely on recombination as their main strategy for adaptation.

Pleolipoviruses have a very narrow host range

The spike protein is crucial for host recognition, interacting with the host cell surface and triggering membrane fusion [79,80]. Interestingly, the phylogenetic reconstruction on the spike protein amino acid sequences (Figure 3, ORF4), displayed a clustering pattern more related to the taxonomic affiliation of the hosts rather than to the virus taxonomy. To explore the selective pressures driving the evolution of the spike protein, we performed an in silico host prediction for our entire dataset of pleolipoviruses using Iphop [61]. Iphop combines multiple approaches to establish virus-host pairs, such as matching CRISPR spacers, protein content and k-mer frequency algorithms (see methods). The prediction yielded significant (Iphop score > 90) results for 125 of the 184 genomes (68%), out of which, 57 virus-host pairs (~30%) were deemed as certain, because they were identified as integrated proviruses or identified by CRISPR spacer matches (Supplementary Table 3). Notably, the vast majority of UVIGs were targeted by one or two CRISPR spacers from a single host organism (Supplementary figure 7), further confirming the proposed host specificity for members of this family [66]. Pleolipoviruses for which a narrow host range was shown experimentally [10,81], also appear very host specific in this analysis. HFPV-1, which was shown experimentally to successfully infect distant-related host organisms, even crossing the family barrier [14], was in accordance with the experimental data an extreme outlier being targeted by 13 different host species.

For those viruses with a predicted host, almost 90% of the hosts belong to three families within the *Halobacteriales* order; the *Haloarculaceae, Haloferacaceae* and the *Natrialbaceae* family (39.2%, 31.2% and 20% respectively, Supplementary Table 3). The most represented genera are *Haloarcula* (28%), *Halorurbum* (16%), *Natrinema* (11.2%) and *Haloferax* (9.6%). All but one of the remaining UVIGs were associated with other genera within the *Halobacteriales* order. Surprisingly, one UVIG (Pink NODE 7523) was associated with the *Flavobacteriaceae* family (Iphop score 91.8), mainly supported by the protein content-based prediction implemented in RaFAH [65]. Interestingly, this UVIG (Pink NODE 7523) was retrieved from Pink Lake metagenomic data, and is related to the divergent gammapleolipovirus-like genomes, for whom no host could be predicted (Figure 4, Supplementary Figure 8). These UVIGs do not share any of the non-structural hallmark genes with any of the currently accepted nor proposed genera, with most of its proteins lacking a functional annotation. A pleolipovirus infecting a bacterial host has not yet been reported, and the morphologically similar bacteria-infecting phage mycoplasma virus L172 is not phylogenetically related to

pleolipoviruses. [66,69,82]. It is therefore unlikely, but not impossible, that this pleolipovirus would infect a bacterial host. We suggest that this and other gamma-like genomes are likely very divergent members within the *Pleolipoviridae* family that form their own individual lineages. In the particular case of the UVIG Pink NODE 7523, it could have evolved through recombining directly with some bacterial-related MGEs, or more likely, emerged as an indirect product of previous horizontal gene transfer (HGT) between bacteria and the archaeal host cell, as HGT has been well documented to occur often between these two domains [83–85].

Host-driven evolution of the structural proteins

To further investigate the impact of the host specificity on the evolution of pleolipoviruses we performed in-silico structural prediction of all retrieved spike proteins using Alphafold 2 [56]. Despite the low levels of similarity at the amino acid sequence among UVIGs, structural comparison of the spike proteins revealed the conservation of the V-shaped fold previously reported for HRPV-2 and HRPV-6 spikes amongst all isolates and UVIGs (Figure 6) [72].

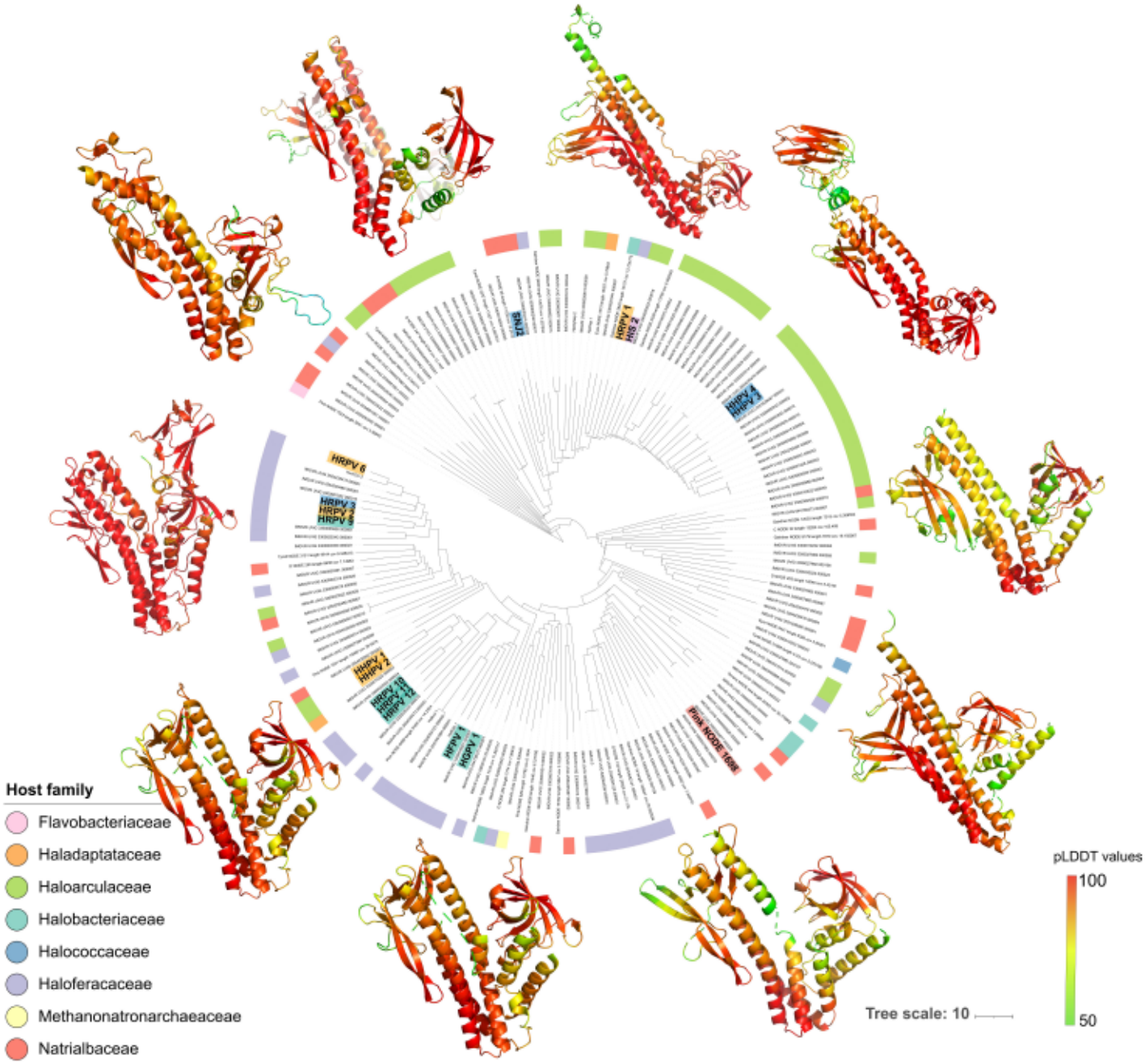


Figure 6: Spike protein structural models. Cladogram of structural models of pleolipovirus spike proteins. Pairwise structural similarities (Z) scores were calculated in DALI. Representative structures for major clades are shown using ribbon representation and colored according to the plDDT values. Colored labels indicate representatives from different genera: *Alphapleolipovirus* (yellow), *Betapleolipovirus* (blue), *Gammapleolipovirus* (magenta), *Deltapleolipovirus* (turquoise) and *Epsilonpleolipovirus* (red). Ring represent the taxonomic affiliation (at family level) of the predicted host for the respective virus.

However, the predicted structures form two distinctive clades unrelated to the pleolipovirus taxonomic classification (Figure 6). In the reported structures for HRPV-2 and HRPV-6 the overall V-fold structure is subdivided in two main domains (N and C terminal) of roughly equal size. The N-terminal domain is again

subdivided in two subdomains: N1, composed of a bundle of 4 α-helices, and N2, composed of an array of beta sheets, with the latter being the proposed receptor binding domain [72]. While the N-terminal N1 subdomain was extremely conserved across all predicted structures, the N2 subdomain was less conserved (several cases of poorly resolved and low plDDT values). Conversely, confidently predicted structures displayed a wider arrange of conformations for N2 (Figure 6), which is likely a result of the monospecific nature of their host range acting as the major evolutionary force.

The structure of the C-terminal domain of the HRPV-2 and HRPV-6 spike proteins display three subdomains (C1-C3) preceding a ~20 amino acid transmembrane α-helix domain that anchors the spike protein to the membrane [72]. While the transmembrane domain was relatively conserved (Supplementary figure 9), the C-terminal domain (C1-C3) appeared rather diverse, as observed for the N2 receptor binding domain. Modeled spike proteins that clustered close with the structures from HRPV-2 and HRPV-6, displayed a similar three-subdomain organization. However, the rest of the spikes displayed mainly a 2-subdomain organization of the C-terminus, with one subdomain being similar to the C3 subdomain (2 to 3 α-helixes) and the second one to the C2 subdomain with beta sheets organized as small beta barrels [86]. The most divergent structures were found in the branch containing mainly pleolipoviruses predicted to infect members of the *Haloarculaceae* family (Figure 6). The modeled spike proteins in this branch exhibit the two short α-helixes C2 subdomain, and the small beta-barrel folded C1 subdomain in close proximity, while an unstructured chain of residues connects the two domains with the transmembrane domain (Supplementary Figure 10).

The overall clustering patterns of the structural dendogram seems to be mostly related to the host taxonomy, with the two main branches relating to the *Haloferacaceae* (classic V-shaped fold) and the *Haloarculaceae* family (shortened C-terminal fold) (Figure 6). Interestingly, the two most variable structural features, i.e. the receptor binding N2 subdomain and the C-terminal domain, are the ones that likely interact closer with the cell surface of the hosts. The archaeal cell is covered by the S-layer glycoprotein, which was shown to be recognized by the N2 subdomain of the spike protein of HRPV-6 and to trigger membrane fusion, the key process during the infection of a new host [80]. However, archaeal S-layers share little sequence similarity even between closely related species. This suggests that the high variability in the N2 subdomains is likely a product of host selective pressures and adaptations to allow initial interaction with a single S-layer protein from a specific host. Additionally, the S-layer likely also interacts with the C-terminal domain of the spike protein while it is anchored to the membrane. The S-layer of *Haloferax volcanii* and other members of the *Haloferacaceae* family was shown to be organized in a dome-shaped manner [87,88], while the triangular archaeon *Haloarcula japonica*, a member of the *Haloarculaceae* family, was proposed to have an arched shaped S-layer structure [88,89]. This suggests, that the organization of the S-layer glycoprotein coat, whether it is a domed for *Haloferacacea*e or an arched organization for *Haloarculaceae*, might be driving the reported

diversity of the C-terminal domain of the spike protein of pleolipoviruses. Altogether, we conclude that the spike protein of pleolipoviruses has evolved driven by the interaction with the S-layer glycoprotein of their specific hosts and it does not represent a reliable marker for the evolution of the *Pleolipoviridae family*. Nonetheless, the host-related clustering patterns observed with the structural comparison have the potential to become a valuable tool to help predict the hosts of yet uncultivated pleolipoviruses and other UVIGs.

Conclusion

In this work, we presented 41 new archaeal pleomorphic virus genomes, which originate from 11 hypersaline lakes in southeast Australia, and mined publicly available datasets to perform a comprehensive study of the *Pleolipoviridae* family, expanding in more than one order of magnitude the current diversity of isolated pleolipoviruses [66]. Similarly to other archaeal viruses, pleolipoviruses display remarkably low sequence similarity, at both nucleotide and protein level, and like haloarchaeal tailed viruses, there seems to be little connection between their geographical origin and their taxonomic organization. Through multiple approaches, we showed that the current taxonomic classification results in several polyphyletic groups, and therefore, propose a new organization for the *Pleolipoviridae* family into 5 different genera: the *Alpha, Beta, Gamma, Delta* and *Epsilonpleolipoviruses*.

