environmental microbiology

Environmental Microbiology (2015) **17**(2), 480–495 doi:10.1111/1462-2920.12528

Metagenomic analysis of the viral community in Namib Desert hypoliths

Evelien M. Adriaenssens,[1] Lonnie Van Zyl,[2] Pieter De Maayer,[1] Enrico Rubagotti,[3] Ed Rybicki,[4] Marla Tuffin[2] and Don A. Cowan[1,3]*

[1]*Centre for Microbial Ecology and Genomics, University of Pretoria, Pretoria, South Africa.*

[2]*Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, Cape Town, South Africa.*

[3]*Genomics Research Institute, University of Pretoria, Pretoria, South Africa.*

[4]*Biopharming Research Unit, Molecular and Cell Biology Department and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.*

Summary

Hypolithic microbial communities are specialized desert communities inhabiting the underside of translucent rocks. Here, we present the first study of the viral fraction of these communities isolated from the hyperarid Namib Desert. The taxonomic composition of the hypolithic viral communities was investigated and a functional assessment of the sequences determined. Phylotypic analysis showed that bacteriophages belonging to the order *Caudovirales*, in particular the family *Siphoviridae*, were most prevalent. Functional analysis and comparison with other metaviromes revealed a relatively high frequency of cell wall-degrading enzymes, ribonucleotide reductases (RNRs) and phage-associated genes. Phylogenetic analyses of *terL* and *phoH* marker genes indicated that many of the sequences were novel and distinct from known isolates, and the class distribution of the RNRs suggests that this is a novel environment. The composition of the viral hypolith fraction containing many *Bacillus*-infecting phages was not completely consistent with Namib hypolith phylotypic surveys of the bacterial hosts, in which the cyanobacterial genus *Chroococcidiopsis* was found to be dominant. This could be attributed to the lack of sequence information about hypolith viruses/bacteria in public databases or the possibility that hypolithic communities incorporate viruses from the surrounding soil.

Received 8 January, 2014; accepted 2 June, 2014. *For correspondence. E-mail don.cowan@up.ac.za; Tel. +27 12 420 5873.

Introduction

The Namib Desert, a coastal zone covering over 130 000 km^2 in South Western Africa, is a well-studied hyperarid desert with an average annual rainfall of 25 mm, and is considered to be one of the oldest deserts in the world (Eckardt *et al*., 2013; Henschel and Lancaster, 2013). At the Gobabeb Desert Research Station, situated on the northern bank of the Kuiseb river bed, 90 km inland of Walvis Bay, relative humidity can drop below 20% during the day, rains are infrequent, and water is mostly only available from fog and dew events (Henschel and Seely, 2008). The gravel plains north of Gobabeb are home to specialized microbial communities such as hypoliths, inhabiting the underside of translucent rocks, which are usually composed of quartz (Bahl *et al*., 2011; Chan *et al*., 2012). Hypoliths are present in all the hot and cold deserts that have been investigated, where they provide shelter from UV irradiation, desiccation and temperature-related stresses (Schlesinger *et al*., 2003; Warren-Rhodes *et al*., 2006; 2007; Pointing *et al*., 2007; Cowan *et al*., 2010; Makhalanyane *et al*., 2013a). With over 35% of Earth's terrestrial surface permanently or seasonably dry, soil and rock-surface communities such as hypoliths are considered to carry out essential roles in desert ecosystems and their importance is highlighted in the UN's Convention to Combat Desertification (Pointing and Belnap, 2012).

The microbial communities of hot desert hypoliths are dominated by cyanobacteria, with members of the genus *Chroococcidiopsis* reported to be the most prevalent (Warren-Rhodes *et al*., 2006; Bahl *et al*., 2011; Lacap *et al*., 2011; Makhalanyane *et al*., 2013b). Recently, a difference in the microbial composition of hypoliths and the surrounding open soil was demonstrated in the Namib Desert (Stomeo *et al*., 2013). For these hypolithic communities, it has been shown that fog effectively replaces rainfall as a source of available water (Warren-Rhodes *et al*., 2013). Microbial research in these hypolith niche habitats has mainly focused on the bacterial fraction, and

to date, no bacteriophages for this environment or infecting the predominant bacterial classes have been reported.

Bacteriophages or bacterial viruses make up the majority of the environmental virus-like particles (VLPs), and are the most abundant biological entities on earth (Breitbart and Rohwer, 2005). Estimations of soil VLP content range from 1.5×10^8 to 4.1×10^9 per gram, depending on soil composition and geographical location, with soils lower in organic matter containing less VLPs (Ashelford *et al*., 2003; Williamson *et al*., 2003; 2005; Srinivasiah *et al*., 2008). Phages are considered major players in the shaping of microbial communities, as evidenced by their abundance and diversity, facilitating horizontal gene transfer and affecting nutrient cycling in many different environments (Hambly and Suttle, 2005). Even though these traits have been mainly investigated in marine environments, the soil phage community has been shown to rapidly respond to changes in environmental conditions (Srinivasiah *et al*., 2008), and we can thus speculate that phages play similar important roles in soil. Considering that hypolithic communities represent dominant sites of productivity in desert environments and are proposed model systems for understanding microbial community assemblages, understanding the impact of the phage contribution in shaping these communities is of importance.

Metagenomic approaches have become the benchmark for research on viral community composition, circumventing the need for culturing steps and filling a considerable void in microbial ecology research (Edwards and Rohwer, 2005; Rosario and Breitbart, 2011; Mokili *et al*., 2012; Willner and Hugenholtz, 2013). The use of metagenomics to survey viral community diversity (metaviromics) also bypasses the additional drawback that there are no universal signature genes present in viruses that can be used as phylogenetic markers to assess their diversity (Rohwer and Edwards, 2002). Indeed, a typical feature of viral metagenomes is the large proportion of unknown sequences or database ORFans, which may encompass up to 90% of the sequence data (reviewed in Mokili *et al*., 2012). In comparison with alternative phage biogeography approaches such as microscopy, genome/amplicon restriction fragment-based fingerprinting or target gene sequencing of specific viral families, metaviromics is the only method that has access to the total diversity present in a habitat (Thurber, 2009). However, methodological biases are still present which prevent the full viral diversity from being sampled (Duhaime and Sullivan, 2012; Solonenko *et al*., 2013). In addition, metaviromic sequence datasets are excellent targets for bioprospecting of novel genes and gene products, and a new field of functional viral metagenomics using these methods is now emerging, investigating among others novel DNA polymerases and endolysins (Schmitz *et al*., 2010; Schoenfeld *et al*., 2010).

The principal focus of environmental (non-human/animal-associated) metaviromic research has been on marine habitats (Suttle, 2005; Breitbart *et al*., 2007; Williamson *et al*., 2008; 2012; Cottrell and Kirchman, 2012; Hurwitz and Sullivan, 2013), and to a lesser extent, extreme environments (Schoenfeld *et al*., 2008; Diemer and Stedman, 2012; Emerson *et al*., 2012; Yoshida *et al*., 2013) and soils (Fierer *et al*., 2007). The only published, hot desert-related metaviromes are those from desert soil from the Joshua Tree National Park (CA, USA) (Fierer *et al*., 2007) and from four perennial ponds in the Sahara desert (Fancello *et al*., 2013). A small volume of sequence data obtained from a soil viral fraction from the Namib Desert showed mostly *Bacillus*-associated and *Siphoviridae* phages (Prestel *et al*., 2008).

In this paper, we present the first dsDNA viral metagenome dataset from a pooled hypolith sample, representing a specific desert niche habitat. Analysis of this viral community will increase the knowledge base of (desert) soil viruses, specifically in these hotspots of microbial productivity, and will complement the microbial community data that have been previously published.

Results and discussion

Electron microscopy

Analysis of the Namib hypolith virus fraction showed mostly virus particles belonging to the order *Caudovirales*, with *Siphoviridae*-type phages most commonly observed, followed by myoviruses and podoviruses, as well as various other VLPs (Fig. 1). The undetermined structures (U) in Fig. 1 might indicate residual cellular material in the suspension.

Metavirome assembly

Contig assembly was performed to circumvent the short read lengths generated with the Illumina platform (average read length 240–250 bp) for open reading frame (ORF) prediction and annotation; however, the possibility of chimeric sequences in the assembly could not be precluded. Assembly with Seqman Ngen and Velvet yielded 4575 contigs larger than 500 bp, with an average length of 1301 bp accounting for a total of 5 950 925 bp. This assembly was uploaded to four different metagenomics pipelines for analysis (see *Experimental procedures*). The RAMMCAP workflow of the CAMERA portal predicted 11 289 ORFs, whereas MetaVir predicted 11 919 genes, VIROME 11 935 ORFs and MG-RAST identified 5830 protein-coding features (Table 1). With VIROME, predicted ORFs were further subdivided into complete

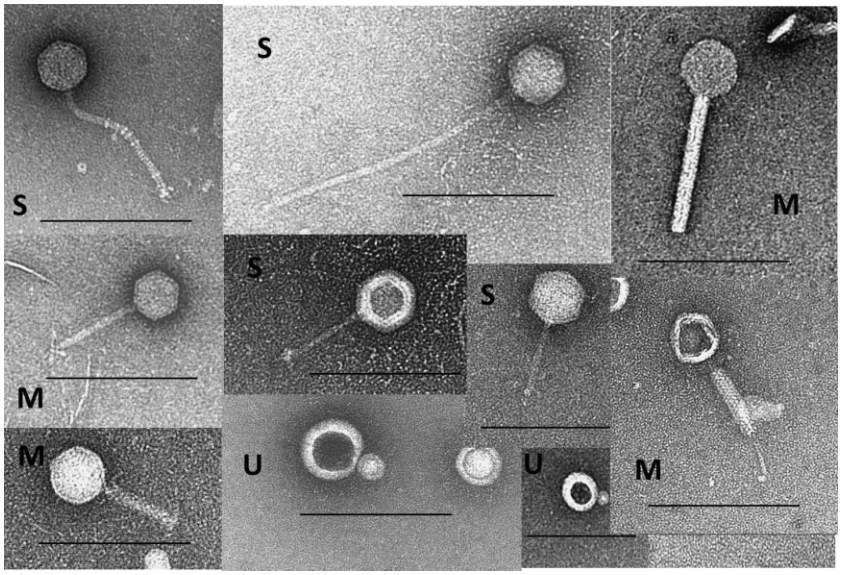

Fig. 1. Transmission electron micrographs of the viral fraction of hypolith scrapings. Scale bars represent 200 nm. Particles were negatively stained with 2% uranyl acetate. Depicted virus particles either belong to the families *Siphoviridae* (S), *Myoviridae* (M) or are of an undetermined (U) shape.

