RUNNING TITLE: Metaviromics of Namib hypoliths

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], Don A Cowan[1,*]

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

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

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

[4] University of Cape Town, South Africa

* Corresponding author: Don A Cowan, Centre for Microbial Ecology and Genomics, Natural Sciences II room 03.12, University of Pretoria, Lynnwood Road, 0028 Pretoria, South Africa

ABSTRACT

Hypolithic microbial communities are specialized desert communities inhabiting the underside of translucent rocks where they are sheltered from harsh environmental conditions. Here, we present the first study of the viral fraction of these communities isolated from the hyperarid Namib Desert (coastal South Western Africa). Using next-generation sequencing of the isolated viral fraction, the diversity and taxonomic composition of hypolith communities was mapped and a functional assessment of the sequences determined. Phylotypic analysis showed that bacteriophages belonging to the order *Caudovirales* with the family *Siphoviridae* were most prevalent. A major fraction of phage types was linked by database homologies to *Bacillus* or *Geobacillus* sp. as a host. Phylogenetic analyses of *terL* and *phoH* marker genes indicated that many of the sequences were novel and distinct from known isolates and environments, an observation supported by the class distribution of identified ribonucleotide reductases. The composition of the viral hypolith fraction was not completely consistent with Namib hypolith phylotypic surveys, in which the cyanobacterial genus *Chroococcidiopsis* was found to be dominant. This could be attributed to lacking sequence information about hypolith viruses/bacteria in public databases or the hypothesis that hypolithic communities actively recruit viruses from the surrounding open soil in which *Bacillaceae*-infecting phages are more commonly found.

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 (Eckardt et al., 2013; Henschel and Lancaster, 2013). At the Gobabeb 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 and water is mostly available from fog and dew events (Henschel and Seely,

2008). The gravel plains north of Gobabeb are home to specialized microbial communities, such as hypolithons, inhabiting the underside of translucent rocks which are found in all major deserts across the world (Bahl et al., 2011; Chan et al., 2012). Hypoliths are present in both hot and cold deserts, where they provide shelter from UV irradiation, dessication and temperature-related stresses (Schlesinger et al., 2003; Warren-Rhodes et al., 2006, 2007; Pointing et al., 2007; Cowan et al., 2010). The microbial communities of hot desert hypoliths are dominated by cyanobacteria, with 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 microbial composition of hypoliths and the surrounding open soil was demonstrated in the Namib Desert (Stomeo et al., 2013). Microbial research in these niche habitats has mainly focused on the bacterial presence and to date, no bacteriophages for this environment or infecting the predominant bacterial classes have been reported.

Metagenomic approaches have become the benchmark for research on microbial community diversity, 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 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 encompassing up to 90% of the sequence data (reviewed in (Mokili et al., 2012)). However, in comparison to alternative phage biogeography approaches such as microscopy, genome/amplicon restriction fingerprinting or target gene sequencing of specific viral families, metaviromics is the only method that can effectively access the total diversity present in a habitat (Thurber, 2009). In addition, metaviromic sequence datasets are excellent new targets for bioprospecting novel genes and gene products, and a new field of functional viral metagenomics using these methods is now emerging (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 amount 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 comprehensive dsDNA viral metagenome dataset from a hot desert niche habitat.

MATERIALS & METHODS

SAMPLE COLLECTION AND PROCESSING

Quartz rocks with established hypolith communities were collected in the Namib Desert near the Gobabeb Research and Training Station (23°33'40" S, 15°02'29" E). The microbial communities were recovered on site and collected in sterile whirl-pack bags (Nasco).

Approximately 0.5kg of hypolith-associated material was suspended in 3L of de-ionized water and homogenized by shaking, then allowed to settle. The supernatant was decanted and the remaining solids suspended in another 3L of de-ionized water and the settling and decanting repeated. The aqueous fraction was centrifuged at low speed (Beckman JA10 rotor - 3000 RPM for 10 min) to remove the largest particles and the supernatant was passed through a 0.22 μm filter (Millipore, Streicup 500ml, Cat. no. SCGPU05RE). The filtrate was then centrifuged to collect phage particles (Beckman JA20 rotor - 19000 RPM for 6 hours). The pellets were collectively resuspended in 3ml TE buffer. This phage suspension was treated with DNAseI

(EN0521) and RNAseA (EN0531) (Fermentas - final concentration of 0.1 µg/ml) at 37°C for 1 hour (DNAseI). The presence of free or background contaminating bacterial DNA was checked by amplifying with 16S RNA gene primers. The phage particles were treated with Proteinase K (Fermentas - final concentration 1 µg/ml) at 55°C for 2 hours. Seventy µl of 20% SDS was then added and the sample was incubated at 37°C for 1 hour. The DNA was extracted with three replicates of phenol:chloroform:isoamylalcohol (25:24:1) phase separation followed by two replicates of chloroform:isoamylalcohol (24:1) phase separation (15ml Sterillin tube, Eppendorf 5810R centrifuge, 5000 RPM for 10min). Precipitation was performed with 1/10 volume of 3M NaOAc (pH 5.2) and 2 x volume 95% ethanol, with overnight incubation at 4°C. Precipitated DNA was recovered by centrifugation at 13000 RPM for 10 minutes and the resulting pellet was resuspended in 30ul of TE buffer. The DNA was further cleaned using the Qiagen Gel Extraction kit (Qiaex II, cat. no. 20021).