Moreover, we showed that the evolutionary history of this family has been repeatedly impacted by recombination events with different groups of mobile MGEs. Particularly features shared with plasmids from diverse origins and environments suggests that these groups of viruses could possibly be found in environments other than previously thought and that pleolipoviruses possibly infect organisms other than haloarchaea. Finally, we observe that the preferred life cycle of pleolipoviruses seems to be productive rather than lysogenic, which further highlights the importance of understanding viral chronic infections and their impact on host metabolism and ecology, which has been scarcely investigated.

Data availability

Raw reads from metagenomic data generated in this study were submitted to ENA-EMBL under project number PRJEB61734.

Acknowledgments

We thank Dr. Renate Dohmen from the Max Planck Computing & Data Facility, Garching, Germany for the infrastructure used for structural prediction with Alphafold2. We thank Daniela Thies and Ingrid Kunze (MPI for Marine Microbiology, Bremen, Germany) for assistance with some of the experiments. We thank Mart Krupovic and Hanna Oksamen for fruitful discussions. Finally, we want to thank the Max-Planck-Institute for Marine Microbiology and the Max-Planck-Society for continuous support.

References

1. Prangishvili D. The Wonderful World of Archaeal Viruses. Annu Rev Microbiol. 2013;67: 565–585. doi:10.1146/annurev-micro-092412-155633

2. Liu Y, Demina TA, Roux S, Aiewsakun P, Kazlauskas D, Simmonds P, et al. Diversity, taxonomy, and evolution of archaeal viruses of the class *Caudoviricetes*. PLoS Biol. 2021;19: 1–26. doi:10.1371/journal.pbio.3001442

3. Koonin E, Dolja V V., Krupovic M, Varsani A, Wolf YI, Yutin N, et al. Global Organization and Proposed Megataxonomy of the Virus World. Microbiol Mol Biol Rev. 2020;84. doi:10.1128/MMBR.00061-19CE

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

5. Iranzo J, Koonin E V., Prangishvili D, Krupovic M. Bipartite Network Analysis of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements. J Virol. 2016;90: 11043–11055. doi:10.1128/jvi.01622-16

6. Iranzo J, Krupovic M, Koonin E V. A network perspective on the virus world. Commun Integr Biol. 2017;10: 1–4. doi:10.1080/19420889.2017.1296614

7. Prangishvili D, Bamford DH, Forterre P, Iranzo J, Koonin E V., Krupovic M. The enigmatic archaeal virosphere. Nat Rev Microbiol. 2017;15: 724–739. doi:10.1038/nrmicro.2017.125

8. Krupovic M, Dolja V V., Koonin E V. Origin of viruses: primordial replicators recruiting capsids from hosts. Nat Rev Microbiol. 2019;17: 449–458. doi:10.1038/s41579-019-0205-6

9. Krupovic M, Koonin E V. Multiple origins of viral capsid proteins from cellular ancestors. Proc Natl Acad Sci. 2017;114: E2401–E2410. doi:10.1073/pnas.1621061114

10. Pietilä MK, Atanasova NS, Manole V, Liljeroos L, Butcher SJ, Oksanen HM, et al. Virion Architecture Unifies Globally Distributed Pleolipoviruses Infecting Halophilic Archaea. J Virol. 2012;86: 5067–5079. doi:10.1128/jvi.06915-11

11. Senčilo A, Paulin L, Kellner S, Helm M, Roine E. Related haloarchaeal pleomorphic viruses contain different genome types. Nucleic Acids Res. 2012;40: 5523–5534. doi:10.1093/nar/gks215

12. Pietilä MK, Roine E, Sencilo A, Bamford DH, Oksanen HM. Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. Arch Virol. 2016;161: 249–256. doi:10.1007/s00705-015-2613-x

13. Pietilä MK, Roine E, Paulin L, Kalkkinen N, Bamford DH. An ssDNA virus infecting archaea: A new lineage of viruses with a membrane envelope. Mol Microbiol. 2009;72: 307–319. doi:10.1111/j.1365-2958.2009.06642.x

14. Alarcón-Schumacher T, Naor A, Gophna U, Erdmann S. Isolation of a virus causing a chronic infection in the archaeal model organism *Haloferax volcanii* reveals antiviral activities of a provirus. Proc Natl Acad Sci. 2022;119: 1–12. doi:https://doi.org/10.1073/pnas.2205037119

15. Kazlauskas D, Varsani A, Koonin E V., Krupovic M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nat Commun. 2019;10: 1–12. doi:10.1038/s41467-019-11433-0

16. Bath C, Cukalac T, Porter K, Dyall-Smith ML. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. Virology. 2006;350: 228–239. doi:10.1016/j.virol.2006.02.005

17. Demina TA, Oksanen HM. Pleomorphic archaeal viruses: the family *Pleolipoviridae* is expanding by seven new species. Arch Virol. 2020;165: 2723–2731. doi:10.1007/s00705-020-04689-1

18. Dyall-Smith M, Pfeiffer F, Chiang P-W, Tang S-L. The Novel Halovirus Hardycor1, and the Presence of Active (Induced) Proviruses in Four Haloarchaea. 2021 [cited 25 Aug 2021]. doi:10.3390/genes12020149

19. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads kenkyuhi hojokin gan rinsho kenkyu jigyo. EMBnet.journal. 2013;17: 10–12. doi:10.14806/ej.17.1.200

20. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012;19: 455–477. doi:10.1089/cmb.2012.0021

21. Hyatt D, Chen G, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal : prokaryotic gene recognition and translation initiation site identification. 2010.

22. Ushnell B. BBMap : A Fast , Accurate , Splice-Aware Aligner. 2014; 3–5.

23. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;2019: 1–13. doi:10.7717/peerj.7359

24. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25: 1043–1055. doi:10.1101/gr.186072.114

25. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35: 725–731. doi:10.1038/nbt.3893

26. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2020;36: 1925–1927. doi:10.1093/bioinformatics/btz848

27. Chen IMA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res. 2019;47: D666–D677. doi:10.1093/nar/gky901

28. Roux S, David P, Chen IA, Palaniappan K, Ratner A, Chu K, et al. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. Nucleic Acids Res. 2020; 1–12. doi:10.1093/nar/gkaa946

29. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. 2015;12. doi:10.1038/nmeth.3176

30. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780. doi:10.1093/molbev/mst010

31. Zhang Z, Wood WI. A profile hidden Markov model for signal peptides generated by HMMER. Bioinformatics. 2003;19: 307–308. doi:10.1093/bioinformatics/19.2.307

32. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. Nucleic Acids Res. 2004;32: 138–141. doi:10.1093/nar/gkh121

33. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2020. doi:10.1038/s41587-020-00774-7

34. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021;9: 1–13. doi:10.1186/s40168-020-00990-y

35. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. You can move, but you can't hide: identification of mobile genetic elements with geNomad. bioRxiv. 2023. doi:https://doi.org/10.1101/2023.03.05.531206

36. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. 2002;30: 2478–2483.

37. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ. 2017;5: e3817. doi:10.7717/peerj.3817

38. Roux S, Adriaenssens EM, Dutilh BE, Koonin E V., Kropinski AM, Krupovic M, et al. Minimum information about an uncultivated virus genome (MIUVIG). Nat Biotechnol. 2019;37: 29–37. doi:10.1038/nbt.4306

39. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. Anal Methods. 2016;8: 12–24. doi:10.1039/c5ay02550h

40. Moraru C, Varsani A, Kropinski AM. VIRIDIC—A novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. Viruses. 2020;12: 1–10. doi:10.3390/v12111268

41. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014;30: 1236–1240. doi:10.1093/bioinformatics/btu031

42. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL 0003 3527 8101, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. 2020;48: 8883–8900. doi:10.1093/nar/gkaa621

43. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018. doi:10.1016/j.jmb.2017.12.007

44. Steinegger M, Mirdita M, Söding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. Nat Methods. 2019;16: 603–606. doi:10.1038/s41592-019-0437-4

45. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47: D309–D314. doi:10.1093/nar/gky1085

46. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38: 5825–5829. doi:10.1093/molbev/msab293

47. Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics. 2017;33: 3396–3404. doi:10.1093/bioinformatics/btx440

48. Bolduc B, Jang H Bin, Doulcier G, You Z-Q, Roux S, Sullivan MB. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect *Archaea* and *Bacteria*. PeerJ. 2017;5: e3243. doi:10.7717/peerj.3243

49. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37: 632–639. doi:10.1038/s41587-019-0100-8

50. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2009.

51. Butts CT. network : A Package for Managing Relational Data in R. 2008;24.

52. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37: 1530–1534. doi:10.1093/molbev/msaa015

53. Hoang DT, Chernomor O, Haeseler A Von, Minh BQ, Vinh LS. UFBoot2 : Improving the Ultrafast Bootstrap Approximation. 2017;35: 518–522. doi:10.5281/zenodo.854445

54. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14: 587–589. doi:10.1038/nmeth.4285

55. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019;47: 256–259. doi:10.1093/nar/gkz239

56. Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Ballard AJ, Cowie A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596. doi:10.1038/s41586-021-03819-2

57. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50: D439–D444. doi:10.1093/nar/gkab1061

58. Announcing the worldwide Protein Data Bank. 2003;10: 2003.

59. L. DW. The PyMOL molecular graphics system. http://www.pymol.org/. 2002 [cited 16 Feb 2023]. Available: https://cir.nii.ac.jp/crid/1570572701247244160.bib?lang=en

60. Holm L. Dali server: structural unification of protein families. Nucleic Acids Res. 2022;50: W210–W215. doi:10.1093/nar/gkac387

61. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, et al. iPHoP: an integrated machine-learning framework to maximize host prediction for metagenome-assembled virus genomes. 2022. Available: https://doi.org/10.1101/2022.07.28.501908

62. Lu C, Zhang Z, Cai Z, Zhu Z, Qiu Y, Wu A, et al. Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. BMC Biol. 2021;19: 1–11. doi:10.1186/s12915-020-00938-6

63. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d2∗ oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res. 2017;45: 39–53. doi:10.1093/nar/gkw1002

64. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics. 2017;33: 3113–3114. doi:10.1093/bioinformatics/btx383

65. Coutinho FH, Zaragoza-Solas A, López-Pérez M, Barylski J, Zielezinski A, Dutilh BE, et al. RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. Patterns. 2021;2. doi:10.1016/j.patter.2021.100274

66. Liu Y, Dyall-Smith M, Oksanen HM. ICTV Virus Taxonomy Profile: *Pleolipoviridae* 2022. J Gen Virol. 2022;103: 1–2. doi:10.1099/jgv.0.001793

67. Medvedeva S, Borrel G, Krupovic M, Gribaldo S. A global virome of methanogenic archaea highlights novel diversity and adaptations to the gut environment. Res Sq. 2023; 1–30. doi:https://doi.org/10.21203/rs.3.rs-2539466/v1 License:

68. Liu J, Shen Q, Id HB. PLOS ONE Comparison of seven SNP calling pipelines for the next-generation sequencing data of chickens. 2022; 1–14. doi:10.1371/journal.pone.0262574

69. Bamford DH, Pietilä MK, Roine E, Atanasova NS, Dienstbier A, Oksanen HM. ICTV virus taxonomy profile: *Pleolipoviridae*. J Gen Virol. 2017;98: 2916–2917. doi:10.1099/jgv.0.000972

70. Demina TA, Pietilä MK, Svirskaitė J, Ravantti JJ, Atanasova NS, Bamford DH, et al. HCIV-1 and other tailless icosahedral internal membrane-containing viruses of the family *Sphaerolipoviridae*. Viruses. 2017;9. doi:10.3390/v9020032

71. Krupovic M, Forterre P. Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. Ann N Y Acad Sci. 2015;1341: 41–53. doi:10.1111/nyas.12675

72. Omari K El, Li S, Kotecha A, Walter TS, Bignon EA, Harlos K, et al. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. doi:10.1038/s41467-019-08728-7