(5789), missing both ends (983), missing start (2564) and missing stop (2599). The lower number of predicted ORFs with MG-RAST was possibly caused by the inability to predict truncated ORFs or due to the difference in gene calling algorithm (see *Experimental procedures*). No rDNA features were found with either VIROME or RAMMCAP, but MG-RAST identified 19 rDNA reads in the raw read datasets, not in the assembly. This can indicate bacterial contamination, previously reported to occur with tangential flow filtration-based concentration methods (Hurwitz *et al*., 2013), or the presence of general transducing phages that package host DNA (including rDNA) randomly upon infection. The presence of small numbers of rDNA features (9–67 features) has been reported previously for metaviromes available on MG-RAST such as the terrestrial hot springs (Schoenfeld *et al*., 2008) or several viromes from the Ocean Viruses project (Angly *et al*., 2006).

Depending on the annotation pipeline used, between 2545 and 6755 ORFs with counterparts in public databases were found, making for a proportion of ORFans in this metavirome between 40.2% and 62.6%. VIROME predicted the largest number of affiliated ORFs, drawing information from Uniref 100 in combination with four annotated databases (KEGG, COG, SEED and ACLAME) and from Metagenomes On-Line (Table 1).

Viral diversity and taxonomic composition

The rarefaction curve computed by MetaVir showed approximately 270 000 sequencing clusters at 90% sequence identity for the 800 000 reads that were uploaded (Fig. 2). Since the rarefaction curve is nearing an asymptote, we could assume a substantial fraction of the viral richness was sampled. For this metavirome, an additional 100 000 sequences is predicted to lead to less than 10 000 extra clusters. At 98% similarity, the number of clusters for this metavirome increased to 340 000 (data not shown). In the current phage taxonomy (ICTV Discussions, talk.ictvonline.org; King *et al*., 2012), isolates are grouped into the same species at nucleotide identity levels of 90% to 95%. Factoring in an average phage

Table 1. Comparison of the automated pipelines used to characterize the Namib hypolith metavirome (assembled contigs).

Feature	MetaVir	VIROME	MG-RAST	CAMERA
# predicted CDS	11 919	11 935	5830	11 289
# affiliated CDS	4462	6755	2545	6741
% ORFans	62.6%	43.4%[a]	56.3%	40.3%
# tRNAs	NA	57	NA	57
# rRNAs	NA	0	0	0
Databases used for CDS annotation	RefSeq virus, pfam	UniRef100, SEED, GO, COG, KEGG, ACLAME, MGOL	GenBank, IMG, KEGG, PATRIC, RefSeq, SEED, SwissProt, TrEMBL, eggNOG, COG, NOG, KOG, subsystems	COG, pfam, TIGRfam

a. VIROME classifies the unaffiliated CDSs in true ORFans and unassigned proteins which have blast hits to the UNIREF100P database at an e value higher than 0.001. We combined these two classes for the % ORFans.
CDS: coding sequence.

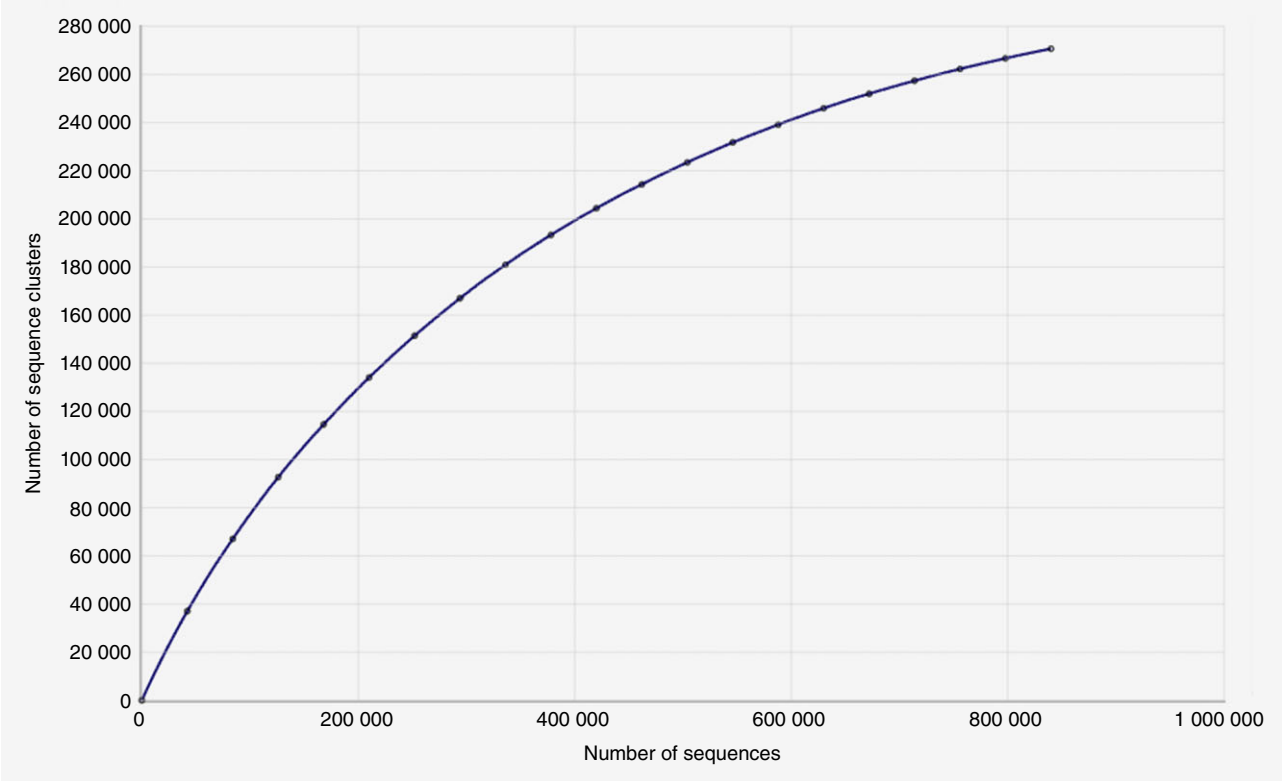

Fig. 2. Rarefaction curve of the Namib hypolith metavirome generated by MetaVir. Clustering was set at 90% similarity.

genome length of 50 000 bp [as used previously for marine genomes (Breitbart *et al*., 2002)] and a read length of 250 bp, we could estimate the lower limit of the viral richness (number of different virus species) in this metavirome to be between 1350 and 1700 [# clusters/(genome length/read length)].

MetaVir was used for analysis of the viral taxonomic composition of the raw reads. This pipeline uses the GAAS tool which normalizes the composition plot against the genome lengths (Angly *et al*., 2009) with BlastX matches (e-value cut-off = 10^{-5}) generated by comparison against the RefSeq complete viral genomes database. With these parameters, 23.77% of the reads produced a significant hit. In assessing the composition of the assembled contigs, 60.99% of the contigs showed similarity to known sequences and so did 37.46% of the predicted genes. MG-RAST and VIROME also computed taxonomic composition plots, but these were heavily biased towards bacterial taxa, as prophage and temperate phage sequences are often classified as bacterial in origin (data not shown). Despite a higher fraction of bacterial signatures in MG-RAST, 43.9% of the raw reads encoded only ORFans (data not shown).

Using the GAAS taxonomic composition plot (Fig. 3), 80% of the reads that produced a significant database hit were recognized as belonging to dsDNA viruses with no RNA stage, 13% as unclassified phages and 7% as ssDNA viruses. The majority of taxonomic hits (48%) were to the *Siphoviridae* family, based on the taxonomy of the viral genomes deposited in the NCBI database. However, the unclassified *Geobacillus* virus E2, making up 6% of the virus fraction, was described as a siphovirus in its original publication (Wang and Zhang, 2008), giving a total of 54% hits to the family *Siphoviridae*. The *Podoviridae* family accounted for 10% of the viral fraction and the *Myoviridae* for 9%. Furthermore, 7% of the sequences shared sequence identity with sequences in the 'unclassified *Caudovirales*' group and 4% in the 'unclassified dsDNA viruses, no RNA stage' grouping. In our case, for the 7% ssDNA viruses, 99% of the reads mapped to a single microvirus, *Enterobacteria* phage phiX174, which can be considered residual contamination of the phiX v3 phage used as a qualitative marker for sequencing (see *Experimental procedures*) and should thus be disregarded in the taxonomic composition. In raw reads number, this microvirus makes up only 0.0014% of the virus fraction, but the GAAS tool compensates for its small genome size. Virus families detected at below 1% abundance (and above 0.05%) include *Tectiviridae*, *Ascoviridae*, *Phycodnaviridae* and *Anelloviridae*.

