Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community

Surendra Vikram[1], Leandro Guerrero[1], Thulani P. Makhalanyane[1], Phuong T. Le[1, 2], Mary Seely[3], Don A. Cowan[1]*

Running Title: Functional capacity in Namib Hypoliths

1. Centre for Microbial Ecology and Genomics, University of Pretoria, South Africa
2. University of Ghent, Ghent, Belgium
3. Desert Research Centre, Namibia

*Corresponding Author: Prof. Don A. Cowan,
Centre for Microbial Ecology and Genomics,
University of Pretoria,
Lynwood Road, 0028 Pretoria,
South Africa

Email: don.cowan@up.ac.za

Abstract

Environmental stressors such as low water activity and temperature extremes impose severe limitations on the productivity of soils in hyperarid deserts. In such ecosystems, macroscopic communities are often restricted to cryptic niche habitats, such as hypoliths (microbial communities found beneath translucent rocks), which are widely distributed in hyperarid desert environments. While hypolithic communities are considered to play a major role in the productivity of hyperarid habitats, the functional guilds implicated in these processes remain unclear. Here, we describe the Illumina-based metagenomic sequencing (± 30 Gb), assembly and analysis of hypolithic microbial communities from the south-west African Namib Desert. Taxonomic analyses using Small Subunit (SSU) phylogenetic markers showed that bacterial phylotypes (93%) dominated the communities, with relatively small proportions of archaea (0.43%) and fungi (5.6%). BlastX analysis against the refseq-viral database showed the presence of double stranded DNA viruses (7.8% contigs), dominated by Caudovirales (59.2%). Analysis of functional genes and metabolic pathways revealed that cyanobacteria were primarily responsible for photosynthesis with the presence of multiple copies of genes for both photosystems I and II, with a smaller but significant fraction of proteobacterial anoxic photosystem II genes. Hypolithic community members demonstrated an extensive genetic capacity for the degradation of phosphonates and mineralization of organic sulfur. Our data suggest that Proteobacterial guilds may be more significant in desert niches than previously recognized, as they showed widespread genetic capacity for mediating key stages in all biogeochemical cycles. Surprisingly, we were unable to show the presence of genes representative of complete nitrogen cycles. The diversity of *nif* genes was low, and the metagenome showed no evidence of other key N-cycling genes. Taken together, our analyses suggest an extensive capacity for carbon, phosphate and sulphate cycling but only limited nitrogen biogeochemistry.

Introduction

The mechanisms controlling primary productivity in soil systems, particularly in hyperarid deserts, remain poorly understood (1). Arid soil ecosystems cover over 40% of terrestrial surfaces and therefore contribute a significant fraction of the global soil carbon budget (2). In hyperarid deserts, where the extreme stochasticity of rainfall events generally results in very low plant primary productivity, it is generally accepted that microbial communities are major contributors to the key processes of ecosystem services (1). Hypoliths, cryptic assemblages found on the ventral surfaces of translucent rocks, are a prominent feature of both hot and cold deserts (3, 4). The cryptic hypolithic habitat is known to modulate some of the extreme environmental stressors associated with hyper-aridity, including temperature extremes, high incidences of UVA/B and low water availability (5-8). These communities provide a model system for understanding the factors that control microbial primary productivity (4).

The majority of published studies on hypoliths have focused on understanding the microbial diversity and ecology of these communities (6, 9-12). Environmental DNA sequence-based analyses have demonstrated that hypolithic microbial communities in hot deserts are dominated by cyanobacterial lineages of the order Pleurocapsales, predominantly members of the genus *Chroococcidiopsis* (10, 12-14). In contrast, hypoliths in cold and polar deserts are dominated by Oscillatorian cyanobacterial morphotypes (3). Hypolithic communities also contain diverse groups of heterotrophic bacteria from the phyla Actinobacteria, Acidobacteria, Proteobacteria and Bacteroidetes, many of which belong to the 'so-called' category of microbial "dark matter" (3, 6, 10, 15). It has been shown that hot desert hypoliths selectively recruit microbial taxa from surrounding soils, and that these cyanobacterial-dominated communities may drive community interactions and system functionality in hyperarid deserts where plant biomass is limited and transitory (12).

Acetylene reduction assays have been used to show that hypolithic communities in cold deserts are a vital input source of nitrogen (16). A more recent study by Chan and colleagues applied microarray analysis of functional genes involved in autotrophy, nitrogen metabolism and stress responses in Antarctic Dry Valley soils (17) and showed, for the first time, that hypoliths harbor high metabolic potential for biogeochemical cycling. However, little is currently known of the breadth of functional capacity in hot desert hypolithic communities and their role in edaphic biogeochemical cycles (4).

Here we report a metagenomic analysis of the functional potential of hot desert (Namib) hypolithic communities. The central Namib Desert, on the south-west coast of continental Africa, is designated as a hyper-arid zone with a mean annual rainfall of approximately 25 mm (18). The northern Namib gravel desert zone is rich in quartz reefs, resulting in extensive contributions of quartz pebbles to the desert pavement (4). In this study, we explore the metagenome-derived community structure, assess the genetic capacity for primary productivity and nutrient cycling (including N, C, and P metabolism) and demonstrate the diversity of genes and pathways which may represent adaptations of taxa in the hypolithic niche to environmental stressors in this hot desert environment.

Results and Discussion

Sequence data

Illumina Hiseq-2000 sequencing of bulked metagenomic DNA from multiple (n = 40) Namib Desert hypolithic biomass samples generated 19.5 billion bp of sequencing data (**Supplementary Materials Table S1**). Primary assembly using Velvet resulted in 2,188,786 contigs with a total assembly size of approximately 634 million bp (**Supplementary**

Materials Table S2). All contigs shorter than 500 bases were culled, and the average size of the remaining contigs was 787 bp (**Supplementary Materials Table S2)**.

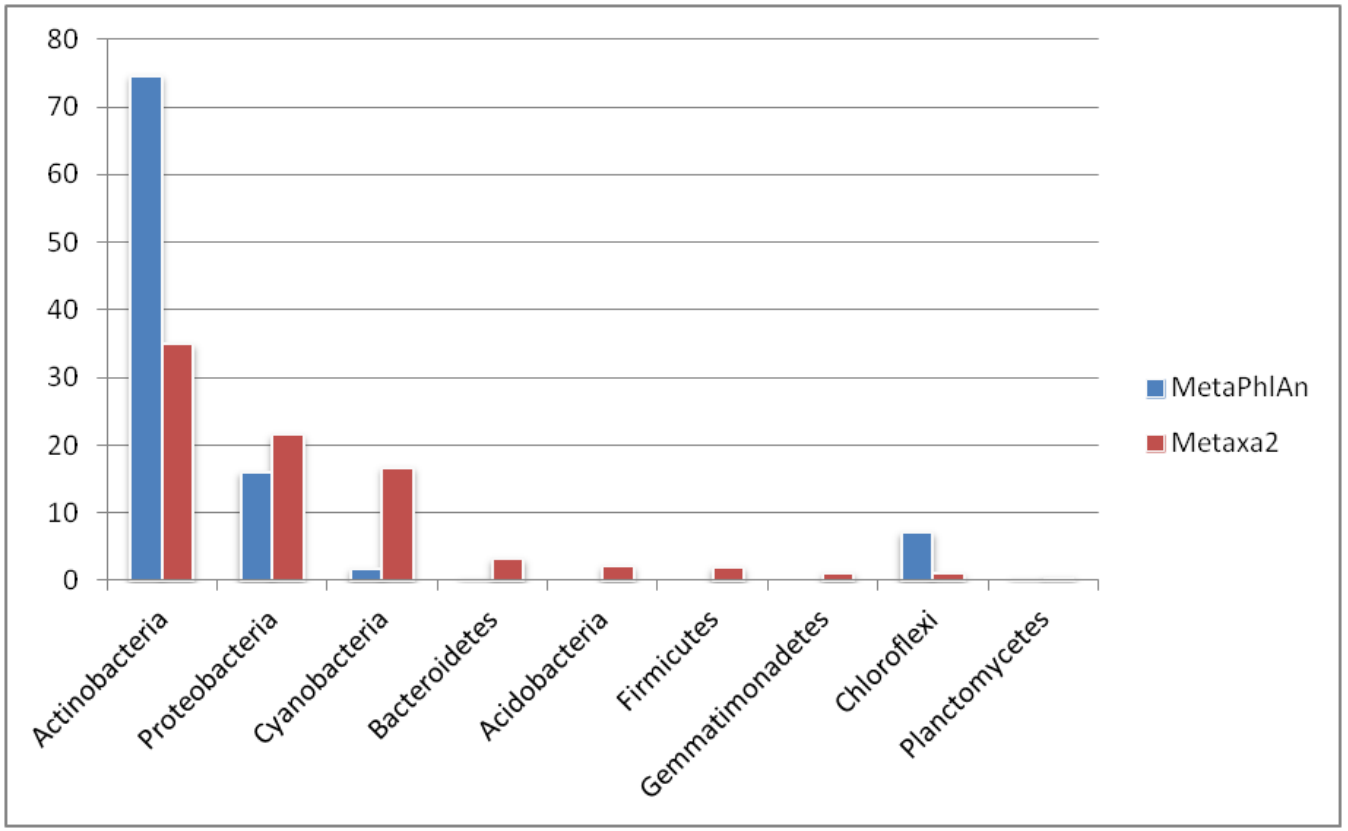
Phylogenetic analysis

We first explored the microbial community composition by analyzing the reads using two approaches; the ribosomal small subunit (SSU) using Metaxa2 (19) and unique clade-specific marker genes using MetaPhlAn (20). SSU analysis suggested that the Namib hypolithic niche contains very high Bacterial diversity (93% of total phylotypic signals) with a significantly lower proportion of Fungi (5.6%) and Archaea (0.43%). The low archaeal and eukaryotic diversity in the hypolith metagenome is consistent with previous phylogenetic surveys, which indicated that these groups are poorly represented in such microenvironments (5, 6, 9, 21, 22).

The taxonomic analysis of SSUs of the bacterial fraction revealed that the phyla Actinobacteria, Proteobacteria and Cyanobacteria were highly represented in the metagenome, while other phyla such as Firmicutes, Chloroflexi, Acidobacteria, Bacteroidetes and Planctomycetes were present as relatively minor contributors to total bacterial diversity (**Fig. 1**). MetaPhlAn analysis of the taxonomic prediction using reads also showed a high abundance of Actinobacteria, Proteobacteria, Chloroflexi and Cyanobacteria, although this method yielded slightly different proportions of the taxonomic groups compared to the Metaxa2 (Fig. 1).