73. Lipps G, Weinzierl AO, Von Scheven G, Buchen C, Cramer P. Structure of a bifunctional DNA primase-polymerase. Nat Struct Mol Biol. 2004;11: 157–162. doi:10.1038/nsmb723

74. Jaakkola ST, Penttinen RK, Vilén ST, Jalasvuori M, Rönnholm G, Bamford JKH, et al. Closely Related Archaeal Haloarcula hispanica Icosahedral Viruses HHIV-2 and SH1 Have Nonhomologous Genes Encoding Host Recognition Functions. J Virol. 2012;86: 4734–4742. doi:10.1128/jvi.06666-11

75. Taylor K, Węgrzyn G. Replication of coliphage lambda DNA. FEMS Microbiol Rev. 1995;17: 109–119. doi:10.1016/0168-6445(95)00077-1

76. Dyall-Smith M, Pfeifer F, Witte A, Oesterhelt D, Pfeiffer F. Complete genome sequence of the model halovirus phiH1 (ΦH1). Genes (Basel). 2018;9: 1–20. doi:10.3390/genes9100493

77. Kala S, Cumby N, Sadowski PD, Hyder BZ, Kanelis V, Davidson AR, et al. HNH proteins are a widespread component of phage DNA packaging machines. Proc Natl Acad Sci U S A. 2014;111: 6022–6027. doi:10.1073/pnas.1320952111

78. Gorlas A, Krupovic M, Forterre P, Geslin C. Living side by side with a virus: Characterization of two novel plasmids from *Thermococcus prieurii*, a host for the spindle-shaped virus TPV1. Appl Environ Microbiol. 2013;79: 3822–3828. doi:10.1128/AEM.00525-13

79. El Omari K, Li S, Kotecha A, Walter TS, Bignon EA, Harlos K, et al. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. Nat Commun. 2019;10. doi:10.1038/s41467-019-08728-7

80. Bignon EA, Chou KR, Roine E. Halorubrum pleomorphic virus-6 Membrane Fusion Is Triggered by an S-Layer Component of Its Haloarchaeal Host. 2022.

81. Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM. Global network of specific virus-host interactions in hypersaline environments. Environ Microbiol. 2012;14: 426–440. doi:10.1111/j.1462-2920.2011.02603.x

82. Dybvig K, Nowak JA, Sladek TL, Maniloff J. Identification of an Enveloped Phage, Mycoplasma Virus L172, that Contains a 14-Kilobase Single-Stranded DNA Genome. J Virol. 1985;53: 384–390. doi:10.1128/microbe.8.242.2

83. Fuchsman CA, Collins RE, Rocap G, Brazelton WJ. Effect of the environment on horizontal gene transfer between bacteria and archaea. PeerJ. 2017;2017: 1–26. doi:10.7717/peerj.3865

84. Mongodin EF, Nelson KE, Daugherty S, DeBoy RT, Wister J, Khouri H, et al. The genome of Salinibacter ruber: Convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc Natl Acad Sci U S A. 2005;102: 18147–18152. doi:10.1073/pnas.0509073102

85. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, et al. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of Thermotoga maritima. Nature. 1999;399: 323–329. doi:10.1038/20601

86. Youkharibache P, Veretnik S, Li Q, Stanek KA, Mura C, Bourne PE. The Small β-Barrel Domain: A Survey-Based Structural Analysis. Structure. 2019;27: 6–26. doi:10.1016/j.str.2018.09.012

87. von Kügelgen A, Alva V, Bharat TAM. Complete atomic structure of a native archaeal cell surface. Cell Rep. 2021;37. doi:10.1016/j.celrep.2021.110052

88. Rodrigues-Oliveira T, Belmok A, Vasconcellos D, Schuster B, Kyaw CM. Archaeal S-layers: Overview and current state of the art. Front Microbiol. 2017;8: 1–17. doi:10.3389/fmicb.2017.02597

89. Horikoshi K, Aono R, Nakamura S. The triangular halophilic archaebacterium Haloarcula japonica strain TR-1. Experientia. 1993;49: 497–502. doi:10.1007/BF01955151

Supplementary Material

Revisiting the evolutionary trajectories and organization of the *Pleolipoviridae* family

Tomas Alarcón-Schumacher[1*], Dominik Lücking[1] and Susanne Erdmann[1*]

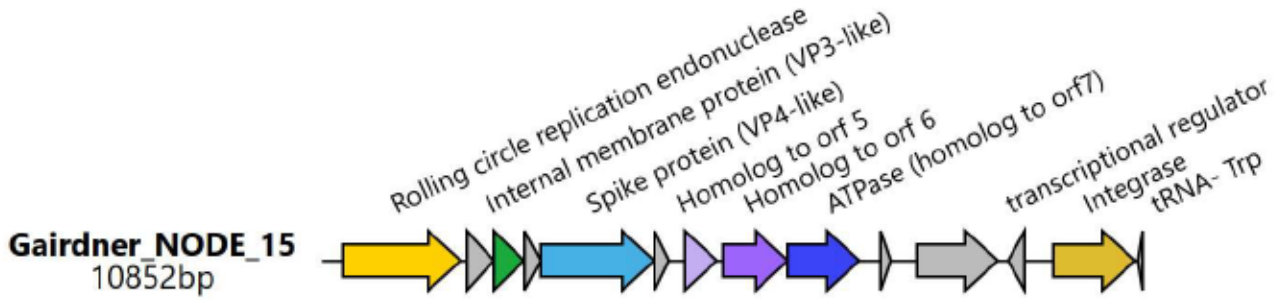
[1] Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany

Supplementary Information

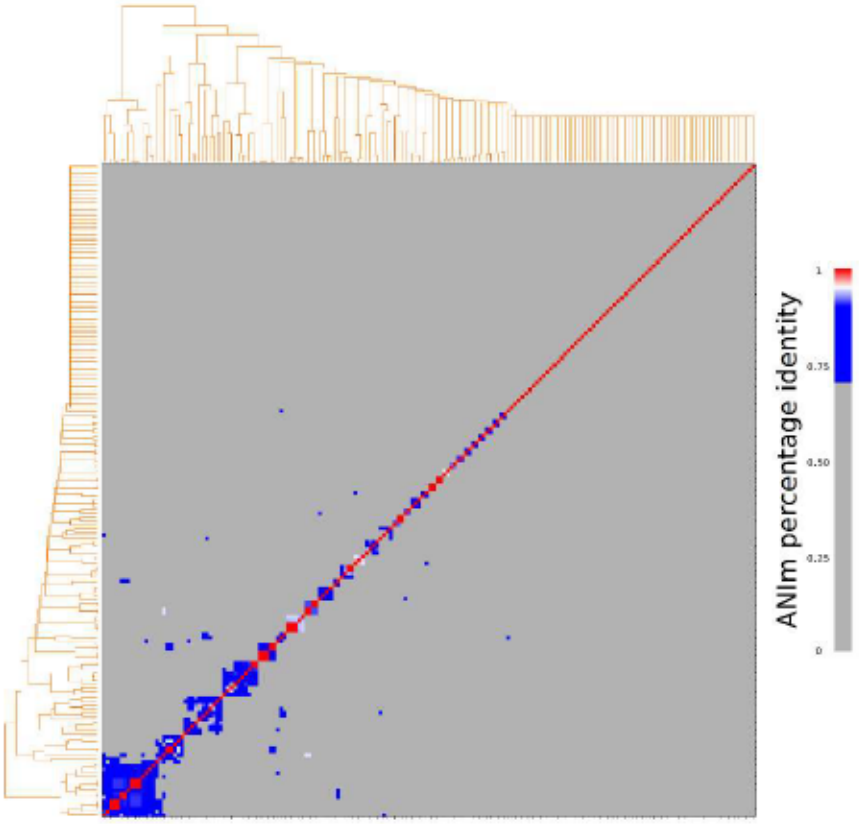
Supplementary Figures

Supplementary Tables

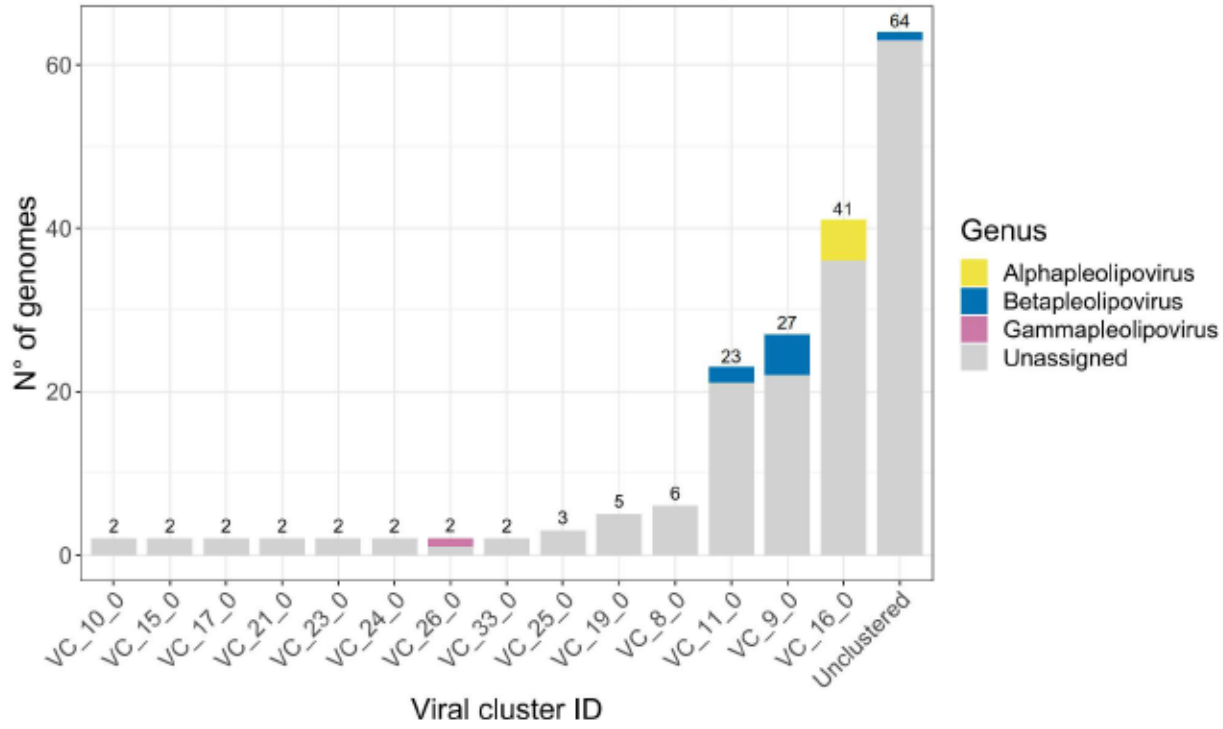
Supplementary Figures



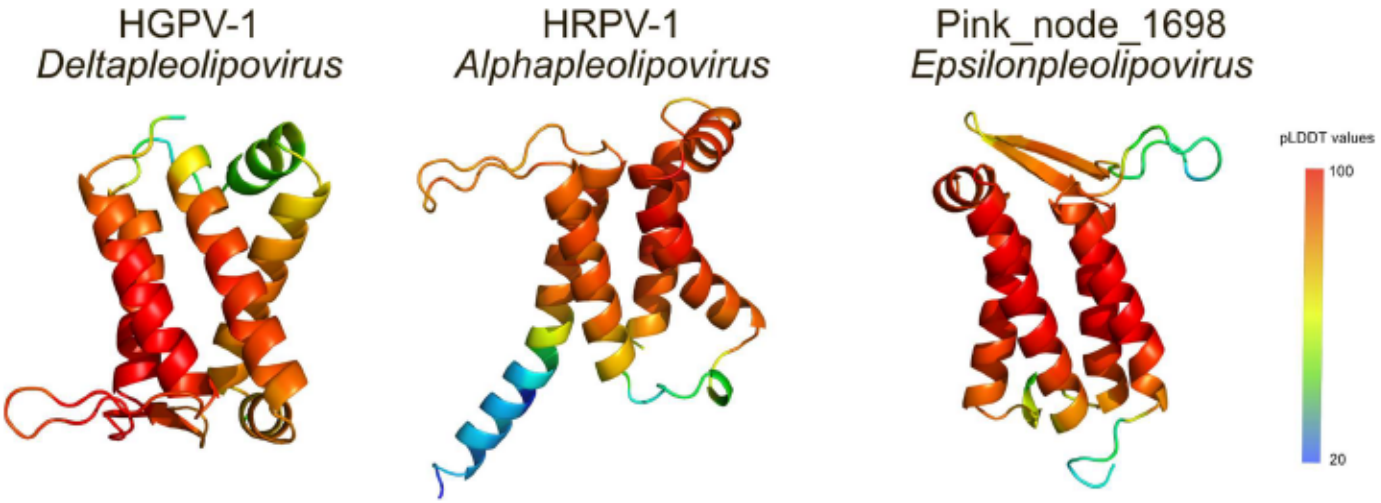
Supplementary Figure 1: Provirus genome. Genomic representation of newly identified Halobaculum pleolipovirus-like provirus. Homologues of conserved genes are colored the same as follows: Rolling circle replication endonuclease (RCRE) (yellow), VP3-like protein (green), Spike protein (light blue), ORF5-like (purple), ORF6-like (magenta), ATPase (blue) and integrase (light brown). The tRNA corresponding to tryptophan depicts the insertion site.