ELECTRON MICROSCOPY

Phage suspensions were prepared as described by Ackermann (Ackermann, 2009). Three µl of each sample was pipetted onto carbon coated 200 mesh copper grids and stained with 2% aqueous uranyl acetate. 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) and the MiSeq Reagent kit V2 (500 cycle), and sequenced using the Illumina MiSeq at the University of the Western Cape, Cape Town, South Africa, generating 2 x 250 bp reads. The raw reads were trimmed and demultiplexed at the sequencing facility, resulting in eight (4 x 2) paired fastq files.

IN SILICO ANALYSES

The trimmed and demultiplexed 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. The unassembled sequences were saved and assembled with Velvet (Zerbino and Birney, 2008) (kmer = 15, coverage cutoff = 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, 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), UniProt (MG-RAST), Uniref100 (VIROME), eggNOG (MG-RAST) and MGOL (VIROME).

The presence of “auxiliary metabolic genes” (AMGs) was confirmed by scanning of the MetaVir contig annotation table output for specific metabolic genes, namely *psbA*, *psbB*, *phoH*, *tal* and *nrd*.

The ribonucleotide reductases (*nrd* genes) found in the metavirome were compared by BLAST analysis against the RNRdb, a curated database of ribonucleotide reductases, 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 et al., 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.

RESULTS & DISCUSSION

ELECTRON MICROSCOPY

Analysis of the Namib hypolith virus fraction showed mostly virus particles belonging to the order *Caudovirales*, with *Siphoviridae* phages most commonly observed, followed by *Myoviridae* and *Podoviridae*, as well as various other virus-like particles (Figure 1). No contaminants of bacteria cellular origin were observed in the suspension.

METAVIROME ASSEMBLY

Contig assembly with Seqman Ngen and Velvet yielded 4,575 contigs larger than 500 bp, with an average length of 1,301 bp accounting for a total of 5,950,925 bp. 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 5,830 protein coding features (Tables & Figures

Table 1). With VIROME, predicted ORFs were further subdivided in complete (5,789), missing both ends (983), missing start (2,564) and missing stop (2,599). No rRNA features were found with either VIROME, RAMMCAP or MG-RAST.

Depending on the annotation pipeline used, between 2,545 and 6,755 ORFs with counterparts in public databases were found. 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 (Tables & Figures

Table 1).

VIRAL DIVERSITY AND TAXONOMIC COMPOSITION

The rarefaction curve computed by MetaVir showed approximately 270,000 sequencing clusters at 90% clustering for the 800k reads that were uploaded (Figure 2). The plateau was not reached, but a significant amount of the viral diversity was sampled. For this metavirome, an additional 100k sequences would lead to less than 10k extra clusters. At 98% similarity for clustering, the number of clusters for this metavirome increased to 340,000 (data not shown). Relating this to current phage taxonomy in which isolates are classified in the same species at a nucleotide identity level of 90 to 95% (ICTV Discussions, talk.ictvonline.org), the sequenced portion of this metavirome comprises between 270,000 and 340,000 estimated different species.

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 BlastP matches (e-value cut-off = 10^{-5}) generated by comparison against the RefSeq complete viral genomes database. With these parameters, 23.77% of the sequences produced a significant hit. Looking at the composition of the assembled contigs, 60.99% of the contigs showed similarity to known sequences and 37.46% of the predicted genes. MG-RAST and VIROME also give a taxonomic distribution output, but this is heavily biased towards bacterial taxa, as prophage and temperate phage sequences are often classified as bacterial in origin (data not shown).

Using the GAAS taxonomic composition plot (Figure 3), 80% of the reads was recognized as belonging to dsDNA viruses with no RNA stage, 13% as unclassified phages and 7% 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 phage *Geobacillus* virus E2, making up 6% of the virus fraction, has been described as a siphovirus in its original publication (Wang and Zhang, 2008), giving a total of 56% 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 homology with sequences in the “unclassified

Caudovirales" group and 4% in the "unclassified dsDNA viruses, no RNA stage" grouping. Within 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 for sequencing and should thus be disregarded in the taxonomic composition. Virus families detected at below one percent abundance (and above 0.05%) include *Tectiviridae*, *Ascoviridae*, *Phycodnaviridae*, and *Anelloviridae*.

The most abundant virus encountered in the sample was most closely related to *Geobacillus* virus E2, followed by relatives of *Bacillus* phages phBC6A51, Spbeta and *Bacillus* virus 1 (Figure 3, Table 2). The most abundant hits comprised many phages that infect members of the phylum Firmicutes. This was an unexpected finding, as Firmicutes have not been identified associated with either hot or cold desert hypolith community structures (Wong et al., 2010; Chan et al., 2012; Stomeo et al., 2013; Makhalanyane et al., 2013a; Makhalanyane et al., 2013b). However, members of the *Bacillaceae* have been readily isolated from hot desert soils across the globe (Roberts and Cohan, 1995; Roberts et al., 1996; Palmisano et al., 2001) and *Geobacillus* phages have been isolated from a number of soil types, with mesophilic soils having the highest phage abundance (Reanney and Marsh, 1973). The 15 most abundant phages (Table 2) are of types known to infect soil-associated bacteria, such as *Geobacillus*, *Bacillus, Paenibacillus, Pseudomonas*, *Listeria* and *Sinorhizobium*.