These results are in agreement with earlier studies, based on PCR amplification of 16S rRNA genes, which found that hypolithic communities in hot deserts were dominated by the Chroococcidiopsis lineages (order Pleurocapsales) followed by phyla Actinobacteria and

Figure 1. Taxonomic classification of metagenomic reads. Classification was performed by filtered high quality reads using MetaPhlAn and Metaxa2 at Phylum level. Bar graph is showing the percent abundance of the different bacterial phyla in hypoliths metagenome.

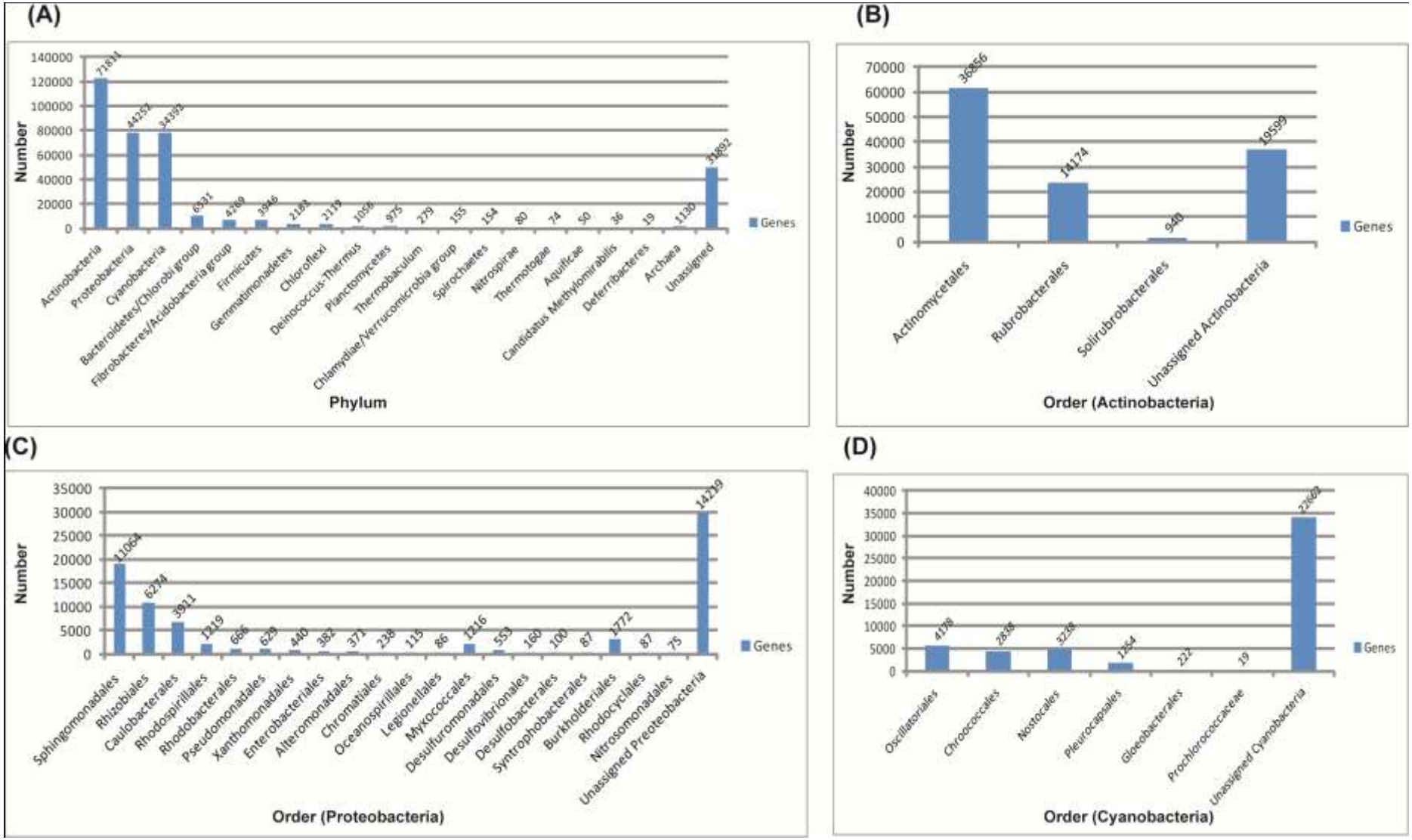


Proteobacteria (6, 12, 23, 24). The results of the taxonomic analysis of reads from the Namib Desert metagenome are generally consistent with the range and relative proportions of phyla present in hypoliths, albeit with slight differences in the relative proportions. The minor inconsistencies between 16S rRNA gene sequence-based diversity from previous studies and the results from the metagenomic analysis are perhaps unsurprising, given the accepted potential for PCR amplification bias (25).

Binning of assembled contigs and functional analysis

Assembled contigs were used to predict ORFs, binned and assigned to taxonomic groups using MyTaxa software which uses the ORF identity in each contig to assign it to the most probable taxon (27). Following contig classification, ORFs were assigned to the following bacterial phyla: Actinobacteria (122362), Proteobacteria (77810), Cyanobacteria (77810), Bacteroidetes (10444), Acidobacteria (6911), Firmicutes (6893), Gemmatimonadetes (3363) and Chloroflexi (3625) (**Fig 2A)**. 36940 ORFs of the most dominant phylum, Actinobacteria, were assigned as unclassified (i.e., no lower order phylogenetic identity) (Fig 2B) and 29805 ORFs of the Proteobacteria were assigned only up to phylum level (**Fig 2C**). Cyanobacterial ORFs were dominated by the order Oscillatoriales (5722 ORFs), with other phyla identified as Nostocales (4808 ORFs), Chroococcales (4360 ORFs), Pleurocapsales (1843 ORFs) and Gloeobacterales (347 ORFs) (**Fig 2D**). A high proportion of the cyanobacterial-assigned contigs (22662: 66%) could not be attributed to specific taxonomic groups. Fewer ORFs (1936) were assigned to archaea (**Fig 2A**), and most archaeal sequences were attributed to the phyla Euryarchaeota (1171 ORFs/648 contigs), Thaumarchaeota (612 ORFs/390 contigs) and Crenarchaeota (146 ORFs/89 contigs). These results were consistent with the results obtained from the classification of reads using Metaxa2. A substantial portion of the total assigned ORFs (80532: 20%) were classified as unknown and were not linked to any known phylum.

Fig 2. Predicted genes (ORF) and contigs in each taxa(A) Abundance of ORF belongs to different bacterial phya. Distribution of ORFs for the three most abundant bacterial phyla **(B)** Actinobacteria **(C)** Proteobacteria, and **(D)** Cyanobacteria. (Number of contigs for each taxa are written on the top of columns.)

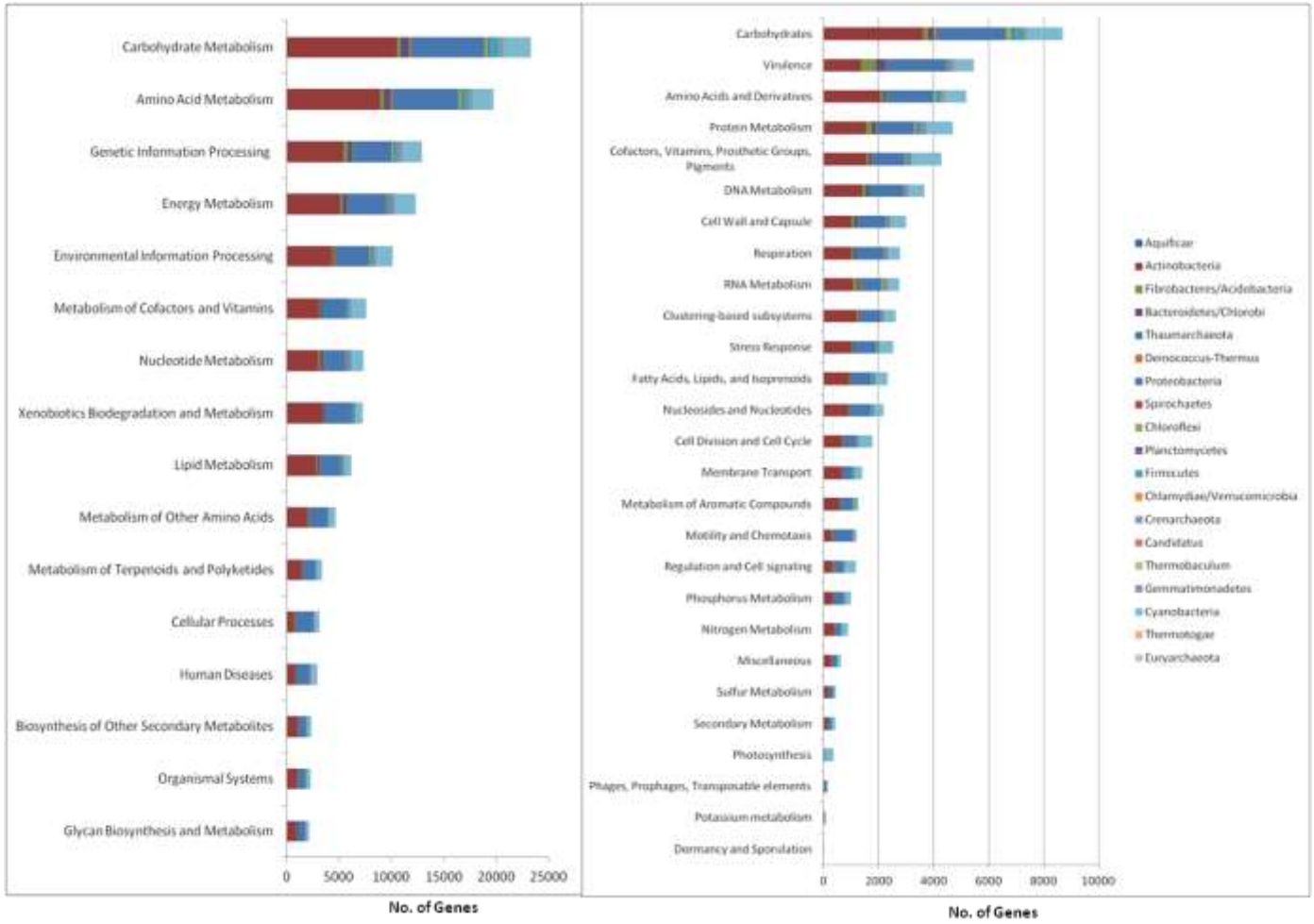


Although viruses and bacteriophages are likely to play a major role in microbial diversity and functionality in soil systems (28), very little is known about phage-host associations and processes in desert soils (29, 30). 21,666 contigs (8.5%) matched (blastx) to the RefSeq virus database and most of the sequences were assigned to dsDNA viruses (**Supplementary Materials Fig S1**). Caudovirales, followed by Phycodnaviridae and Mimiviridae, respectively, were the most abundant orders. This result is consistent with findings from a recent hot desert hypolith metavirome sequence analysis (31) and supports the conclusion that these members of the Caudovirales are widespread in hot deserts. Notably, the proportion of the three families (Myoviridae, Siphoviridae and Podoviridae) in the Caudovirales have only slightly different values from those reported by Adriaenssens and coworkers (31) **(Supplementary Materials Table S3)**, perhaps a surprising result given the known biases of the multiple displacement amplification protocol (32) underlying this metaviromics study.

