Reviews in Mineralogy & Geochemistry
Vol. 75 pp. 649-675, 2013

The Deep Viriosphere: Assessing the Viral Impact on Microbial Community Dynamics in the Deep Subsurface

Rika E. Anderson[1], William J. Brazelton[1,2], John A. Baross[1]

[1]School of Oceanography and Astrobiology Program
University of Washington
Seattle, Washington 98195, U.S.A.

[2]Department of Biology
East Carolina University,
Greenville, North Carolina 27858, U.S.A.

jbaross@u.washington.edu

INTRODUCTION

All regions of Earth's biosphere that we have studied—the waters of Earth's oceans, the soil beneath our feet, and even the air we breathe—teem with viruses. Viral particles are among the smallest biological entities on the planet, with the average viral particle measuring about 100 nm in length: a size so small that five thousand viruses, lined end to end, would fit across the thickness of a human fingernail. What they lack in size, though, they compensate with sheer abundance. If we were to line up all the viruses in the ocean, they would stretch across the diameter of the Milky Way galaxy one hundred times (Suttle 2007). Those viruses are responsible for up to 10^{23} infections per second in the oceans (Suttle 2007). With each new infection, viruses can have a profound impact on their hosts: they can alter the structure of a microbial population, break up cellular biomass into its constituent organic matter, or introduce new genes into their hosts. Through this activity, viruses play a role in top-down as well as bottom-up processes, and can potentially alter the course of evolution.

The importance of viruses in the surface oceans is now well recognized, and research is increasingly dedicated to improving our understanding of their role in important marine processes. The viral role in the deep subsurface, however, is rarely considered. Deep within the crust and sediment below the ocean, viruses may play a profound role in altering biogeochemical cycles, structuring microbial diversity, and manipulating genetic content. Yet many questions remain unanswered: Are certain species or strains in the deep subsurface more susceptible to viral infection than others? What role do viruses play in driving natural selection and evolution in the deep biosphere? Is it more common for viruses to persist as protein-bound virion particles, or do they more commonly incorporate their genomes into that of their hosts? What impact do viruses have on their hosts while incorporated as stable symbionts? Can viruses provide their hosts with the keys to survival in the extreme environments of our planet?

The following chapter seeks to address these topics by exploring the nature of the relationship between viruses and their microbial hosts across a range of environments within the deep subsurface biosphere. We begin with a review of viral diversity by briefly describing the diversity of viral morphologies, nucleic acid types, and genetic content, and provide an overview of viral life cycles. We briefly discuss what is known of the viral impact on microbial biogeochemistry, microbial population structure and diversity, and on genetic content and expression patterns. We then apply these concepts to the deep subsurface, a region where unique

1529-6466/13/0075-0020$00.00 DOI: 10.2138/rmg.2013.75.20

attributes such as low nutrient and energy levels, enclosed pore spaces, and fluid flux may combine to produce an environment in which viruses play a significant role in manipulating the genetic landscape of deep subsurface microbial communities. We discuss completed and ongoing work that seeks to address some of these issues. Finally, we ask whether these viruses may have been involved in the origin of life in the subsurface. Viruses of the deep may play an important role in altering the evolutionary trajectory of their microbial hosts, and in doing so they complicate the concepts of parasitism and symbiosis in the microbial world, both now and in life's deep past. Ultimately, it is possible that the smallest biological entities on the planet have their most profound influence in its deepest realms, both now and in Earth's early history.

DIVERSITY IN THE VIRAL WORLD

Viruses infect all three domains of life and in doing so they adopt a wide variety of morphologies, lifestyle strategies, and genetic materials. These differences in viral types can have important implications for the nature of the virus-host relationship, and for the ways in which viruses can manipulate microbial community structure and evolution. By understanding the types of viruses that predominate in a given system, we can predict the nature of their impact on the host community. Here, we provide a brief overview of different viral types and life cycles, and then describe what types of viruses we might expect to predominate in the deep subsurface, given the environmental conditions.

Viral life cycles

Viruses infecting archaea and bacteria assume two different lifestyle strategies, each with significant implications for the viral relationship with the host and for the nature of virus-host co-evolution. Here, we provide a very simplified overview of viral life cycles; these are illustrated schematically in Figure 1. In the *lytic cycle*, viral particles attach to the outside of the host and inject their genetic material into the host cytoplasm. This genetic material then mounts a takeover of cell machinery for immediate synthesis of viral particles, which accumulate within the cell until it bursts, or lyses, releasing the viral particles into the surrounding medium, ready to infect a new host (Fig. 1A). Viruses employing the *lysogenic cycle*, in contrast, incorporate their genome into the host genome upon infection. Incorporated viral genomes are known as "prophage" or "proviruses," and can lie latent within the cellular genome for many generations. A glossary (Table 1) defines these terms and others used throughout the manuscript. Cells can maintain one or several prophage,

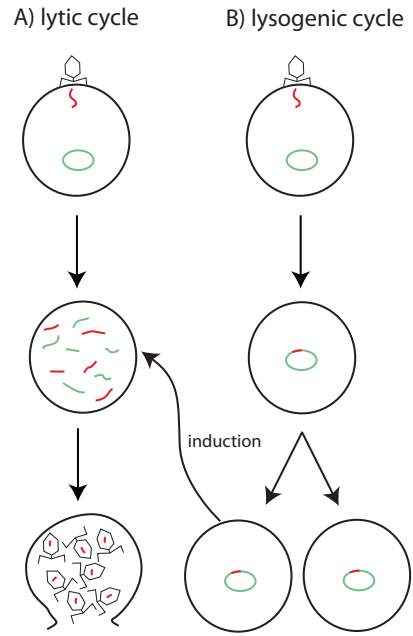

Figure 1. Schematic and generalized depiction of the lytic and lysogenic cycles of phage. A) Lytic cycle, in which a virus lands on the cellular membrane, injects genetic material into the cytoplasm, resulting in viral takeover of cellular machinery. New viral capsids are synthesized, packaged with viral genomes, and lyse the cell. B) Lysogenic cycle, in which a virus lands on the cellular membrane and injects its genetic material into the cytoplasm, and then integrates into the host genome. This "prophage" lies latent for several generations, then enters the lytic cycle in response to an induction event. See text for details.

Table 1. Glossary of terms used throughout the manuscript.

Term	Description
16S small subunit ribosomal RNA	A polynucleotide of about 1500 base pairs in length that makes up part of the small subunit of the prokaryotic ribosome; the highly conserved 16S ribosomal RNA sequence is used to determine evolutionary relationships between organisms.
Archaea	One of the three domains of life (in addition to Bacteria and Eukarya). Also single-celled, archaea are often found living in extreme or energy-limited environments, though they are also found in temperate environments such as the open ocean. Distinguishable from bacteria through genetic differences (including 16S rRNA sequence), a lack of peptidoglycan in the cell wall, differences in lipid composition, and many other biochemical differences.
Bacteria	One of the three domains of life (in addition to Archaea and Eukarya). These single-celled organisms have been found in nearly every environment investigated on the planet, with tremendously high diversity of morphology and metabolism.
Capsid	The protein coat that encapsulates viral genetic material.
Cytoplasm	Cellular contents within the cellular membrane.
DNA polymerase	An enzyme that catalyzes the polymerization of deoxyribonucleotides into a single strand. The sequence is used in some cases by viral ecologists to distinguish between viral types.
Lysogeny/lysogenic cycle	Viral life cycle in which the viral genetic material is integrated into the genome of the host.
Metagenome/ metagenomics	Method used by microbial ecologists to characterize a microbial community by collecting and sequencing genetic material directly from the environment, yielding thousands of sequencing reads currently ranging from ~50-1000 base pairs in length. Target organisms can be archaea, bacteria, viruses, or even single-celled eukaryotes.
Myoviridae, Podoviridae, Siphoviridae	Three families within the *Caudovirales* order, which are categorized according to morphology. All *Caudovirales* are characterized by a head-tail morphology and are extremely common in our oceans.
Phage	Short for bacteriophage ("phage" is Greek for "to devour.") Viruses that infect the bacteria; sometimes applied to archaeal viruses as well.
Phenotype	The physical manifestation of an organism's genetic material (or genotype).
Prokaryote	Organisms that lack membrane-bound organelles (generally including the archaea and the bacteria).
Prophage	Term for viral genetic material that is in the lysogenic state, integrated in a host chromosome.
Recombination	The process by which genetic material from two separate organisms are brought together.
Syntrophy	A relationship between two organisms in which they combine metabolic capabilities to derive energy from a net reaction that neither could metabolize independently.
Transduction	Virally-mediated horizontal gene transfer.
Virus	A parasitic element with a DNA or RNA genome that relies on a host to replicate, but has an extracellular state in which the genetic material is contained within a protein and/or lipid coat. Infects all three domains of life.

which can sometimes provide immunity from superinfection by other viruses. Viral genes can be expressed while integrated into the host genome, and can thereby influence the cellular phenotype. Generally, these viruses are induced, or triggered to enter the lytic cycle, in response to an environmental stimulus. At this point, the viral genome removes itself from the host genome and takes over the cellular machinery to create new viral particles, which then lyse the cell to begin the infection cycle anew (Fig. 1B). We believe it to be likely that the lysogenic cycle predominates among viruses in the deep biosphere, for reasons we will discuss below.

Viral sizes and morphologies

Viruses can range in size from 20 nm to well over 800 nm, and adopt myriad shapes, genome sizes, and replication strategies. Most viruses possess genomes ranging between a few to ~100 kilobases (kb), but recently the giant amoeba-infecting Mimivirus was discovered to possess a genome of 1,185 kb, and the virus structure itself is larger than some of the smallest cells (La Scola et al. 2003; Raoult et al. 2004). On the other end of the spectrum, the tiny Sputnik virus possesses a genome of only 18 kb, and parasitizes not a cell, but the Mimivirus itself (La Scola et al. 2008). RNA viruses are often among the smallest of the viruses, with some RNA viruses possessing genomes of only about 2 kb. Giant viruses continue to be discovered in various biomes of the globe (Fischer et al. 2010), and much remains to be learned about their lifestyles, replication mechanisms, and their evolutionary and ecological impacts on their hosts.

Viruses of the archaea and bacteria, our focus here, are represented by a wide variety of morphologies, including filamentous, icosahedral, and head-tail viruses. Many of the archaeal viruses possess particularly unusual shapes that have only recently been discovered. Examples of typical archaeal and bacterial virus morphologies are depicted in Figure 2. The most commonly observed phages (bacterial viruses) in the oceans are the head-tail viruses (Suttle 2005), all of which have linear double-stranded DNA genomes. Among the dsDNA viruses, morphology can give an indication of lifestyle and host range. In the marine realm the most abundant viruses are from the *Podoviridae* family, which have short, non-contractile tails and tend to infect only a narrow range of hosts, usually only particular strains within a species (Suttle 2005). In contrast, the members of the *Myoviridae* family, with contractile tails, and the *Siphoviridae*, with long non-contractile tails, tend to have a broader host range. Consequently, environments dominated by *Myoviridae* or *Siphoviridae* are more likely to be sites of interspecies viral infections.

However, viruses are not limited to the use of double-stranded DNA. Viruses also use single-stranded DNA (ssDNA) as their genetic material, and ssDNA viruses are increasingly found to be important members of the marine viral community. A recent study found that *Microviridae,* a family of ssDNA viruses, is one of the most common viral types in marine waters (Angly et al. 2006). Viruses also use RNA as their genetic material, and can be double- or single-stranded with plus or minus sense RNA strands. RNA viruses have been found to be important constituents of the marine ecosystem (Culley et al. 2003, 2006), infecting members across the trophic levels, from bacteria to whales. Retroviruses are one type of RNA virus that use an enzyme called reverse transcriptase to produce DNA from their RNA genomes, and then integrate this DNA into the genome of the host. Retroviruses also occur in both double-stranded and single-stranded forms. Interestingly, while retroviruses are common in eukaryotes, none have yet been found to naturally infect either the archaea or the bacteria.