Cyanophages were expected to be abundant based on the known dominance of Cyanobacteria in hypolithon bacterial communities (Warren-Rhodes et al., 2006; Makhalanyane et al., 2013b). To test their presence, the contig annotations (MetaVir) were investigated for cyanophage resemblance (Supplementary Table 1). All but two of the cyanophages in the NCBI virus genome database, of which the majority consists of marine cyanophages, could be mapped to the hypolith metavirome, yet at a low significance level. The same pattern was visible in the MG-RAST taxonomic analysis of the virus-related sequences, which compares the sequences against more databases, and showed only one cyanophage, 9515-10a, in its Best Hit or Representative

Hit classification output (data not shown). In the VIROME analysis, which utilizes BlastP analysis against the Uniref100p database to determine the best match, the same phylotypic composition was observed. *Bacillus* phages were the most common, with relatively few cyanophages identified from one or two ORFs only in the taxonomic analysis output; these included *Synechococcus* phage syn5 and *Phormidium* phage Pf-WMP3 in the *Podoviridae* family, *Synechococcus* phage P-SSM2 in the *Myoviridae* family and *Synechococcus* phage S-CBS2 for the *Siphoviridae*. From these findings we can hypothesize that either cyanophages make up a negligible fraction of the hypolith metavirome or that they are significantly distinct from their marine counterparts to the extent that with the current homology searches, they are not recognized as cyanophages.

A closer analysis of the genes the metavirome reads or predicted ORFs map to (Table 2, column 6), identified many conserved proteins. Numerous reads mapped to the tail tape measure proteins (TMP) found in *Siphoviridae* and *Myoviridae* phages. This is consistent with the electron microscopy analysis that these families were most commonly observed (Figure 1). Other conserved proteins were DNA replication related (primase, helicase, polymerase) or lysis proteins (lysin, amidase). In consequence, coverage of the entire length of these abundant genomes was uneven and on average lower than expected based on the total number of mapped reads (Table 2, ratio of predicted coding sequences (CDS) hits over reads).

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 gene with the most hits in the hypolith virome (as calculated by MetaVir) was the terminase large subunit TerL, present in phages of the *Caudovirales* family. 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 (Supplementary Figure 1), showing that most of the hypolith TerL sequences clustered separately from those of cultured tailed phages, and indicated the novelty of the metavirome. Some sequences did cluster with

known phages, 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 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 3,804 hits, pfam (CAMERA) with 2,222 hits and GO (VIROME) with 3,389 hits. Almost half of the hits in the subsystems functional annotation were directly phage related (Figure 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 signaling', and 'DNA metabolism' as the dominant annotations. In these categories, many proteins were phage-related, such as DNA polymerases, helicases, ribonucleotide reductases 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.

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

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, these AMGs can be involved in photosynthesis (*psbA* and *psbB*), carbon turnover (*talC*), phosphate uptake (*phoH*) or nucleotide metabolism (*nrd* genes) (Sullivan et al., 2006; 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. The presence of these AMGs, associated with nutrient-limited conditions, combined with the absence of photosynthesis-related AMGs suggests that nutrient stress is the most important stress in this hypolith community.

Eighteen complete *phoH* genes and 23 partial sequences were identified in the metavirome by MetaVir. A PhyML phylogenetic tree was generated (Figure 5), in which the topology clearly shows that the majority of the hypolith PhoH protein sequences (in red) cluster separately from those of complete phage genomes, supported by high bootstrap values. The marine cyanophages (in blue) form a distinct clade, unrelated to the hypolith sequences. This again supports the hypothesis that 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 inhabiting this environment. Additionally, this exemplifies the novelty of this hypolith metaviromic dataset. 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 this type of hypolith (Makhalanyane et al., 2013b).

Given that no photosynthesis-related host-derived genes were identified in this metavirome, these findings support the hypothesis of Williamson and colleagues that *phoH* acquisition is more beneficial for phages than that of *pst* genes (Williamson et al., 2008). However, the role of *phoH,* which is frequently used as a marker gene in marine phage diversity, in phosphate uptake

or starvation remains unclear (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 ribonucleotide reductases (RNRs) were identified in the hypolith metavirome. 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 and III RNR genes were also found at low 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 metaviromes analyzed (Dwivedi et al., 2013), again reflecting the novelty of the biome described in this study. 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 findings, that siphoviruses are most common in the Namib hypolith community, and suggests that the host community is aerobic. Furthermore, the hypothesis that *nrdJ* offers a competitive advantage to phages in nutrient-limited 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 ten contigs between 20 and 40 kb in size, were generated during assembly, with the largest contig being 108 kb. Only one contig represented a complete phage genome, which 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 (Figure 6). Of the 75 coding sequences (CDS), 12 showed similarity to siphoviruses and seven to myoviruses, indicating that this phage most probably belongs to the order *Caudovirales*. The genome size is consistent with it being a member of the family *Siphoviridae*. MetaVir classified this contig as being related to

Bacillus phage Cherry, using lowest common ancestor affiliation. 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 (see Figure 2).

ORF prediction of this contig with fgenesV0 (www.softberry.com) identified 62 ORFs while MetaGene (Noguchi et al., 2006), using the CAMERA portal, predicted 72 ORFs. Manual verification of the predicted genes 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 not shown). 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 on hot desert hypolithic community composition have been performed using qPCR, tRFLPs, amplicon sequencing or clone library sequencing with 16S rRNA and ITS primer sets (Warren-Rhodes et al., 2006; Wong et al., 2010; Stomeo et al., 2013; Makhalanyane et al., 2013b). 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 this metavirome is, at first sight, not fully consistent with the results of 16S rRNA gene sequence analyses performed on samples from the same habitat, with the metavirome being dominated by *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. Firstly, there are no known cultured phages, which infect *Chroococcidiopsis* sp. or other bacterial species dominant in hypoliths. As a result, the metavirome sequences could not be linked to known phages. This is related to the method employed by MetaVir to determine the taxonomic composition. The program used best hits generated by BlastP comparison against cultured phage genomes, which will have a bias towards the most frequently isolated phages available in public database (RefSeq complete viral genomes). Subsequently, in the taxonomic composition the presence of cyanophages is severely underestimated compared to that of *Bacillus* phages.