Supplementary Figure 2: Pleolipovirus genome similarity. Average nucleotide identity (ANI) values were calculated with PyANI using MUMner as alignment method. Values are displayed as percentage identity (ANIm).

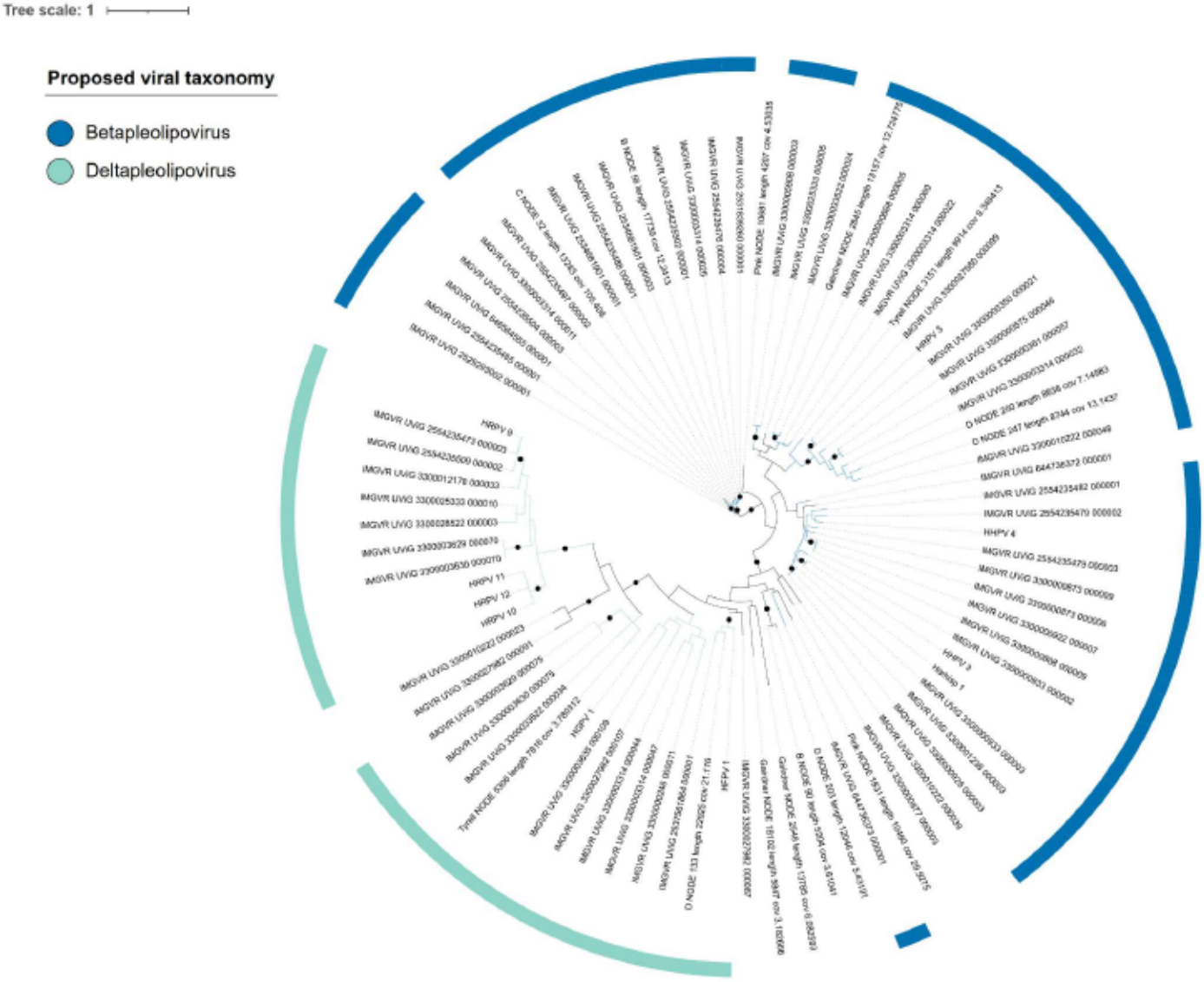


Supplementary figure 3: Viral clusters. Pleolipovirus-like genomes were clustered based on their shared protein content with vConTACT2. Numbers on top of bars indicate the number of pleolipovirus genomes in the respective cluster. Colored labels represent isolated representatives from each genus: *Alphapleolipovirus* (yellow), *Betapleolipovirus* (blue), *Gammapleolipovirus* (magenta). Genomes classified as singletons, outliers and belonging to overlapping clusters were grouped under the category "Unclustered".

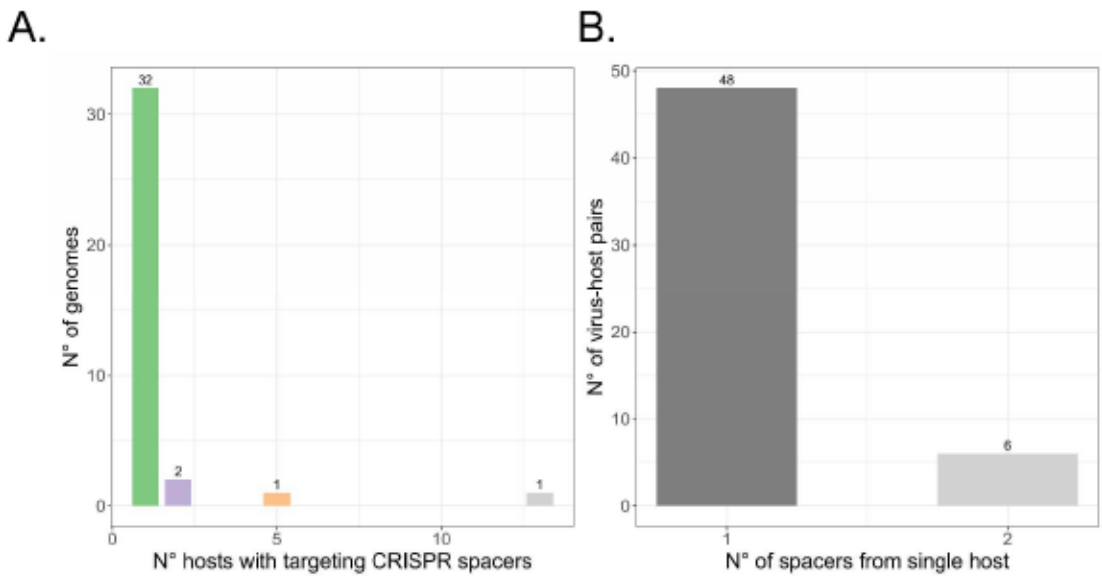


Supplementary figure 4: Internal membrane protein comparison. Structure prediction of the internal membrane protein type I (ORF 2 in Haloferax pleomorphic virus 1) generated with AlphaFold2. Representative structures the genera *Alphapleolipovirus* (HRPV-1), *Deltapleolipovirus* (HGPV-1) and *Epsilonpleolipovirus* (Pink_Node_1698) are shown using ribbon representation and colored according to the plDDT values.

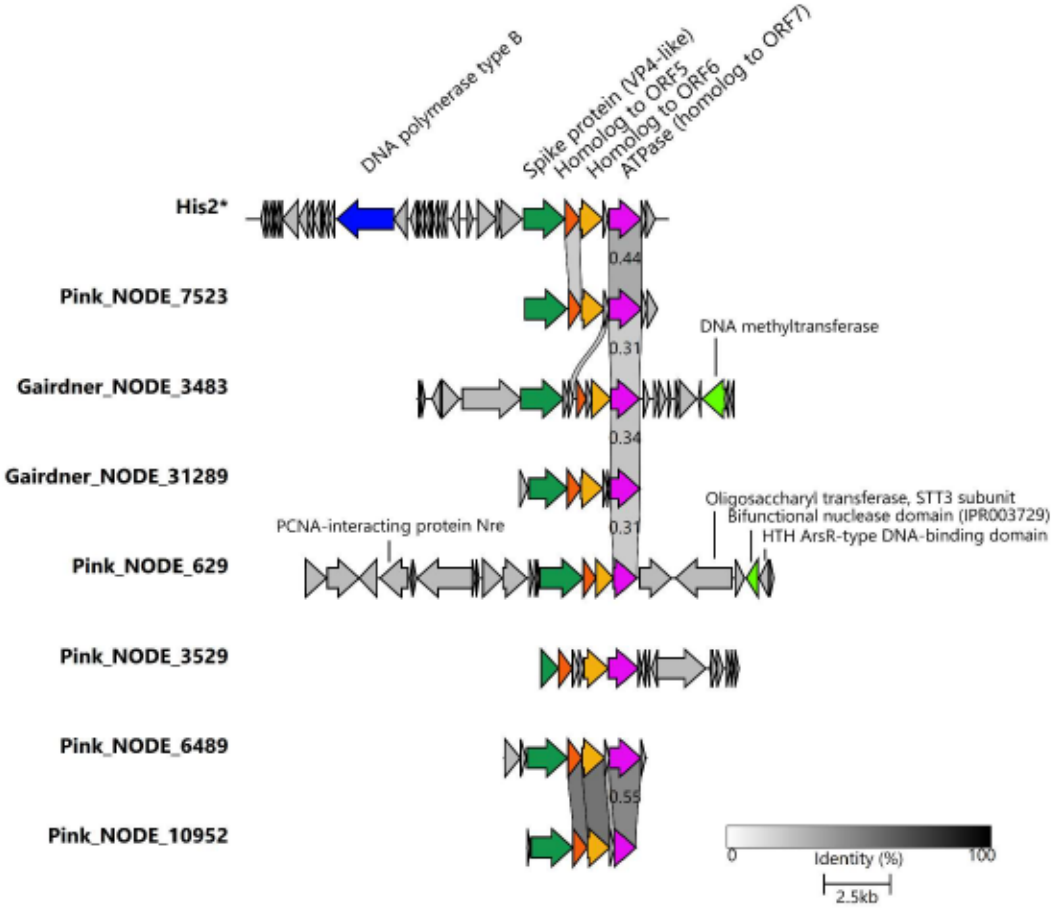
Supplementary figure 5: Phylogeny conserved Rep Beta. Phylogenetic tree reconstruction of the proposed replication-like proteins of betapleolipoviruses. Sequences were obtained from the 184 pleolipovirus-like genomes dataset generated in this study. Tree was constructed with iqtree with 10.000 ultrafast bootstrap. Supported branches (SH-aLRT >= 80 and ultrafast bootstrap >= 95) are demarcated with black circles. Scale bar represents the number of substitutions every 100 amino acids.