While much is known about the morphologies and nucleic acid types of bacteriophages, very little is known about archaeal viruses. The few archaeal viruses isolated thus far have morphologies vastly different from those seen in bacterial viruses (Pina et al. 2011; Prangishvili et al. 2006) (Fig. 2). Some archaeal viruses possess a never-before-seen ability to change their morphology outside of the host, extruding tails on each end of an initially lemon-shaped viral capsid after release from the host (Häring et al. 2005). The unusual viral shapes encountered among the archaeal viruses are occasionally accompanied by unique release mechanisms from

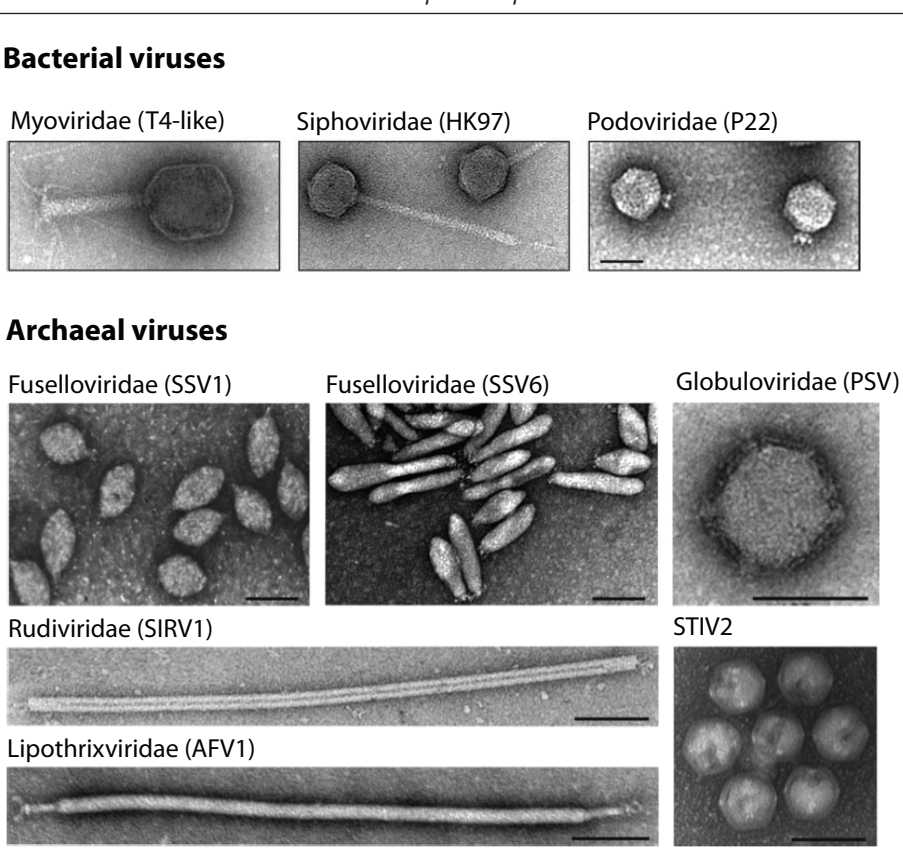

Figure 2. Transmission electron micrographs of bacterial and archaeal viruses. Scale bars for bacterial viruses are 50 nm; scale bars for archaeal viruses are 100 nm. Reprinted with permission from American Society of Microbiology from Krupovic et al. (2011).

the host cell, such as the formation of pyramid-like structures in archaeal membranes that serve as virus outlet sites (Bize et al. 2009; Brumfield et al. 2009). Most archaeal viruses studied to date are double-stranded DNA viruses, with only a single ssDNA archaeal virus discovered thus far (Pietilä et al. 2009). However, these results almost certainly reflect the nature of the detection techniques that have been utilized thus far and not the true diversity of archaeal viruses in natural environments. Considering the similarities between the eukaryotic and archaeal transcription apparatus, the discovery of archaeal RNA viruses and retroviruses seems imminent and may have great potential for yielding important insights into viral evolution. In the deep subsurface biosphere, where archaea constitute a larger proportion of the community than in surface oceans (Biddle et al. 2006), archaeal viruses may dominate, and further study may reveal as yet unknown morphologies or life strategies. Finally, a recent metagenomics study in an acidic, high-temperature lake in Lassen Volcanic Park, USA, uncovered a viral genome sequence suggesting recombination between an RNA and a DNA virus (Diemer and Stedman 2012). While the host of this particular virus was most likely eukaryotic, this study points to the possibility of such recombination events, which may occur between bacterial or archaeal RNA and DNA viruses as well.

Genetic diversity

An important question in viral ecology is the degree to which viral types are restricted to a given environment, or whether there is movement across biomes. In this sense viruses represent a further test of the null hypothesis of microbial biogeography: "Everything is everywhere, but the environment selects" (Baas Becking 1934; O'Malley 2007). One of the great challenges in assessing viral diversity and biogeography is the lack of a universal "barcoding" gene, analogous to the 16S small ribosomal subunit among the archaea and bacteria, which might be used to compare across all groups. Therefore, other techniques are used to assess virus biogeography. Steward et al. (2000) compared the relative genome sizes of viruses using pulsed-field gel electrophoresis, and found that certain genome size classes are found in many different marine environments. Similarly, Breitbart et al. (2004) investigated the environmental distribution of the T7 phage DNA polymerase gene, and found that the same sequences were found in a wide variety of diverse biomes, indicating a ubiquity of similar viruses across diverse environmental types. In this scenario, viral diversity is high locally, but viral types are distributed globally. The observation of globally-distributed viral types implies extensive movement among biomes and potential infection of (and sharing genes between) a wide array of hosts (Breitbart and Rohwer 2005).

Other studies present a contrasting picture of viral biogeography. For example, genomic analysis of a thermophilic virus of *Sulfolobus* revealed that viruses and their hosts tend to be spatially restricted in hot springs (Held and Whitaker 2009). Metagenomic analysis of viruses in stromatolites and thrombolites found a similar geographic restriction (Desnues et al. 2008), and metagenomic characterization of viruses in soil found distinctions between viral assemblages in soil samples and those in marine or fecal samples (Fierer et al. 2007). On a larger scale, metagenomic studies have revealed that while certain types of viruses, such as the myoviruses, were ubiquitous across sample sites, others, such as podoviruses and siphoviruses, had more site-specific distributions (Williamson et al. 2008a). Thus, an opposing paradigm suggests that distinct groups of viruses are tied closely to specific hosts, resulting in spatial restriction (Thurber 2009). While further study will provide greater insight into this story, it seems that some viral types are globally distributed, while others are much more spatially restricted. Furthermore, spatial distribution is likely to be determined by host specificity, but this relationship is mostly unexplored. Future work aimed at distinguishing between widely distributed viral types and more locally restricted (and presumably more host-specific) viral types may give insight into which viral types are most likely to facilitate gene flow between biomes.

In this context, the viruses of the deep subsurface represent an interesting case. It might be expected that viruses in the deep subsurface, on the one hand, should have reduced mobility as a result of being restricted within a sediment or rock matrix, and therefore have limited and patchy geographic distribution. This limited range may be particularly the case in sedimented regions away from the ridge axis. On the other hand, fluid flux within the subsurface in regions closer to the ridge axis, as well as allochthonous input from above, might facilitate movement of hosts and therefore of viruses from one locality to the next (Anderson et al. 2011a). It seems entirely possible that some viral types are restricted to particular regions of the subsurface, while others are more ubiquitous across the deep biosphere, perhaps in biogeographic correlation with their hosts. Further study will be required to resolve these questions.

VIRAL IMPACTS ON HOST ECOLOGY AND EVOLUTION

Viruses are a peculiarly potent force in that they can impact host community structure through both bottom-up and top-down control, and they can influence host genetic content through horizontal gene transfer and lysogenic conversion. Here, we provide a brief overview

of what is known thus far of the viral impact on host microbial communities, with the aim of better understanding the ecological and evolutionary dynamics of the deep subsurface habitat.

Bottom-up effects: the biogeochemical impact

Through lysis of microbial hosts, viruses convert biomass to dissolved and particulate organic matter (Proctor and Fuhrman 1990). Estimates show that viral lysis removes approximately 20-40% of prokaryotic biomass in the ocean daily, though quantifying mortality rates due to viral lysis is difficult (Suttle 2007). This rapid turnover has tremendous biogeochemical implications, as viral lysis converts organic matter from biomass into the pool of dissolved organic matter (DOM), redirecting it from higher trophic levels and effectively short-circuiting the microbial loop. This phenomenon has been dubbed the "viral shunt," and has the effect of stimulating bacterial production by providing a source of DOM and thus stimulating respiration (Suttle 2007). Moreover, the "viral shunt" is thought to stimulate the ocean's biological pump by accelerating sinking rates of lysed cells or transforming bacterial biomass into dissolved organic matter, though it is unclear what percentage of this lysed material is recalcitrant or labile (Jiao et al. 2010). This is depicted schematically in Figure 3A.

The impact of the viral shunt on the deep biosphere naturally depends on virus-to-cell ratios, which impact the rate of infection. As this ratio varies according to depth and location, it is difficult to calculate the net impact of viruses on prokaryotic mortality in the deep subsurface. Danovaro et al. (2008) showed that viruses become the predominant source of prokaryotic mortality as depth increases in the sediments; in continental margin sediments off of Chile, it was estimated that viruses were responsible for mortality of 38-144% of bacterial net production (Middelboe et al. 2006). In mud volcanoes, viruses account for up to 33% of cells killed daily, and also contributed an estimated 49 mg carbon per square meter per day—a substantial contribution to the total carbon budget (Corinaldesi et al. 2011). Thus it can be expected that viruses in the deep biosphere will have a significant, if poorly constrained, impact on microbial mortality and, by extension, biogeochemical cycles. The extent of viral impact will also necessarily depend upon the predominant life cycle of viruses in the subsurface: if lysis predominates, the virus to cell ratio will be the most important factor in determining the importance of viruses in microbial mortality and trophic cycling; whereas if lysogeny predominates, the viral impact on mortality will also be dependent on the frequency and pattern of induction events within each environment.

Top-down effects: altering population structure

As predators, viruses also control population structure from the top-down; in the deep subsurface, where other predators such as grazers are likely to be absent, viruses may constitute the sole inducer of cell mortality, aside from natural decay. The question that then arises is how the dynamics of viral host range, lifestyle and infection frequency can alter the structure of host microbial communities.