In the PhoH phylogenetic tree (Figure 4), two sequences related to *Xenococcus* sp. were identified, giving an indication of the presence of currently unidentified Pleurocapsales-infecting phages in this metavirome. To test whether the unknown fraction of sequence reads belongs to hypolithon cyanobacteria-infecting phages, both bacteria and phages would need to be isolated and the latter's genomes sequenced. In addition, the presence of most sequenced marine cyanophages in the blast results (Supplementary Table 1), could indicate that the related sequences belong to hypolithic cyanophages and that they are, as mentioned before, significantly different from their marine counterparts.

Another possible explanation is that the contigs currently linked to Firmicutes phages do not belong to phages actually infecting this bacterial type. Evidence for this hypothesis can be found in the annotation of the larger contigs, which showed little overall sequence similarity with the phages they are linked to taxonomically.

A final hypothesis could be that the hypolithic microbial community has recruited phages from the surrounding open soil. This concept is supported by previous phylogenetic surveys (Makhalanyane et al., 2013b) which showed that 80% of bacterial operational taxonomic units (OTUs) found in the hypoliths were also found in the surrounding soil. The most abundant

bacterial taxon in Namib Desert open soil was the phylum Actinobacteria (Makhalanyane et al., 2013b). This phylum is represented in the hypolith metavirome by a number of *Mycobacterium, Rhodococcus* and *Propionibacterium* siphoviruses, albeit at low abundance and low coverage (data not shown). Only one previous study has investigated the phage composition in desert soil (Joshua Tree NP, CA, USA) and reported that the most abundant phage types infected common soil bacteria including the genus *Mycobacterium* (Fierer et al., 2007). In general, soils seem to be dominated by five bacterial phyla, Acidobacteria, Actinobacteria, Alphaproteobacteria, Beta/Gammaproteobacteria and Bacteroidetes (Lauber et al., 2009). Of these, only the Proteobacteria and Actinobacteria have a large number of well-described phages (NCBI), leading to a database bias towards these phage types.

CONCLUSION & 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 of which the order *Caudovirales* was predominant as well as an abundance of *Bacillus*-infecting phages. This is consistent with the phages previously described from the Namib Desert (Prestel et al., 2008). The prevalence of Firmicutes-infecting phages as opposed to the expected cyanophages is thought to be caused either by a database bias towards the former group of phages or by phage recruitment from the surrounding desert soil. The database issue could be resolved by further research, including full metagenomic sequencing of hypoliths and surrounding soil and culturing of hypolith-associated bacteria and their bacteriophages.

REFERENCES

Ackermann,H.-W. (2009) Basic phage electron microscopy. In, Clokie,M.R.J. and Kropinski,A.M. (eds), *Bacteriophages: Methods and Protocols*. Humana Press, New York, NY, USA, pp. 113–126.

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.

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

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., Craig Cary,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.

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**: W465–469.

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.

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**: 6309–20.

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,Noah, 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.

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.

Goldsmith,D.B., Crosti,G., Dwivedi,B., McDaniel,L.D., Varsani,A., Suttle,C. a, et al. (2011) Pho Regulon Genes in Phage: Development of phoH as a Novel Signature Gene for Assessing Marine Phage Diversity. *Appl Environ Microbiol* **77**: 7730–7739.

Guindon,S., Delsuc,F., Dufayard,J.-F., and Gascuel,O. (2009) Estimating maximum likelihood phylogenies with PhyML. In, Posada,D. (ed), *Bioinformatics for DNA sequence analysis*. Humana Press, Clifton, NJ, USA, 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.

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.

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.

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.

Lauber,C.L., Hamady,M., Knight,R., and Fierer,N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75**: 5111–20.

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

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., Marla Tuffin,I., and Cowan,D.A. (2013) Evidence for successional development in Antarctic hypolithic bacterial communities. *ISME J* **online fir**:

Makhalanyane,T.P., Valverde,A., Lacap,D.C., Pointing,S.B., Tuffin,M.I., and Cowan,D.A. (2013) 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.

Palmisano,M.M., Nakamura,L.K., Duncan,K.E., Istock,C. a, and Cohan,F.M. (2001) Bacillus sonorensis sp. nov., a close relative of Bacillus licheniformis, isolated from soil in the Sonoran Desert, Arizona. *Int J Syst Evol Microbiol* **51**: 1671–1679.

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.

Reanney,D.C. and Marsh,S.C.N. (1973) The ecology of viruses attacking *Bacillus stearothermophilus* in soil. *Soil Biol Biochem* **5**: 399–408.

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

Roberts,M.S. and Cohan,F.M. (1995) Recombination and migration rates in natural populations of Bacillus subtilis and Bacillus mojavenis. *Evolution (N Y)* **49**: 1084–1091.

Roberts,M.S., Nakamura,L.K., and Cohan,F.M. (1996) Bacillus vallismortis sp. nov., a close relative of Bacillus subtilis, isolated from soil in Death Valley, California. *Int J Syst Bacteriol* **46**: 470–475.

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.

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**: 7181–7.

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.

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**: 4164–74.

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. *PNAS* **110**: 1053–1058.

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., 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**: 2935–51.

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.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**: D546–551.

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 U S A* **108**: E757–64.

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**: 61–6.

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., 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.

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.

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.

Willner,D. and Hugenholtz,P. (2013) From deep sequencing to viral tagging: Recent advances in viral metagenomics. *Bioessays*.

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**: 427–39.

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**: 730–9.

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**: 124–8.

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**: D391–4.

TABLES & FIGURES

TABLE 1: AUTOMATED PIPELINE COMPARISON OF THE NAMIB HYPOLITH METAVIROME.