In order to better understand the functional potential of the microorganisms represented in this metagenome, we used MEGAN to assign functions to the predicted ORFs (33). Our analysis showed that of a total of 396,495 genes, an estimated 118,983 (~30%) were successfully assigned to the KEGG orthology (KO numbers) (**Fig. 3A**). A further 57,365 (~14%) were annotated to biological SEED subsystem proteins using the refseq protein database **(Fig. 3B)**. For comparison, in a study of the human gut microbiome metagenome, 47% of the genes were assigned to the KEGG orthology (34).

The most abundant phyla (Actinobacteria, Proteobacteria and Cyanobacteria) were selected for an analysis of shared KEGG pathway modules. From a total of 580 KEGG pathway modules, 358 were incomplete in all three phyla. Interestingly, 83 modules were shared

Figure 3. Classification of the genes by MEGAN. Bar graph is showing the number of genes assigned to the each phylum **(A)** KEGG pathway and **(B)** SEED subsystem

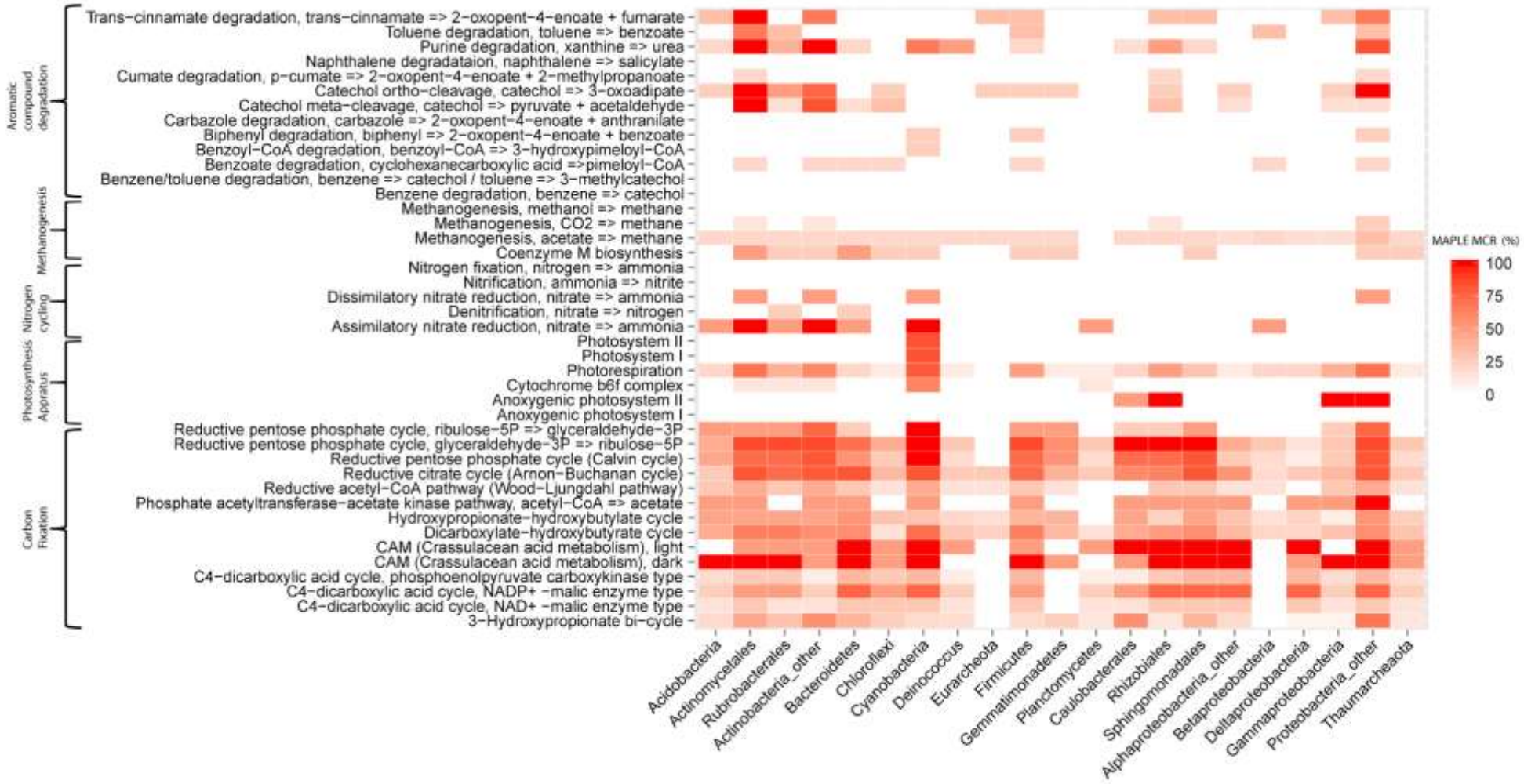


among these phyla, which suggest that these modules represent core metabolic pathways and are essential for organismal survival. Forty-five pathway modules, shared by the phyla Actinobacteria and Proteobacteria, were mainly involved in heterotrophic metabolism and stress response regulatory systems (**Supplementary Materials Fig S2 & Table S4**). Proteobacteria and Cyanobacteria shared nine pathway modules, including carbon fixation and nitrogen fixation, while Actinobacteria and Cyanobacteria also shared nine pathway modules, assigned to metal transport (**Supplementary Materials Fig S1 & Table S4**). Desert microbial communities have previously been shown to possess multiple genes involved in metal acquisition and the maintenance of metal homeostasis (35). The presence of a high number of common pathway modules implicated in metal homeostasis suggests that these may be essential for survival in hot desert edaphic environments.

Primary productivity: photosynthesis and carbon metabolism

Photosynthetic microorganisms are keystone taxa in hypolithic systems (12), and may be the dominant primary producers for long periods in hyper-arid environments (6, 13, 15, 23). We used the MAPLE server, which evaluates KEGG pathway modules based on KAAS assignment of KEGG orthology terms to specific genes/proteins and calculates the module completion ratios (MCR) for each pathway. Results from MAPLE-MCR analysis suggested that photosystem-I and II modules were complete for the phylum Cyanobacteria, with potentially photosynthetically functional members of the orders Oscillatoriales, Chroccocales, Nostocales and Pleurocapsales **(Fig 2D)**. In addition, anoxygenic photosytem-II genes were present in Proteobacteria, including members of the Deltaproteobacteria, Rhizobiales and unclassified Proteobacteria, although no evidence of the anoxigenic photosystem-I pathway could be found for these taxa **(Fig. 4)**. SEED subsystem and KEGG pathway analyses failed to identify any photosynthesis genes belonging to any other non-cyanobacterial or non-

Figure 4. Heat map showing the percentage module completion ration (MCR) for the aromatic compound degradation, energy metabolism and Photosystem apparatus. Module completion ratio for the pathways was calculated by MAPLE server using KEGG database.



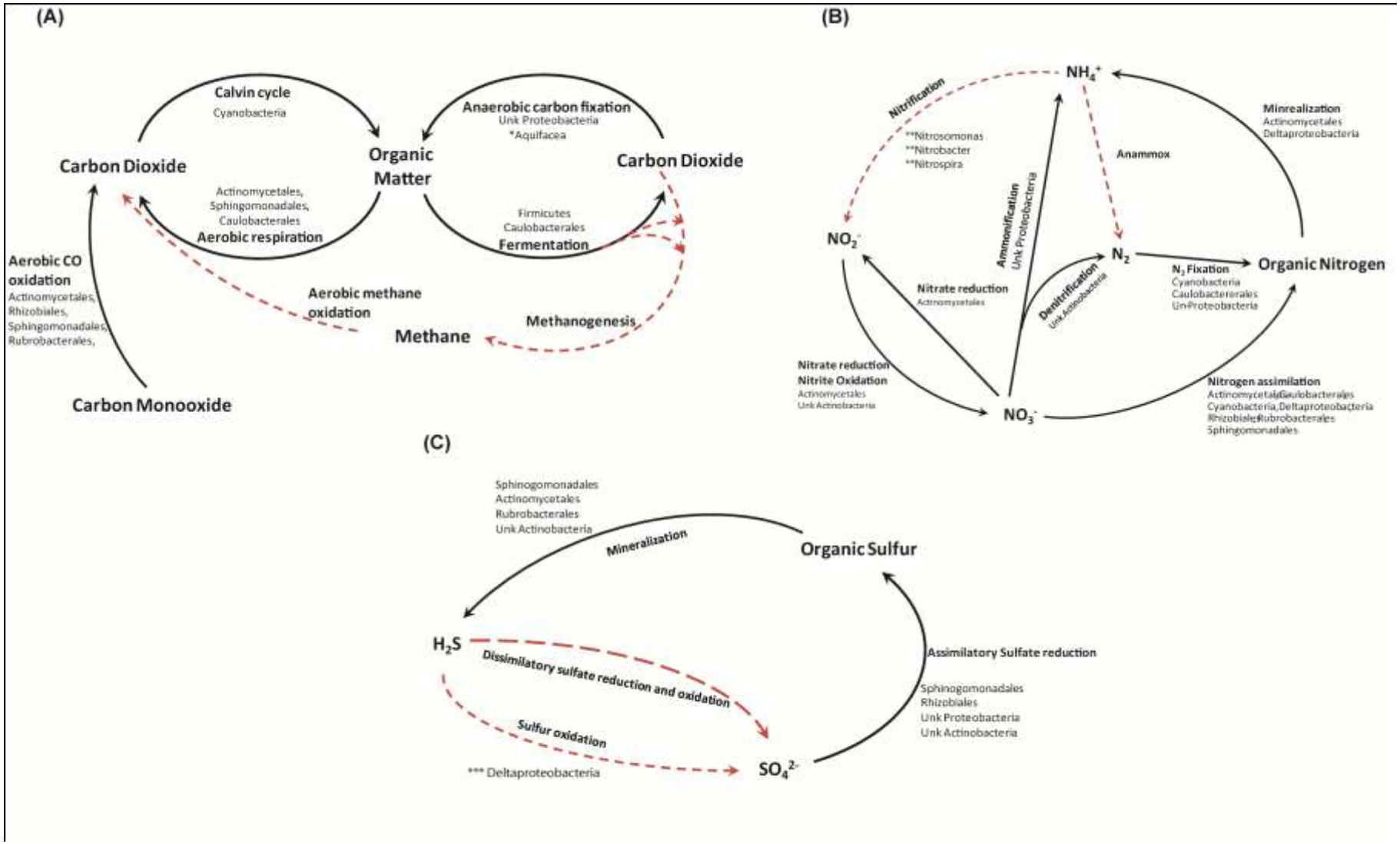
proteobacterial phototrophs, such as Chloroflexi, despite the identification of 220 contigs belonging to the genus *Chloroflexus* (known for its phototrophic metabolism).