The Namib hypolith viral composition was compared with selected metaviromes (Supporting Information

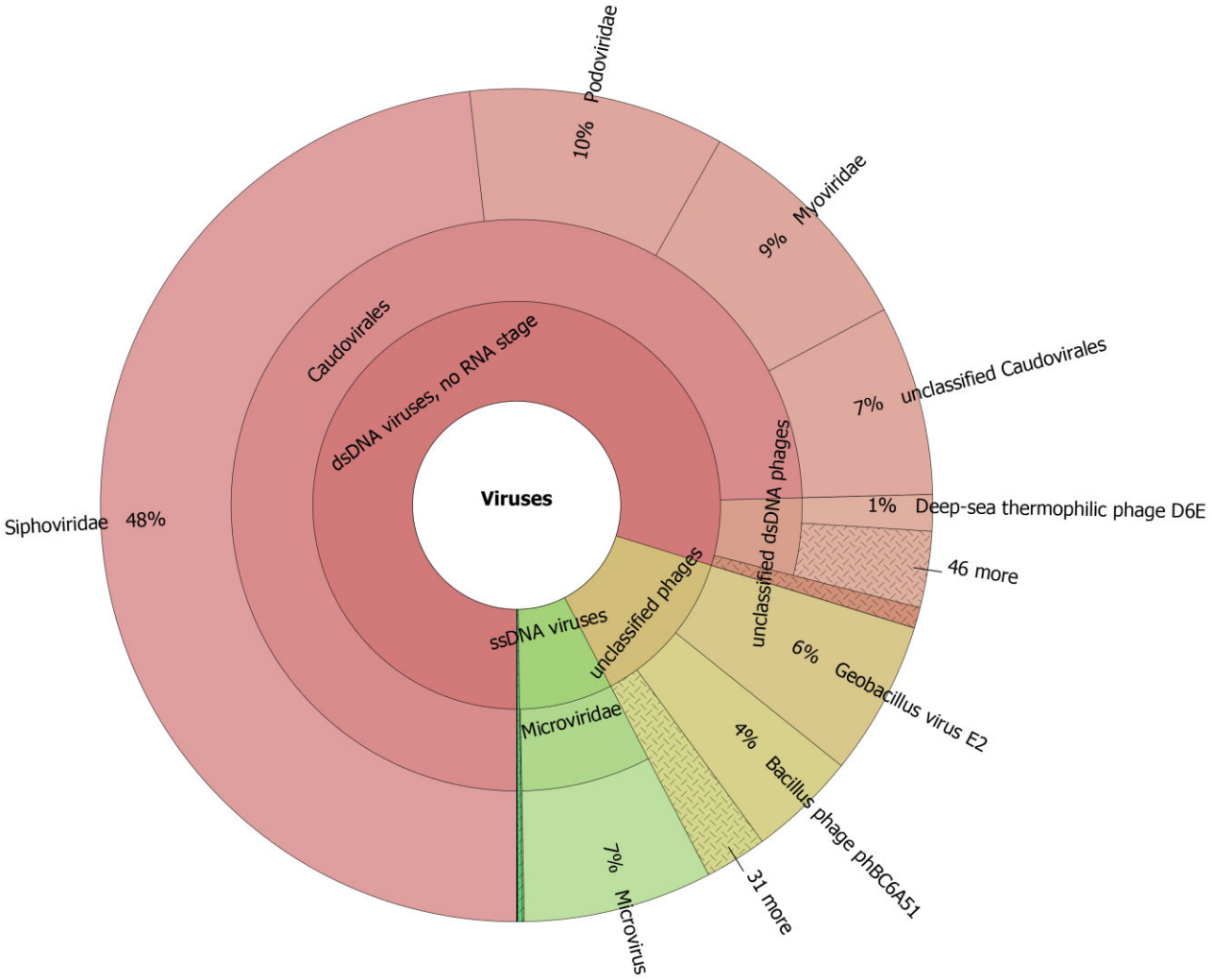

Fig. 3. Taxonomic composition of the Namib hypolith metavirome. Composition type was GAAS (genome length normalization) with a threshold of 10^{-5} on e value, as generated by MetaVir. The most abundant virus taxa are indicated in red, second yellow, then green and blue (at below 1%).

Table S1) spanning different biomes, including freshwater [French lakes (Roux *et al*., 2012), Antarctic Lake (López-Bueno *et al*., 2009), Saharan Ponds (Fancello *et al*., 2013)], hypersaline [solar salterns (Rodriguez-Brito *et al*., 2010)], hyperthermophilic [Yellowstone hot springs (Schoenfeld *et al*., 2008)] and marine [Pacific Ocean Virome (Hurwitz and Sullivan, 2013), Indian Ocean (Williamson *et al*., 2012)]. The prevalence of *Caudovirales* as the dominant taxon was observed in most metaviromes except for three of the freshwater datasets (Lake Pavin, Lake Bourget and Antarctic Lake) and two extreme environments (saltern high and Octopus hot spring). The former frequencies ranged from 31% to 89% of the known virus fraction, whereas the freshwater lakes were dominated by ssDNA viruses and the saltern and hot spring by unclassified dsDNA viruses. The higher presence of ssDNA viruses might have been caused by the use of phi29 multiple displacement amplification (MDA) of the metagenomic DNA during library construction, which has been reported to be biased towards ssDNA (Kim *et al*., 2008). However, the Saharan ponds and solar saltern samples also used MDA and were dominated by dsDNA viruses. Comparing the families within the *Caudovirales*, podovirus signatures represented the major fraction in marine environments, while siphovirus reads were predominant in non-marine environments, including our hypolith dataset. As of yet, no explanation for this division could be offered. The rare virus families present in the hypolith dataset were also found at similar frequencies in most of the other metaviromes.

Cyanophages were expected to be abundant based on the known dominance of Cyanobacteria in hypolith bacterial communities (Warren-Rhodes *et al*., 2006; Makhalanyane *et al*., 2013b). To test their presence, the

contig annotations (MetaVir) were investigated for cyanophage resemblance (Supporting Information Table S2). All but two of the cyanophages in the NCBI virus genome database, the majority of which consists of marine cyanophages, could be mapped to the hypolith metavirome, yet at significance levels lower than the best Blast hit. From these findings, we can hypothesize that either cyanophages make up a small fraction of the hypolith metavirome, or that they are significantly distinct from their marine counterparts to the extent that, with current similarity searches, they are not recognized as cyanophages. In this respect, it is important to note that cyanobacteria found in terrestrial environments differ significantly from those used as model organisms in marine studies such as *Synechococcus* and *Prochlorococcus* sp. (i.e. the bacteria belong to different orders) which will have an effect on the phage population (Schlesinger *et al*., 2003; Wong *et al*., 2010).

An alternative approach to investigating the diversity and taxonomy of phages is to assign the phylogeny of certain signature genes predicted in the metavirome. The marker gene with the most hits in the hypolith virome (as calculated by MetaVir) was the terminase large subunit *terL*, present in phages of the order *Caudovirales*. The phylogeny of this gene can give an indication of the type of DNA packaging mechanism utilized by certain phage groups (Sullivan *et al*., 2009). A PhyML phylogenetic tree was generated from the contig sequences (Supporting Information Fig. S1) and showed that most of the hypolith metavirome *terL* sequences clustered separately from those of cultured tailed phages. Some sequences did cluster with known phages; these were either siphoviruses or myoviruses, which are known to employ headful packaging mechanisms with *pac* sites. None of the hypolith sequences clustered with *cos* site phages or T4-like phages using random headful packaging.

Functional analysis

The putative functions of the annotated ORFs from the assembled contig dataset were predicted using VIROME, MG-RAST and CAMERA. The database searches resulting in the most functional hits were those against SEED (subsystems approach of MG-RAST) with 3804 hits, pfam (CAMERA) with 2222 hits and GO (VIROME) with 3389 hits. Almost half (49%) of the hits in the subsystems functional annotation belonged to the subsystem 'phages, prophages, transposable elements, plasmids' (Fig. 4) with phage structural, integration/excision and DNA metabolism-related proteins most commonly identified. The other SEED functional categories showed 'nucleotides and nucleosides', 'regulation and cell signalling' and 'DNA metabolism' as the dominant annotations. In these categories, many proteins could be phage-related (of possible cellular origin), such as DNA polymerases, helicases, ribonucleotide reductases (RNRs) and peptidoglycan-degrading enzymes. These hits were also found in the pfam databases, with nucleic acid binding and DNA replication families being the most common protein families identified, followed by peptidoglycan-degrading or hydrolase enzymes. Comparisons against the GO database identified the largest number of hits for proteins with hydrolase, transferase and nucleic acid binding activities. In this database, a large number of hits relating to cellular, nitrogen and macromolecular metabolic processes were also found.