	MetaVir	VIROME	MG-RAST	CAMERA
# predicted CDS	11,919	11,935	5,830	11,289
# affiliated CDS	4,462	6,755	2,545	6,741
# 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

TABLE 2: METAVIR READ RECRUITMENT AND GENOME COVERAGE DATA: VIRUS DATABASE HITS OF PREDICTED GENES OF THE METAVIROME IN NUMBER OF READS AND PREDICTED CODING SEQUENCES (CDS). THE TOP 15 MOST COMMONLY IDENTIFIED PHAGES ARE SHOWN.

Phage	Taxonomy[a]	# reads in tax compos.[b]	# reads hits[b]	# predicted CDS hits	CDS with >10% of total hits
Geobacillus virus E2	*Siphoviridae*	10804	9336	326	TMP, minor structural protein, N-acetylmuramoyl-L-alanine amidase, hypothetical protein
Bacillus phage phBC6A51	Unclassified phage (prophage)	10678	8439	297	Hypothetical protein, scaffolding protein
Bacillus phage SPbeta	*Spbetalikevirus*, *Siphoviridae*	8182	6488	261	Hypothetical protein, transglycosylase
Bacillus virus 1	Unclassified *Caudovirales*	3952	3372	238	N-acetylmuramoyl-L-alanine amidase,
Bacillus phage phi105	*Siphoviridae, Lambdalikevirus*	3051	2619	79	Recombinase, hypothetical protein
Pseudomonas phage vB_Pae-Kakheti25	*Siphoviridae*	3040	2611	46	-
Lactococcus phage 1706	*Siphoviridae*	3210	2614	53	Helicase
Deep-sea thermophilic phage D6E	*Myoviridae*	3119	2520	174	Lysin, lipid A transport ATP binding
Burkholderia phage KL1	*Siphoviridae*	2971	2563	40	Helicase
Paenibacillus phage phiIBB_Pl23	Unclassified *Caudovirales*	2961	2511	90	Helicase
Listeria phage LP-125	*Myoviridae*	2485	2309	44	nrdA, nrdB
Sinorhizobium phage PBC5	Unclassified *Caudovirales*	2467	2226	41	Hypothetical protein
Enterococcus phage BC-611	"Sap6likevirus", *Siphoviridae*	2415	1831	165	DNA primase, transcriptional regulator, replicative DNA helicase, hypothetical protein, DNA polymerase
Bacillus phage PBC1	*Siphoviridae*	2146	1880	49	TMP, thymidylate synthase, hypothetical protein
Bacillus phage BtCS33	*Siphoviridae*	2131	1796	168	TMP, HTH transcriptional regulator

[a] Taxonomy is presumed classification from the original publications, ICTV, NCBI and taxonomy proposals submitted to ICTV (genera between ").

[b] # Reads hits takes only the Best Blast Hit into account while in the taxonomic composition of the raw reads, a read can be assigned to multiple taxa.

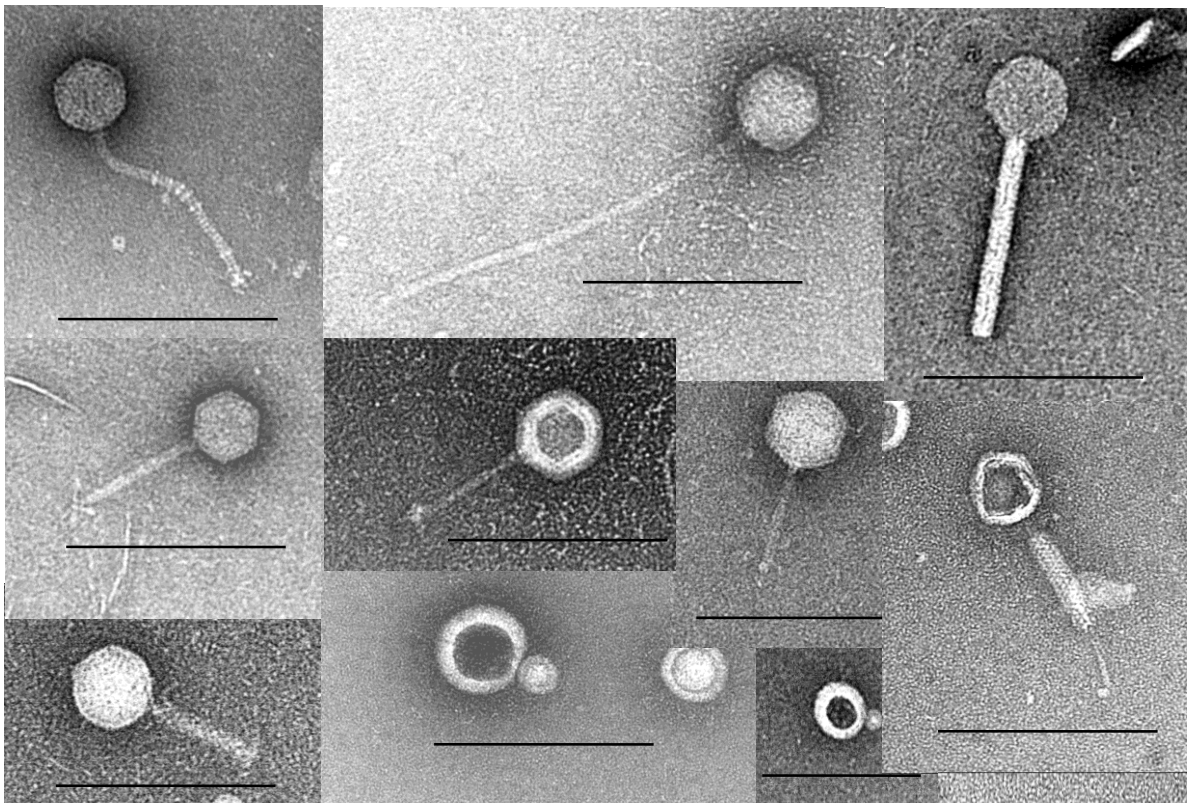

FIGURE 1: TRANSMISSION ELECTRON MICROGRAPHS OF THE VIRAL FRACTION OF HYPOLITH SCRAPINGS. SCALE BARS REPRESENT 200 NM. PARTICLES WERE NEGATIVELY STAINED WITH 2% URANYL ACETATE.