One of the most influential ideas related to viral control of population structure is the notion of "kill the winner" (Thingstad and Lignell 1997). Several authors have observed that viral infection rates are dependent upon cell density and growth rate (e.g., Middelboe 2000); as most viruses have a fairly limited host range, this dependence implies that if a particular microbial group becomes dominant in a population, those cells are most susceptible to viral infection as a result of their increased density. This concept is depicted schematically in Figure 3A. Consequently, viruses may act as a homogenizing agent on the diversity of microbial communities, effectively maintaining high species evenness. Studies have shown that viruses are instrumental in the termination of certain types of plankton blooms (Bratbak et al. 1993). Moreover, Rodriguez-Valera et al. (2009) found that regions with the greatest variability within a given species' genome were regions coding for surface receptors, which are potential phage-recognition targets. They argue that viruses maintain high diversity in a system through

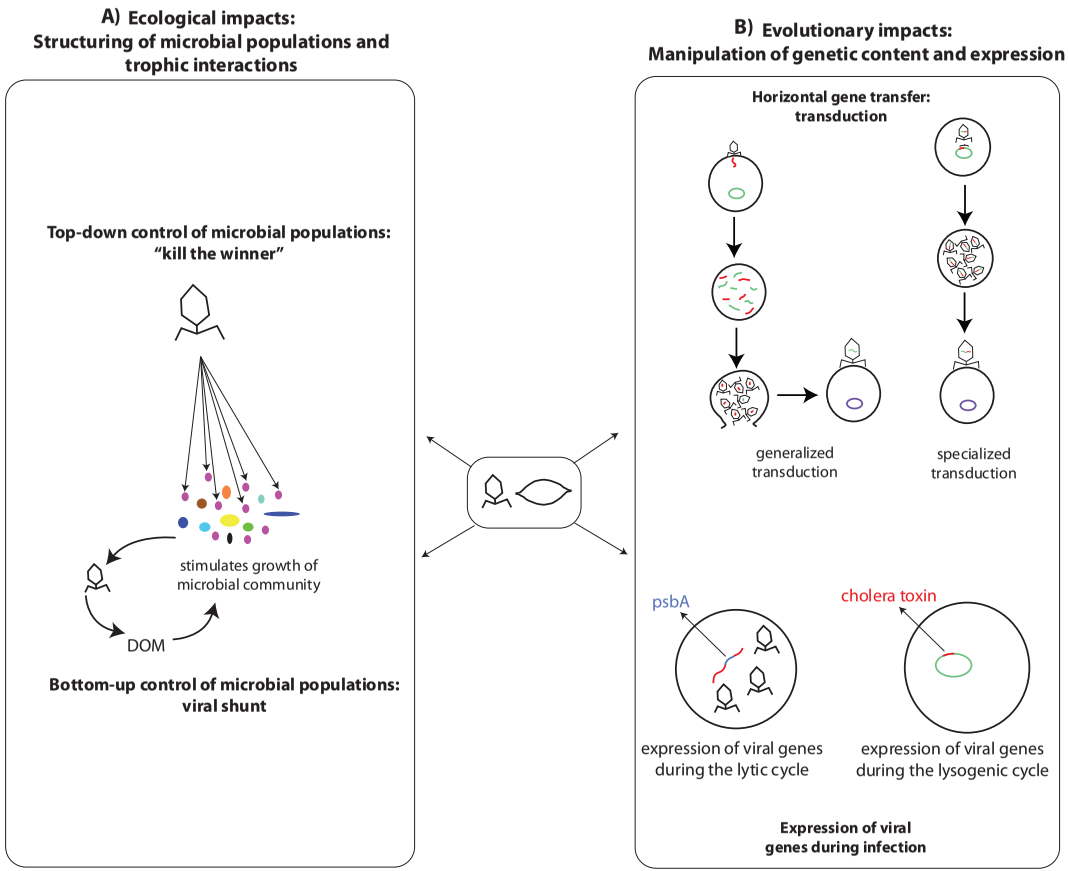

Figure 3. Schematic summary of the effects of archaeal and bacterial viruses on hosts. A) Ecological impacts in which viruses play a role in structuring microbial diversity and trophic interactions. Through "kill the winner" type dynamics (see text for details), viruses maintain greater evenness within a population by reducing the abundance of the most abundant strains. Through the viral shunt, viruses lyse cells and thus contribute to the pool of dissolve organic matter (DOM), which may stimulate growth of the microbial community. B) Evolutionary impacts in which viruses play a role in manipulating genetic content and expression of their hosts. Viruses can alter the genetic content of their hosts through horizontal gene transfer (called transduction). Lytic viruses do this through the process of generalized transduction, in which individual viral capsids are packaged with host genetic material instead of viral genetic material, which is then transferred to the next host. Lysogenic viruses can also do this through the process of specialized transduction, in which enzymes removing the viral genome from the host genome accidentally remove some host genetic material adjacent to the prophage. This genetic material is also packaged into a viral capsid and transmitted to the next host. Viruses can also express viral genes during the process of infection: during the lytic cycle, viruses can express genes to support cellular activity while viral capsids and genomes are synthesized (such as psbA); they can also do this during the lysogenic cycle, while integrated as prophage (such as cholera toxin).

kill-the-winner-like purges of ecotypes carrying the same surface receptors, which they coined the "constant-diversity" (CD) model. These viral purges contrast with the natural selection purges in the theory of "periodic selection," in which occasional changes in environmental conditions drastically reduce diversity by eliminating all groups not adapted to those conditions (Cohan 2002). Thus, viruses can contribute to ecosystem stability by maintaining high levels of diversity, even though they are agents of mortality.

Should viruses be a potent force in the deep subsurface, these impacts on population structure should not be discounted, and the impact is likely to vary depending on the nature of the environment. In stagnant sediments with little fluid flux, for example, environmental conditions may be fairly stable, and thus the CD model posited by Rodriguez-Valera et al. may be a primary mechanism for maintaining diversity among strains in subsurface communities. However, in more dynamic environments, such as hydrothermal vent systems, community diversity may be structured through a synergistic combination of periodic selective sweeps through environmental change as well as CD dynamics through viral predation.

Viral manipulation of genetic content and expression

We have reviewed several processes by which viruses influence the evolution of biological communities throughout the globe: viruses are agents of cell mortality and nutrient recycling, they stimulate co-evolution with their hosts through the virus-host arms race, and they play a hand in the structuring of communities and therefore in the generation of new ecotypes and species. Additionally, viruses are known to manipulate genetic content and expression through horizontal gene transfer and lysogenic conversion. Through these mechanisms, viruses may facilitate adaptation to specific niches within a given ecosystem and thereby exert profound impacts on the evolution of their hosts.

Transduction. Viruses facilitate horizontal gene transfer through the process of transduction, which occurs when a virus introduces foreign genetic material into a host during the course of infection. This transfer can occur during the course of lytic infection in a process known as *generalized transduction*, depicted schematically in Figure 3B. During the lytic cycle, a virus degrades host DNA and synthesizes viral particles. In this process, host DNA can be accidentally incorporated into a new virus capsid. The resulting *transducing particles* can then infect a new host, introducing genetic material from a previous host into a new one, which can then recombine into the genome of the new host. In this process, almost any region of genetic material may be transferred from the donor cell to the recipient.

In the process of *specialized transduction*, lysogenic viruses excise their genomes incorrectly, incorporating a small region of adjacent host genetic material into the viral genome. This mechanism is shown schematically in Figure 3B. Combined, generalized and specialized transduction can have a significant impact on the genetic content of viral hosts: one study estimated that up to 10^{14} transduction events occur per year in Tampa Bay Estuary alone (Jiang and Paul 1998).

A more recent discovery may increase the estimated rates for transduction even further. Gene transfer agents, or GTAs, are viral-like transducing particles, most likely defective phages, which seem to have been usurped by the host to facilitate the process of horizontal gene transfer. While most have been found in Alphaproteobacteria such as *Rhodobacter* (Lang and Beatty 2000) or *Brachyspira* (Matson et al. 2005), GTAs have also been found in methanogens and other groups (Stanton 2007) and may be widespread throughout the archaeal and bacterial domains. A recent study suggested that GTA transduction rates may be over one million times higher than previously reported viral transduction rates in the marine environment (McDaniel et al. 2010). Because of their small size (Matson et al. 2005) GTA particles should be well-represented in "viral" metagenomes, but positive identifications of GTAs are difficult due to the scarcity of sequenced GTAs thus far (Kristensen et al. 2009). Nevertheless, GTAs may

constitute a crucial source of genetic innovation in all biomes of the planet, including the deep subsurface. Isolation of strains encoding GTAs may be necessary to enable metagenomic identifications and to increase our knowledge of the scope of their impact.

Expression of genes during the course of infection. Viruses can also carry genes that are expressed during the course of infection. These genes often serve to improve host fitness, which in turn improves virus fitness while the virus is dependent on the host. Some of these genes are expressed by the virus during the lytic cycle, presumably as a means to support host machinery while viral genomes and capsids are replicating within the cell. The most well-known examples of this effect are photosynthesis genes expressed by cyanophage infecting *Prochlorococcus* and *Synechococcus* (Mann et al. 1993). Genes encoding the photosystem core reaction center D1 are expressed during the course of infection in *Prochlorococcus* phage, and it was proposed that these genes improve phage fitness by supporting host photosynthesis during infection (Lindell et al. 2005).

Lysogenic viruses can also manipulate host genetic content while integrated as prophage in the cell. As prophage, lysogenized viruses depend on the host for survival over a longer term, and thus benefit from improving host fitness while integrated in the host genome. In some lysogenic viruses, selection has favored the maintenance of genes known as "fitness factors"—genes that are encoded and expressed by prophage that alter host phenotype and enhance fitness. One of the most well-known examples of this process is the production of cholera toxin by a filamentous bacteriophage integrated into the genomes of virulent *Vibrio cholera*e strains (Waldor and Mekalanos 1996). Studies have shown that infection by a prophage can drastically alter the phenotypic range for a given species (Vidgen et al. 2006). In some cases, phage can also alter host phenotype by suppressing certain metabolic capabilities; it has been suggested that these phage can act to slow host metabolism to shut down unnecessary pathways and conserve energy in environments with low nutrient or energy resources (Paul 2008)—conditions that are expected to be quite common in many deep subsurface ecosystems.

Thus, in addition to influencing host evolution through top-down or bottom-up control of microbial communities, viruses can directly manipulate host genetic content in multiple ways: through generalized or specialized transduction, or through expression of genes either during the lytic cycle or as prophage. Next, we highlight how the unique attributes of the deep subsurface biosphere may create an environment ripe for viral manipulation of host genetic content.

VIRAL MANIPULATION OF THE DEEP SUBSURFACE BIOSPHERE

As we have reviewed above, the viral impact on the microbial ecology of the deep subsurface is potentially profound but largely unexplored. The deep biosphere is not a homogenous biome (Colwell and D'Hondt 2013; Schrenk et al. 2013), and as such, the potential role of viruses is likely to vary depending on the region, particularly with distance from ridge axes, as fluid flux declines and sedimentation increases. Here we examine the potential roles of viruses in these regions, with a focus on the marine subsurface.

Hydrologically active regions of the subsurface

Much of the ocean crust experiences fluid flux to a certain degree; it is estimated that at least 60% of the ocean crust is hydrologically active (Edwards et al. 2011). The volume of fluid fluxing through the crust is at its highest near active hydrothermal systems at mid-ocean ridges, but does not immediately dissipate. The degree of fluid flux varies depending on the sediment cover, as sediments tend to restrict fluid flow (Edwards et al. 2005). This dynamic situation is illustrated schematically in Figure 4. Most fluid flux occurs through connected channels in ocean crust, such as around breccia zones and around pillow basalts or flow boundaries

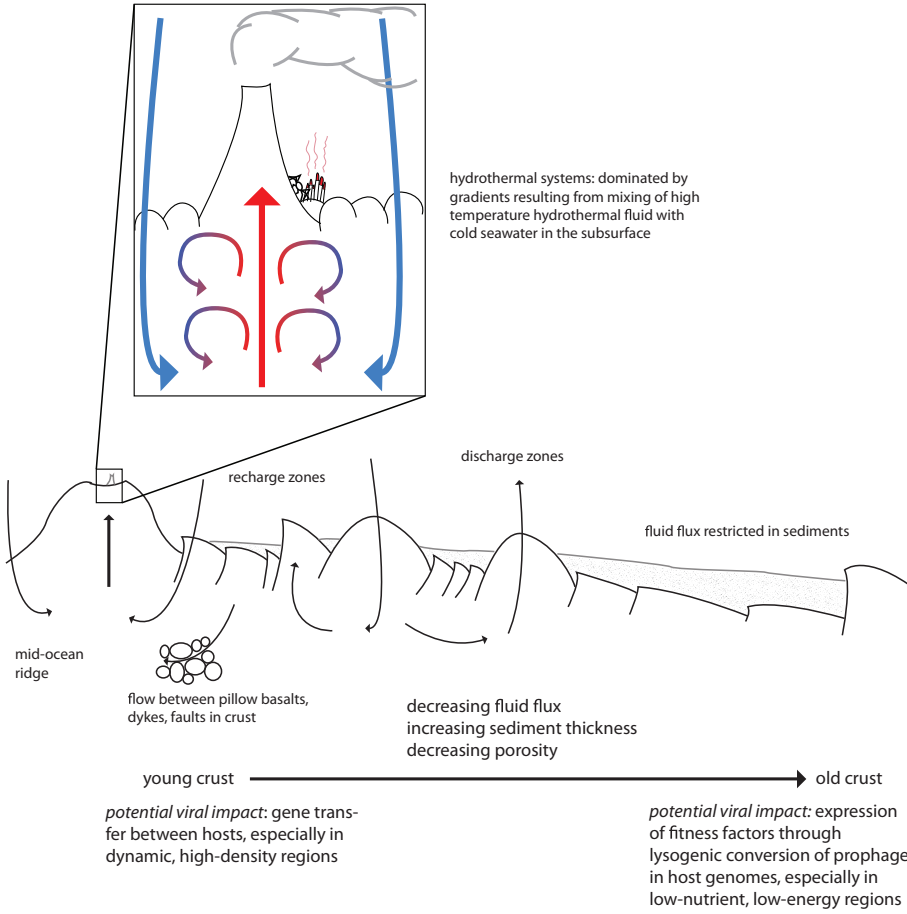

Figure 4. Schematic depicting fluid flux and porosity in the marine subsurface. Arrows depict fluid flux across the seawater/subsurface interface, and throughout the subsurface. Inset shows detail of fluid flux and mixing of seawater and high-temperature hydrothermal fluid in mid-ocean ridge hydrothermal systems, where high-temperature hydrothermal fluid (red) rises from high-temperature water rock reactions deeper in the subsurface and mixes with colder seawater (blue). Crust ages as it moves away from the mid-ocean spreading ridge.