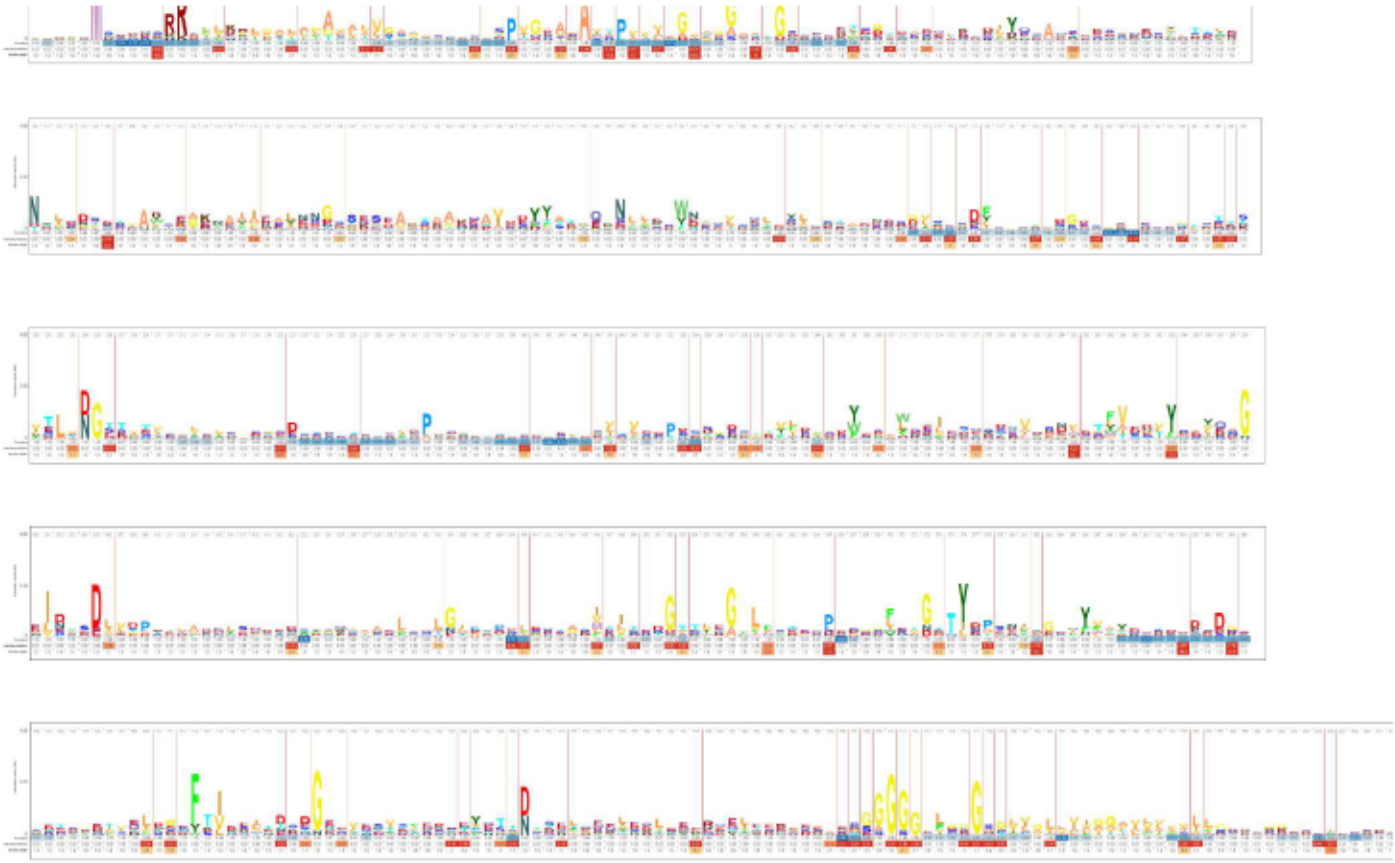
Supplementary figure 6: Phylogeny conserved ORF8. Phylogenetic tree reconstruction of ORF8-like proteins. Sequences were obtained from the 184 pleolipovirus-like genomes dataset generated in this study. ORF8 was designated according to the annotation of Haloferax pleomorphic virus 1 (HFPV-1). Tree was constructed with iqtree with 10.000 ultrafast bootstrap. Supported branches (SH-aLRT >= 80 and ultrafast bootstrap >= 95) are demarcated with black circles. Scale bar represents the number of substitutions every 100 amino acids.



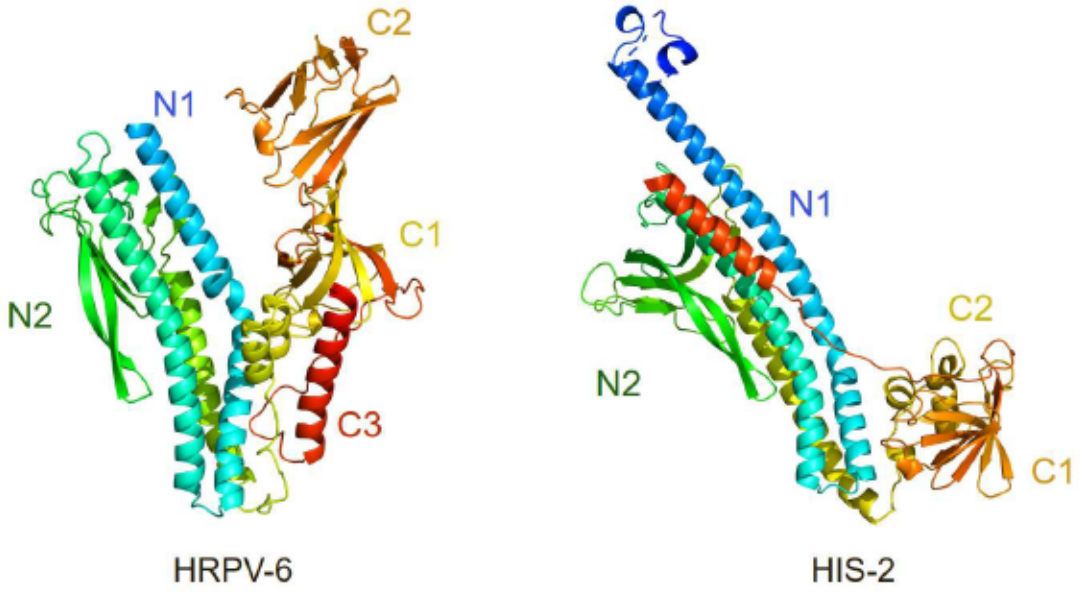
Supplementary figure 7: CRISPR spacers targeting. CRISPRs spacers were queried using blastn against the 184 pleolipovirus genomes from this study to assess the potential host range. **A.** Number of pleolipovirus genomes targeted by different hosts. **B.** Number of spacers from a specific host CRISPR array(s) for each one of the identified virus-host pairs. Numbers on top of the bars indicate the number of genomes and virus-host pairs respectively for A and B.



Supplementary figure 8: The gamma-like pleolipovirus. Genomic alignment of genomes related to the gammapleolipovirus genus. Similarity values (blastp) are indicated by grayscale shading. Homologues of conserved genes are colored the same as follows: Spike protein (dark green), ORF5-like (red), ORF6-like (orange), ATPase (magenta), type B DNA polymerase (blue), and DNA methyl transferase (light green). Complete genomes are highlighted with (*) symbol.



Supplementary figure 9: HMM logo spike protein. Multiple sequence alignment of the Spike protein. Sequences were obtained from the 184 pleolipovirus-like genomes dataset generated in this study and aligned using MAFFT v.7. Height of the letter depict the relative amino acid frequencies for each position of the alignment.



Supplementary figure 10: Spike protein structural comparison. Structure prediction of the spike protein of the *Alphapleolipovirus* HRPV-6 and the *Gammapleolipovirus* HIS-2 type I generated with AlphaFold2. Representative structures for major clades are shown using ribbon representation and colored using rainbow scheme from the N-terminus (blue) to the C-terminus (red). C1-C3 and N1-N2 indicate the protein subdomains.

Supplementary Tables

Supplementary Table 1: Pleolipovirus-like elements from Australian hypersaline lakes. Excel file of Supplementary Table 1 summarizes the completeness and quality assessment performed with Checkv for all pleolipovirus-like contigs detected in the metagenomic data generated in this study (minimum contig length 2kb). Average values of length, viral and total gene count, host genes, completeness and contamination are indicated at the bottom of each respective column. Database version 1.4 (Aug 27, 2022) was used.

Contig id	Contig length	Provirus	Proviral length	Gene count	Viral genes	Host genes	Checkv quality	Completeness	Completeness method	Contamination
becking_node_22801_length_2479_cov_2.62789	2479	yes	1411	4	2	1	low-quality	7.34	hmm-based (lower-bound)	43.08
becking_node_3020_length_6891_cov_15.2989	6891	yes	1680	10	2	3	low-quality	8.74	hmm-based (lower-bound)	75.62
frome_node_10076_length_5670_cov_8.711665	5670	yes	3132	8	3	1	low-quality	25.61	aai-based (medium-confidence)	44.76
frome_node_32810_length_3161_cov_3.311333	3161	yes	2511	6	2	1	low-quality	13.06	hmm-based (lower-bound)	20.56
frome_node_9743_length_5764_cov_5.160799	5764	yes	3863	7	4	1	low-quality	37.34	aai-based (medium-confidence)	32.98
gairdner_node_117469_length_2456_cov_1.588505	2456	yes	972	5	1	1	low-quality	5.05	hmm-based (lower-bound)	60.42
gairdner_node_117612_length_2454_cov_4.587328	2454	yes	1218	3	2	1	low-quality	6.33	hmm-based (lower-bound)	50.37
gairdner_node_12003_length_7013_cov_5.306554	7013	yes	5714	8	4	1	low-quality	29.71	hmm-based (lower-bound)	18.52
gairdner_node_133377_length_2311_cov_1.919770	2311	yes	1554	3	1	1	low-quality	8.08	hmm-based (lower-bound)	32.76
gairdner_node_161349_length_2110_cov_2.225791	2110	yes	1163	3	2	1	low-quality	6.05	hmm-based (lower-bound)	44.88
gairdner_node_47416_length_3769_cov_2.951804	3769	yes	2903	5	3	1	low-quality	15.1	hmm-based (lower-bound)	22.98
gairdner_node_54453_length_3532_cov_3.622663	3532	yes	2910	5	3	1	low-quality	15.13	hmm-based (lower-bound)	17.61
gairdner_node_99137_length_2660_cov_3.892514	2660	yes	895	2	1	1	low-quality	4.65	hmm-based (lower-bound)	66.35
hardy_node_3119_length_7100_cov_4.11682	7100	yes	6439	8	4	1	low-quality	39.85	aai-based (medium-confidence)	9.31
hardy_node_5076_length_5415_cov_3.31884	5415	yes	1322	6	1	4	low-quality	6.87	hmm-based (lower-bound)	75.59
hardy_node_6155_length_4868_cov_2.31186	4868	yes	2283	6	3	2	low-quality	11.87	hmm-based (lower-bound)	53.1
kenyon_node_30379_length_2118_cov_16.8396	2118	yes	1369	3	2	1	low-quality	9.81	aai-based (medium-confidence)	35.36
pink_node_35867_length_2230_cov_2.83494	2230	yes	1436	4	1	1	low-quality	7.47	hmm-based (lower-bound)	35.61
torrens_node_37947_length_3258_cov_2.351858	3258	yes	2238	4	2	1	low-quality	11.64	hmm-based (lower-bound)	31.31
torrens_node_644_length_20303_cov_30.778892	20303	yes	8615	18	5	4	low-quality	44.8	hmm-based (lower-bound)	57.57
tyrell_node_10395_length_5740_cov_4.475638	5740	yes	1809	10	1	4	low-quality	9.41	hmm-based (lower-bound)	68.48
tyrell_node_20807_length_4199_cov_5.124276	4199	yes	791	4	1	1	low-quality	4.11	hmm-based (lower-bound)	81.16

tyrell_node_27598_length_3684_cov_2.657757	3684	yes	3149	5	2	1	low-quality	16.37	hmm-based (lower-bound)	14.52
tyrell_node_49934_length_2789_cov_4.146306	2789	yes	2141	5	2	1	low-quality	11.13	hmm-based (lower-bound)	23.23
tyrell_node_50790_length_2766_cov_3.779786	2766	yes	1471	4	2	1	low-quality	7.65	hmm-based (lower-bound)	46.82
tyrell_node_83240_length_2185_cov_1.699061	2185	yes	1597	3	2	1	low-quality	8.3	hmm-based (lower-bound)	26.91
Averages	4497.12		2484.08	5.73	2.23	1.46		14.29		41.92

Supplementary Table 2: Metagenomic assembled genomes from Australian hypersaline lakes. Excel file of Supplementary Table 2 summarizes the taxonomic affiliation and the quality assessment of metagenomic assembled genomes (MAGs) retrieved from the metagenomic data generated in this study. Completeness and contamination analysis were performed with Checkm. High and medium-quality MAGs were then selected for further taxonomic classification with the GTDB-Tk toolkit v1.4.0. High-quality and medium quality MAGs with completion > 85% are displayed.