The SEED functional categories were used to compare the hypolith metavirome functional composition with those of six publically available viromes [Octopus hot spring (Schoenfeld *et al*., 2008), Joshua Tree desert (Fierer *et al*., 2007), Matapeake Bay (K.E. Wommack, unpublished), GWAD Pacific rise (K.E. Wommack, unpublished), LJ26S POV (Hurwitz and Sullivan, 2013) and M6O1K Indian Ocean (Williamson *et al*., 2012)] using the comparison function of VIROME. Eight functional categories were identified that were present at a significantly higher frequency in the Namib hypolith annotation than in any of the other metaviromes (Table 2) and no categories were present at significantly lower frequency. Three of these categories, 'murein hydrolases', 'recycling of peptidoglycan amino acids' and 'clustering-based subsystems Cbss-393121.3.peg.2760' are involved in the degradation of the bacterial cell wall, with the most commonly identified protein a N-acetyl-muramoyl-L-alanine-amidase. This type of amidase has been included recently in several patents describing lytic enzymes used as antibacterials [Fischetti *et al*., 2013 (US8580553 B2); Grallert and Forchheim, 2013 (US8492519 B2); Loessner *et al*., 2012 (US20120171188 A1)], making this hypolithic environment an ideal target for antibacterial enzyme bioprospecting. Three categories of phage-related proteins were also more frequent in the Namib, 'phage tail 2', 'r1t-Like streptococcal phages' and 'Listeria phi-A118-like prophages', all encoding several structural proteins and the latter two several phage enzymes as well. This might indicate a higher frequency of the r1t and A118-like phage types in the hypolith metavirome without making assumptions about the host. RNRs, involved in dNTP synthesis, were also more frequent and are discussed further in this section. The remaining two categories contain cell division-related and phosphate metabolism-associated proteins from prokaryotes, plasmids or phages with no clearly defined function.

Only 3% of the functionally annotated genes were classified in the subsystem 'virulence, disease and defense' by MG-RAST (Fig. 4). The corresponding coding sequences (CDSs) were further investigated and identified as either hypothetical proteins from known

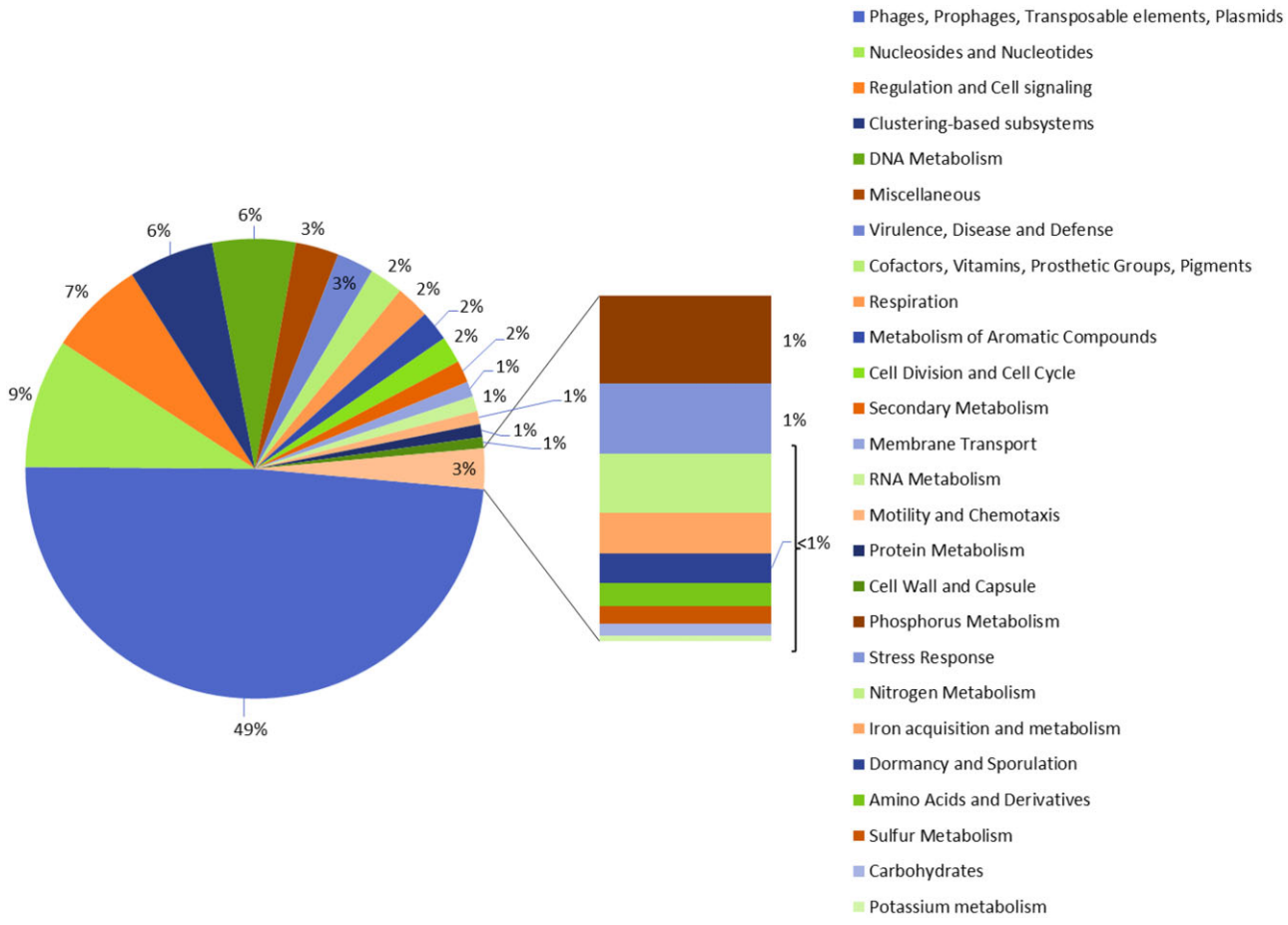

Fig. 4. Composition of predicted functional genes of the Namib hypolith contigs. The CDSs were compared with the SEED database using subsystems in MG-RAST. The predicted subsystems in the legend are listed according to their relative occurrence (from high to low) and in the pie chart starting at the bottom in a clockwise direction. The expanded region shows the functional classes represented at low abundance, 1% and below (in legend from phosphorus metabolism to potassium metabolism).

pathogenic bacteria or phage-related proteins such as integrases and replication proteins (data not shown). All contigs were blasted against an online virulence, toxin and resistance gene database, MvirDB (Zhou *et al*., 2007), but no relevant CDSs were identified in the hypolith metavirome, leading us to assume that the CDSs in this functional category are more involved in phage virulence than human-pathogenic virulence.

The presence of 'auxiliary metabolic genes' (AMGs) in phages, which are presumed to assist in rate-limiting or key steps in host metabolism, has been described previously (Breitbart *et al*., 2007). In marine cyanophages, where these AMGs have been mainly studied, they can be involved in photosynthesis (*psbA* and *psbB*), carbon turnover (*talC*), phosphate uptake (*pstS*), nucleotide metabolism (*nrd* genes) or of unknown function (*phoH*) (Sullivan *et al*., 2006; Williamson *et al*., 2008; Goldsmith *et al*., 2011; Thompson *et al*., 2011; Dwivedi *et al*., 2013). The hypolith metavirome was investigated for the presence of these AMGs and both *phoH* and several different classes of *nrd* genes were found.

Eighteen complete *phoH* genes and 23 partial sequences were identified in the metavirome by MetaVir. The topology of the PhyML phylogenetic tree generated (Fig. 5) clearly shows that the majority of the hypolith PhoH amino acid sequences (in red) cluster separately from those of complete phage genomes (deposited in the NCBI database), supported by high bootstrap values. The three long branches in the tree were caused by the inclusion of partial genes sequences of the hypolith metavirome that do not span the length of the alignment. The marine cyanophages (in blue), for which *phoH* is used as a marker gene (Goldsmith *et al*., 2011), form a distinct clade, unrelated to the hypolith sequences. This supports the hypothesis that putative hypolith cyanophages are unrelated to marine cyanophages, and at the same time illustrates the lack of sequence information on non-marine cyanophages and their hosts that

Table 2. SEED categories and associated genes from the Namib hypolith metavirome present at significantly higher frequencies than in selected metaviromes as annotated by VIROME[a].

SEED functional category	Predicted genes in Namib hypolith metavirome	Putative function
Murein hydrolases	N-acetyl-muramoyl-L-alanine-amidase, tail fiber, carboxypeptidase	Bacterial cell wall degradation
Recycling of peptidoglycan amino acids	N-acetyl-muramoyl-L-alanine-amidase, tail fiber	Bacterial cell wall degradation
Clustering-Based Subsystems Cbss-393121.3.peg.2760	Phage N-acetyl-muramoyl-L-alanine-amidase, (metalloendo)peptidase, sporulation protein	Bacterial cell wall degradation
Bacterial rna-metabolizing zn-dependent hydrolases	cell division protein, FtsK/SpoEIII protein	DNA transfer
Phage tail proteins 2	tail tape measure protein, major tail protein	Phage particle assembly
R1t-Like streptococcal phages	structural proteins, homing endonucleases, phage-associated proteins, endonuclease, endopeptidase, hypothetical protein	Phage particle assembly
Listeria phi-a118-like prophages	putative tail or baseplate protein, minor/major capsid protein, terS, replication protein, lysin, tmp, hypothetical protein, ssDNA binding protein, repressor, antirepressor	Phage particle assembly
Phosphate metabolism	hypothetical protein, phoH	Unknown function
Ribonucleotide reduction	Ribonucleotide reductases	dNTP synthesis

a. The Namib hypolith metavirome was compared with the following public metaviromes: Octopus hot spring (Schoenfeld *et al*., 2008), Joshua Tree Desert (Fierer *et al*., 2007), Matapeake (K.E. Wommack, unpublished), GWAD (K.E. Wommack, unpublished), LJ26S POV (Hurwitz *et al*., 2013), M6O1K Indian Ocean (Williamson *et al*., 2012).

inhabit this environment. Bacterial PhoH sequences from the NCBI database (in green) were also included in the analysis, showing that two hypolith contigs clustered with *Xenococcus* sp. PC7305. This cyanobacterial isolate has been shown to be closely related to *Chroococcidiopsis* sp. PC6712 (Shih *et al*., 2013), a member of the most prevalent bacterial genus in Namibian desert hypoliths (Makhalanyane *et al*., 2013b).