(Fisher and Becker 2000). Seawater flows through seamounts, ridge flanks and recharge zones away from the ridge axis, with residence times ranging from days to years, depending on the location (Johnson and Pruis 2003). Thus a substantial portion of the ocean subsurface biosphere is exposed to dynamic fluid flux.

Connectivity of habitats within the subsurface via fluid flux from one region to the next has the potential to bring members of separate microbial communities, including archaea, bacteria, and their viruses, into contact with each other. This connectivity, in turn, could facilitate gene flow between each of these communities. Cells could exchange genes through conjugation, whereby DNA exchange occurs by direct cell-to-cell contact. However, conjugation requires two cells to be in close proximity and physiologically equipped for this process to occur. Cells can also acquire novel genetic material through transformation, whereby free DNA from lysed cells is taken up into a competent cell. The availability of free DNA varies depending on DNA

degradation rates, which tend to be lower in marine sediments (Lorenz and Wackernagel 1994). It is also possible that DNA is stabilized in surface-attached biofilm communities, which are discussed below. In general, though, viruses that have the ability to travel between hosts and to protect genetic material within a protein capsid may represent one of the most important vectors of gene transfer in these environments.

It should be noted that while much of the focus of this review is on regions of the marine subsurface, viruses have been observed to be present in terrestrial subsurface systems as well. While not much is known about viruses in the terrestrial subsurface, observations have been made of viruses in deep granitic groundwater at abundances ranging from 10^5-10^7 per milliliter (Kyle et al. 2008). These viruses seem to be somewhat diverse morphologically, though the viruses isolated thus far have a fairly narrow host range (Eydal et al. 2009). Viruses are thus a potentially important ecological and evolutionary force in the terrestrial subsurface as well, yet much remains unknown about their role.

Hydrothermal vent systems. Hydrothermal vent systems are found at mid-ocean ridge spreading centers or seamounts, and are driven by high-temperature water-rock reactions that occur when seawater comes into close contact with a magma chamber. These basalt-hosted systems are characterized by high-temperature, low-pH fluids that are enriched in reduced compounds, transition metals, sulfide, CO_2, helium, methane, and hydrogen, and are depleted in magnesium (Von Damm 1990; Schrenk et al. 2013). Upon reaching the ocean-crust interface, these fluids precipitate sulfide minerals to create sulfide chimneys that play host to complex microbial communities, which in turn form the trophic basis of macrofaunal communities hosting worms, mussels, crabs, and shrimp. One of the defining characteristics of these systems is the dominance of chemical, physical, and mineralogical gradients that shape the structure of the microbial communities inhabiting these systems (Baross and Hoffman 1985; Schrenk et al. 2003). As seawater mixes with hydrothermal fluid, this results in temperatures that range from 2˚ to 400˚C, acidities that range from 2 to 8, and chemical composition that spans the spectrum of reduced hydrothermal fluid to oxidized seawater. As a result, vents play host to a wide range of archaeal and bacterial species: hyperthermophiles, thermophiles, and psychrophiles, including both heterotrophs and autotrophs.

Few studies have been conducted on viruses inhabiting these systems. Ortmann and Suttle (2005) found that the abundance of viral-like particles in diffuse flow fluids was approximately 10^6 per milliliter, about ten times that of cells, a ratio that is also typical of surface seawater. Williamson et al. (2008b) found that a higher abundance of viral-like particles were induced from hydrothermal vent microbial communities exposed to a mutagen compared to those from the upper water column, suggesting that lysogeny is a more predominant lifestyle at vents than in the upper water column. Metagenomics has revealed that the marine vent viral assemblage has the potential to infect a wide variety of bacterial and archaeal hosts from a range of thermal regimes, reflecting the gradient-dominated nature of the environment (Anderson et al. 2011b). Together, these studies suggest that many different archaeal and bacterial groups may have prophage integrated into their genomes that potentially introduce novel genetic material or express fitness factors. If this is the case, then genomic analyses of subsurface archaea and bacteria that contain prophage should be a fairly efficient, though clearly biased, approach for exploring the diversity of subsurface viruses. Ideally, such analyses would be coupled with a metagenomic census of free viral particles (e.g., Anderson et al. 2011b) in order to compare lysogenic and lytic viruses. Clearly, much more research remains to be done to better understand the nature of viral roles in the subsurface.

Deeply buried sediments

Regions of the deep subsurface with more restricted fluid flux, particularly in regions with high sedimentation such as on continental margins, present a drastically different set of conditions for microbial inhabitants. Within the sediment matrix, viral mobility may be reduced,

resulting in a potentially lower host contact rate. This reduced contact rate would be the case especially if cell abundances are low in deeply buried sediments, such as in the sediments of the South Pacific Gyre (D'Hondt et al. 2009). On the other hand, viruses that form small but hardy particles may be less affected by restrictions on fluid flux than other mechanisms of genetic exchange, which would accentuate the importance of viruses in the ecology and evolution of these isolated communities. The challenges faced by microbial communities inhabiting these regions include limitations on nutrient and energy levels, particularly in deeply buried sediments, where organic carbon and potential oxidizing agents are scarce (Jørgensen and D'Hondt 2006), and metabolic rates have been shown to be extremely low (Røy et al. 2012). The low activity and long doubling times of cells in these regions likely provides further resistance to viral infection, which is largely dependent on the density and activity of the host (Fuhrman 2009). As with cellular abundances, viral abundance decreases with depth in the sediments. Middelboe et al. (2011) quantified viral abundance with depth on the eastern margin of the Porcupine Seabight and observed approximately 10^8 VLPs/cm^{-3} at 4 meters below the seafloor (mbsf), to about 10^6 VLPs/cm^{-3} at 96 mbsf. However, another study by Engelhardt et al. (2012) found that the virus-to-cell ratio increased with depth in the sediments, potentially indicating continued viral production at these depths in sediments, rather than long-term viral preservation, as had been suggested previously.

Lysogeny is expected to be a common viral lifestyle in the deep subsurface, resulting from selection for a viral lifestyle that limits the necessity for finding hosts in a sparse soil matrix and in harsh environmental conditions. Work by Engelhardt et al. (2011) has demonstrated that nearly half of the bacterial isolates tested from a deep-sea sediment core harbored prophage, and a subsequent study found that all deeply buried isolates of a common deep-sea species, *Rhizobium radiobacter,* were lysogenic (Engelhardt et al. 2012). The ubiquity of lysogeny could have interesting implications for cellular survival in these energy-limited systems, where archaea and bacteria are likely to be under strong selection pressure to harness alternate forms of energy when they are available, and to minimize energy use when energy sources are limiting. Previous studies have shown that certain lysogenic phage can actively repress host metabolic genes, and therefore repress wasteful host metabolic processes when conditions are not favorable (Paul 2008), a trait that would be particularly useful in the energy-limited subsurface. Further work in deeply buried sediments will reveal what genes are expressed by these lysogenized phages, perhaps revealing that the relationship between virus and host transcends the parasitic, becoming instead a mutualistic symbiosis.

Viral impacts on surface-attached communities

Regardless of whether a deep subsurface habitat is hydrologically active or not, most of the inhabitants probably live as biofilms (i.e., communities attached to some sort of hard surface, which could be hard rock, vent chimney deposit, or sediment). Generally, biofilms have much higher cell density than the surrounding medium, so there is potential for biofilms to be hotspots of viral activity. Viruses are known to accumulate in biofilms growing in drinking water systems (Skraber et al. 2005), and viral lysis is frequent during biofilm development in *Staphylococcus aureus* (Resch et al. 2005). Metagenomic sequencing of biofilms from an acid mine drainage site recovered complete viral genomes and extensive evidence that bacterial genomes are continually influenced by viral infection (Andersson and Banfield 2008). Viral genes are highly expressed in *Pseudomonas aeruginosa* biofilms (Whiteley et al. 2001), and viral-mediated cell death is a normal component of biofilm development (Webb et al. 2003). Beyond these basic detections of viruses and viral genes in biofilm habitats, however, surprisingly little is known about the molecular mechanisms and ecological impacts of virus-biofilm interactions, especially in the subsurface.

The polysaccharide-rich extracellular matrix of biofilms is probably a barrier against infection, but it is clear that viruses can penetrate the barrier, in some cases via enzymatic

digestion (Weinbauer 2004). Another complication is the recent finding that a human virus can generate its own biofilm-like matrix (Thoulouze and Alcover 2011). The prevalence of viruses encased within extracellular matrices is entirely unexplored in subsurface ecosystems, and it is likely that such surface-attached viral populations can evade detection and depress counts of viral-like particles in fluid samples. Therefore, interpretations of viral abundance and activity data from subsurface fluid samples must consider how well the fluid samples represent the rocks and sediments that provide habitat for most bacteria, archaea, and viruses in the subsurface.

In addition to high cell density, many biofilm communities also have high genetic and phenotypic diversity, resulting in complex interactions among many species on microscopic spatial scales (Stoodley et al. 2002). One potential consequence is that viruses with high host specificity may have greater difficulty finding their host in a tightly-packed, diverse biofilm, resulting in a large total number of viruses, each capable of infecting only a tiny proportion of the diverse biofilm community. This scenario is one possible explanation of the "infectivity paradox": the observation that many habitats have high viral abundance but low infectivity (Weinbauer 2004). Preliminary data (Filippini et al. 2006) suggest that biofilms exemplify the infectivity paradox, but no such studies have been conducted in the deep subsurface.

Many studies have demonstrated the importance of lateral gene transfer in biofilm communities (Molin and Tolker-Nielsen 2003), and in some cases, viruses have been identified as the agents of transfer (Whiteley et al. 2001; Webb et al. 2003). It is clear that in biofilms, gene transfer is not a rare curiosity but a fundamental aspect of biofilm formation and development, notably as a mechanism for a phenomenon known as "phenotype switching." In *Staphylococcus epidermidis* biofilms, for example, genomic insertion of a mobile genetic element results in a stable population of variant cells unable to produce the biofilm matrix (Ziebuhr et al. 1999). The effect is reversible because the inserted DNA is frequently excised, restoring biofilm production. Other species also exhibit reversible phenotype switching associated with biofilm formation, and viruses have been implicated in at least one case (Webb et al. 2004). The evolutionary dynamics of such processes have not been explored experimentally, but one simulation study predicted that the coexistence of multiple phenotypes in a biofilm community can be promoted by continual gene transfer. If two phenotypes are linked, as in a syntrophic partnership, the fitness of each member is dependent on the fitness of the other. Therefore, natural selection of such cells living in a dense community could result in complex inter-species relationships that are dependent on (potentially viral-mediated) transfer of genetic content. In summary, we expect future research to reveal that biofilms in subsurface habitats exemplify the concept described above that viral activity in the subsurface is likely to have complex and varied evolutionary consequences that extend beyond just cell mortality.