MAG id	Classification gtdb-tk	Completeness	Contamination
eyresouth_bin_33	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__Natronomonas;s__	98.3	3.19
eyre_bin_18	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__Natronomonas;s__	97.8	3.19
gairdner_bin_91	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haladaptataceae;g__Halorussus;s__	97.63	0.43
pink_bin_33	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Wenzhouxiangellaceae;g__GCA-2722315;s__	97.28	2.52
bumbago_bin_30	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__;f__;g__;s__	94.51	5.49
gairdner_bin_97	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Nitriliruptorales;f__Nitriliruptoraceae;g__PWLR01;s__	94.44	0.85
torrens_bin_87	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__Natronomonas;s__	93.94	2.13
pink_bin_17	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__Natronomonas;s__	93.4	1.41
pink_bin_37	d__Archaea;p__Halobacteriota;c__Methanonatronarchaeia;o__Methanonatronarchaeales;f__PZKD01;g__PZKD01;s__	92.62	0.98
frome_bin_11	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloferacaceae;g__Halogeometricum;s__	92.53	1.96
pink_bin_51	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloferacaceae;g__J07HR59;s__	92.46	2.33
tyrell_bin_20	d__Bacteria;p__Deferribacterota;c__Deferribacteres;o__Deferribacterales;f__Flexistipitaceae;g__Flexistipes;s__	92.19	1.29
hardy_bin_42	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__;g__;s__	91.73	0
hardy_bin_47	d__Archaea;p__Halobacteriota;c__Methanosarcinia;o__Methanosarcinales;f__Methanosarcinaceae;g__Methanohalobium;s__Methanohalobium evestigatum	91.58	0.65
eyre_bin_21	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__;g__;s__	91.4	1.83
pink_bin_30	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__Halovenus;s__	91.21	1.6
hardy_bin_45	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__PXRE01;g__PXRE01;s__	90.61	3.22
tyrell_bin_61	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Salinarchaeaceae;g__Salinarchaeum;s__	90.6	1.42
kenyon_bin_14	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__;s__	90.46	2.6
kenyon_bin_40	d__Archaea;p__Hadarchaeota;c__Hadarchaeia;o__;f__;g__;s__	90.4	4.8
eyre_bin_10	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__;s__	90.21	5.04
frome_bin_38	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__;g__;s__	90.13	1.83
gairdner_bin_42	d__Archaea;p__Halobacteriota;c__Methanonatronarchaeia;o__;f__;g__;s__	89.98	1.96
gairdner_bin_89	d__Bacteria;p__Lindowbacteria;c__T1Sed10-126;o__T1Sed10-126;f__;g__;s__	89.83	0
eyre_bin_57	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__XJ16;f__Halofilaceae;g__Halofilum;s__	89.77	1.84

pink_bin_7	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__SW-7-71-33;g__;s__	89.68	4.65
torrens_bin_51	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__CBA1134;s__	89.53	0.85
gairdner_bin_15	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloferacaceae;g__Halobaculum;s__	89.28	2.38
kenyon_bin_31	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloferacaceae;g__Halorubrum; s__Halorubrum sp000496175	88.9	6.5
eyresouth_bin_24	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__PXRE01;g__PXRE01;s__	88.79	3.23
torrens_bin_22	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__;f__;g__;s__	88.46	4.4
eyresouth_bin_35	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Halobacteriaceae;g__Halodesulfurarchaeum;s__	88.26	4.38
eyresouth_bin_6	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__;g__;s__	88.08	2.63
bumbago_bin_32	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__;g__;s__	88.06	0.65
hardy_bin_33	d__Archaea;p__Hadarchaeota;c__Hadarchaeia;o__;f__;g__;s__	88	5.6
frome_bin_31	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Halococcaceae;g__Halococcus;s__	87.86	2.2
tyrell_bin_34	d__Bacteria;p__Desulfobacterota;c__Desulfobacteria;o__Desulfobacterales;f__Desulfobacteraceae;g__;s__	87.74	0
becking_bin_5	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloferacaceae;g__A07HB70;s__A07HB70 sp000496195	87.6	0.98
kenyon_bin_3	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloarculaceae;g__;s__	87.5	1.7
eyresouth_bin_34	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA5335;f__UBA5335;g__;s__	87.37	0.27
frome_bin_62	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__SKSH01;g__SKSH01;s__	87.28	2.67
frome_bin_14	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Mycobacteriales;f__SCTD01;g__SCTD01;s__	86.78	1.08
gairdner_bin_147	d__Archaea;p__Thermoproteota;c__Korarchaeia;o__;f__;g__;s__	86.22	0.93
becking_bin_30	d__Archaea;p__Hadarchaeota;c__Hadarchaeia;o__;f__;g__;s__	85.98	4.67
gairdner_bin_93	d__Archaea;p__Hadarchaeota;c__Hadarchaeia;o__;f__;g__;s__	85.98	0.93
tyrell_bin_19	d__Bacteria;p__Deferribacterota;c__Deferribacteres;o__Deferribacterales;f__Flexistipitaceae;g__Flexistipes;s__	85.79	0
tyrell_bin_93	d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__;f__;g__;s__	85.71	4.95
gairdner_bin_41	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__Haloferacaceae;g__Halopenitus;s__	85.67	0.95
kenyon_bin_26	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__SW-7-71-33;g__;s__	85.47	2.77
pink_bin_54	d__Bacteria;p__Planctomycetota;c__Brocadiae;o__SM23-32;f__SM23-32;g__;s__	85.26	8.6
hardy_bin_27	d__Archaea;p__Halobacteriota;c__Halobacteria;o__Halobacteriales;f__SW-7-71-33;g__;s__	85.02	3.58

Supplementary Table 3: Pleolipovirus genomes compiled data. Supplementary Table 3 summarizes the integrated data from the analysis performed in this study for each high-confidence pleolipovirus genome identified in samples from Australian hypersaline lakes metagenomes. Integrated completeness assessment is the result of the combination of the following criteria: Genomes were then considered complete if they presented terminal repeats (TRs), were identified as complete integrated proviruses, or if they displayed CheckV completeness value = 100. The Proposed Genus taxonomy column integrated the phylogenetic reconstruction of multiple marker genes (see main text results and discussion). Host prediction was performed with Iphop, and only results with a minimum confidence score of 90 are displayed. Topology values and main viral groups (single-stranded or double-stranded DNA) were predicted with Virsorter2.

Virus id	Contig length	Gene count	Viral genes	Integrated completion	Proposed taxonomy	Host genus	Confidence score	Geographic origin	Group virsorter2 (vs2)
b_node_56_length_17739_cov_12.2413	17739	29	9	complete	Beta	Natrinema	90.4	Lake Crosbie	dsDNAphage
b_node_90_length_5204_cov_3.61041	5204	9	5	Partial	Beta		0	Lake Crosbie	dsDNAphage
bumbago_node_3572_length_6235_cov_3.70081	6235	7	5	Partial	Alpha		0	Lake Bumbunga	dsDNAphage
c_node_32_length_13283_cov_105.408	13283	16	7	Partial	Beta	Haloterrigena	93.1	Lake Crosbie	dsDNAphage
d_node_133_length_22925_cov_21.176	22925	28	14	complete	Delta	Haloferax	96.7	Lake Becking	dsDNAphage
d_node_203_length_12046_cov_5.43191	12046	18	5	Partial	Beta	Natrinema	93.5	Lake Becking	dsDNAphage
d_node_247_length_8744_cov_13.1437	8744	11	7	complete	Beta	Natrinema	94.5	Lake Becking	dsDNAphage
d_node_250_length_8638_cov_7.14983	8638	8	7	complete	Beta	Natrinema	93.9	Lake Becking	dsDNAphage
d_node_266_length_7774_cov_7.06918	7774	10	7	Partial	Alpha	Haloferax	90.1	Lake Becking	ssDNA
eyre_node_1573_length_15027_cov_5.79842	15027	20	5	complete	Alpha	Halorussus	90.9	Lake Eyre	dsDNAphage
eyre_node_5421_length_8326_cov_3.83001	8326	13	4	Partial	Unassigned		0	Lake Eyre	dsDNAphage
frome_node_3245_length_9832_cov_9.545771	9832	10	4	Partial	Epsilon	Haloarcula	94.5	Lake Frome	dsDNAphage
gairdner_node_10822_length_7318_cov_11.667217	7318	12	5	complete	Alpha		0	Lake Gairdner	dsDNAphage
gairdner_node_12003_length_7013_cov_5.306554	5714	7	4	Partial	Epsilon		0	Lake Gairdner	dsDNAphage
gairdner_node_15_length_128941_cov_25.552636	10852	13	5	complete	Alpha	Halobaculum	92.1	Lake Gairdner	dsDNAphage
gairdner_node_16940_length_6025_cov_4.241206	6025	8	4	Partial	Alpha	Halomicroarcula	92.2	Lake Gairdner	dsDNAphage
gairdner_node_18102_length_5847_cov_3.182666	5847	8	6	Partial	Delta		0	Lake Gairdner	dsDNAphage
gairdner_node_21431_length_5420_cov_4.609506	5420	12	4	Partial	Epsilon		0	Lake Gairdner	dsDNAphage
gairdner_node_2548_length_13785_cov_6.082593	13785	22	7	Partial	Betapleolipovirus		0	Lake Gairdner	dsDNAphage
gairdner_node_2845_length_13157_cov_12.724775	13157	15	7	Partial	Beta	Halobacterium	91.6	Lake Gairdner	dsDNAphage
gairdner_node_31289_length_4562_cov_7.348791	4562	7	4	Partial	Unassigned		0	Lake Gairdner	ssDNA
gairdner_node_3483_length_12076_cov_7.357042	12076	13	5	Partial	Unassigned		0	Lake Gairdner	dsDNAphage
gairdner_node_4036_length_11340_cov_6.727780	11340	16	9	Partial	Alpha		0	Lake Gairdner	dsDNAphage

gairdner_node_9179_length_7870_cov_19.192067	7870	12	8	complete	Delta		0	Lake Gairdner	ssDNA
gairdner_node_9457_length_7768_cov_3.119279	7768	10	4	Partial	Epsilon	Halorussus	91.6	Lake Gairdner	dsDNAphage
hardy_node_3119_length_7100_cov_4.11682	6439	7	4	Partial	Delta	Halapricum	90.7	Lake Hardy	dsDNAphage
pink_node_10681_length_4207_cov_4.53035	4207	5	5	Partial	Beta		0	Pink Lake	dsDNAphage
pink_node_10952_length_4159_cov_2.44298	4159	5	4	Partial	Unassigned		0	Pink Lake	dsDNAphage
pink_node_1698_length_10852_cov_6.06659	10852	19	5	complete	Epsilon		0	Pink Lake	dsDNAphage
pink_node_1831_length_10460_cov_29.9275	10460	21	7	complete	Beta		0	Pink Lake	dsDNAphage
pink_node_3529_length_7511_cov_5.60703	7511	12	4	Partial	Unassigned		0	Pink Lake	dsDNAphage
pink_node_4856_length_6343_cov_3.26606	6343	11	4	Partial	Epsilon	Natrinema	94.5	Pink Lake	dsDNAphage
pink_node_629_length_17795_cov_6.1394	17795	18	4	Partial	Unassigned		0	Pink Lake	dsDNAphage
pink_node_6489_length_5459_cov_14.2324	5459	8	5	Partial	Unassigned		0	Pink Lake	ssDNA
pink_node_7523_length_5051_cov_3.56845	5051	7	4	Partial	Unassigned	Flavobacterium	91.8	Pink Lake	ssDNA
torrens_node_644_length_20303_cov_30.778892	8615	9	5	Partial	Epsilon		0	Lake Torrens	dsDNAphage
tyrell_node_11156_length_5565_cov_5.212704	5565	8	4	Partial	Unassigned		0	Lake Tyrell	dsDNAphage
tyrell_node_21668_length_4123_cov_5.570796	4123	5	5	Partial	Beta		0	Lake Tyrell	ssDNA
tyrell_node_2337_length_11321_cov_5.697231	11321	15	4	Partial	Unassigned		0	Lake Tyrell	dsDNAphage
tyrell_node_3151_length_9914_cov_9.348413	9914	11	6	complete	Beta		0	Lake Tyrell	dsDNAphage
tyrell_node_5306_length_7816_cov_3.780312	7816	11	7	Partial	Delta	Haloarcula	91.4	Lake Tyrell	dsDNAphage

Supplementary Table 4: Functional prediction Epsilonpleolipovirus. Excel file of Supplementary Table 4 summarizes the functional prediction of the proteins from the proposed *Epsilonpleolipovirus* genus. Prediction was performed with interproscan v5 and functional categories were assigned using eggnog-mapper (minimum e-value < 10^5).