Given that no photosynthesis- or nutrient-stress-related host-derived genes (*psb pst, tal* genes) were identified in this metavirome, we could assume that these genes provide no additional benefit for the viral community in this environment or that their absence is merely a reflection of the host gene content. The presence of *phoH* at frequencies higher than in other metaviromes could suggest that this gene is more universally valuable in this environment than the other AMGs. The role of *phoH* in phosphate uptake or starvation remains unclear with expression being either upregulated or downregulated in different bacteria (Sullivan *et al*., 2010; Goldsmith *et al*., 2011). Recently, it was found that transcript levels in phages do not rise under phosphate limiting conditions, suggesting an alternate role for this gene (Zeng and Chisholm, 2012).

A total of 123 putative RNRs were identified in the hypolith metavirome. The RNR genes are divided in three classes based on oxygen dependency, with class I genes oxygen dependent, class II genes oxygen independent and class III genes oxygen intolerant, which means that the class distribution can give an indication of the host environment (Nordlund and Reichard, 2006; Dwivedi *et al*., 2013). Most of the RNR genes (60 predicted ORFs) belonged to class II RNRs which is represented by *nrdJ*, followed by class Ia RNRs in which *nrdA* and *nrdB* had an almost equal number of ORF hits (21 and 19 respectively) (Nordlund and Reichard, 2006). Class Ib (*nrdH* = 3, *nrdI* = 2, *nrdE* = 11, *nrdF* = 6) and III RNR genes (*nrdD* = 2) were also found at lower frequency, with the exception of the class III *nrdG* gene. This metavirome RNR composition, with all classes present and class II most abundant, has not yet been described in any other metaviromes analysed (Dwivedi *et al*., 2013), reflecting the novelty of the hypolith biome. The presence of both classes I and II at much higher frequency than class III suggests that the host community is aerobic, or facultatively anaerobic. Other metaviromes with class II RNRs at the highest frequency include a freshwater microbialite, a saltern at high salinity and a coral metavirome, but in these three metaviromes, at least one of the other classes of RNRs is absent (Dwivedi *et al*., 2013). The soil metaviromes (Fierer *et al*., 2007) on the other hand, contained all classes but at similar frequencies. It was also observed that RNR classes of podoviruses are host-dependent (Dwivedi *et al*., 2013; Holmfeldt *et al*., 2013), but soil phage isolation studies would be required to draw conclusions about the hypolith metavirome hosts. In a comparison of RNR distribution in completely sequenced phage genomes sorted by environmental source, host oxygen requirement and phage family (Dwivedi *et al*., 2013), class II RNRs were most common in *Siphoviridae* isolated from the soil environment. This is consistent with our finding that siphoviruses are most common in the Namib hypolith community, and provides additional evidence that the host community is aerobic (Dwivedi *et al*., 2013). Furthermore, the hypothesis that *nrdJ* offers a competitive advantage to phages in nutrient-limited

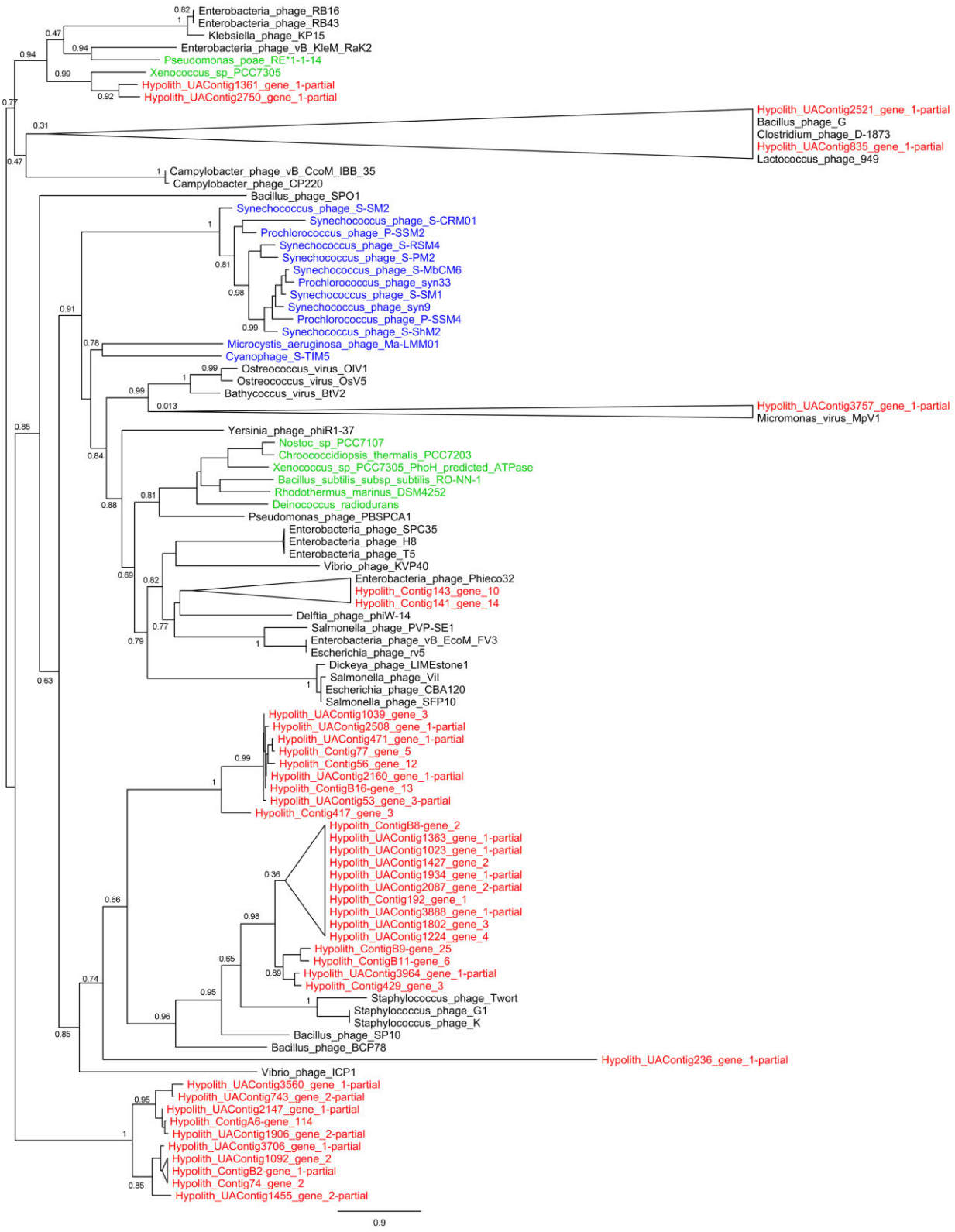

Fig. 5. PhoH phylogenetic tree. PhyML phylogenetic tree (maximum likelihood) of PhoH amino acid sequences of the Namib hypolith metavirome (red), selected phages with marine cyanophages in blue and selected bacteria (green). Sequences were aligned with MUSCLE on the phylogeny.fr server and visualized with FigTree. Nodes with poor bootstrap support (on 100 bootstraps) were collapsed. Scale bar indicates the number of substitutions per site.

environments (Dwivedi *et al*., 2013) is supported by the RNR composition of the hypolith metavirome.

Contig analysis

Seven contigs larger than 40 kb, as well as 10 contigs between 20 and 40 kb in size, were generated during assembly, with the largest contig being 108 kb. Only one contig represented a putative complete phage genome that was either circularly permuted or had terminal repeats longer than the average read length. The unique sequence of this contig was 48 632 bp long, encoding 75 ORFs as predicted by MetaVir with MetaGeneAnnotator (Fig. 6). Of the 75 coding sequences (CDSs), 12 showed similarity to siphoviruses and 7 to myoviruses, indicating that this phage almost certainly belongs to the order *Caudovirales*. The genome size is consistent with it being a member of the order *Caudovirales*, most likely part of the *Siphoviridae* family (King *et al*., 2012). MetaVir classified this contig, using lowest common ancestor affiliation, as being related to *Bacillus* phage Cherry. However, no definitive classification can be made from these data, taking into account the low number of functional gene predictions and the lack of signature genes for specific phage families. For example, in the phage genome in Fig. 6, the terminase large subunit places the phage in the order *Caudovirales* but no other signature genes are annotated which would narrow down its classification into a specific family or even genus. This terL sequence was also inserted in the phylogenetic tree and showed a distant relation to a clade of siphoviruses infecting Firmicutes (Supporting Information Fig. S1, Hypolith_contig1_gene34).

The ORF prediction capacities of several virus-optimized gene prediction programs were compared using Contig 1. MetaGeneAnnotator (included in the MetaVir pipeline) predicted 75 ORFs, while FgenesV0 (www.softberry.com) and MetaGene (in the CAMERA portal) identified 62 and 72 ORFs respectively. Manual verification of the predicted genes for the presence of a ribosome binding site, small ORF lengths and minimal intergene region lengths showed that MetaVir produced the most accurate output. FgenesV0 failed to predict many of the small hypothetical proteins that are common in viral genomes and both FgenesV0 and MetaGene tended to predict genes within ORFs on the opposing strand without valid ribosome binding sites (data not shown).

Taxonomic delineation of the other large contigs was not definitive. For example, contig32 (74 003 bp) was identified by MetaVir as an unclassified dsDNA phage, related to the deep-sea thermophilic phage D6E. In its annotation, however, several myovirus-related genes could be found, such as baseplate proteins, a tail tube and a tail sheath protein (data available from the MetaVir server). The annotation pipelines offer a taxonomic classification, but after examination of the predicted genes, it was clear that classification beyond the level of family, or even order, for this environment is difficult.