Tools for analysis: viral metagenomics in the deep subsurface

Given the current state of knowledge about viruses in the deep subsurface, how can we gain further insight into the role they play in manipulating geochemical cycles, altering diversity, and influencing the course of evolution in their hosts? One method by which we can probe the viral world is through metagenomics, in which a sample of community DNA is extracted and sequenced directly from the environment. While metagenomics in the microbial realm has traditionally focused on asking "who is there?" and "what are they doing?", viral metagenomics presents a unique set of challenges. Viruses are generally separated from the microbial fraction through size fractionation, which may exclude large viral particles or include small cells, so contamination is an issue of concern. Moreover, one of the primary challenges facing viral metagenomics is the large proportion of unknown sequences. The average percentage of viral metagenomic sequences with no match to existing databases ranges from about 60 to over 90%, depending on the read length (i.e., Breitbart et al. 2002; Angly et al. 2006; Desnues et al. 2008; Anderson et al. 2011b; Rosario and Breitbart 2011).

The vast number of sequences with no match to existing databases presents a challenge to viral ecologists seeking to understand who viruses infect, what impacts they have on their hosts, and what types of genes they encode and transfer.

One goal of viral ecology in any environment is the identification of which archaeal or bacterial groups play host to those viruses. This information is key to understanding how viruses may impact a given microbial community. If only certain groups are most susceptible to viral attack, this selectivity may have further implications for microbial population structure or biogeochemistry. Some information about potential hosts can be gleaned by identification of known viral groups: *Rudiviridae* and *Fuselloviridae*, for example, are only known to infect the archaea. As we have noted, classification of viral metagenomic sequences is tremendously challenging, and even if it is successful, only limited information is gained because many families of viruses infect wide ranges of hosts.

One method that has been used to identify potential hosts of a viral assemblage is to identify clustered regularly interspaced palindromic repeats (CRISPRs), an immune system used by archaea and bacteria to combat invasive genetic material, including viruses and plasmids (Barrangou et al. 2007; Brouns et al. 2008; Sorek et al. 2008; van der Oost et al. 2009; Horvath et al. 2010; Labrie et al. 2010; Marraffini and Sontheimer 2010). CRISPRs are found on bacterial and archaeal genomes and are structured as a series of short repeats of about 20-50 bp in length, interspersed by slightly longer sequences, called "spacers," that are 25-75 bp in length. These spacers are synthesized to match a short region of foreign DNA, such as a virus or a plasmid that has invaded the cell. This short spacer region is then inserted into the CRISPR locus. Subsequently, if foreign DNA containing a region with a close match to a CRISPR sequence invades the cell in the future, an immune response is mobilized via CRISPR-associated, or *Cas*, genes that work in conjunction with small RNAs derived from the CRISPR spacers to bind and cleave invading DNA (Garneau et al. 2010; Jore et al. 2011).

CRISPRs have great potential for the study of viral ecology and viral-host coevolution because each locus can be treated as a library of previous viral infections for a given strain. Previous studies have used CRISPRs as an ecological tool to track the impacts of viruses over time and space within and between microbial populations (Tyson and Banfield 2008; Held and Whitaker 2009). We have previously used CRISPRs as a tool to identify potential hosts of the viral assemblage of hydrothermal vent diffuse flow fluids by constructing a CRISPR spacer database to query viral metagenomes and identify potential hosts based on matches of metagenomic reads to known CRISPR sequences in archaeal and bacterial genomes (Anderson et al. 2011b). This analysis indicated that viruses in the hot subsurface near vent systems have the potential to infect several different taxa of archaea and bacteria from a wide range of thermal regimes. Expanding this analysis to viral assemblages from other regions of the subsurface would indicate whether viral assemblages throughout the deep subsurface are similarly broad in host range, or whether only certain groups tend to be susceptible to viral infection.

Another outstanding question regarding viral roles in the subsurface is the degree to which viruses mediate horizontal gene transfer, or manipulate archaeal and bacterial genomes through incorporation as prophage. One way to address this question is to examine viral and cellular metagenomes for sequences potentially associated with mobile elements, such as those that encode transposases, integrases, and recombinases. The presence of abundant genes that encode such enzymes provides one piece of evidence that the organisms in a particular community actively exchange genes. Table 2 shows the percentage of reads in a set of viral metagenomes that had a match to a transposase, recombinase, or integrase domain (maximum e-value: 10^{-5}). Each of the viral metagenomes listed in the table was sequenced with pyrosequencing technology, and each of the reads for these metagenomes ranged between approximately 90-300 base pairs. The hydrothermal vent viral metagenome contained one of

Table 2. Percent of reads in a set of viral metagenomes with a match to a transposase, recombinase, or integrase protein domain. tblastn searches were conducted with a set of Pfam seed sequences matching transposase, recombinase, and integrase domains as the query. Hits were designated as a "match" if they had an e-value below 10^{-5}.

Metagenome	Number of reads	Number of hits	% of reads with a hit	Reference
Antarctic Lake summer	30515	66	0.22	Lopez-Bueno et al. 2009
Hydrothermal vent	231246	227	0.098	Anderson et al. 2011b
Arctic Ocean	688590	605	0.088	Angly et al. 2006
Bay of British Columbia	138347	81	0.059	Angly et al. 2006
Antarctic Lake spring	31691	13	0.041	Lopez-Bueno et al. 2009
Gulf of Mexico	263908	106	0.040	Angly et al. 2006
Microbialites	621110	246	0.040	Desnues et al. 2008
Coral	36354	14	0.039	Dinsdale et al. 2008
Sargasso Sea	399343	17	0.0043	Angly et al. 2006

the highest proportions of reads matching a potential component of a mobile element, perhaps implying that viruses in the vent viral assemblage actively transduce genes or integrate their genomes into that of their hosts. This trend is consistent with the hypothesis that lysogeny is a prevalent lifestyle in hydrothermal systems and that viruses actively transduce genetic material from one host to the next. Interestingly, the authors of the paper describing the Antarctic Lake virome noted that there was a distinct shift from ssDNA viruses in the spring to dsDNA viruses in the summer (Lopez-Bueno et al. 2009). They suggest that under cover of the ice, metabolic rates are restricted, and thus the only actively replicating phages may be lytic ssDNA viruses with small genomes. In contrast, in the summer, larger dsDNA phages (which presumably were integrated as prophage in the winter and spring) were able to enter the lytic phase and therefore would have been collected in the viral size fraction sampled for the metagenome. This transition is reflected in the relative abundance of mobile elements in each of the metagenomes: the summer fraction, in which the lysogenic phage have entered the lytic phase and are therefore captured in the sample, contains a higher abundance of sequences potentially associated with mobile elements. It would therefore be quite interesting to examine a viral metagenome from deeply buried sediments, where we also hypothesize lysogeny to be common, and calculate the relative abundance of mobile elements. Induction of the sample prior to sampling would cause lysogenized viruses to enter the lytic cycle, so that they can be sampled in the viral size fraction (Anderson et al. 2011a).

Similarly, an analysis of viral metagenomes from a range of different environments indicates that DNA ligases are also particularly abundant in the viral assemblage from diffuse flow hydrothermal fluids (Anderson et al. 2011a), with over ten times the percentage of reads encoding a DNA ligase compared to the metagenome with the next highest abundance. As ligases are known to play a role in repairing double-stranded breaks, it is possible that these ligases play a role in facilitating horizontal gene transfer in the subsurface, though further research will be required to support or disprove this hypothesis.

If genes are in fact being transferred by viruses in these environments, a subsequent question to ask is the extent of gene diversity within the viral gene pool: does the viral assemblage contain a high diversity of protein-encoding genes that can be potentially transferred, or is the pool of genes relatively small? One way to address this question is to construct rarefaction

curves comparing the relative gene diversity of viral metagenomes. Figure 5 depicts rarefaction curves of protein-encoding genes in viral metagenomes from several different environments. ORFs derived from reads in these metagenomes identified by FragGeneScan in the MG-RAST pipeline (Meyer et al. 2008) were clustered to 40% similarity using UCLUST (Edgar 2010). Rarefaction curves were generated with mothur (Schloss et al. 2009). The slopes of these curves indicate that most of the marine viral assemblages have fairly high diversity compared to the Antarctic Lake metagenomes and two of the microbialite metagenomes, perhaps reflecting the diversity of the hosts they infect in each of these environments. Further work will elucidate the functions of the genes encoded by these viral assemblages, and comparative metagenomics between viral and cellular fractions may provide insight into which types of genes are enriched in the viral fraction compared to the cellular fraction, perhaps indicating which genes are selected to be maintained in the viral gene pool.

In summary, metagenomics holds great potential for illuminating the virus-host relationship, and further sequencing of both viral and cellular metagenomes from regions of the deep subsurface will contribute much to our understanding of which organisms are most susceptible to viral infection, whether the lytic or lysogenic lifestyle is more common in the subsurface, and the nature of the role viruses play in facilitating horizontal gene transfer in the deep subsurface. The isolation of virus-host systems from organisms in pure culture would also provide further insights into the nature of the host-virus relationship in the subsurface. Virus-host systems could uncover new morphologies or escape mechanisms, particularly among the archaeal viruses, or new relationships between the virus and the host. Experiments with virus-host systems could reveal new insights into the nature of viral manipulation of the host by identifying which genes are expressed, upregulated or downregulated when prophage are integrated into a host genome.

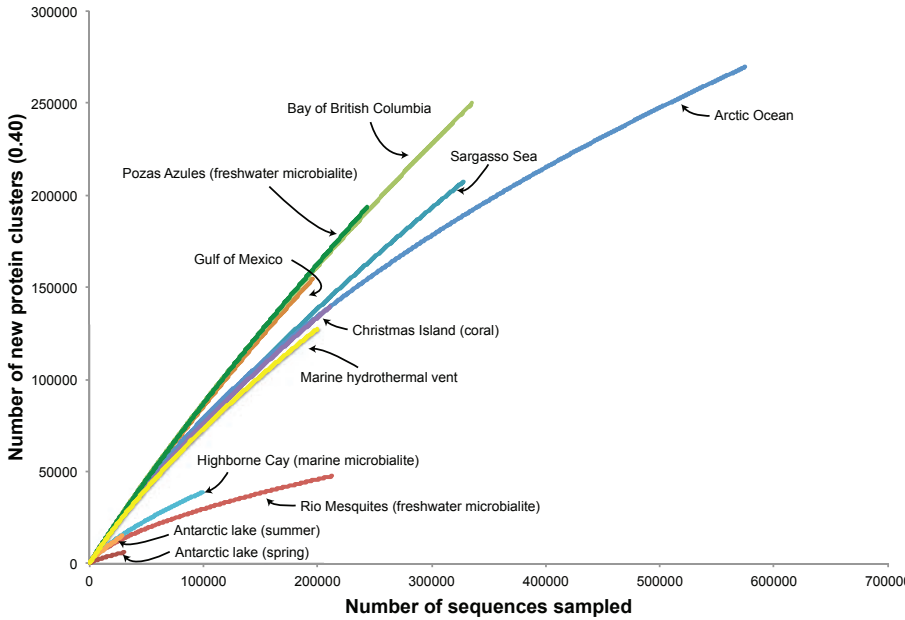

Figure 5. Rarefaction curves of clusters of translated open reading frames from viral metagenomic sequences from around the globe. ORFs were clustered to 0.40 similarity using UCLUST (Edgar 2010). Rarefaction curves were generated in mothur (Schloss et al. 2009).

VENTS, VIRUSES AND THE ORIGIN OF LIFE

The evidence appears to be clear that viruses can have a substantial impact on the evolution of their hosts, and have most likely been doing so for billions of years. But for how long has this mutual evolutionary relationship persisted? When and how did viruses originate? The question is particularly relevant here because the deep subsurface, and hydrothermal systems in particular, are often considered the most ancient continuously inhabited ecosystems on the planet (Reysenbach and Shock 2002), and indeed, are often thought to have been an important setting for the origin of life on Earth (e.g., Baross and Hoffman 1985; see Hazen 2005).