KEGG pathway analyses of the key photosynthetic enzymes chlorophyll synthase (*chlG*) and bacteriochlorophyll synthase (*bchG*) showed sequences with homology to unclassified Cyanobacteria and Proteobacteria (genera *Methylobacterium*, *Rhodopseudomonas* and *Brevundimonas*). Interestingly, Methylobacteria have been previously identified as widespread colonists of hypolithons in both the Atacama and Namib Deserts (15, 23). MCR data and analysis of functional genes showed the presence of the complete Calvin-Benson cycle attributed to the phyla Cyanobacteria and Proteobacteria (**Fig. 4** and **Fig.5**).

We identified subunits of the gene *acl* (one copy of *acl*A and two copies of *acl*B), which encode the key enzyme (ATP citrate lyase), required for the reductive TCA cycle (rTCA cycle) were all assigned to phylum Aquificae (**Fig 5**). We could not identify the other two key genes (oxoglutarate synthase and fumarate reductase) involved in this cycle. However, these *acl* genes have only been reported from prokaryotes (using the rTCA cycle) and can thus be considered as 'indicator genes' for this pathway (36, 37). We suggest that the Namib Desert hypolithic community may harbor novel taxa affiliated to the phylum Aquificae which may have the capacity to drive anaerobic carbon fixation (**FIG 5**). The proposed capacity for both aerobic and anaerobic carbon fixation in hot desert soil communities may be a consequence of the limited C in these systems (38).

ORFs with homology to genes encoding formyltetrahydrofolate synthetase (FTHFS), a key enzyme in the Wood-Ljungdahl anaerobic acetogenesis pathway (17), were related to those of Actinobacteria, Proteobacteria, Gemmatimonadetes and Firmicutes. The key enzyme

FIGURE 5. Schematic representation of the biogeochemical cycling pathways (based on the analysis of marker genes described by Llorens-Mares et al., 2015) **(A)** Carbon cycling; *Aquifacea potential anaerobic carbon fixation step based on the presence of key enzyme ATP citrate lyase in the phylum **(B)** Nitrogen cycling; **Potential nitrifying bacteria contigs were found in the metagenome but genes for the nitrification were not identified **(C)** Sulfur cycling; ***Deltaproteobacteria *soxB* marker gene for the sulfur oxidation. The dotted lines are representing the absence of the marker genes in the metagenome.



responsible for methanogenesis (methyl-coenzyme M reductase (*mcr*A)) was not detected in any of the archaeal contigs, possibly due to the low abundance of methanogens in the largely aerobic desert soils (4, 39).

The extent of carbon fixation by members of the Proteobacteria, which are ubiquitous soil colonists, may have previously been underestimated in hypoliths. Proteobacteria possess the capacity for anoxigenic photosynthesis, but this contribution to C fixation is often largely ignored in comparison to cyanobacterial oxygenic photosynthesis (17). In cold desert systems, Cyanobacteria and Proteobacteria both appear to drive carbon fixation (17), and the identification of the relevant genes in the Namib Desert hypoliths metagenome suggest that similar processes may occur in hot deserts soils.

Genes for the heterotrophic utilization of complex carbohydrates (such as starch, cellulose, pectin and xylan) were largely associated with Actinobacteria (*Rubrobacterales*) (**Supplementary Materials Table 5**). Aromatic compound degradation genes were also identified: genes encoding the *ortho*- and *meta*-catechol ring cleavage enzymes (catechol-1, 2-dioxygenase and catechol-2, 3-dioxygenase) with similarity to those of the order *Actinomycetales*, unclassified Actinobacteria and unclassified Proteobacteria were identified (**Fig. 4**). This suggests that Actinomycetales in the hypolithic consortia may play a key role in detoxification of naturally occurring aromatic organics (40).

Nitrogen fixation and metabolism

Hyperarid desert environments are typically nitrogen limited (1), thereby enhancing the importance of diazotrophic microorganisms. Surprisingly, the hypolith metagenome sequence dataset showed very few *nif*H genes, encoding the first and rate-limiting step in the nitrogen

cycle (41). MAPLE server analysis yielded no *nif*H genes, while only one *nifH* gene variant, belonging to the phylum Cyanobacteria, was identified via MEGAN analysis. However a *hmmer* search performed against the pre-aligned *nifH* gene database (42) yielded at least five copies of the *nif*H gene, belonging to the phyla Cyanobacteria and Proteobacteria (Alphaproteobacteria and unclassified Proteobacteria) **(FIG 5B and Supplementary Materials Table S6).** This finding is congruent to previous studies, which have shown that heterocystous cyanobacteria were largely responsible for nitrogen fixation in depauperate edaphic systems (43).

A recent study reported the presence of ammonia-oxidizing bacteria in semi-arid soils (44). However, genes implicated in nitrification (ammonia monooxygenase (*amo*)) could not be detected in our metagenomic contigs. It has been suggested that the relative abundance of these genes may be related to 'rain events' (45). Our samples were collected during the late summer, following a period of months with zero precipitation, which may explain the absence of these genes from our metagenome and suggests that nitrification processes may be severely constrained for extended periods in hyper-arid soils. Genes for nitrate reduction and nitrite oxidation (*nar*GH/*nxr*AB) were identified and showed homology to those previously identified from Actinomycetales and unclassified Actinobacteria (**FIG 5B**). We also found signatures for genes implicated in denitrification (*nor*B), primarily affiliated to members of the phylum Actinobacteria **(FIG 5B)**. Nitrate reduction (*napA*) and ammonification (*nrfA*) genes were mostly affiliated to Actinomycetales and unclassified Proteobacteria, respectively (**FIG 5B**). We also identified the capacity for nitrogen (and ammonia) assimilation, based on the presence of marker genes such as glutamate synthase (*gltA* and *gltB*), assimilatory nitrate reductase (*nas*B) and glutamine synthase (*glnA*)). These genes were linked to a wide range of taxa, including members of the Actinomycetales, Rubrobacterales, Cyanobacteria,

unclassified Actinobacteria, Caulobacterales, Deltaproteobacteria, Acidobacteria, and Firmicutes (**FIG 5B**). However, genes for the anaerobic ammonium oxidation (Anammox) pathway, which converts ammonia directly into free nitrogen, were not identified. This finding is in contrast to other terrestrial systems (46, 47), particularly nutrient rich agricultural soils. *AmoA*-containing microorganisms are more common in aquatic or marine habitats, but the *amoA* gene also been identified in hypersaline microbial mats associated with desert springs (48). Genes for anammox have also been detected in the Antarctic hypolith microbial communities (17), which are known to have higher water contents than the largely aerobic 'Dry Valley' desert soils (49). The absence of anammox genes in the Namib hypolith metagenome may also reflect the limited capacity for anaerobic niches in hot desert soils. ~~It has also been shown that anammox rates in biological soil crusts (BSCs) from the Colorado Plateau were below detection rates (50).~~

The combined results from this analysis suggest that N cycling processes may be severely truncated in Namib Desert hyperarid soil niche communities. Denitrification rates in biological soil crusts have also been found to be low, despite the availability of NO_3^- in desert soils (51). Based on the genetic capacity for diazotrophy, we speculate that hypolithons may have a similar role in hyperarid desert systems. In low nitrogen availability environments such as deserts, nitrification is probably restricted to a limited number of low abundance taxa (52). A limited number of contigs assigned to nitrifying taxa such as *Nitrosomonas* (56 contigs/44702 bases), *Nitrobacter* (65 contigs/49915 bases) and *Nitrospira* (80 contigs/66667 bases) were identified, but the low number of sequences implicated in nitrification processes supports a view that these communities harbor a low genetic capacity for both nitrification and denitrification.

Phosphorus and sulfur metabolism

Biologically available phosphorus (P) in soils is mainly derived from rock weathering or from the decomposition of organic matter (53). In deserts, phosphorus-solubilizing bacteria (PSBs) release phosphorus from soil as orthophosphate anions (54). Gluconic acid and 2-ketogluconic acid biosynthesis in the periplasm of Gram-negative bacteria is known to be important for phosphate solubilization activity in soils (55). Gluconic acid biosynthesis is mainly carried out by the enzyme glucose dehydrogenase (GCD) in the presence of the cofactor pyrroloquinoline quinone (PQQ) (56). We identified copies of the *gcd* gene, with homology to members of the orders Rhizobiales, Solibacteriales and Xanthomonadales, and Proteobacteria and Bacteroidetes phyla, respectively, and suggest that these bacteria might be involved in phosphate solubilization in Namib Desert soil communities.

Phosphonates are an alternate source of phosphorus for microorganisms in desert environments, and are produced by protozoa, flagellates, coelenterates, mollusks, fungi and some bacteria (including Actinobacteria, *Pseudomonas* and *Bacillus*) (57). Although phosphonates are widely available in the environment, only microorganisms have the ability to degrade these compounds (58). We identified *phn* genes in the hypolith metagenome which may be implicated in the utilization of alkylphosphonate and phosphonates, ascribed to a wide range of taxa including Alphaproteobacteria (order Rhizobiales, Sphingomonadales and unclassified Alphaproteobacteria), Betaproteobacteria, Gammaproteobacteria and unknown Proteobacteria, Firmicutes, Chloroflexi, Planctomycetes, Cyanobacteria and Actinobacteria (Rubrobacterales, Actinomycetales and unclassified Actinobacteria). We suggest that the presence of diverse *phn* genes in the hypolithon indicates that bacterial utilization of phosphate from phosphonates and alkylphosphonates may be a key factor in 'P' turnover.

Another important and very common enzyme in phosphate metabolism is alkaline phosphatase ("Alp"), involved in the release of inorganic phosphate (Pi) from the both small and polymeric organic substrates including DNA and proteins (59). 'Alp' genes in this metagenome were associated with Rhizobiales, Caulobacterales, Sphingomonadales, Cyanobacteria, Chloroflexi and Firmicutes, all of which are known to play a role in plant P nutrition (60).