Implications of the virus complement on the hypolith community structure

Previous studies of the hot desert hypolithic microbial community composition have been performed using qPCR, terminal Restriction Fragment Length Polymorphism, amplicon sequencing and clone library sequencing with 16S rRNA and Internal Transcribed Spacer region primer sets (Warren-Rhodes *et al*., 2006; Wong *et al*., 2010; Makhalanyane *et al*., 2013b; Stomeo *et al*., 2013). While these techniques give a good representation of the microbial community composition and diversity, they offer no information on the viral composition. The composition of the metavirome sequenced here is, at first sight, not fully consistent with the results of 16S rRNA gene sequence analyses performed on samples from the same habitat; the metavirome is dominated by presumed *Bacillus* and *Geobacillus*-infecting phages, while the most dominant genus of bacteria identified in Namib hypolith communities is *Chroococcidiopsis* from the order Pleurocapsales of the Phylum Cyanobacteria (Warren-Rhodes *et al*., 2007; Wong *et al*., 2010; Makhalanyane *et al*., 2013b).

There are several hypotheses that can explain these results. First, there are no known cultured phages which infect *Chroococcidiopsis* sp. or other bacterial species dominant in hypoliths, making this a database issue,

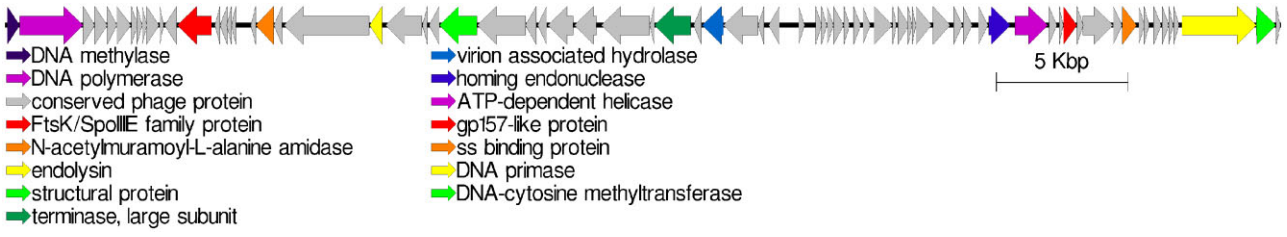

Fig. 6. Predicted full phage genome (contig 1). Arrows indicated the protein coding sequences (CDSs) as predicted by MetaVir. Colored arrows had putative functions assigned by BlastP similarity. The figure was generated using Easyfig (Sullivan *et al*., 2011).

keeping in mind that over 40% of the predicted CDSs are ORFans. Under this hypothesis, the expected (cyano)phages show too little sequence similarity to cultured phages to be identified. Related to this, it is possible that the contigs currently linked to *Bacillus* phages infect different hosts, as they are assigned based on only a few conserved CDSs. A final hypothesis could be that the hypolithic microbial community is seeded from the surrounding open soil and has incorporated some of the phages present there. This concept is supported by previous phylogenetic surveys in Namibia (Makhalanyane *et al*., 2013b), which showed that 80% of bacterial operational taxonomic units found in the hypoliths were also found in the surrounding soil.

Conclusion and perspectives

Analysis of the metavirome of Namib Desert hypoliths has revealed that this is a novel, soil-related virome. The majority of the sequence reads were classified as unknown, with only 24% having known virus counterparts among which members of the order *Caudovirales* were predominant. This is consistent with the phages previously described from the Namib Desert (Prestel *et al*., 2008) as well as other metavirome compositions from a range of different biomes. However, using a quantitative transmission electron microscopy approach on marine samples, it was recently discovered that non-tailed viruses dominate that environment, contradicting the metaviromic data (Brum *et al*., 2013). This calls to question which virus types are present in the unknown fraction of the metavirome and whether they are the dominant group. When comparing the functional composition of the hypolith metavirome, this habitat seemed to be richer in cell wall-degrading enzymes, offering new perspectives for bioprospecting of antibacterials.

The prevalence of presumptive Firmicutes-infecting phages as opposed to the expected cyanophages is thought to be caused by a database bias towards the former group of phages or possibly by phage incorporation from the surrounding desert soil. For evidence-based linking of viruses to their hosts, novel techniques such as viral tagging and phageFISH have been developed recently (Deng *et al*., 2012; Allers *et al*., 2013) and incorporation of these techniques in future studies will greatly reduce the database/bioinformatics biases currently present in environmental studies.

The unique patterns observed in this Namib hypolith metavirome (taxonomic composition, functional diversity and RNR distribution) will aid the interpretation of individual patterns from future soil metaviromic studies and the incorporation of metaviromes from extreme environments into current datasets will enhance future environmental analyses. Further research could focus on sequencing the full hypolith and surrounding soil metagenomes, as well as by culturing of hypolith-associated bacteria and their bacteriophages, whereas the metabolically active fraction of hypoliths could be determined using transcriptomics studies.

Experimental procedures

Sample collection and processing

Quartz rocks with established hypolith communities were collected by hand (gloved) in the Namib Desert near the Gobabeb Desert Research and Training Station (23°33′40″S, 15°02′29″E). The ecology of hypoliths in this area has been previously described (Warren-Rhodes *et al*., 2013). Roughly 80 rocks were scraped at depths ranging from ± 10 cm to ± 20 cm over 100 m^2. The soil was dry in appearance and in accordance with other hypolith studies, all material cemented to the rock when extracted was considered as part of the hypolith community (Khan *et al*., 2011). The microbial communities were recovered on site with gloved hands, scraped off with sterile spatulas and collected in sterile whirlpack bags (Nasco).

Approximately 0.5 kg of hypolith-associated material was suspended in 3 l of de-ionized water, homogenized by shaking and then allowed to settle. The suspension was made by first grinding the material to a fine powder using a mortar and pestle, then vigorously shaken manually for 10 min. It was left to stand for 30 min so that phage particles could diffuse freely into the suspension and this process was repeated 10 times. The supernatant was decanted and the remaining solids suspended in another 3 l of de-ionized water and the settling and decanting processes repeated. The aqueous fraction was centrifuged at low speed to remove the largest particles and the supernatant was passed through a 0.22 μm filter (Millipore, Streicup 500 ml, Cat. no. SCGPU05RE). The initial low speed spin (1590 × *g* for 10 min) to get rid of large debris was performed using a Beckman JA10 rotor using autoclaved 500 ml Nalgene PPCO bottles (cat. no. 3120-9500) and the supernatant was decanted into a larger autoclaved vessel. To collect phage particles, centrifugation of this supernatant was performed in a Beckman JA20 rotor at 43667 × *g* for 6 h in autoclaved 30 ml Nalgene PPCO (cat. no. 3119-0030) tubes. The individual pellets were resuspended in 3 ml successively. In other words, the first pellet was resuspended in 3 ml and the liquid was then transferred to the next tube, the pellet resuspended properly, then transferred to the next tube and so on until all pellets were resuspended. This phage suspension was treated with DNase I (EN0521) and RNase A (EN0531) (Fermentas – final concentration of 0.1 μg ml^{-1}) at 37°C for 1 h (DNase I). The presence of free or background contaminating bacterial DNA was checked by polymerase chain reaction amplification of the 16S RNA gene [primers E9F: 5′-GAGTTTGATCCTGGCTCAG-3′; U1510R: 5′-GGTTACC TTGTTACGACTT-3′ (Reysenbach and Pace, 1995; Hansen *et al*., 1998)]. The phage particles were treated with Proteinase K (Fermentas – final concentration 1 μg ml^{-1}) at 55°C for 2 h. Seventy microlitres of 20% SDS was then added and the sample was incubated at 37°C for 1 h. The DNA was extracted with three rounds of phenol : chloroform :

isoamylalcohol (25:24:1) phase separation followed by two replicates of chloroform : isoamylalcohol (24:1) phase separation (15 ml Sterillin tube, Eppendorf 5810R centrifuge, 5000 r.p.m. for 10 min). Precipitation was performed with 1/10 volume of 3 M NaOAc (pH 5.2) and 2× volume 95% ethanol, with overnight incubation at 4°C. Precipitated DNA was recovered by centrifugation at 13 000 r.p.m. for 10 min and the resulting pellet was resuspended in 30 µl of TE (Tris/EDTA) buffer. The DNA was further purified using the Qiagen Gel Extraction kit (Qiaex II, cat. no. 20021).

Electron microscopy

Phage suspensions were prepared according to the ammonium acetate method as described by Ackermann (2009). Three microlitre of each sample taken from the 30 µl concentrate was pipetted onto carbon coated 200 mesh copper grids and stained with 2% aqueous uranyl acetate for 30 s. The samples were viewed using a LEO 912 Omega TEM (Zeiss, Oberkochen, Germany) at 120 kV. Images were collected using a ProScan CCD camera.

Sequencing

Library preparation of the hypolith viral DNA was performed with the Nextera XT kit (Illumina) with a 10% phiX v3 spike as per the manufacturer's instructions (Preparation Guide, Part #15031942 Rev A May 2012) and the MiSeq Reagent kit V2 (500 cycle). One ng of uncloned, unamplified viral DNA was used to prepare one NexteraXT library. This was repeated four times and the resultant libraries were sequenced using the Illumina MiSeq at the University of the Western Cape, Cape Town, South Africa, generating 2×250 bp reads. The raw reads were trimmed (bases with a Q-score less than 36 were trimmed from the 3′end) and demultiplexed at the sequencing facility, resulting in eight (4×2) paired fastq files (accessible at the Sequence Read Archive Bioproject PRJNA229525, accession number SRX385198).