Hydrothermal vents and the deep subsurface: key settings in the origin of life

On the Hadean Earth, about 4 billion years ago, hydrothermal vent systems would have been present in perhaps even greater abundance than they are today. Residual heat of formation would have resulted in a volcanically and seismically more active planet, with longer mid-ocean ridges and more plate tectonic activity (Hargraves 1986), resulting in a higher incidence of water-rock reactions at the bottom of the ocean. Both basalt-hosted and peridotite-hosted hydrothermal systems are likely to have been present in the Hadean Earth. Metal-sulfide minerals in basalt-hosted systems, including pyrite, have been implicated as key catalysts in several important prebiotic reactions in which CO or CO_2 is fixed into simple organic compounds (Wächtershäuser 1988a, 1988b, 1990; Cody 2004). Peridotite-hosted systems, formed off-axis and powered by serpentinization through the interaction of seawater with peridotite, are characterized by lower-temperature fluids with high pH and form large calcium carbonate structures (Kelley et al. 2001; Schrenk et al. 2013).These vents may have also acted as a source for key organic compounds generated in the process of serpentinization, including formate, acetate, methane, organic sulfur compounds, and larger hydrocarbons (Heinen and Lauwers 1996; Proskurowski et al. 2008; Lang et al. 2010). Moreover, the calcium carbonate porous structures formed in these systems may have acted as a concentrating mechanism for early prebiotic compounds (Baaske et al. 2007).

One of the most appealing aspects of hydrothermal vents as a setting for the origin of life is the formation of geological, physical, and chemical gradients in these systems (Baross and Hoffman 1985). These gradients provide a wide range of environmental conditions within a relatively small physical space with fluid flow between them, facilitating the occurrence of multiple chemical processes across a multiplicity of environmental conditions in parallel. These gradients extend beyond hydrothermal vent fields themselves to other regions of the deep subsurface. The minerals catalyzing reactions in one region, such as at basalt-hosted hydrothermal vents, would have differed from those in other regions of the subsurface, such as at peridotite-hosted systems or in sedimented regions. As mentioned above, much of the ocean crust is linked by fluid flux, which moves at different flow rates and volumes depending on the depth and degree of porosity in the crust. Thus, compounds synthesized in one region of the ocean crust, whether at a hydrothermal system or more distal to a mid-ocean ridge, could be transferred from one region of the subsurface to the next.

In this sense, the deep subsurface may have acted as a natural laboratory for the origin of life, in which multiple "experiments" could have been carried out in tandem. Later, the products of these natural experiments could have been combined to form an autocatalytic network. Several studies have suggested that the chemiosmotic gradients at vent sites, combined with enclosed pore spaces and organic syntheses, could have resulted in the first autocatalytic networks (Koonin and Martin 2005; Martin and Russell, 2007; Martin et al. 2008; Lane et al. 2010). Martin, Russell and others describe a model in which a chemiosmotic potential is generated across the membrane of an iron-sulfide bubble, which they presume would form in an anoxic Hadean ocean. The potential could then have been harnessed to yield a protometabolism based on the reduction of CO_2 by H_2.

The question that arises is how the first self-replicating entities formed and evolved in these settings. Koonin and Martin (2005) suggest that self-replicating networks could have formed within the walls of these iron-sulfide compartments. In their scenario, each component of the network consisted of a selfish RNA molecule encoding one or a few proteins, with the original selection pressures favoring rapid self-replication. The authors refer to these replicating entities as "virus-like RNA molecules," which Koonin then elaborated upon in a later publication detailing the "Virus World" (Koonin et al. 2006). In this model, the authors describe a scenario in which viruses emerged early from the various replicating entities and networks that formed part of the RNA world. These scenarios suggest that viruses may have played a primary role at the earliest stages of life's evolution.

The viral role in the origin of life

Viruses have not always been considered to be primordial elements. Historically, three theories were put forward regarding the origin of viruses: first, that viruses were originally parasitic cells that evolved into a viral-like form (the "reduction hypothesis"); second, that viruses were rogue genetic elements from cells that developed a protein capsid to survive in an extracellular state (the "escape hypothesis"); and third, that viruses originated in parallel with cells (the "virus first hypothesis") (Prangishvili et al. 2006). The last hypothesis, however, has been gaining favor as scientists have found that viruses infecting different domains of life share certain "hallmark genes" that are missing from cellular genomes, perhaps pointing to an early origin that predates the divergence of the three domains of life (Koonin et al. 2006). Others have suggested that DNA as a genetic material first arose in a virus, which later spread to the cellular world (Forterre 2006).

The tremendous diversity of viruses, though, greatly complicates an elucidation of their origin. Viruses encompass one portion of a spectrum of mobile genetic elements, which range in size and complexity from simple introns and transposons, to GTAs, to RNA viruses, viroids, and satellite viruses, to dsDNA viruses and the giant Mimivirus described above. These elements may not share a common origin, yet in many cases, many virus-like elements share genes that are not found in the cellular world (Koonin et al. 2006), or share structural features in their protein capsids (Bamford et al. 2005). Many of these shared attributes transcend domains, leading many to consider viruses to be ancient.

The attribute that all viruses share is their dependence on a host for the purposes of replication: in a word, these are parasites. Consideration of the role of parasites in the origin of life is not a new concept. In an RNA-protein world, or even a pre-RNA world, parasites could have undermined replication networks, as they could take resources from these replication networks (or "hosts," in a sense) without benefiting them, and thus destroy the cycle (Maynard-Smith and Szathmáry 1999). In these networks, elements are linked such that each element replicates another element in the cycle. Parasites emerge when a mutant of one of the elements is preferentially replicated, but does not replicate another element in the cycle (Fig. 6A). It has been suggested that containing replication cycles within a compartment, or at least confining them to a surface, may circumvent this problem by placing the selective pressure not on the individual elements within a replication cycle, but on the cycle as a whole (Maynard-Smith and Szathmáry 1999). This compartmentalization effectively provides a basis for heredity and competition between individuals, which are required for natural selection to occur. Moreover, spatial structuring of the environment may reduce the spread of parasites from one hypercycle to the next (Boerlijst and Hogeweg 1991). However, spatial structuring may prevent "sharing" of new functions through horizontal gene transmission (Poole 2009).

Yet as we have discussed here, parasites can at times improve the fitness of the host they depend upon. Just as modern viruses can express fitness factors to boost the fitness of their host, the same may have applied in life's early evolution: for example, if a selfish element were to

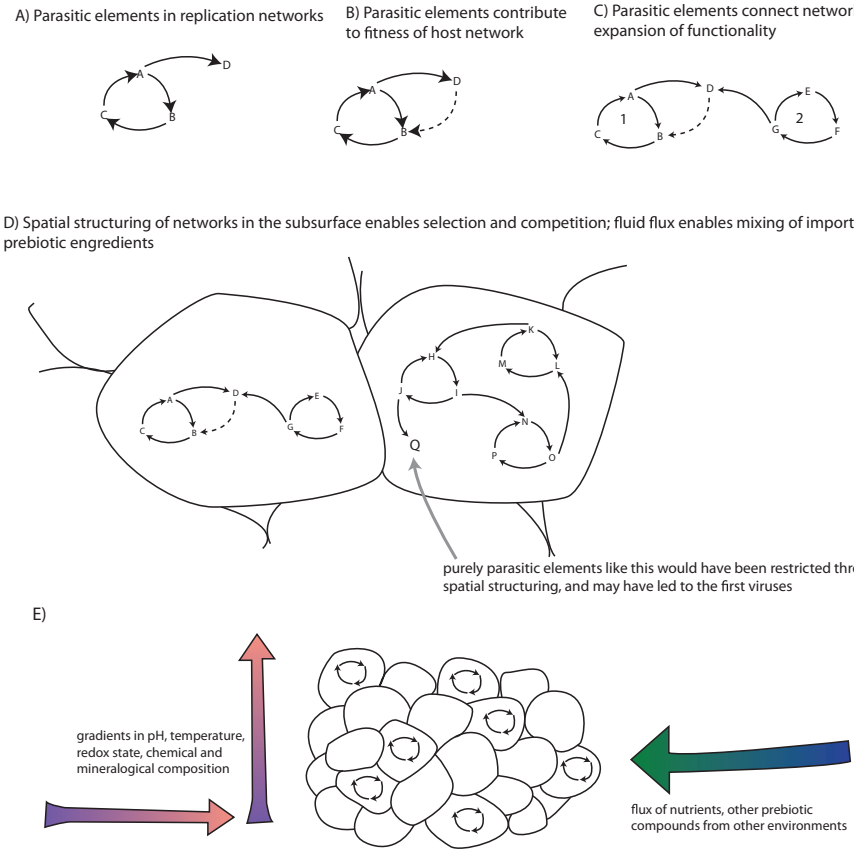

Figure 6. Role of parasitic elements in early replication cycles. A) Cooperative replication network. Element D is a parasite to the network because it uses network resources, but it does not contribute to network fitness. B) Selection may favor elements that are able to contribute to the fitness of the host network. Here, element D contributes to the stability of element B, thus improving network fitness. This feedback improves the fitness of D as well. C) If element D is also replicated by element G in another network, this replication could link the two networks, thereby increasing network functionality. D) Spatial structuring through restriction to mineral surfaces, or enclosure in pore spaces, could restrict the degree to which parasitic elements (like element Q, considered "parasitic" here because it does not contribute to the parent network or any other networks) spread between networks. E) Scenario in which replicating networks operate in the subsurface, with diverse replicating networks defined by gradients in environmental conditions, and fed by an influx of prebiotic compounds through fluid flux in the subsurface.

contribute toward the overall fitness of a given replication cycle, such as through stabilizing another element in the cycle, this change would improve the fitness of the whole cycle and therefore the fitness of the element as well (Fig. 6B). Indeed, this process may have been the means by which new functions were added to replication networks. Just as modern viruses can contribute novel genetic material through transduction or expression of fitness-boosting genes, ancient parasites may have increased the functionality of the networks they were part of by linking together disparate networks. Rather than (or in addition to) presenting a problem for early replication networks, early viruses may have provided a means by which to increase their functionality. Viral-like particles or selfish elements may have acted as a means to share genes between networks, and ultimately may have allowed early genomes to expand (Fig. 6C).

In this sense the meaning of words like "parasite" or "mutualist" become blurred, as selection at this level may favor varying degrees of parasitism or mutualism; in the RNA world, selection operated at the level of both the individual elements and at the level of whole networks.

This scenario is consistent with previously published ideas about the origin of life that may or may not have explicitly specified roles for viruses. For example, the idea that all life today evolved from primitive cells in an early biofilm-like community evolving "through prolific genetic exchange with other 'precells' in the community, perhaps involving structures resembling transposable genetic elements and viral-like particles" (Baross and Hoffman 1985) was inspired by the initial discovery of prolific subsurface life evident at hydrothermal vents. Woese (1998, 2002) has developed in detail the idea of a communal ancestor in which horizontal gene transfer is the primary driver of evolution and viral-like elements could very well have been one of the mediators of this gene transfer.

However, as with modern viruses, these parasitic elements likely would have had a wide host range, in which some could spread rapidly to other networks, whereas others were more restricted. Modern viruses also exhibit a range of virulence, in which some were almost entirely parasitic whereas others are almost entirely mutualistic, and we envision prebiotic selfish elements to have had similar characteristics. In this sense, indiscriminate horizontal gene transfer in the communal ancestor may have been disruptive by facilitating the spread of the more virulent, entirely deleterious parasites (Poole 2009). It is also unclear how the communal ancestor would have evolved "as a unit" (Woese 1998) without competition or selection with other units. We therefore envision a scenario in which spatial structuring of the environment facilitates selection at the level of an entire network, rather than on individual elements. This structuring would also serve to restrict the movement of wide-ranging, deleterious parasites. Iron-sulfur bubbles or pores in hydrothermal systems could have acted as a structuring mechanism prior to the emergence of lipid membranes (Russell and Hall 1997; Koonin and Martin 2005; Martin and Russell 2007) (Fig. 6D).