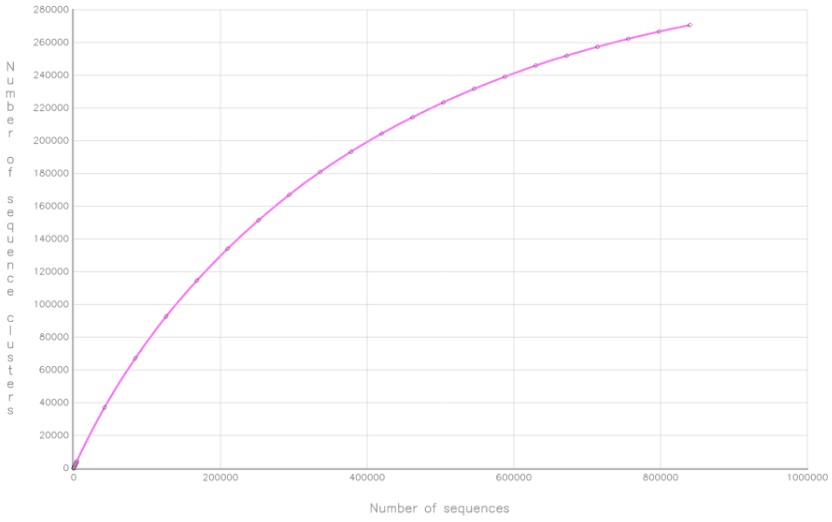

FIGURE 2: RAREFACTION CURVE OF THE NAMIB HYPOLITH METAVIROME GENERATED BY METAVIR. CLUSTERING PERCENTAGE WAS SET AT 90%.

FIGURE 3: TAXONOMIC COMPOSITION OF THE NAMIB HYPOLITH METAVIROME (METAVIR OUTPUT). COMPOSITION TYPE WAS GAAS (GENOME LENGTH NORMALIZATION) AND THE THRESHOLD WAS SET AT 10^{-5} ON E VALUE. THE MOST ABUNDANT VIRUS TAXA ARE INDICATED IN RED, SECOND YELLOW, THEN GREEN, BLUE AND PURPLE.

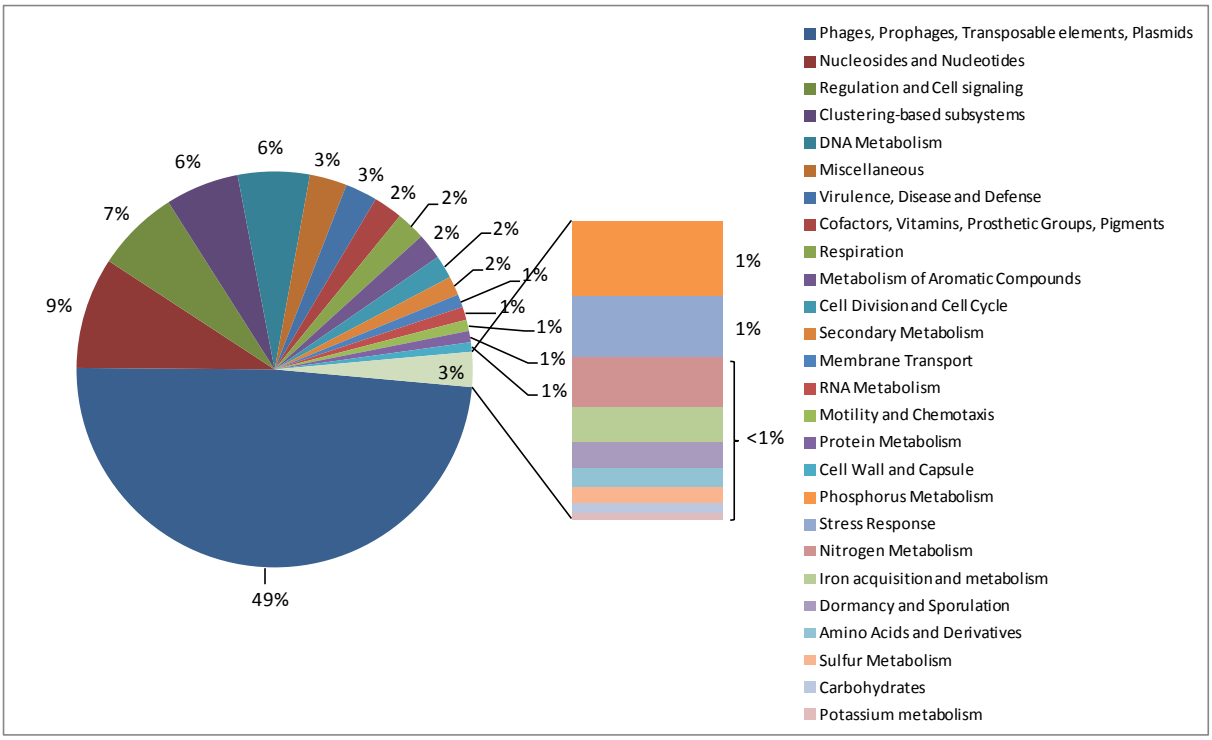

FIGURE 4: COMPOSITION OF PREDICTED FUNCTIONAL GENES OF THE NAMIB HYPOLITH CONTIGS AGAINST THE SEED DATABASE USING SUBSYSTEMS (MG-RAST).