Protein id	Acc. N°	Functional prediction	Score	Evalue
frome_node_3245_length_9832_cov_9.545771_4	TIGR01409	TAT_signal_seq: Tat (twin-arginine translocation) pathway signal sequence	45	7.00E-05
gairdner_node_12003_length_7013_cov_5.306554_3	PF10518	TAT (twin-arginine translocation) pathway signal sequence	58	1.10E-05
gairdner_node_21431_length_5420_cov_4.609506_4	PF02697	Putative antitoxin	37	1.10E-06
gairdner_node_9457_length_7768_cov_3.119279_6	PF18742	REase_DpnII-MboI	351	1.50E-40
gairdner_node_9457_length_7768_cov_3.119279_7	PF05707	Zonular occludens toxin (Zot)	261	1.60E-04
imgvr_uvig_3300000868_000004_15	TIGR04126	PGF_CTERM: PGF-CTERM archaeal protein-sorting signal	1033	2.60E-09
imgvr_uvig_3300000868_000004_15	TIGR04207	halo_sig_pep: surface glycoprotein signal peptide	35	2.40E-13
imgvr_uvig_3300000868_000004_20	PF01402	Ribbon-helix-helix protein, copG family	41	2.80E-06
imgvr_uvig_3300000868_000004_24	PF00589	Phage integrase family	214	5.50E-10
imgvr_uvig_3300000868_000004_30	PF00239	Resolvase, N terminal domain	154	2.80E-24
imgvr_uvig_3300000868_000004_35	PF05707	Zonular occludens toxin (Zot)	264	1.30E-04
imgvr_uvig_3300000873_000003_28	PF09250	Bifunctional DNA primase/polymerase, N-terminal	269	2.20E-07
imgvr_uvig_3300000873_000003_33	PF04343	Protein of unknown function, DUF488	130	8.40E-09
imgvr_uvig_3300005273_000010_1	PF08706	D5 N terminal like	159	2.90E-05
imgvr_uvig_3300005273_000010_1	TIGR01613	primase_Cterm: phage/plasmid primase, P4 family, C-terminal domain	418	1.20E-42
imgvr_uvig_3300005273_000010_10	PF05707	Zonular occludens toxin (Zot)	265	3.20E-05
imgvr_uvig_3300005273_000010_11	PF02416	mttA/Hcf106 family	56	1.10E-15
imgvr_uvig_3300005273_000010_11	TIGR01411	tatAE: twin arginine-targeting protein translocase, TatA/E family	50	3.20E-18
imgvr_uvig_3300005273_000010_7	PF13620	Carboxypeptidase regulatory-like domain	231	1.20E-12
imgvr_uvig_3300005273_000010_7	PF13620	Carboxypeptidase regulatory-like domain	144	2.20E-10
imgvr_uvig_3300010222_000037_10	PF00801	PKD domain	168	4.40E-06
imgvr_uvig_3300010222_000037_13	PF13620	Carboxypeptidase regulatory-like domain	103	3.10E-08
imgvr_uvig_3300010222_000037_21	PF09250	Bifunctional DNA primase/polymerase, N-terminal	289	8.30E-07
imgvr_uvig_3300010222_000037_5	PF08281	Sigma-70, region 4	78	5.90E-11
imgvr_uvig_3300027982_000068_3	PF13620	Carboxypeptidase regulatory-like domain	121	3.40E-11
imgvr_uvig_3300027982_000068_7	PF00801	PKD domain	179	3.90E-08
imgvr_uvig_3300033522_000019_1	PF08281	Sigma-70, region 4	268	2.20E-07

imgvr_uvig_3300033522_000019_15	PF09250	Bifunctional DNA primase/polymerase, N-terminal	300	1.50E-06
pink_node_1698_length_10852_cov_6.06659_18	PF13392	HNH endonuclease	77	4.30E-09
torrens_node_644_length_20303_cov_30.778892_1	PF02899	Phage integrase, N-terminal SAM-like domain	91	1.10E-05
torrens_node_644_length_20303_cov_30.778892_1	PF00589	Phage integrase family	306	6.70E-10
torrens_node_644_length_20303_cov_30.778892_6	PF05707	Zonular occludens toxin (Zot)	269	5.30E-06

8. General Discussion and Outlook

This thesis significantly contributes to the understanding of how pleomorphic viruses establish chronic infections on their haloarchaeal hosts, the impact of these viruses on host metabolism and the diversity and evolutionary trajectories of this particular group of viruses. The following discussion intends to place the findings of this work into the broader theoretical and empirical body of haloarchaeal biology and the nature of chronic virus infections.

Viruses are intracellular genetic parasites that rely on their host metabolic machinery to ensure their replication. Genetic parasites such as viruses, plasmids and a wide range of selfish semi-autonomous genetic elements are an intrinsic part of the genetic pool of every organism, and they are present in all domains of life. Theoretical modeling, as well as empiric evidence suggests that the emergence of such genetic parasites is inevitable in prokaryotic system, as their complete suppression would involve the loss of the ability to perform horizontal gene transfer, a key evolutionary and survival mechanism of these organisms (Koonin, Wolf, and Katsnelson 2017). The effects that these elements can have on their host ranges from almost imperceptible in the case of the simplest replicons, to cell death and population extinction by lytic viruses.

Traditionally, prokaryotic infecting viruses have been classified in a dichotomic manner using a mechanistic perspective. This view separates viruses based on their overall effect on host physiology, with lytic viruses killing their host and lysogenic viruses replicating along with the cell. However, nature is rarely dichotomic and the discovery of chronic infecting viruses in bacteria and archaea has changed how we assess the impact of viruses at the cellular and community level. While chronic infecting viruses are common in eukaryotes, their presence on unicellular non-compartmentalized prokaryotic cells remains puzzling, as virus and host must have undergone a co-evolutionary process that allowed coexistence of viral and host replication or possibly, the establishment of stable long-term infections. For the majority of other mobile genetic elements this coexistence is not a great issue, as they usually carry some mechanisms to ensure their persistence on the system (such as toxin/antitoxin systems) or provide a selective advantage to their host by expanding their genetic pool or enhancing recombination (Koonin et al. 2019). Intriguingly, in the case of viruses performing non-lytic life cycles, the benefit, if existing at all, is often hard to identify, as selective advantages can be dependent on particular conditions or interactions, which are not predictable. Previous studies of chronic infecting filamentous phages of *E. coli*, showed that when phage genomes were manipulated to transmit only vertically (no budding), those strains with the least effect to host growth rates were positively selected, but upon reinstating horizontal transmission, the positive selection was lost (Bull, Molineux, and Rice 1991). This highlights the complexity of interactions and hurdles that chronic virus-host systems must face to establish

long-term infections. In this work, we have isolated and described the virus HFPV-1, that performs a chronic infection on its host *H. volcanii*. Remarkably, this virus was able to replicate its genome at unprecedented high rates, with hundreds to a thousand of virus genome copies per copy of host chromosome, while having little impact on host growth rates.

It has been modeled that the presence of persistent parasitic genetic elements correlates with the rate of horizontal gene transfer (HGT), with a particular mobile genetic element being more likely to persist within a population that experiences higher rates of HGT (Iranzo et al. 2016). Interestingly, haloarchaeal organisms seem to possess flexible genomes and seem to be relatively promiscuous when it comes to gene exchange (Papke et al. 2004). The haloarchaea were proposed to have evolve from a methanogenic ancestor that acquired many genes from the bacteria domain (Nelson-Sathi et al. 2012) and it has been documented that they often can acquire genes from a wider range of environments (Rhodes et al. 2011). Members of this clade frequently display polyploidy (multiple copies of the main chromosome), and numerous small and large plasmids (secondary chromosomes). Furthermore, they are prone to inter- and intragenomic homologous recombination, resulting in dissimilar rRNA operons within a single cell (Boucher et al. 2004), spontaneous generation of new chromosomal elements (Ausiannikava et al. 2018) and an overall great genomic diversity and speciation (Papke et al. 2015; Tschitschko et al. 2018). It is not farfetched then to hypothesize that this intrinsic genome flexibility and high rates of HGT serve as an advantageous scenery for the development of chronic infections, providing an explanation to the unusually high fitness of HFPV-1 and other pleolipoviruses.

Notably, despite being isolated from distant and unrelated geographical regions, with *H. volcanii* isolated from the Dead Sea and HFPV-1 from sediments of Lake Tyrrell, Australia, the interaction between HFPV-1 and its host prove to be extremely stable and long term under laboratory conditions. Meanwhile, no selective advantage for infected cultures could be observed under the studied conditions nor the emergence of resistant cells. On the contrary, we observed that infection with HFPV-1 at higher temperatures results in the inhibition of the host morphology transition to motile rod-shapes. In the scenario of controlled single-species conditions, this did not lead to a major reduction of host fitness, as only slight growth retardation was observed. However, it is unmistakably that in a broader community context, the inability to transition into the motile form, and to respond to chemotactic signals molecules will represent a major disadvantage for the host. Nevertheless, it must be noted, that at the higher temperatures that trigger the cell-shape transition (45 °C), the virus is also experiencing a reduction of its fitness, which is likely dues to viral proteins being poorly adapted to the optimal temperature for host growth. The latter could imply, that in nature these viruses would more successfully replicate within hosts that grow at the fringes of their ecological niche, which in turn would explain why members of the pleolipovirus are positively correlated with increasing salinities (Guixa-Boixareu et al. 1996).

While we did not observed that infection provided advantageous traits under any condition, it should not be excluded that this can occur. Among the advantages that viruses can provide are superinfection exclusion mechanisms, which inhibit infection by additional incoming viruses, which could at least partly explain the frequent detection of proviruses in microbial genomes (Bondy-Denomy and Davidson 2014). However, in this ongoing conflict often referred as an arms race between virus and host follows a continuous trend of developing adaptations and counter adaptations to cope with each other (Hampton, Watson, and Fineran 2020). These results in a perpetual coevolution, were the parties involved must keep up with each other innovations, yet on average their fitness remains unchanged in what has been regarded as the "Red Queen hypothesis" (Brockhurst et al. 2014; Van Valen 1973). Interestingly, in the case of *H. volcanii* and HFPV-1, we observe that proviral regions are being specifically downregulated (Alarcón-Schumacher et al. 2022). These proviral regions are similar to pleolipoviruses, and they have been propose to provide a selective advantage for the host. In particular, the region Halfvol3, encodes a protein with homology to a repressor of the phage PhiH1 (Dyall-Smith et al. 2018; Ken and Hackett 1991), and the deletion of one of these region leads to a detrimental effect on host fitness under infection with HFPV-1. This suggests that there is an interplay between HFPV-1 and the residing provirus, which limits the negative impact on host's fitness while allowing HFPV-1 successful replication and the production of relatively high extracellular titers. However, the absence of this provirus region tilts the equilibrium towards a more virulent viral infection, which in a broader community context would likely represent a strong negative selective pressure for *H. volcanii*. Thus, this suggests that the development of a stable and long-term chronic infection, HFPV-1 likely requires the presence of the provirus, which I hypothesize, would ensure the maintenance of a relative equilibrium. Intriguingly, while the host range of HFPV-1 includes organisms that do not contain resident proviruses, we did not observed any other case of severe growth retardation (Alarcón-Schumacher et al. 2022). The latter suggests that HFPV-1 replication and success is not dependent on the presence of evolutionary related proviruses, but rather this is a particular interaction with *H. volcanii*, further suggesting that the outcome of infection is dependent on the different molecular interactions and coevolutionary processes between the virus and each particular host. Nevertheless, the possibility that the ability of HFPV-1 to specifically downregulated viral elements is not restricted to resident proviruses should not be completely ruled out. Thus, I hypothesize that the infection by HFPV-1 could provide a selective advantage upon infection with additional viruses or other MGEs that could have a larger detrimental impact on host fitness, however further experimental evidence for this should be provided.