On a larger scale in the prebiotic world, there would have been extensive fluid flux in the subsurface, and this transport could have served as a conduit for nutrients or products of prebiotic reactions from other environments. Subsurface flow would have connected these networks and facilitated some degree of gene sharing between them, as well as provided them with important prebiotic precursors (Fig. 6E). Gradients in temperature, pH, chemical and mineralogical composition through the subsurface or in hydrothermal structures would have generated diversity in these replicating networks, creating variation in the population and facilitating selection among them. In this sense, the environment may have fostered the earliest stages of natural selection, and these early viral-like or "selfish" elements may have been important in facilitating gene transfer between these networks, allowing them to grow and change.

Regardless of the degree to which horizontal gene transfer occurred, subpopulations within the ancestral community became more resistant to genetic exchange with other subpopulations over time, which may have arisen as a defense against parasitic genetic elements: the first viruses. The crossing of this "Darwinian threshold" from one ancestral community to many independent cells marked the origin of speciation and the emergence of the life forms we know today (Woese 2002), and it is possible that viruses or viral-like elements were intimately involved in this critical stage in the evolution of life.

CONCLUSION

There is clearly much work that remains to be done to understand the nature of the viral impact on biogeochemical cycling, microbial community structure, and evolution in the deep subsurface. Yet the few available details provide tantalizing hints that the role of viruses in the

deep subsurface could be profound on many levels. Viral infection may significantly impact biogeochemical cycling in the subsurface through lysis of cellular biomass, releasing nutrients and compounds that would otherwise be entrained in biomass. Viruses are also known to alter the structure of the microbial communities they infect, potentially increasing overall diversity through lysis of cells that become most abundant in a given region. Through the process of lysogeny and transduction, viruses may manipulate the genomes and expression of the hosts they infect throughout the subsurface, effectively resulting in a mutualistic, symbiotic relationship between host and virus that transcends traditional notions of viruses as parasites. Indeed, this role of virus as parasite, as mutualist, and as a sharer of information through gene transfer may be a fundamental underpinning of life in the deep subsurface that extends back in time to the dawn of life itself. Finally, the hot subsurface environments associated with hydrothermal systems harbor many of the most deeply rooted microorganisms on the universal phylogenetic tree of life; most are hyperthermophilic archaea. Very little is known about the viruses that are associated with these microorganisms. Given their antiquity, including their primordial setting, it is possible they harbor viruses and virus-like particles that could lead to a better understanding of the origin of viruses and their role in the early evolution of life.

REFERENCES

Anderson RE, Brazelton WJ, Baross JA (2011a) Is the genetic landscape of the deep subsurface biosphere affected by viruses? Front Extreme Microbiol 2:00219 doi: 10.339/fmicb.2011.00219

Anderson RE, Brazelton WJ, Baross JA (2011b) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. FEMS Microbiol Ecol 77:120-133, doi: 10.1111/j.1574-6941.2011.01090.x/full

Andersson A, Banfield J (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320:1047-1050, doi: 10.1126/science.1157358

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F (2006) The marine viromes of four oceanic regions. PLoS Biol 4:e368, doi: 10.1371/journal.pbio.0040368

Baas Becking LGM (1934) Geobiologie of Inleiding tot de Milieukunde. Van Stockum & Zoon, The Hague

Baaske P, Weinert FM, Duhr S, Lemke KH, Russell MJ, Braun D (2007) Extreme accumulation of nucleotides in simulated hydrothermal pore systems. Proc Natl Acad Sci USA 104:9346-9351, doi: 10.1073/pnas.0609592104

Bamford DH, Grimes JM, Stuart DI (2005) What does structure tell us about virus evolution? Curr Opin Struc Biol 15:655-663, doi: 10.1016/j.sbi.2005.10.012

Baross JA, Hoffman SE (1985) Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. Origins Life Evol B 15:327–345

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709-1712, doi: 10.1126/science.1138140

Biddle JF, Lipp JS, Lever MA, Lloyd KG, Sørensen KB, Anderson R, Fredricks HF, Elvert M, Kelly TJ, Schrag DP, Sogin ML, Brenchley JE, Teske A, House CH, Hinrichs K-U (2006) Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. Proc Natl Acad Sci USA 103:3846-3851, doi: 10.1073/pnas.0600035103

Bize A, Karlsson EA, Ekefjärd K, Quax TEF, Pina M, Prevost M-C, Forterre P, Tenaillon O, Bernander R, Prangishvili D (2009) A unique virus release mechanism in the Archaea. Proc Natl Acad Sci USA 106:11306-11311, doi: 10.1073/pnas.0901238106

Boerlijst MC, Hogeweg P (1991) Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. Physica D 48:17-28, doi: 10.1016/0167-2789(91)90049-F

Bratbak G, Egge J, Heldal M (1993) Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. Mar Ecol-Prog Ser 93:39-48, doi: 10.3354/meps093039

Breitbart M, Miyake JH, Rohwer F (2004) Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol Lett 236:249-256, doi: 10.1111/j.1574-6968.2004.tb09654.x

Breitbart M, Rohwer F (2005) Here a virus, there a virus, everywhere the same virus? Trends Microbiol 13:278-284. doi: 16/j.tim.2005.04.003

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F (2002) Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci USA 99:14250-14255, doi: 10.1073/pnas.202488399

Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ. H, Snijders AP. L, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321:960-964, doi: 10.1126/science.1159689

Brumfield SK, Ortmann AC, Ruigrok V, Suci P, Douglas T, Young MJ (2009) Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus Sulfolobus turreted icosahedral virus. J Virol 83:5964-5970, doi: 10.1128/JVI.02668-08

Cody GD (2004) Transition metal sulfides and the origins of metabolism. Annu Rev Earth Planet Sci 32:569-599, doi: 10.1146/annurev.earth.32.101802.120225

Cohan FM (2002) What are bacterial species? Annu Rev Microbiol 56:457-487, doi: 10.1146/annurev.micro.56.012302.160634

Colwell FS, D'Hondt S (2013) Nature and extent of the deep biosphere. Rev Mineral Geochem 75:547-574

Corinaldesi C, Dell'anno A, Danovaro R (2011) Viral infections stimulate the metabolism and shape prokaryotic assemblages in submarine mud volcanoes. ISME J 6:1250-1259, doi: 10.1038/ismej.2011.185

Culley AI, Lang AS, Suttle CA (2003) High diversity of unknown picorna-like viruses in the sea. Nature 424:1054-1057, doi: 10.1038/nature01886

Culley AI, Lang AS, Suttle CA (2006) Metagenomic analysis of coastal RNA virus communities. Science 312:1795-1798, doi: 10.1126/science.1127404

D'Hondt S, Spivack AJ, Pockalny R, Ferdelman TG, Fischer JP, Kallmeyer J, Abrams LJ, Smith DC, Graham D, Hasiuk F, Schrum H, Stancin AM (2009) Subseafloor sedimentary life in the South Pacific Gyre. Proc Natl Acad Sci USA 106:11651-11656, doi: 10.1073/pnas.0811793106

Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C, Weinbauer M (2008) Major viral impact on the functioning of benthic deep-sea ecosystems. Nature 454:1084-1087, doi: 10.1038/nature07268

Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M, Liu H, Furlan M, Wegley L, Chau B, others, Ruan Y, Hall D, Angly FE, Edwards RA, Li L, Thurber RV, Reid RP, Siefert J, Souza V, Valentine DL, Swan BK, Breitbart M, Rohwer F (2008) Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature 452:340–343, doi: 10.1038/nature06735

Diemer GS, Stedman KM (2012) A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. Biol Direct 7:13, doi: 10.1186/1745-6150-7-13

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran M, Nelson K, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F (2008) Functional metagenomic profiling of nine biomes. Nature 452:629-632, doi: 10.1038/nature06810

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-1, doi: 10.1093/bioinformatics/btq461

Edwards KJ, Bach W, McCollom TM (2005) Geomicrobiology in oceanography: microbe-mineral interactions at and below the seafloor. Trends Microbiol 13:449–456

Edwards KJ, Wheat CG, Sylvan JB (2011) Under the sea: microbial life in volcanic oceanic crust. Nature Rev Microbiol 9:703-712, doi: 10.1038/nrmicro2647

Engelhardt T, Sahlberg M, Cypionka H, Engelen B (2011) Induction of prophages from deep-subseafloor bacteria. Environ Microbiol 3:459-465, doi: 10.1111/j.1758-2229.2010.00232.x

Engelhardt T, Sahlberg M, Cypionka H, Engelen B (2012) Biogeography of Rhizobium radiobacter and distribution of associated temperate phages in deep subseafloor sediments. ISME J, advance online publication, doi: 10.1038/ismej.2012.92

Eydal HSC, Jägevall S, Hermsson M, Pedersen K (2009) Bacteriophage lytic to *Desulfovibrio aespoeensis* isolated from deep groundwater. ISME J 3:1139-1147, doi: 10.1038/ismej.2009.66

Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, Jackson RB (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl Environ Microb 73:7059-7066, doi: 10.1128/AEM.00358-07

Filippini M, Buesing N, Bettarel Y, Sime-Ngando T, Gessner MO (2006) Infection paradox: high abundance but low impact of freshwater benthic viruses. Appl Environ Microbiol 72:4893-4898, doi: 10.1128/AEM.00319-06

Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010) Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci USA 107:19508

Fisher A, Becker K (2000) Channelized fluid flow in oceanic crust reconciles heat-flow and permeability data. Nature 403:71-74, doi: 10.1038/47463

Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 117:5-16, doi: 10.1016/j.virusres.2006.01.010

Fuhrman JA (2009) Microbial community structure and its functional implications. Nature 459:193-199, doi: 10.1038/nature08058

Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468:67-71

Hargraves RB (1986) Faster spreading or greater ridge length in the Archean? Geology 14:750, doi: 10.1130/0091-7613(1986)14<750:FSOGRL>2.0.CO;2

Häring M, Vestergaard G, Rachel R, Chen L, Garrett RA, Prangishvili D (2005) Virology: independent virus development outside a host. Nature 436:1101-1102, doi: 10.1038/4361101a

Hazen RM (2005) Genesis: The Scientific Quest for Life's Origins. Joseph Henry Press, Washington, DC

Heinen W, Lauwers AM (1996) Organic sulfur compounds resulting from the interaction of iron sulfide, hydrogen sulfide and carbon dioxide in an anaerobic aqueous environment. Origins Life Evol Biosph 26:131-150, doi: 10.1007/BF01809852

Held NL, Whitaker RJ (2009) Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. Environ Microbiol 11:457-466, doi: 10.1111/j.1462-2920.2008.01784.x

Horvath P, Barrangou R, Hovarth P (2010) CRISPR/Cas, the Immune System of Bacteria and Archaea. Science 327:167-170, doi: 10.1126/science.1179555

Jiang SC, Paul JH (1998) Gene transfer by transduction in the marine environment. Appl Environ Microbiol 64:2780-2787

Jiao N, Herndl GJ, Hansell DA, Benner R, Kattner G, Wilhelm SW, Kirchman DL, Weinbauer MG, Luo T, Chen F, Azam F (2010) Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. Nature Rev Microbiol 8:593-599, doi: 10.1038/nrmicro2386

Johnson HP, Pruis MJ (2003) Fluxes of fluid and heat from the oceanic crustal reservoir. Earth Planet Sci Lett 216:565-574, doi: 10.1016/S0012-821X(03)00545-4

Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul Ü, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders APL, Dickman MJ, Doudna JA, Boekema EJ, Heck AJR, van der Oost J, Brouns SJJ, Pul Ü, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snigders APL, Dickman MG, Doudna JA, Boekma EJ, Heck AJR, van der Oost J, Brouns SJJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature Struct Mol Biol 18:529-536, doi: 10.1038/nsmb.2019