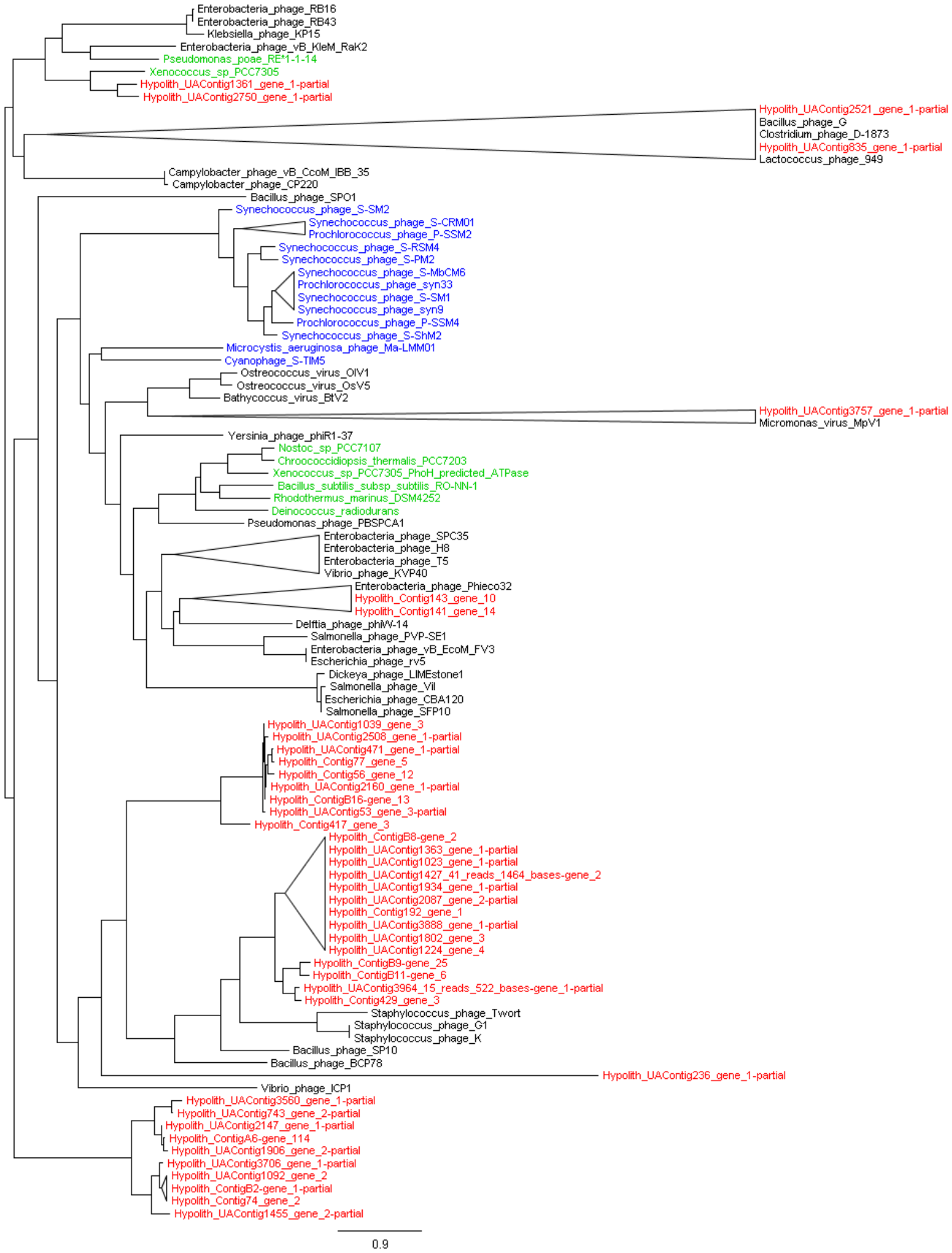

FIGURE 5: PHYML PHYLOGENETIC TREE OF PHOH AMINO ACID SEQUENCES OF THE NAMIB HYPOLITH METAVIROME (RED), SELECTED PHAGES WITH MARINE CYANOPHAGES IN BLUE AND SELECTED BACTERIA (GREEN). SEQUENCES WERE ALIGNED WITH MUSCLE ON THE

PHYLOGENY.FR SERVER AND VISUALIZED WITH FIGTREE. NODES WITH LESS THAN 40% BOOTSTRAP SUPPORT WERE COLLAPSED (ON 100 BOOTSTRAPS). SCALE BAR INDICATES THE NUMBER OF SUBSTITUTIONS PER SITE.