Additionally, while some of them encode one of the major viral hallmark genes, the rolling-circle replication endonuclease (RCRE), they are mostly unrelated to other viruses and their evolutionary origin remains unknown. Interestingly, the major capsid protein (spike proteins) presents a unique protein fold that is not related to other viral capsid proteins such as single and double-jelly roll, HK97 or alpha-helix hairpins.

Moreover, it has been proposed that the spike proteins of pleolipoviruses are a member of an undescribed and likely ancient clade of membrane fusion protein (El Omari et al. 2019). The lack of any similar proteins in the cellular or viral proteomes uncovered to date, along with the large diversity of spike protein structures described in our work suggests that this protein might have been acquired from an extinct lineage early in the evolution of life. Even more, the fact that recent evidence points toward the fact that pleolipoviruses are not restricted to hypersaline environments, but that they rather infect additional lineages of archaea in moderate environments, reinforced the idea that they are indeed a relatively ancient lineage. Thus, it should not be ruled out that this protein family might be even a remnant or descendent of a primordial membrane fusion mechanism, that has been lost in existing extant groups.

In this thesis, we have further contributed to uncover the diversity of uncultivated pleolipoviruses from hypersaline environments. During the study of these newly identified viruses, it has become clear that, similar to other viruses, they present modular genomes with a conserved structural module and an entangled history of recombination with different MGEs that lead to the acquisition of different replication modules. Furthermore, the comprehensive analyses of pleolipovirus genomes lead to the identification of coherent phylogenetic markers for this group, while also prompting us to propose the reorganization of the *Pleolipoviridae* family into five genera: the *Alpha, Beta, Gamma, Delta* and *Epsilonpleolipovirus*. Interestingly, the majority of reported pleolipoviruses from the Australian salt lakes were not integrated into their host chromosomes, but rather seem to exist as episomal elements, suggesting that they are likely performing active infections. Nonetheless, the vast majority of the reported diversity of pleolipoviruses correspond to uncultivated virus genomes (UViGs), and further cultivation efforts could provide a key to understand virus-host interaction and their effect on host fitness an ecology. Furthermore, the large evolutionary distances between these clades suggest that we are likely just beginning to uncover the tip of the iceberg in terms of pleolipovirus diversity.

Our understanding of how viral takeover or coexistence is mediated between virus and host is still scarce. While HFPV-1 appears to have an extremely broad host range, this seems to be a very unique trait, as all other pleolipoviruses are able to infect only their host of isolation, or in the case on UViGs, they are commonly targeted by spacers present in only a single. Weather pleolipoviruses have indeed such a narrow host range requires more isolation and characterization efforts, although it should not come as surprising, as they coevolve through a very tight relationship with their respective hosts. However, it remains puzzling how HFPV-1 is able to infect beyond genus and even family threshold. The identification of specific host receptors or the entry and exit mechanisms would greatly contribute to understanding membrane fusion and recognition in prokaryotic systems and the development of HFPV-1 as a genetic tool.

Lastly, given the increasing range of organisms affected by pleolipovirus, the changes to the metabolism induced during chronic infections, together with the key ecological roles that many archaeal clades play on different environments, studying how these viruses' impact host physiology and interactions becomes imperative to understand their role within microbial communities and on biogeochemical cycles.

References

Alarcón-Schumacher, Tomas, Adit Naor, Uri Gophna, and Susanne Erdmann. 2022. "Isolation of a Virus Causing a Chronic Infection in the Archaeal Model Organism *Haloferax Volcanii* Reveals Antiviral Activities of a Provirus." *Proceedings of the National Academy of Sciences* 119 (35): 1–12. https://doi.org/https://doi.org/10.1073/pnas.2205037119.

Ausiannikava, Darya, Laura Mitchell, Hannah Marriott, Victoria Smith, Michelle Hawkins, Kira S. Makarova, Eugene V. Koonin, Conrad A. Nieduszynski, and Thorsten Allers. 2018. "Evolution of Genome Architecture in Archaea: Spontaneous Generation of a New Chromosome in *Haloferax Volcanii*." *Molecular Biology and Evolution* 35 (8): 1855–68. https://doi.org/10.1093/molbev/msy075.

Bondy-Denomy, Joseph, and Alan R. Davidson. 2014. "When a Virus Is Not a Parasite: The Beneficial Effects of Prophages on Bacterial Fitness." *Journal of Microbiology* 52 (3): 235–42. https://doi.org/10.1007/s12275-014-4083-3.

Boucher, Yan, Christophe J. Douady, Adrian K. Sharma, Masahiro Kamekura, and W. Ford Doolittle. 2004. "Intragenomic Heterogeneity and Intergenomic Recombination among Haloarchaeal rRNA Genes." *Journal of Bacteriology* 186 (12): 3980–3990. https://doi.org/10.1099/mic.0.2007/009175-0.

Brockhurst, Michael A., Tracey Chapman, Kayla C. King, Judith E. Mank, Steve Paterson, and Gregory D.D. Hurst. 2014. "Running with the Red Queen: The Role of Biotic Conflicts in Evolution." *Proceedings of the Royal Society B: Biological Sciences* 281 (1797). https://doi.org/10.1098/rspb.2014.1382.

Bull, J. J., I. J. Molineux, and W. R. Rice. 1991. "Selection of Benevolence in a Host-Parasite System." *Evolution* 45 (4): 875–82. https://doi.org/10.1111/j.1558-5646.1991.tb04356.x.

Dyall-Smith, Mike, Felicitas Pfeifer, Angela Witte, Dieter Oesterhelt, and Friedhelm Pfeiffer. 2018. "Complete Genome Sequence of the Model Halovirus PhiH1 (ΦH1)." *Genes* 9 (10): 1–20. https://doi.org/10.3390/genes9100493.

Guixa-Boixareu, Núria, Juan I. Calderón-Paz, Mikal Heldal, Gunnar Bratbak, and Carlos Pedrós-Alió. 1996. "Viral Lysis and Bacterivory as Prokaryotic Loss Factors along a Salinity Gradient." *Aquatic Microbial Ecology* 11 (3): 215–27. https://doi.org/10.3354/ame011215.

Hampton, Hannah G., Bridget N.J. Watson, and Peter C. Fineran. 2020. "The Arms Race between Bacteria and Their Phage Foes." *Nature* 577 (7790): 327–36. https://doi.org/10.1038/s41586-019-1894-8.

Iranzo, Jaime, Pere Puigbo, Alexander E. Lobkovsky, Yuri I. Wolf, and Eugene V. Koonin. 2016.

"Inevitability of Genetic Parasites." *Genome Biology and Evolution* 8 (9): 2856–69. https://doi.org/10.1093/gbe/evw193.

Ken, R., and N. R. Hackett. 1991. "*Halobacterium Halobium* Strains Lysogenic for Phage ΦH Contain a Protein Resembling Coliphage Repressors." *Journal of Bacteriology* 173 (3): 955–60. https://doi.org/10.1128/jb.173.3.955-960.1991.

Koonin, Eugene V., Kira S. Makarova, Yuri I. Wolf, and Mart Krupovic. 2019. "Evolutionary Entanglement of Mobile Genetic Elements and Host Defence Systems: Guns for Hire." *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0172-9.

Koonin, Eugene V., Yuri I. Wolf, and Mikhail I. Katsnelson. 2017. "Inevitability of the Emergence and Persistence of Genetic Parasites Caused by Evolutionary Instability of Parasite-Free States." *Biology Direct* 12 (1): 1–12. https://doi.org/10.1186/s13062-017-0202-5.

Nelson-Sathi, Shijulal, Tal Dagan, Giddy Landan, Arnold Janssen, Mike Steel, James O. McInerney, Uwe Deppenmeier, and William F. Martin. 2012. "Acquisition of 1,000 Eubacterial Genes Physiologically Transformed a Methanogen at the Origin of Haloarchaea." *Proceedings of the National Academy of Sciences of the United States of America* 109 (50): 20537–42. https://doi.org/10.1073/pnas.1209119109.

Omari, Kamel El, Sai Li, Abhay Kotecha, Thomas S. Walter, Eduardo A. Bignon, Karl Harlos, Pentti Somerharju, et al. 2019. "The Structure of a Prokaryotic Viral Envelope Protein Expands the Landscape of Membrane Fusion Proteins." *Nature Communications* 10 (1). https://doi.org/10.1038/s41467-019-08728-7.

Papke, R. Thane, Paulina Corral, Nikhil Ram-Mohan, Rafael R. de la Haba, Cristina Sánchez-Porro, Andrea Makkay, and Antonio Ventosa. 2015. "Horizontal Gene Transfer, Dispersal and Haloarchaeal Speciation." *Life* 5 (2): 1405–26. https://doi.org/10.3390/life5021405.

Papke, R. Thane, Jeremy E. Koenig, Francísco Rodríguez-Valera, and W. Ford Doolittle. 2004. " Frequent Recombination in a Saltern Population of *Halorubrum* ." *Science* 306 (5703): 1928–29. https://doi.org/10.1126/science.1103289.

Rhodes, Matthew E., John R. Spear, Aharon Oren, and Christopher H. House. 2011. "Differences in Lateral Gene Transfer in Hypersaline versus Thermal Environments." *BMC Evolutionary Biology* 11 (1). https://doi.org/10.1186/1471-2148-11-199.

Tschitschko, Bernhard, Susanne Erdmann, Matthew Z Demaere, Simon Roux, Pratibha Panwar, Michelle A Allen, Timothy J Williams, et al. 2018. "Genomic Variation and Biogeography of Antarctic

Haloarchaea," 1–16. https://doi.org/10.1186/s40168-018-0495-3.

Valen, L Van. 1973. "A New Evolutionary Law." *Evol. Theory* 1: 1–30.

9. Addendum

Universität Bremen | Fachbereich 02 | Postfach 33 04 40, 28334 Bremen

Universität Bremen
Fachbereich 2
Prüfungsamt Chemie
z. Hd. Frau Frauke Ernst
Leobener Straße

28359 Bremen
Deutschland

Prüfungsamt
Chemie

Frauke Ernst
Geschäftsstelle
Fachbereich 02
Leobener Str. / NW2
D-28359 Bremen

Verwaltungspavillon 06
Tel. 0421 218-62802
Fax 0421 218-9862802
frauke.ernst@uni-bremen.de
www.fb2.uni-bremen.de

Versicherung an Eides Statt

Name, Vorname	Alarcon Schumacher, Tomas Benjamin
Matrikel-Nr.	3166621
Straße	Hastedter osterdiech 172b
Ort, PLZ	Bremen, 28207

Ich, _____ (Vorname, Name)

versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.
Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.
Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

_____
Ort, Datum / Unterschrift

Declaration on the contribution of the candidate to a multi-author article/manuscript, which is included as a chapter in the submitted doctoral thesis

Name of the candidate: Tomas Benjamin Alarcon Schumacher

Title of the Thesis: Virus-host interactions during chronic infections & Diversity and evolution of Pleolipoviruses

Contribution of the candidate in percentage of the total workload (up to 100% for each of the following categories):

Task	Chapter 1	Chapter 2	Chapter 3
Experimental concept and design	ca. 70%	ca. 90%	ca. 90%
Experimental work and acquisition of (experimental) data	ca. 80%	ca. 60%	ca. 70%
Data analysis and interpretation	ca. 80%	ca. 80%	ca. 90%
Preparation of figures and tables	ca. 90%	ca. 70%	ca. 90%
Drafting of the manuscript	ca. 80%	ca. 90%	ca. 90%

We declare that all authors of this article/manuscript have been informed on and did not object to the listed contributions of the candidate as author of this article/manuscript

Date:

Signatures:

(Candidate) (Supervisor)