Genes for assimilatory sulfate reduction (*cys*C, *cys*N and *cys*D) were found in the Rhizobiales, Sphingomonadales, unclassified Proteobacteria and unclassified Actinobacteria. Genes for the mineralization of organic sulfur compounds were detected (**Supplementary Materials Fig S3**), with high homology to those of Actinobacteria (Actinomycetales, Rubrobacterales) and Sphinogomonadales (**Fig. 5C**). Although Namib Desert soils are SO_4^{2-} rich (64), genes for the anaerobic process of dissimilatory sulfate reduction and sulfide oxidation (*apr*A, *apr*B and *dsr*A) were not detected in the metagenomic contigs. However, using a conserved domain search (CDD), we identified one partial *sox*B gene assigned to Deltaproteobacteria and *sox*YZ genes in a contig assigned to Alphaproteobacteria (genus *Methylobacterium*). The Sox enzyme system has four principal complexes (*sox*XA, *sox*YZ, *sox*B and *sox*CD) encoding enzymes which catalyze the oxidation of hydrogen sulfite, thiosulfate, elemental sulfur and sulfite to sulfur intermediates or sulfate (61). The *soxB* gene is typically used as a marker gene for the sox system in the environmental bacteria (62). Sox enzymes are commonly associated with the facultative chemolithotrophic Alphaproteobacteria and the *soxB* gene has been found in the chemolithotrophic *Thiobacillus*-like Betaproteobacteria in agriculture soil (63). In the Namib Desert soils, high sulphate concentrations (3242.5 mg/kg) (64) and the presence of SOX system in the hypolith metagenome is suggestive of the presence of chemolithoautotrophic metabolism.

Conclusion

Metagenome sequence data can be validly used to assess the functional capacity of microorganisms in poorly studied environments (67). The analysis of metagenome sequence data from Namib desert hypolithic extracts has provided an expanded overview of the taxonomic and functional diversity of hypolithic microbial communities. Our analysis has shown that these communities are predominantly bacterial, but provides evidence of the presence of archaea and eukaryotes, albeit in much lower proportions. While we identified viral sequences affiliated to Caudovirales, Phycodnaviridae and Mimiviridae, the influence of viruses on the diversity of hypolithic systems remains unknown and complementary studies focused on this particular group are urgently required. Our data analyses also provide evidence of novel and unclassified taxa predominantly affiliated to Actinobacteria, Proteobacteria and Cyanobacteria. The analysis of functional gene diversity has implicated a large diversity of genes affiliated with these taxa in primary productivity, with members of the Proteobacteria and Actinobacteria potentially implicated in chemolithotrophic metabolism (P and S) in the desert environment. Overall, our data support the concept that Actinobacteria may be significant in driving productivity in soils, as indicated by the presence of numerous genes and modules implicated in heterotrophic carbon utilization, aromatic compound degradation and (to a much lesser extent) N cycling.

Edaphic ecosystems are key elements of climate-feedback models, due to their extensive capacity for the release and absorption of greenhouse gases (65). Cryptic niches, such as hypoliths, constitute substantial components of desert edaphic ecosystems (1, 30) and may be important drivers of gas exchange and geochemical cycling processes in desert soil ecosystems. Our data suggest that hypolithic communities have a high capacity for C

fixation, as evidenced by the substantial presence of key photosynthetic genes. In contrast, our metagenome sequence analyses suggest a severely limited capacity for N cycling. Given the known interplay between the C and N cycle (66), it is uncertain what the potential imbalance between C and N turnover processes in hypolithic communities means in terms of 'system stability' and ecosystem services.

Materials and Methods

Sequencing and assembly of the hypolith metagenome

Hypolith samples (n=50) were collected from the Namib Desert (S 23°32.031', E 15°01.813') in April 2010 (11). Samples were first processed for the isolation of total DNA and purified metagenomic DNA samples pooled for sequencing. Sequencing of metagenomic DNA was carried out with Illumina Hiseq-2000 using paired-end technology (2 x 101 bases). The metagenomic DNA was sheared into fragments of 300 bases and recovered from agarose gels. Adapters were ligated to the ends of the DNA fragments for bridge amplification and sequencing. The short paired-end reads were used to assess the quality of sequencing data using an in-house custom python script. The reads having ambiguous base (N) and average quality score less than 25 were removed using a custom python script. Assembly of the contigs was performed by Velvet v1.2.10 at hash length (k) 51 (68).

Bioinformatic analysis of metagenome

Metagenomic data were used for all taxonomic and functional gene analyses. ORFs were predicted from the contigs using the program *MetaGeneMark* (69). First, high quality reads were used for the taxonomic assessment by screening for small subunit (SSU) rRNAs with Metaxa2 (19) and for phylogenetic marker genes with metagenomic phylogenetic analysis

(*MetaPhAn*) software (20). Metaxa2 software extracts the SSU sequences from larger sequence datasets and assigns them for archaeal, bacterial, nuclear eukaryote, mitochondrial or chloroplast origins while *MetaPhlAn* predefines unique clade-specific marker genes as species-specific name tags (20). Next, assembled contigs longer than 500 bases were further selected for the binning process using *MyTaxa* (27). Functional annotation of ORFs was based on KEGG pathways and SEED subsystems using a *blastp* search (E-value cutoff at 1e-5) against the NCBI refseq protein database: results were further analysed using MEGAN v5.0.3 (70). KEGG modules were analysed by the Metabolic And Physiological Potential Evaluator (MAPLE) web server (33). The Module Completion Ratio (MCR) for each phylum was calculated using the bi-directional best hit (BBH) algorithm. Venn diagrams were computed by analysing the 100% complete KEGG pathway modules for the three most dominant phyla. Marker genes for the analysis of the Carbon, Nitrogen and Sulfur metabolism were selected and analysed as described by the Llorens-Mares (71).

Nucleotide accession number: The high quality paired end short reads were deposited at the NCBI under the Bioproject ID: PRJNA290687 and SRA accession number is SRR2124832.

Acknowledgements

This study would not have been possible without the financial support from the National Research Foundation of South Africa (SV, LG, TPM, DAC), the University of Pretoria Genomics Research Institute (TPM, DAC) and the Centre for High Performance Computing, Cape Town, South Africa for support computational resources.

References

1. **Pointing SB, Belnap J.** 2012. Microbial colonization and controls in dryland systems. Nat Rev Microbiol **10:**551-562.
2. **Trumper K, Ravilious C, Dickson B.** 2008. Carbon in drylands: desertification, climate change and carbon finance. A UNEP-UNDPUNCCD Technical Note for Discussions at CRIC 7 **Istanbul, Turkey:**1-12.
3. **Chan YK, Lacap DC, Lau MCY, Ha KY, Warren-Rhodes KA, Cockell CS, Cowan DA, McKay CP, Pointing SB.** 2012. Hypolithic microbial communities: between a rock and a hard place. Environ Microbiol **14:**2272-2282.
4. **Makhalanyane TP, Valverde A, Gunnigle E, Frossard A, Ramond J-B, Cowan DA.** 2015. Microbial ecology of hot desert edaphic systems. FEMS Microbiol Rev doi:10.1093/femsre/fuu011.
5. **Warren-Rhodes KA, Rhodes KL, Boyle LN, Pointing SB, Chen Y, Liu S, Zhuo P, McKay CP.** 2007. Cyanobacterial ecology across environmental gradients and spatial scales in China's hot and cold deserts. FEMS Microbiol Ecol **61:**470-482.
6. **Warren-Rhodes KA, Rhodes KL, Pointing SB, Ewing SA, Lacap DC, Gomez-Silva B, Amundson R, Friedmann EI, McKay CP.** 2006. Hypolithic cyanobacteria, dry limit of photosynthesis, and microbial ecology in the hyperarid Atacama Desert. Microb Ecol **52:**389-398.
7. **Cowan DA, Pointing SB, Stevens MI, Cary SC, Stomeo F, Tuffin IM.** 2011. Distribution and abiotic influences on hypolithic microbial communities in an Antarctic Dry Valley. Polar Biol **34:**307-311.
8. **Schlesinger WH, Pippen JS, Wallenstein MD, Hofmockel KS, Klepeis DM, Mahall BE.** 2003. Community composition and photosynthesis by photoautotrophs under quartz pebbles, southern Mojave Desert. Ecology **84:**3222-3231.
9. **Pointing SB, Chan Y, Lacap DC, Lau MC, Jurgens JA, Farrell RL.** 2009. Highly specialized microbial diversity in hyper-arid polar desert. P Natl Acad Sci USA **106:**19964-19969.
10. **Makhalanyane TP, Valverde A, Birkeland NK, Cary SC, Tuffin IM, Cowan DA.** 2013. Evidence for successional development in Antarctic hypolithic bacterial communities. ISME J **7:**2080-2090.
11. **Stomeo F, Valverde A, Pointing SB, McKay CP, Warren-Rhodes KA, Tuffin MI, Seely M, Cowan DA.** 2013. Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. Extremophiles **17:**329-337.
12. **Valverde A, Makhalanyane TP, Seely M, Cowan DA.** 2015. Cyanobacteria drive community composition and functionality in rock-soil interface communities. Mol Ecol doi:10.1111/mec.13068.
13. **Bahl J, Lau MC, Smith GJ, Vijaykrishna D, Cary SC, Lacap DC, Lee CK, Papke RT, Warren-Rhodes KA, Wong FK, McKay CP, Pointing SB.** 2011. Ancient origins determine global biogeography of hot and cold desert cyanobacteria. Nat Commun **2:**163.
14. **Smith JJ, Tow LA, Stafford W, Cary C, Cowan DA.** 2006. Bacterial diversity in three different Antarctic Cold Desert mineral soils. Microb Ecol **51:**413-421.
15. **Lacap DC, Warren-Rhodes KA, McKay CP, Pointing SB.** 2011. Cyanobacteria and chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. Extremophiles **15:**31-38.
16. **Cowan DA, Sohm JA, Makhalanyane TP, Capone DG, Green TGA, Cary SC, Tuffin IM.** 2011. Hypolithic communities: important nitrogen sources in Antarctic desert soils. Environ Microbiol Rep **3:**581-586.