In silico *analyses*

The processed reads were loaded into Seqman Ngen® (DNASTAR, Madison, WI, USA) with the following parameters: kmer = 21, no read trimming and a minimum of 100 reads per contig; other parameters were at the default settings. The unassembled sequences were saved and assembled with Velvet (Zerbino and Birney, 2008) (kmer = 15, coverage cut-off = 3). The Ngen and Velvet assemblies were merged and autoblasted with BioEdit (Hall, 1999) to manually extend contigs.

The contigs from the above assembly were uploaded to four automated annotation pipelines available online (from which the analyses presented in this paper can be accessed), two specifically designed for viral metagenomes [MetaVir (Roux *et al*., 2011) and VIROME (Wommack *et al*., 2012)] and two general metagenomic webservers [MG-RAST (Meyer *et al*., 2008) and the RAMMCAP workflow of CAMERA (Li, 2009; Sun *et al*., 2011)]. Raw reads were also uploaded to MetaVir and MG-RAST. For the former, the eight fastq files were converted to fasta format using the Fastq2fasta program at bio.chpc.ac.za/ER. These were then merged into one file containing 946 094 reads. With this file, a reference assembly against the microvirus phiX v3, used in the Illumina Miseq quality control, was performed at 99% identity, which removed 104 636 reads. The remaining 841 458 reads were uploaded to the server. Taxonomic composition was assessed with MetaVir (on the reads), which uses the GAAS tool (Angly *et al*., 2009), with MG-RAST which combines annotation from all database sources and with VIROME using the top Uniref 100 BLAST hits. For ORF prediction of the contigs, MetaVir and VIROME use MetaGeneAnnotator (Noguchi *et al*., 2008), MG-RAST uses FragGeneScan (Rho *et al*., 2010), and the ORF prediction algorithm chosen for RAMMCAP was MetaGene (Noguchi *et al*., 2006). The predicted genes were scanned against the following databases for functional annotation: RefSeqVirus (MetaVir), ACLAME (VIROME), pfam (MetaVir, RAMMCAP), TIGRfam (RAMMCAP), GO (MG-RAST, VIROME), SEED (MG-RAST, VIROME), NCBI nr (MG-RAST), COG (RAMMCAP, VIROME), KEGG (MG-RAST, VIROME, RAMMCAP), UniProt (MG-RAST), Uniref100 (VIROME), eggNOG (MG-RAST) and MGOL (VIROME).

A comparison of the SEED functional categories was carried out on the VIROME webserver using the comparison tab. Significant differences in frequencies between the Namib hypolith metavirome and the selected publically available metaviromes were calculated by uploading the VIROME output pairwise onto the Metastats webserver (http://metastats.cbcb.umd.edu/detection.html) at $P < 0.05$ and 100 permutations (White *et al*., 2009).

The presence of AMGs was confirmed by scanning of the MetaVir contig annotation table output for specific metabolic genes, namely *psbA*, *psbB*, *phoH*, *tal* and *nrd*. The RNRs (*nrd* genes) found in the metavirome were compared by BLAST analysis against the RNRdb, a curated database of RNRs, to determine the class (Lundin *et al*., 2009).

Phylogenetic analyses were performed using the phylogenetic tree computation tool on the MetaVir server, described in detail in Roux and colleagues (2011). Briefly, selected amino acid sequences were aligned with MUSCLE (Edgar, 2004) and trees with 100 bootstraps were generated with PhyML (Guindon *et al*., 2009). The output was visualized with FigTree (Rambaud, 2007). For the PhoH tree, amino acid sequences were downloaded from NCBI, alignment was performed using MUSCLE (Edgar, 2004) and tree rendering with PhyML 3.0 (Guindon *et al*., 2010) on the phylogeny.fr server (Dereeper *et al*., 2008) without curation.

Acknowledgements

The authors gratefully acknowledge financial contributions from the National Research Foundation of South Africa and the University of Pretoria Institutional Research Theme in Genomics. EMA is funded by the Vice-Chancellor's Postdoc Fellowship of the University of Pretoria. We would also like to thank the reviewers for their helpful suggestions and hypotheses. The authors declare that no competing interests exist.

References

Ackermann, H.-W. (2009) Basic phage electron microscopy. In *Bacteriophages: Methods and Protocols*. Clokie, M.R.J.,

and Kropinski, A.M. (eds). New York, NY, USA: Humana Press, pp. 113–126.

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., *et al*. (2013) Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15:** 2306–2318.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., *et al*. (2006) The marine viromes of four oceanic regions. *PLoS Biol* **4:** e368.

Angly, F.E., Willner, D., Prieto-Davó, A., Edwards, R.A., Schmieder, R., Vega-Thurber, R., *et al*. (2009) The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput Biol* **5:** e1000593.

Ashelford, K.E., Day, M.J., and Fry, J.C. (2003) Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol* **69:** 285–289.

Bahl, J., Lau, M.C.Y., Smith, G.J.D., Vijaykrishna, D., Cary, S.C., Lacap, D.C., *et al*. (2011) Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nat Commun* **2:** 163.

Breitbart, M., and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* **13:** 278–284.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., *et al*. (2002) Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* **99:** 14250–14255.

Breitbart, M., Thompson, L.R., Suttle, C.A., and Sullivan, M.B. (2007) Exploring the vast diversity of marine viruses. *Oceanography* **20:** 135–139.

Brum, J.R., Schenck, R.O., and Sullivan, M.B. (2013) Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME J* **7:** 1738–1751.

Chan, Y., Lacap, D.C., Lau, M.C.Y., Ha, K.Y., Warren-Rhodes, K.A., Cockell, C.S., *et al*. (2012) Hypolithic microbial communities: between a rock and a hard place. *Environ Microbiol* **14:** 2272–2282.

Cottrell, M., and Kirchman, D. (2012) Virus genes in Arctic marine bacteria identified by metagenomic analysis. *Aquat Microb Ecol* **66:** 107–116.

Cowan, D.A., Pointing, S.B., Stevens, M.I., Cary, C.S., Stomeo, F., and Tuffin, I.M. (2010) Distribution and abiotic influences on hypolithic microbial communities in an Antarctic Dry Valley. *Polar Biol* **34:** 307–311.

Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. (2012) Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *mBio* **3:** e00373–12.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., *et al*. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36:** W465–W469.

Diemer, G.S., and Stedman, K.M. (2012) A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol Direct* **7:** 13.

Duhaime, M.B., and Sullivan, M.B. (2012) Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* **434:** 181–186.

Dwivedi, B., Xue, B., Lundin, D., Edwards, R.A., and Breitbart, M. (2013) A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol Biol* **13:** 33.

Eckardt, F.D., Soderberg, K., Coop, L.J., Muller, A.A., Vickery, K.J., Grandin, R.D., *et al*. (2013) The nature of moisture at Gobabeb, in the central Namib Desert. *J Arid Environ* **93:** 7–19.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Edwards, R.A., and Rohwer, F. (2005) Viral metagenomics. *Nat Rev Microbiol* **3:** 504–510.

Emerson, J.B., Thomas, B.C., Andrade, K., Allen, E.E., Heidelberg, K.B., and Banfield, J.F. (2012) Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Appl Environ Microbiol* **78:** 6309–6320.

Fancello, L., Trape, S., Robert, C., Boyer, M., Popgeorgiev, N., Raoult, D., and Desnues, C. (2013) Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J* **7:** 359–369.

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., *et al*. (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* **73:** 7059–7066.

Fischetti, V.A., Schuh, R., and Nelson, D. (2013) Phage-associated lytic enzymes for treatment of *Bacillus anthracis* and related conditions. US patent US8580553 B2.

Goldsmith, D.B., Crosti, G., Dwivedi, B., McDaniel, L.D., Varsani, A., Suttle, C.A., *et al*. (2011) Development of phoH as a novel signature gene for assessing marine phage diversity. *Appl Environ Microbiol* **77:** 7730–7739.

Grallert, H., and Forchheim, M. (2013) Protease-stable, cell wall-lysing enzymes. US patent US8492519 B2.

Guindon, S., Delsuc, F., Dufayard, J.-F., and Gascuel, O. (2009) Estimating maximum likelihood phylogenies with PhyML. In *Bioinformatics for DNA Sequence Analysis*. Posada, D. (ed.). Clifton, UK: Humana Press, pp. 113–137.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59:** 307–321.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hambly, E., and Suttle, C.A. (2005) The viriosphere, diversity, and genetic exchange within phage communities. *Curr Opin Microbiol* **8:** 444–450.

Hansen, M.C., Tolker-Nielsen, T., Givskov, M., and Molin, S. (1998) Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiol Ecol* **26:** 141–149.

Henschel, J.R., and Lancaster, N. (2013) Gobabeb – 50 years of Namib Desert research. *J Arid Environ* **93:** 1–6.

Henschel, J.R., and Seely, M.K. (2008) Ecophysiology of atmospheric moisture in the Namib Desert. *Atmos Res* **87:** 362–368.

Holmfeldt, K., Solonenko, N., Shah, M., Corrier, K., Riemann, L., Verberkmoes, N.C., and Sullivan, M.B. (2013) Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci USA* **110:** 12798–12803.

Hurwitz, B.L., and Sullivan, M.B. (2013) The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* **8:** e57355.

Hurwitz, B.L., Deng, L., Poulos, B.T., and Sullivan, M.B. (2013) Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ Microbiol* **15:** 1428–1440.

Khan, N., Tuffin, M., Stafford, W., Cary, C., Lacap, D.C., Pointing, S.B., and Cowan, D. (2011) Hypolithic microbial communities of quartz rocks from Miers Valley, McMurdo Dry Valleys, Antarctica. *Polar Biol* **34:** 1657–1668.