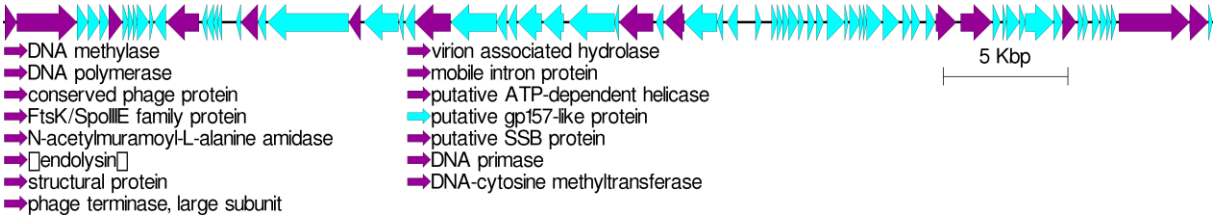

FIGURE 6: CONTIG 1 WITH CDS (ARROWS) AS PREDICTED WITH METAVIR. CDS IN PURPLE HAVE PUTATIVE FUNCTIONS ASSIGNED. THE FIGURE WAS GENERATED USING EASYFIG (SULLIVAN ET AL., 2011)**.**

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY TABLE 1: LIST OF ALL CYANOPHAGES IN THE NCBI VIRAL GENOMES DATABASE AND THE NUMBER OF HITS WITH METAVIR OF THE HYPOLITH METAVIROME TO THESE PHAGES.

Organism	RefSeq accession number	Namib hypolith MV hits	Taxonomy
Cyanophage 9515-10a	NC_016657	4	Podoviridae, Autographivirinae
Cyanophage KBS-M-1A	NC_020836	10	unclassified dsDNA phages
Cyanophage KBS-P-1A	NC_020865	7	unclassified dsDNA phages
Cyanophage KBS-S-2A	NC_020854	11	unclassified dsDNA phages
Cyanophage MED4-117	NC_020857	3	unclassified dsDNA phages
Cyanophage MED4-184	NC_020847	3	unclassified dsDNA phages
Cyanophage MED4-213	NC_020845	8	unclassified dsDNA phages
Cyanophage NATL1A-7	NC_016658	3	Podoviridae, Autographivirinae
Cyanophage NATL2A-133	NC_016659	9	Podoviridae, Autographivirinae
Cyanophage P-RSM1	NC_021071	17	Myoviridae
Cyanophage P-RSM6	NC_020855	17	Myoviridae
Cyanophage P-SSP2	NC_016656	7	Caudovirales
Cyanophage PSS2	NC_013021	74	Siphoviridae
Cyanophage S-TIM5	NC_019516	36	Myoviridae
Cyanophage SS120-1	NC_020872	17	Caudovirales
Cyanophage Syn30	NC_021072	11	unclassified dsDNA phages
Microcystis aeruginosa phage Ma-LMM01	NC_008562	69	Myoviridae
Phormidium phage Pf-WMP3	NC_009551	28	Podoviridae
Phormidium phage Pf-WMP4	NC_008367	33	Podoviridae
Planktothrix phage PaV-LD	NC_016564	62	Podoviridae
Prochlorococcus phage P-GSP1	NC_020878	8	unclassified dsDNA phages
Prochlorococcus phage P-HM1	NC_015280	7	Myoviridae
Prochlorococcus phage P-HM2	NC_015284	42	Myoviridae
Prochlorococcus phage P-RSM4	NC_015283	13	Myoviridae
Prochlorococcus phage P-SSM2	NC_006883	24	Myoviridae, Tevenviruses, T4likevirus
Prochlorococcus phage P-SSM3	NC_021559		Myoviridae
Prochlorococcus phage P-SSM4	NC_006884	25	Myoviridae, Tevenviruses, T4likevirus
Prochlorococcus phage P-SSM7	NC_015290	37	Myoviridae
Prochlorococcus phage P-SSP10	NC_020835	4	Podoviridae
Prochlorococcus phage P-SSP3	NC_020874	7	unclassified dsDNA phages
Prochlorococcus phage P-SSP7	NC_006882	11	Podoviridae, Autographivirinae
Prochlorococcus phage Syn1	NC_015288	11	Caudovirales
Prochlorococcus phage Syn33	NC_015285	36	Caudovirales
Synechococcus phage P60	NC_003390	40	Podoviridae, Autographivirinae
Synechococcus phage S-CAM1	NC_020837	14	Myoviridae
Synechococcus phage S-CBS1	NC_016164	6	Siphoviridae
Synechococcus phage S-CBS2	NC_015463	86	Siphoviridae
Synechococcus phage S-CBS3	NC_015465	7	Siphoviridae
Synechococcus phage S-CBS4	NC_016766	28	Siphoviridae
Synechococcus phage S-CRM01	NC_015569	38	Myoviridae
Synechococcus phage S-IOM18	NC_021536		Myoviridae

Synechococcus phage S-MbCM6	NC_019444	13	Myoviridae, Tevenviruses, T4likevirus
Synechococcus phage S-PM2	NC_006820	23	Myoviridae, Tevenviruses, T4likevirus
Synechococcus phage S-RIP1	NC_020867	5	Podoviridae
Synechococcus phage S-RIP2	NC_020838	7	Podoviridae
Synechococcus phage S-RSM4	NC_013085	32	Myoviridae, Tevenviruses, T4likevirus
Synechococcus phage S-SKS1	NC_020851	17	Siphoviridae
Synechococcus phage S-SM1	NC_015282	12	Myoviridae
Synechococcus phage S-SM2	NC_015279	23	Myoviridae
Synechococcus phage S-SSM4	NC_020875	17	Myoviridae
Synechococcus phage S-SSM5	NC_015289	15	Myoviridae
Synechococcus phage S-SSM7	NC_015287	34	Myoviridae
Synechococcus phage S-ShM2	NC_015281	5	Myoviridae
Synechococcus phage Syn19	NC_015286	13	Myoviridae
Synechococcus phage Syn5	NC_009531	5	Podoviridae, Autographivirinae
Synechococcus phage metaG-MbCM1	NC_019443	8	Myoviridae, Tevenviruses, T4likevirus
Synechococcus phage syn9	NC_008296	10	Myoviridae, Tevenviruses, T4likevirus
Total hits of metavirome to cyanophages		1112	