17. **Chan Y, Van Nostrand JD, Zhou J, Pointing SB, Farrell RL.** 2013. Functional ecology of an Antarctic Dry Valley. P Natl Acad Sci USA **110:**8990-8995.
18. **Henschel JR, Lancaster N.** 2013. Gobabeb-50 years of Namib Desert research. JArid Environ **93:**1-6.
19. **Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DG, Nilsson RH.** 2015. Metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. Mol Ecol Resour doi:10.1111/1755-0998.12399.
20. **Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C.** 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods **9:**811-814.
21. **Wong FK, Lacap DC, Lau MC, Aitchison JC, Cowan DA, Pointing SB.** 2010. Hypolithic microbial community of quartz pavement in the high-altitude tundra of central Tibet. Microb Ecol **60:**730-739.
22. **Wood SA, Rueckert A, Cowan DA, Cary SC.** 2008. Sources of edaphic cyanobacterial diversity in the Dry Valleys of Eastern Antarctica. ISME J **2:**308-320.
23. **Makhalanyane TP, Valverde A, Lacap DC, Pointing SB, Tuffin MI, Cowan DA.** 2013. Evidence of species recruitment and development of hot desert hypolithic communities. Environ Microbiol Rep **5:**219-224.
24. **Pointing SB, Warren-Rhodes KA, Lacap DC, Rhodes KL, McKay CP.** 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.
25. **Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E.** 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res **21:**494-504.
26. **Prakash T, Taylor TD.** 2012. Functional assignment of metagenomic data: challenges and applications. Brief Bioinform **13:**711-727.
27. **Luo C, Rodriguez RL, Konstantinidis KT.** 2014. MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res **42:**e73.
28. **Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, Wommack KE.** 2008. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. Res Microbiol **159:**349-357.
29. **Weinbauer MG, Rassoulzadegan F.** 2004. Are viruses driving microbial diversification and diversity? Environ Microbiol **6:**1-11.
30. **Makhalanyane TP, Valverde A, Gunnigle E, Frossard A, Ramond JB, Cowan DA.** 2015. Microbial ecology of hot desert edaphic systems. FEMS Microbiol Rev doi:10.1093/femsre/fuu011.
31. **Adriaenssens EM, Van Zyl L, De Maayer P, Rubagotti E, Rybicki E, Tuffin M, Cowan DA.** 2014. Metagenomic analysis of the viral community in Namib Desert hypoliths. Environ Microbiol doi:10.1111/1462-2920.12528.
32. **Thurber RV.** 2011. Methods in Viral Metagenomics. Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats **2:**15.
33. **Takami H, Taniguchi T, Moriya Y, Kuwahara T, Kanehisa M, Goto S.** 2012. Evaluation method for the potential functionome harbored in the genome and metagenome. BMC Genomics **13:**699.
34. **Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le**

Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Jian M, Zhou Y, Li Y, Zhang X, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464:**59-65.

35. **Xu Z, Hansen MA, Hansen LH, Jacquiod S, Sorensen SJ.** 2014. Bioinformatic approaches reveal metagenomic characterization of soil microbial community. PLoS One **9:**e93445.
36. **Hugler M, Huber H, Molyneaux SJ, Vetriani C, Sievert SM.** 2007. Autotrophic CO2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: evidence for two ways of citrate cleavage. Environ Microbiol **9:**81-92.
37. **Fuchs G.** 1989. Alternative pathways of autotrophic CO2 fixation. Autotrophic bacteria**:**365-382. *In* H. G. Schlegel (ed.), Biology of autotrophic bacteria. Science Tech, Madison, WI.
38. **Novis PM, Whitehead D, Gregorich EG, Hunt JE, Sparrow AD, Hopkins DW, Elberling B, Greenfield LG.** 2007. Annual carbon fixation in terrestrial populations of Nostoc commune (Cyanobacteria) from an Antarctic dry valley is driven by temperature regime. Glob Change Biol **13:**1224-1237.
39. **Angel R, Soares MI, Ungar ED, Gillor O.** 2010. Biogeography of soil archaea and bacteria along a steep precipitation gradient. ISME J **4:**553-563.
40. **Packter NM.** 1992. Microbial degradation of natural products: Edited by G Winkelmann. pp 420. VCH, Weinheim, Germany. 1992. £93: ISBN 3-527-28354-4. Biochem Educ **20:**191-192.
41. **Dixon R, Kahn D.** 2004. Genetic regulation of biological nitrogen fixation. Nat Rev Microbiol **2:**621-631.
42. **Gaby JC, Buckley DH.** 2011. A global census of nitrogenase diversity. Environ Microbiol **13:**1790-1799.
43. **Yeager CM, Kornosky JL, Housman DC, Grote EE, Belnap J, Kuske CR.** 2004. Diazotrophic community structure and function in two successional stages of biological soil crusts from the Colorado Plateau and Chihuahuan Desert. Appl Environ Microb **70:**973-983.
44. **Banning NC, Maccarone LD, Fisk LM, Murphy DV.** 2015. Ammonia-oxidising bacteria not archaea dominate nitrification activity in semi-arid agricultural soil. Sci Rep **5**.
45. **Orlando J, Alfaro M, Bravo L, Guevara R, Carú M.** 2010. Bacterial diversity and occurrence of ammonia-oxidizing bacteria in the Atacama Desert soil during a “desert bloom” event. Soil Biol Biochem **42:**1183-1188.
46. **Humbert S, Tarnawski S, Fromin N, Mallet MP, Aragno M, Zopfi J.** 2010. Molecular detection of anammox bacteria in terrestrial ecosystems: distribution and diversity. ISME J **4:**450-454.
47. **Zhu G, Wang S, Wang Y, Wang C, Risgaard-Petersen N, Jetten MS, Yin C.** 2011. Anaerobic ammonia oxidation in a fertilized paddy soil. ISME J **5:**1905-1912.
48. **Abed RM, De Beer D, Stief P.** 2015. Functional-structural analysis of nitrogen-cycle bacteria in a hypersaline mat from the Omani desert. Geomicrobiol J **32:**119-129.
49. **Cowan DA, Pointing SB, Stevens MI, Craig Cary S, Stomeo F, Tuffin IM.** 2010. Distribution and abiotic influences on hypolithic microbial communities in an Antarctic Dry Valley. Polar Biol **34:**307-311.
50. **Strauss SL, Day TA, Garcia-Pichel F.** 2012. Nitrogen cycling in desert biological soil crusts across biogeographic regions in the Southwestern United States.

Biogeochemistry **108:**171-182.
51. **Johnson SL, Neuer S, Garcia-Pichel F.** 2007. Export of nitrogenous compounds due to incomplete cycling within biological soil crusts of arid lands. Environ Microbiol **9:**680-689.
52. **Bothe H, Jost G, Schloter M, Ward BB, Witzel K.** 2000. Molecular analysis of ammonia oxidation and denitrification in natural environments. FEMS Microbiol Rev **24:**673-690.
53. **Delgado-Baquerizo M, Maestre FT, Gallardol A, Bowker MA, Wallenstein MD, Quero JL, Ochoa V, Gozalo B, Garcia-Gomez M, Soliveres S, Garcia-Palacios P, Berdugo M, Valencia E, Escolar C, Arredondol T, Barraza-Zepeda C, Bran D, Carreiral JA, Chaiebll M, Conceicao AA, Derak M, Eldridge DL, Escudero A, Espinosa CI, Gaitan J, Gatica MG, Gomez-Gonzalez S, Guzman E, Gutierrez JR, Florentino A, Hepper E, Hernandez RM, Huber-Sannwald E, Jankju M, Liu JS, Mau RL, Miriti M, Monerris J, Naseri K, Noumi Z, Polo V, Prina A, Pucheta E, Ramirez E, Ramirez-Collantes DA, Romao R, Tighe M, Torres D, Torres-Diaz C, Ungar ED, et al.** 2013. Decoupling of soil nutrient cycles as a function of aridity in global drylands. Nature **502:**672-676.
54. **Rodriguez H, Fraga R.** 1999. Phosphate solubilizing bacteria and their role in plant growth promotion. Biotechnol Adv **17:**319-339.
55. **Chhabra S, Brazil D, Morrissey J, Burke JI, O'Gara F, D ND.** 2013. Characterization of mineral phosphate solubilization traits from a barley rhizosphere soil functional metagenome. Microbiology Open **2:**717-724.
56. **Shen YQ, Bonnot F, Imsand EM, RoseFigura JM, Sjolander K, Klinman JP.** 2012. Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone. Biochemistry **51:**2265-2275.
57. **White AK, Metcalf WW.** 2007. Microbial metabolism of reduced phosphorus compounds. Annu Rev Microbiol **61:**379-400.
58. **Kononova SV, Nesmeyanova MA.** 2002. Phosphonates and their degradation by microorganisms. Biochemistry-Moscow+ **67:**184-195.
59. **Kobori H, Sullivan CW, Shizuya H.** 1984. Heat-labile alkaline phosphatase from Antarctic bacteria: Rapid 5' end-labeling of nucleic acids. Proc Natl Acad Sci USA **81:**6691-6695.
60. **Gyaneshwar P, Kumar GN, Parekh L, Poole P.** 2002. Role of soil microorganisms in improving P nutrition of plants, p 133-143. *In* Food Security in Nutrient-Stressed Environments: Exploiting Plants' Genetic Capabilities. Springer.
61. **Ghosh W, Dam B.** 2009. Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS Microbiol Rev **33:**999-1043.
62. **Headd B, Engel AS.** 2013. Evidence for niche partitioning revealed by the distribution of sulfur oxidation genes collected from areas of a terrestrial sulfidic spring with differing geochemical conditions. Appl Environ Microbiol **79:**1171-1182.
63. **Tourna M, Maclean P, Condron L, O'Callaghan M, Wakelin SA.** 2014. Links between sulphur oxidation and sulphur-oxidising bacteria abundance and diversity in soil microcosms based on soxB functional gene analysis. FEMS Microbiol Ecol **88:**538-549.
64. **Gombeer S, Ramond JB, Eckardt FD, Seely M, Cowan DA.** 2015. The influence of surface soil physicochemistry on the edaphic bacterial communities in contrasting terrain types of the Central Namib Desert. Geobiology **13:**494-505.
65. **Schimel DS, Braswell B, Holland EA, McKeown R, Ojima D, Painter TH, Parton WJ, Townsend AR.** 1994. Climatic, edaphic, and biotic controls over storage and

turnover of carbon in soils. Global Biogeochem Cycl **8:**279-293.
66. **Thornton PE, Lamarque JF, Rosenbloom NA, Mahowald NM.** 2007. Influence of carbon-nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability. Global biogeochemical cycles **21**.
67. **Cowan D, Ramond J, Makhalanyane T, De Maayer P.** 2015. Metagenomics of extreme environments. Curr Opin Microbiol **25:**97-102.
68. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:**821-829.
69. **Zhu W, Lomsadze A, Borodovsky M.** 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res **38:**e132.
70. **Huson DH, Weber N.** 2013. Microbial community analysis using MEGAN. Methods Enzymol **531:**465-485.
71. **Llorens-Mares T, Yooseph S, Goll J, Hoffman J, Vila-Costa M, Borrego CM, Dupont CL, Casamayor EO.** 2015. Connecting biodiversity and potential functional role in modern euxinic environments by microbial metagenomics. ISME J doi:10.1038/ismej.2014.254.

Supplementary Materials

Supplementary Materials Table S1: Sequencing and filtering result of the Namib hypolith metagenome

Total number of raw reads (paired end)	387093778
Total number of bases in raw reads (paired end)	39096471578
Total number of high quality reads (Paired reads)	305073640
Total number of high quality bases (Paired reads)	29960891804

Supplementary Materials Table S2: Assembly statistics of the Namib hypolith metagenome by Velvet

	Total Assembly	**Contigs longer than 500 bases**
Total Number of Sequences	2188786	253243
Total number of bases	634 Million	198 million
N50 of the assembly (bp)	327	767
Average Sequence length (bp)	289.73	785.45
GC (%)	62.62	60.43
Ns (%)	1.42	3.06

Supplementary Materials Table S3: KEGG Pathway modules exclusively found in the three most dominant phyla.