Kim, K.-H., Chang, H.-W., Nam, Y.-D., Roh, S.W., Kim, M.-S., Sung, Y., *et al*. (2008) Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl Environ Microbiol* **74:** 5975–5985.

King, A.M.Q., Adams, M.J., Carstens, E.B., and Lefkowitz, E.J. (eds) (2012) *Virus Taxonomy*, 9th edn. London, UK: Elsevier.

Lacap, D.C., Warren-Rhodes, K.A., McKay, C.P., and Pointing, S.B. (2011) Cyanobacteria and chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. *Extremophiles* **15:** 31–38.

Li, W. (2009) Analysis and comparison of very large metagenomes with fast clustering and functional annotation. *BMC Bioinformatics* **10:** 359.

Loessner, M., Schmelcher, M., Grallert, H., and Bretfeld, F. (2012) Artificial Peptidoglycan lysing enzymes and peptidoglycan binding proteins. US patent US20120171188 A1.

López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., and Alcamí, A. (2009) High diversity of the viral community from an Antarctic lake. *Science* **326:** 858–861.

Lundin, D., Torrents, E., Poole, A.M., and Sjöberg, B.-M. (2009) RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences deposited to Genbank. *BMC Genomics* **10:** 589.

Makhalanyane, T.P., Valverde, A., Birkeland, N.-K., Cary, S.C., Tuffin, M.I., and Cowan, D.A. (2013a) Evidence for successional development in Antarctic hypolithic bacterial communities. *ISME J* **7:** 2080–2090.

Makhalanyane, T.P., Valverde, A., Lacap, D.C., Pointing, S.B., Tuffin, M.I., and Cowan, D.A. (2013b) Evidence of species recruitment and development of hot desert hypolithic communities. *Environ Microbiol Rep* **5:** 219–224.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Mokili, J.L., Rohwer, F., and Dutilh, B.E. (2012) Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* **2:** 63–77.

Noguchi, H., Park, J., and Takagi, T. (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34:** 5623–5630.

Noguchi, H., Taniguchi, T., and Itoh, T. (2008) MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* **15:** 387–396.

Nordlund, P., and Reichard, P. (2006) Ribonucleotide reductases. *Annu Rev Biochem* **75:** 681–706.

Pointing, S.B., and Belnap, J. (2012) Microbial colonization and controls in dryland systems. *Nat Rev Microbiol* **10:** 551–562.

Pointing, S.B., Warren-Rhodes, K.A., Lacap, D.C., Rhodes, K.L., and McKay, C.P. (2007) Hypolithic community shifts occur as a result of liquid water availability along environmental gradients in China's hot and cold hyperarid deserts. *Environ Microbiol* **9:** 414–424.

Prestel, E., Salamitou, S., and DuBow, M.S. (2008) An examination of the bacteriophages and bacteria of the Namib desert. *J Microbiol* **46:** 364–372.

Rambaud (2007) *FigTree version 1.4.0*. URL http://tree.bio.ed.ac.uk/.

Reysenbach, A., and Pace, N. (1995) Reliable amplification of hyperthermophilic archaeal 16S rRNA genes by the polymerase chain reaction. In *Archaea: A Laboratory Manual*. Robb, F., and Place, A. (eds). New York, NY, USA: Cold Spring Harbor Laboratory Press, pp. 101–107.

Rho, M., Tang, H., and Ye, Y. (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38:** e191.

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., *et al*. (2010) Viral and microbial community dynamics in four aquatic environments. *ISME J* **4:** 739–751.

Rohwer, F., and Edwards, R. (2002) The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* **184:** 4529–4535.

Rosario, K., and Breitbart, M. (2011) Exploring the viral world through metagenomics. *Curr Opin Virol* **1:** 289–297.

Roux, S., Faubladier, M., Mahul, A., Paulhe, N., Bernard, A., Debroas, D., and Enault, F. (2011) Metavir: a web server dedicated to virome analysis. *Bioinformatics* **27:** 3074–3075.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., *et al*. (2012) Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS ONE* **7:** e33641.

Schlesinger, W.H., Pippen, J.S., Wallenstein, M.D., Hofmockel, K.S., Klepeis, D.M., and Mahall, B.E. (2003) Community composition and photosynthesis by photoautotrophs under quartz pebbles, Southern Mojave desert. *ESA Ecol* **84:** 3222–3231.

Schmitz, J.E., Schuch, R., and Fischetti, V.A. (2010) Identifying active phage lysins through functional viral metagenomics. *Appl Environ Microbiol* **76:** 7181–7187.

Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M., and Mead, D. (2008) Assembly of viral

metagenomes from Yellowstone hot springs. *Appl Environ Microbiol* **74:** 4164–4174.
Schoenfeld, T., Liles, M., Wommack, K.E., Polson, S.W., Godiska, R., and Mead, D. (2010) Functional viral metagenomics and the next generation of molecular tools. *Trends Microbiol* **18:** 20–29.
Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., *et al*. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA* **110:** 1053–1058.
Solonenko, S.A., Ignacio-Espinoza, J.C., Alberti, A., Cruaud, C., Hallam, S., Konstantinidis, K., *et al*. (2013) Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics* **14:** 320.
Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K.E. (2008) Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res Microbiol* **159:** 349–357.
Stomeo, F., Valverde, A., Pointing, S.B., McKay, C.P., Warren-Rhodes, K.A., Tuffin, M.I., *et al*. (2013) Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. *Extremophiles* **17:** 329–337.
Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., and Chisholm, S.W. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4:** e234.
Sullivan, M.B., Krastins, B., Hughes, J.L., Kelly, L., Chase, M., Sarracino, D., and Chisholm, S.W. (2009) The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ Microbiol* **11:** 2935–2951.
Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P.R., *et al*. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* **12:** 3035–3056.
Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27:** 1009–1010.
Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., *et al*. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res* **39:** D546–D551.
Suttle, C. (2005) Viruses in the sea. *Nature* **437:** 356–361.
Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J., and Chisholm, S.W. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci USA* **108:** E757–E764.
Thurber, R.V. (2009) Current insights into phage biodiversity and biogeography. *Curr Opin Microbiol* **12:** 582–587.
Wang, Y., and Zhang, X. (2008) Characterization of a novel portal protein from deep-sea thermophilic bacteriophage GVE2. *Gene* **421:** 61–66.
Warren-Rhodes, K.A., Rhodes, K.L., Pointing, S.B., Ewing, S.A., Lacap, D.C., Gómez-Silva, B., *et al*. (2006) Hypolithic cyanobacteria, dry limit of photosynthesis, and microbial ecology in the hyperarid Atacama Desert. *Microb Ecol* **52:** 389–398.
Warren-Rhodes, K.A., Rhodes, K.L., Boyle, L.N., Pointing, S.B., Chen, Y., Liu, S., *et al*. (2007) Cyanobacterial ecology across environmental gradients and spatial scales in China's hot and cold deserts. *FEMS Microbiol Ecol* **61:** 470–482.
Warren-Rhodes, K.A., McKay, C.P., Boyle, L.N., Wing, M.R., Kiekebusch, E.M., Cowan, D.A., *et al*. (2013) Physical ecology of hypolithic communities in the central Namib Desert: the role of fog, rain, rock habitat, and light. *J Geophys Res Biogeosciences* **118:** 1451–1460.
White, J.R., Nagarajan, N., and Pop, M. (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* **5:** e1000352.
Williamson, K.E., Wommack, K.E., and Radosevich, M. (2003) Sampling natural viral communities from soil for culture-independent analyses. *Appl Environ Microbiol* **69:** 6628–6633.
Williamson, K.E., Radosevich, M., and Wommack, K.E. (2005) Abundance and diversity of viruses in six Delaware soils. *Appl Environ Microbiol* **71:** 3119–3125.
Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., *et al*. (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* **3:** e1456.
Williamson, S.J., Allen, L.Z., Lorenzi, H.A., Fadrosh, D.W., Brami, D., Thiagarajan, M., *et al*. (2012) Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS ONE* **7:** e42047.
Willner, D., and Hugenholtz, P. (2013) From deep sequencing to viral tagging: recent advances in viral metagenomics. *Bioessays* **35:** 436–442.
Wommack, K.E., Bhavsar, J., Polson, S.W., Chen, J., Dumas, M., Srinivasiah, S., *et al*. (2012) VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand Genomic Sci* **6:** 427–439.
Wong, F.K.Y., Lacap, D.C., Lau, M.C.Y., Aitchison, J.C., Cowan, D.A., and Pointing, S.B. (2010) Hypolithic microbial community of quartz pavement in the high-altitude tundra of central Tibet. *Microb Ecol* **60:** 730–739.
Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T., and Takai, K. (2013) Metagenomic analysis of viral communities in (hado)pelagic sediments. *PLoS ONE* **8:** e57271.
Zeng, Q., and Chisholm, S.W. (2012) Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Curr Biol* **22:** 124–128.
Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821–829.
Zhou, C.E., Smith, J., Lam, M., Zemla, A., Dyer, M.D., and Slezak, T. (2007) MvirDB – a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res* **35:** D391–D394.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. PhyML phylogenetic tree of TerL amino acid sequences of the Namib hypolith metavirome (red) with selected phages from the RefSeq database as constructed by MetaVir. Nodes with poor bootstrap support (on 100) were collapsed. Scale bar indicates the number of substitutions per site.

Table S1. Comparison of the Namib hypolith metavirome (HY) taxonomic composition (GAAS) with selected publically available viromes as performed with MetaVir.

Table S2. List of all cyanophages in the NCBI viral genomes database and the number of contig hits with MetaVir (Blastp e value cut-off 10^{-5}) of the hypolith metavirome to these phages.