Jørgensen BB, D'Hondt S (2006) Ecology. A starving majority deep beneath the seafloor. Science 314:932-934, doi: 10.1126/science.1133796

Kelley DS, Karson JA, Blackman DK, Früh-Green GL, Butterfield DA, Lilley MD, Olson EJ, Schrenk MO, Roe KK, Lebon GT, Rivizzigno P (2001) An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 degrees N. Nature 412:145-149, doi: 10.1038/35084000

Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. Trends Genet 21:647-654, doi: 10.1016/j.tig.2005.09.006

Koonin EV, Senkevich TG, Dolja VV (2006) The ancient Virus World and evolution of cells. Biol Direct 1:29, doi: 10.1186/1745-6150-1-29

Kristensen DM, Mushegian AR, Dolja VV, Koonin EV (2009) New dimensions of the virus world discovered through metagenomics. Trends Microbiol 18:11-19

Krupovic M, Prangishvili D, Hendrix RW, Bamford DH (2011) Genomics of bacterial and archaeal viruses: Dynamics within the prokaryotic virosphere. Microbiol Mol Biol Rev 75:610-635, doi: 10.1128/MMBR.00011-11

Kyle JE, Eydal HSC, Ferris FG, Pedersen K (2008) Viruses in granitic groundwater from 69 to 450m depth of the Äspö hard rock laboratory, Sweden. ISME J 2:571-574, doi: 10.1038/ismej.2008.18

La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, Birtles R, Claverie J-M, Raoult D (2003) A giant virus in amoebae. Science 299:2033, doi: 10.1126/science.1081867

La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E, Raoult D (2008) The virophage as a unique parasite of the giant mimivirus. Nature 455:100-104, doi: 10.1038/nature07218

Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. Nature Rev Microbiol 8:317-327, doi: 10.1038/nrmicro2315

Lane N, Allen JF, Martin W (2010) How did LUCA make a living? Chemiosmosis in the origin of life. BioEssays 32:271-280, doi: 10.1002/bies.200900131

Lang AS, Beatty JT (2000) Genetic analysis of a bacterial genetic exchange element: The gene transfer agent of *Rhodobacter capsulatus*. Proc Natl Acad Sci USA 97:859-864, doi: 10.1073/pnas.97.2.859

Lang SQ, Butterfield DA, Schulte M, Kelley DS, Lilley MD (2010) Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. Geocim Cosmochim Act 74:941-952, doi: 10.1016/j.gca.2009.10.045

Li W, Godzik A (2006) cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659, doi: 10.1093/bioinformatics/btl158

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. Nature 438:86-89, http://dx.doi.org/10.1038/nature04111

Lopez-Bueno A, Tamames J, Velazquez D, Moya A, Quesada A, Alcami A (2009) High diversity of the viral community from an Antarctic lake. Science 326:858

Lorenz MG, Wackernagel W (1994) Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Mol Biol Rev 58:563-602

Mann NH, Cook A, Millard A, Bailey S, Clokie M (1993) Bacterial photosynthesis genes in a virus. Environ Microbiol 59:3736–3743

Marraffini LA, Sontheimer EJ (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463:568-571, doi: 10.1038/nature08703

Martin W, Baross JA, Kelley D, Russell MJ (2008) Hydrothermal vents and the origin of life. Nature Rev Microbiol 6:805-814, doi: 10.1038/nrmicro1991

Martin W, Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. Philos Trans R Soc London Ser B 362:1887-1925, doi: 10.1098/rstb.2006.1881

Matson EG, Thompson MG, Humphrey SB, Zuerner RL, Stanton TB (2005) Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. J Bacteriol 187:5885-5892, doi: 10.1128/JB.187.17.5885-5892.2005

Maynard-Smith J, Szathmáry E (1999) The Origins of Life: From the Birth of Life to the Origin of Language. Oxford University Press, New York

McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH (2010) High frequency of horizontal gene transfer in the oceans. Science 330:50

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386, doi: 10.1186/1471-2105-9-386

Middelboe M (2000) Bacterial growth rate and marine virus-host dynamics. Microbial Ecol 40:114-124, doi: 10.1007/s002480000050

Middelboe M, Glud RN, Wenzhöfer F, Oguri K, Kitazato H (2006) Spatial distribution and activity of viruses in the deep-sea sediments of Sagami Bay, Japan. Deep-Sea Res Part 1 53:1-13, doi: 16/j.dsr.2005.09.008

Molin S, Tolker-Nielsen T (2003) Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. Curr Opin Biotech 14:255-261, doi: 10.1016/S0958-1669(03)00036-3

O'Malley MA (2007) The nineteenth century roots of "everything is everywhere". Nature Rev Microbiol 5:647-651, doi:10.1038/nrmicro1711

Ortmann AC, Suttle CA (2005) High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. Deep-Sea Res Part 1 52:1515–1527

Paul JH (2008) Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? ISME J 2:579-589

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

Pina M, Bize A, Forterre P, Prangishvili D (2011) The archaeoviruses. FEMS Microbiol Rev 35:1035-1054, doi: 10.1111/j.1574-6976.2011.00280.x

Poole AM (2009) Horizontal gene transfer and the earliest stages of the evolution of life. Res Microbiol 160:473-480, doi: 10.1016/j.resmic.2009.07.009

Prangishvili D, Forterre P, Garrett RA (2006) Viruses of the Archaea: a unifying view. Nature Rev Microbiol 4:837-848

Proskurowski G, Lilley MD, Seewald JS, Früh-Green GL, Olson EJ, Lupton JE, Sylva SP, Kelley DS (2008) Abiogenic hydrocarbon production at lost city hydrothermal field. Science 319:604-607, doi: 10.1126/science.1151194

Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie J-M (2004) The 1.2-megabase genome sequence of Mimivirus. Science 306:1344-1350, doi: 10.1126/science.1101485

Resch A, Fehrenbacher B, Eisele K, Schaller M, Götz F (2005) Phage release from biofilm and planktonic *Staphylococcus aureus* cells. FEMS Microbiol Lett 252:89-96, doi: 10.1016/j.femsle.2005.08.048

Reysenbach A-L, Shock E (2002) Merging genomes with geochemistry in hydrothermal systems. Science 296:1077-1082, doi: 10.1126/science.1072483

Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, Mira A (2009) Explaining microbial population genomics through phage predation. Nature Rev Microbiol 7:828-836, doi: 10.1038/nrmicro2235

Rosario K, Breitbart M (2011) Exploring the viral world through metagenomics. Curr Opin Virol 1:289-297, doi: 10.1016/j.coviro.2011.06.004

Røy H, Kallmeyer J, Adhikari RR, Pockalny R, Jørgensen BB, D'Hondt S (2012) Aerobic microbial respiration in 86-milion-year-old deep-sea red clay. Science 336:922-925, doi: 10.1126/science.1219424

Russell MJ, Hall AJ (1997) The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. J Geol Soc London 154:377-402, doi: 10.1144/gsjgs.154.3.0377

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537-7541, doi: 10.1128/AEM.01541-09

Schrenk MO, Kelley DS, Delaney JR, Baross JA (2003) Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. Appl Environ Microbiol 69:3580-3592, doi: 10.1128/AEM.69.6.3580-3592.2003

Schrenk MO, Brazelton WJ, Lang SQ (2013) Serpentinization, carbon, and deep life. Rev Mineral Geochem 75:575-606

Skraber S, Schiven J, Gantzer C, de Roda Husman AM (2005) Pathogenic viruses in drinking-water biofilms: A public health risk? Biofilms 2:105-117

Sorek R, Kunin V, Hugenholtz P (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nature Rev Microbiol 6:181-186, doi: 10.1038/nrmicro1793

Stanton TB (2007) Prophage-like gene transfer agents–Novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. Anaerobe 13:43-49

Steward GF, Montiel JL, Azam F Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. Linmol Oceanogr 45:1697-1706

Stoodley P, Sauer K, Davies DG, Costerton JW (2002) Biofilms as complex differentiated communities. Annu Rev Microbiol 56:187-209, doi: 10.1146/annurev.micro.56.012302.160705

Suttle CA (2005) Viruses in the sea. Nature 437:356–361, doi: 10.1038/nature04160

Suttle CA (2007) Marine viruses--major players in the global ecosystem. Nature Rev Microbiol 5:801-812, doi: 10.1038/nrmicro1750

Thingstad T, Lignell R (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat Microb Ecol 13:19-27, doi: 10.3354/ame013019

Thoulouze M-I, Alcover A (2011) Can viruses form biofilms? Trends Microbiol 19:257-262, doi: 10.1016/j.tim.2011.03.002

Thurber RV (2009) Current insights into phage biodiversity and biogeography. Curr Opin Microbiol 12:582-587, doi: 10.1016/j.mib.2009.08.008

Tyson GW, Banfield JF (2008) Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol 10:200-207, doi: 10.1111/j.1462-2920.2007.01444.x

van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJJ (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci 34:401-407, doi: 10.1016/j.tibs.2009.05.002

Vidgen M, Carson J, Higgins M, Owens L (2006) Changes to the phenotypic profile of *Vibrio harveyi* when infected with the *Vibrio harveyi* myovirus-like (VHML) bacteriophage. J Appl Microbiol 100:481-487, doi: 10.1111/j.1365-2672.2005.02829.x

Von Damm KL (1990) Seafloor hydrothermal activity: Black smoker chemistry and chimneys. Annu Rev Earth Planet Sci 18:173-204, doi: 10.1146/annurev.ea.18.050190.001133

Wächtershäuser G (1988a) Pyrite Formation, the first energy source for life: a hypothesis. Syst Appl Microbiol 10:207-210, doi: 10.1016/S0723-2020(88)80001-8

Wächtershäuser G (1988b) Before enzymes and templates: theory of surface metabolism. Microbiol Rev 52:452-484

Wächtershäuser G (1990) Evolution of the first metabolic cycles. Proc Natl Acad Sci USA 87:200-204, doi: 10.1073/pnas.87.1.200

Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272:1910-1914, doi: 10.1126/science.272.5270.1910

Webb JS, Givskov M, Kjelleberg S (2003) Bacterial biofilms: prokaryotic adventures in multicellularity. Curr Opin Microbiol 6:578-585, doi: 10.1016/j.mib.2003.10.014

Webb JS, Lau M, Kjelleberg S (2004) Bacteriophage and phenotypic variation in Pseudomonas aeruginosa biofilm development. J Bacteriol 186:8066-8073, doi: 10.1128/JB.186.23.8066-8073.2004

Weinbauer MG (2004) Ecology of prokaryotic viruses. FEMS Microbiol Rev 28:127-181, doi: 10.1016/j.femsre.2003.08.001

Whiteley M, Bangera MG, Bumgarner RE, Parsek MR, Teitzel GM, Lory S, Greenberg EP (2001) Gene expression in *Pseudomonas aeruginosa* biofilms. Nature 413:860-864, doi: 10.1038/35101627.

Williamson SJ, Cary SC, Williamson KE, Helton RR, Bench SR, Winget D, Wommack KE (2008a) Lysogenic virus–host interactions predominate at deep-sea diffuse-flow hydrothermal vents. ISME J 2:1112–1121

Williamson SJ, Rusch DB, Yooseph S, Halpern AL, Heidelberg KB, Glass JI, Andrews-Pfannkoch C, Fadrosh D, Miller CS, Sutton G, Frazier M, Venter JC (2008b) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. PloS One 3:e1456, doi: 10.1371/journal.pone.0001456

Woese C (1998) The universal ancestor. Proc Natl Acad Sci USA 95:6854-6859, doi: 10.1073/pnas.95.12.6854

Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci USA 99:8742-8747, doi: 10.1073/pnas.132266999

Ziebuhr W, Krimmer V, Rachid S, Lossner I, Gotz F, Hacker J (1999) A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. Mol Microbiol 32:345-356, doi: 10.1046/j.1365-2958.1999.01353.x