Actinobacteria	
M00033	Ectoine biosynthesis, aspartate => ectoine
M00038	Tryptophan metabolism, tryptophan => kynurenine => 2-aminomuconate
M00141	C1-unit interconversion, eukaryotes
M00530	Dissimilatory nitrate reduction, nitrate => ammonia
M00546	Purine degradation, xanthine => urea
M00551	Benzoate degradation, benzoate => catechol / methylbenzoate => methylcatechol
M00198	Putative sn-glycerol-phosphate transport system
M00205	N-Acetylglucosamine transport system
M00211	Putative ABC transport system
M00216	Multiple sugar transport system
M00224	Fluoroquinolones transport system
M00233	Glutamate transport system
M00267	PTS system, N-acetylglucosamine-specific II component
M00274	PTS system, mannitol-specific II component
M00342	Bacterial proteasome
M00491	Putative arabinogalactan oligomer transport system
M00603	Putative aldouronate transport system
M00443	SenX3-RegX3 (phosphate starvation response) two-component regulatory system
M00446	RstB-RstA two-component regulatory system
M00460	MprB-MprA (maintenance of persistent infection) two-component regulatory system
M00461	MtrB-MtrA (osmotic stress response) two-component regulatory system
M00473	UhpB-UhpA (hexose phosphates uptake) two-component regulatory system
M00475	BarA-UvrY (central carbon metabolism) two-component regulatory system
M00478	DegS-DegU (multicellular behavior control) two-component regulatory system
M00479	DesK-DesR (membrane lipid fluidity regulation) two-component regulatory system
M00481	LiaS-LiaR (cell wall stress response) two-component regulatory system
Proteobacteria	
M00045	Histidine degradation, histidine => N-formiminoglutamate => glutamate
M00063	CMP-KDO biosynthesis
M00064	ADP-L-glycero-D-manno-heptose biosynthesis
M00535	Isoleucine biosynthesis, pyruvate => 2-oxobutanoate
M00247	Putative ABC transport system
M00253	Sodium transport system
M00265	PTS system, glucose-specific II component
M00324	Dipeptide transport system
M00333	Type IV secretion system
M00335	Sec (secretion) system
M00339	RaxAB-RaxC type I secretion system
M00349	Microcin C transport system
M00597	Anoxygenic photosystem II
M00600	alpha-1,4-Digalacturonate transport system

M00607	Glycerol transport system
M00462	PrrB-PrrA (intracellular multiplication) two-component regulatory system
M00463	TrcS-TrcR two-component regulatory system
M00497	GlnL-GlnG (nitrogen regulation) two-component regulatory system
M00498	NtrY-NtrX (nitrogen regulation) two-component regulatory system
M00499	HydH-HydG (metal tolerance) two-component regulatory system
M00500	AtoS-AtoC (cPHB biosynthesis) two-component regulatory system
M00504	DctB-DctD (C4-dicarboxylate transport) two-component regulatory system
M00512	CckA-CtrA/CpdR (cell cycle control) two-component regulatory system
M00517	RpfC-RpfG (cell-to-cell signaling) two-component regulatory system
M00519	YesM-YesN two-component regulatory system
M00520	ChvG-ChvI (acidity sensing) two-component regulatory system
M00523	RegB-RegA (redox response) two-component regulatory system
Cyanobacteria	
M00034	Methionine salvage pathway
M00053	Pyrimidine deoxyribonuleotide biosynthesis, CDP/CTP => dCDP/dCTP,dTDP/dTTP
M00116	Menaquinone biosynthesis, chorismate => menaquinone
M00145	NAD(P)H:quinone oxidoreductase, chloroplasts and cyanobacteria
M00161	Photosystem II
M00163	Photosystem I
M00321	Bicarbonate transport system
M00322	Neutral amino acid transport system
M00323	Urea transport system
M00438	Nitrate/nitrite transport system
M00587	Arginine/lysine/histidine/glutamine transport system
M00465	ManS-ManR (manganese homeostasis) two-component regulatory system
M00466	NblS-NblR (photosynthesis) two-component regulatory system
M00467	SasA-RpaAB (circadian timing mediating) two-component regulatory system
M00506	CheA-CheYBV (chemotaxis) two-component regulatory system
M00508	PixL-PixGH (positive phototaxis) two-component regulatory system
M00509	WspE-WspRF (chemosensory) two-component regulatory system
M00510	Cph1-Rcp1 (light response) two-component regulatory system
Actinobacteria-Proteobacteria	
M00149	Succinate dehydrogenase, prokaryotes
M00151	Cytochrome bc1 complex respiratory unit
M00178	Ribosome, bacteria
M00191	Thiamine transport system
M00196	Multiple sugar transport system
M00197	Putative fructooligosaccharide transport system
M00208	Glycine betaine/proline transport system
M00215	D-Xylose transport system
M00218	Fructose transport system
M00220	Rhamnose transport system
M00238	D-Methionine transport system

M00255	Lipoprotein-releasing system
M00259	Heme transport system
M00273	PTS system, fructose-specific II component
M00280	PTS system, glucitol/sorbitol-specific II component
M00313	indolepyruvate:ferredoxin oxidoreductase
M00319	Manganese/zinc/iron transport system
M00439	Oligopeptide transport system
M00605	Glucose/mannose transport system
M00445	EnvZ-OmpR (osmotic stress response) two-component regulatory system
M00447	CpxA-CpxR (envelope stress response) two-component regulatory system
M00450	BaeS-BaeR (envelope stress response) two-component regulatory system
M00454	KdpD-KdpE (potassium transport) two-component regulatory system
M00471	NarX-NarL (nitrate respiration) two-component regulatory system
M00493	AlgZ-AlgR (alginate production) two-component regulatory system
M00515	FlrB-FlrC (polar flagellar synthesis) two-component regulatory system
M00524	FixL-FixJ (nitrogen fixation) two-component regulatory system
Acinobacteria-Cyanobacteria	
M00126	Tetrahydrofolate biosynthesis, GTP => THF
M00338	Cysteine biosynthesis, homocysteine + serine => cysteine
M00242	Zinc transport system
M00243	Manganese/iron transport system
M00245	Cobalt/nickel transport system
M00246	Nickel transport system
M00252	Lipooligosaccharide transport system
M00582	Energy-coupling factor transport system
M00601	Putative chitobiose transport system
Proteobacteria-Cyanobacteria	
M00118	Glutathione biosynthesis, glutamate => glutathione
M00165	Reductive pentose phosphate cycle (Calvin cycle)
M00167	Reductive pentose phosphate cycle, glyceraldehyde-3P => ribulose-5P
M00175	Nitrogen fixation, nitrogen => ammonia
M00193	Putative spermidine/putrescine transport system
M00232	General L-amino acid transport system
M00320	Lipopolysaccharide export system
M00452	CusS-CusR (copper tolerance) two-component regulatory system
M00457	TctE-TctD (tricarboxylic acid transport) two-component regulatory system
Actinobacteria-Proteobacteria-Cyanobacteria	
M00001	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate
M00002	Glycolysis, core module involving three-carbon compounds
M00003	Gluconeogenesis, oxaloacetate => fructose-6P
M00004	Pentose phosphate pathway (Pentose phosphate cycle)
M00005	PRPP biosynthesis, ribose 5P => PRPP
M00006	Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P
M00007	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P

M00010	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate
M00015	Proline biosynthesis, glutamate => proline
M00018	Threonine biosynthesis, aspartate => homoserine => threonine
M00019	Valine/isoleucine biosynthesis, pyruvate => valine / 2-oxobutanoate => isoleucine
M00021	Cysteine biosynthesis, serine => cysteine
M00022	Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate
M00023	Tryptophan biosynthesis, chorismate => tryptophan
M00026	Histidine biosynthesis, PRPP => histidine
M00028	Ornithine biosynthesis, glutamate => ornithine
M00035	Methionine degradation
M00048	Inosine monophosphate biosynthesis, PRPP + glutamine => IMP
M00049	Adenine ribonucleotide biosynthesis, IMP => ADP,ATP
M00050	Guanine ribonucleotide biosynthesis IMP => GDP,GTP
M00052	Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP
M00083	Fatty acid biosynthesis, elongation
M00086	beta-Oxidation, acyl-CoA synthesis
M00087	beta-Oxidation
M00091	Phosphatidylcholine (PC) biosynthesis, PE => PC
M00096	C5 isoprenoid biosynthesis, non-mevalonate pathway
M00098	Acylglycerol degradation
M00115	NAD biosynthesis, aspartate => NAD
M00119	Pantothenate biosynthesis, valine/L-aspartate => pantothenate
M00120	Coenzyme A biosynthesis, pantothenate => CoA
M00123	Biotin biosynthesis, pimeloyl-ACP/CoA => biotin
M00125	Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD
M00127	Thiamine biosynthesis, AIR => thiamine-P/thiamine-2P
M00133	Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine
M00135	GABA biosynthesis, eukaryotes, putrescine => GABA
M00166	Reductive pentose phosphate cycle, ribulose-5P => glyceraldehyde-3P
M00168	CAM (Crassulacean acid metabolism), dark
M00169	CAM (Crassulacean acid metabolism), light
M00176	Assimilatory sulfate reduction, sulfate => H2S
M00307	Pyruvate oxidation, pyruvate => acetyl-CoA
M00364	C10-C20 isoprenoid biosynthesis, bacteria
M00432	Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate
M00527	Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine
M00531	Assimilatory nitrate reduction, nitrate => ammonia
M00549	Nucleotide sugar biosynthesis, glucose => UDP-glucose
M00554	Nucleotide sugar biosynthesis, galactose => UDP-galactose
M00555	Betaine biosynthesis, choline => betaine
M00565	Trehalose biosynthesis, D-glucose 1P => trehalose
M00570	Isoleucine biosynthesis, threonine => 2-oxobutanoate => isoleucine
M00572	Pimeloyl-ACP biosynthesis, BioC-BioH pathway, malonyl-ACP => pimeloyl-ACP
M00153	Cytochrome d ubiquinol oxidase

M00155	Cytochrome c oxidase, prokaryotes
M00157	F-type ATPase, prokaryotes and chloroplasts
M00183	RNA polymerase, bacteria
M00185	Sulfate transport system
M00188	NitT/TauT family transport system
M00189	Molybdate transport system
M00190	Iron(III) transport system
M00207	Putative multiple sugar transport system
M00209	Osmoprotectant transport system
M00210	Putative ABC transport system
M00212	Ribose transport system
M00221	Putative simple sugar transport system
M00222	Phosphate transport system
M00223	Phosphonate transport system
M00236	Putative polar amino acid transport system
M00237	Branched-chain amino acid transport system
M00239	Peptides/nickel transport system
M00240	Iron complex transport system
M00244	Putative zinc/manganese transport system
M00248	Putative antibiotic transport system
M00250	Lipopolysaccharide transport system
M00254	ABC-2 type transport system
M00256	Cell division transport system
M00258	Putative ABC transport system
M00260	DNA polymerase III complex, bacteria
M00299	Spermidine/putrescine transport system
M00336	Twin-arginine translocation (Tat) system
M00436	Sulfonate transport system
M00359	Aminoacyl-tRNA biosynthesis, eukaryotes
M00360	Aminoacyl-tRNA biosynthesis, prokaryotes
M00362	Nucleotide sugar biosynthesis, prokaryotes
M00434	PhoR-PhoB (phosphate starvation response) two-component regulatory system

Supplementary Materials Table S4: Distribution of family *Caudovirales* in the Namib metagenome (comparison with current and previous study)

Reference	**Unclassified Caudovirales (%)**	**Myoviridae (%)**	**Siphoviridae (%)**	**Podoviridae (%)**
Adrianssens et. al., 2014	15.71	19.83	58.36	6.07
This Study	11.43	31.94	43.83	12.78

Supplementary Materials Table S5: Occurrence of complex carbohydrates degradation genes in the metagenome

Bacterial Taxa	Number of Genes Present in each taxa				
	Arabinofuranosidase	**Endoglucanase**	**Xylose Isomerase**	**Mannosidase**	**Exoglucanase**
Actinobacteria					
Beutenbergia cavernae	2	Nd	Nd	Nd	Nd
Kineococcus radiotolerans	2	Nd	Nd	Nd	Nd
Actinoplanes sp. N902-109	1	Nd	Nd	Nd	Nd
Kribbella flavida	2	Nd	1	3	Nd
Micromonospora sp.	1	Nd	Nd	1	Nd
Catenulispora acidiphila	Nd	1	Nd	Nd	Nd
Amycoplastopsis mediterranei	Nd	1	Nd	1	1
Cellulomonas sp.	Nd	1	Nd	Nd	Nd
Pseudonocardia dioxanivorans	Nd	Nd	2	2	Nd
Actinosynnema mirum	Nd	Nd	Nd	Nd	1
Rubrobacter xylanophilus	Nd	23	5	2	Nd
Conexibacter woesei	Nd	Nd	2	2	Nd
Unclassified Actinobacteria	6	9	8	5	1
Bacteroidetes/Chlorobi					
Flavobacterium johnsoniae	1	Nd	Nd	Nd	Nd
Unclassified Bacteroidetes/Chlorobi	1	Nd	Nd	1	Nd
Chlamydiae/ Verrucomicrobia					
Opitutus terrae	Nd	Nd	Nd	2	Nd
Chloroflexi					
Rosiflexus sp.	Nd	Nd	Nd	1	Nd
Roseiflexus castenholzii	Nd	Nd	Nd	2	Nd
Unclassified Chloroflexi	Nd	Nd	2	1	Nd
Cyanobacteria					
Microleus sp. PCC 7113	Nd	1	Nd	Nd	Nd
Fibrobacteres/Acidobacteria					
Candidatus Koribacter versatilis	Nd	Nd	Nd	Nd	Nd
Unclassified Fibrobacter/Acidobacteria	2	Nd	Nd	1	Nd
Firmicutes					
Geobacillus sp.	1	Nd	Nd	Nd	Nd
Unclassified Firmicutes	1	Nd	Nd	1	Nd
Proteobacteria					
Brucella sp	1	Nd	1	1	Nd
Sinorhizobium meliloti	Nd	Nd	1	Nd	Nd
Novosphingobium sp.	Nd	Nd	Nd	1	Nd
Rhizobium tropici	Nd	Nd	Nd	1	Nd
Burkholderia glumae	1	Nd	Nd	Nd	Nd
Pseudoxantomonas sp.	1	Nd	Nd	Nd	Nd
*Xanthomonas*sp.	1	Nd	Nd	Nd	Nd
Pseudomonas fluorescens	Nd	Nd	1	1	Nd
Unclassified Rhizobiales	1	Nd	Nd	Nd	Nd
Unclassified Proteobacteria	Nd	Nd	1	Nd	Nd
Gemmatimonadetes					
Gemmatimonas aurantica	Nd	1	Nd	Nd	Nd
Planctomycetes					
Planctomyces sp.	Nd	Nd	1	Nd	Nd
Unclassified Planctomycetes	Nd	1	1	Nd	Nd
Sphingomonadales	Nd	Nd	1	Nd	Nd
Thermobaculum					
Thermobaculum terrenum	Nd	3	Nd	Nd	Nd
Thermotogae					
Thermotoga sp.	1	Nd	Nd	Nd	Nd

Note:
Nd= not detected

Supplementary Materials Table S6: Occurrence of Catechol *ortho* and *meta* cleavage genes in the metagenome

Bacteria Taxa	**Catechol 1,2 dioxygenase**	**Catechol 2,3 dioxygenase**
Actinobacteria		
Arthrobacter phenanthrenivorans	Nd	2
Arthrobacter sp.	Nd	1
Conexibacter woesei	Nd	1
Pseudonocardia dioxanivorans	2	1
Rhodobacter xylanophilus	Nd	1
Rhodococcus sp.	1	Nd
Streptomyces cattleya	Nd	1
Streptomyces sp.	Nd	1
Unclassified Actinobacteria	1	Nd
Proteobacteria		
Xnathobacter autotrophicus	1	Nd
Sinorhizobium fredii	Nd	1
Bradyrhizobium japonicum	Nd	1
Unclassified Proteobacteria	3	Md

Note:
Nd= not detected

Supplementary Materials Table S7: Genes identified using hmmer search against *nifH* nucleotide database.

Gene ID	**Evalue**	**Bit Score**	**Phylum**
gene_id_95870	2.10E-41	146	Cyanobacteria
gene_id_230526	6.30E-37	131.1	Cyanobacteria
gene_id_172577	1.30E-82	282.5	Cyanobacteria
gene_id_133822	1.40E-48	169.8	Proteobacteria
gene_id_116375	3.80E-39	138.5	Alphaproteobacteria

Supplementary Materials Table S8: Occurrence of the quinoprotein glucose dehydrogenase (involved in phosphate solubilization) in the metagenome

Bacterial Taxa	**Occurrence of Quinoprotein Glucose dehydrogenase**
Proteobacteria	
Sinorhizobium fredii	1
Xanthomonas euvesicatoria	1
Sphingomanadaceae	2
Unclassified Proteobacteria	3
Acidobacteria	
Candidatus Solibacter usitatus	3
Granulicells mallensis	1
Bacteroidetes	
Cyanophagaceae	1
Unclassified Bacteroidetes	3
Planctomycetes	
Planctomytaceae	1

Supplementary Materials table S9: Occurrence of organic sulfur assimilation pathway genes in the metagenome

Bacterial Taxa	Alkane Sulfonate metabolism	Taurine utilization	Utilization of glutathione as a sulfur source	Alkanesulfonates utilization
Actinobacteria				
Rubrobacter xylanophilus	15	Nd	25	1
Mycobacterium smegmatis	2	Nd	Nd	Nd
Mycobacterium sp.	1	Nd	Nd	Nd
Streptomyces violaceusniger	2	Nd	Nd	Nd
Streptomyces coelicolor	4	Nd	Nd	Nd
Amycolaptopsis mediterranei	2	Nd	Nd	Nd
Kribbella flavida	1	Nd	Nd	Nd
Saccharothrix espanaensis	1	Nd	Nd	1
Saccharomonospora viridis	1	Nd	Nd	Nd
Thermobispora bispora	1	Nd	Nd	1
Arthrobacter sp.	Nd	Nd	1	Nd
Rhodocossus sp.	Nd	Nd	1	Nd
Unclassified Actinobacteria	11	Nd	3	5
Proteobacteria				
Hyphomicrobium sp. MC1	1	Nd	Nd	Nd
Methylobacterium sp.	1	Nd	1	1
Chelativorans sp. BNC1	2	Nd	Nd	Nd
Caulobacter sp.	4	1	2	Nd
Burkholderia sp.	2	2	Nd	Nd
Azorhizobium caulinodans	1	Nd	Nd	Nd
Bradyrhizobium sp.	1	Nd	5	1
Brucella sp.	2	Nd	Nd	Nd
Rhodobacter spaeroides	2	Nd	Nd	Nd
Dinoroseobacter shibae	3	Nd	Nd	Nd
Rhodopseudomonas palustris	1	Nd	Nd	Nd
Caulobacter vibrioides	Nd	1	Nd	Nd
Agrobacterium	Nd	Nd	1	Nd
Pseudomonas putida	Nd	Nd	1	Nd
Mesorhizobium sp.	Nd	Nd	1	Nd
Xanthomonas compestris	Nd	Nd	1	Nd
Rhizobium leguminosarum	Nd	Nd	1	Nd
Unclassified Proteobacteria	16	Nd	13	2
Cyanobacteria				
Nostoc punctiforme	3	Nd	1	Nd
Gleocapsa sp. PCC 7428	1	Nd	Nd	Nd
Cyanothece sp. PCC 7822	1	Nd	Nd	1
Microcystis aeruginosa	1	Nd	Nd	Nd
Anabaena sp.	3	Nd	Nd	3
Cyanothece sp.	Nd	Nd	1	Nd
Unclassified Cyanobacteria	11	3	Nd	4
Firmicutes				
Bacillus coagulans	1	Nd	Nd	1
Bacteroidetes/Chlorobi				
Haliscomendobacter hydrossis	1	Nd	Nd	Nd
Parabacteroides distasonis	1	Nd	Nd	Nd
Echinicola vietnamensis	1	Nd	Nd	Nd
Dyadobacter fermentans	1	Nd	Nd	Nd
Unclassified Bacteroidetes/Chlorobi	1	Nd	Nd	Nd
Planctomycetes	1			
Rhodopirellula baltica		Nd	Nd	Nd
Gemmatimonadetes				
Gemmatimonas aurantica	Nd	Nd	1	Nd
Fibrobacter/Acidobacteria				
Candidatus Solibacter usitatus	Nd	Nd	3	Nd
Candidatus Chloroacidobacterium thermophilum	Nd	Nd	1	Nd
DeinocccusNdThermus				
Truepera radiovitrix	Nd	Nd	1	Nd

Note:
Nd= not detected

Supplementary Materials Figure S1. MEGAN classification of Namib contigs using refseq-virus database. Number of assigned contigs are written after the semicolon