Comparative metagenomics unveils functions and genome features of microbialite-associated communities along a depth gradient

Aurélien Saghaï, Yvan Zivanovic, David Moreira, Karim Benzerara, Paola Bertolino, Marie Ragon, Rosaluz Tavera, Ana Isabel López-Archilla, Purificación López-García

Published in final edited form as:
Environ Microbiol. 2016 December ; 18(12): 4990–5004. doi:10.1111/1462-2920.13456.

Comparative metagenomics unveils functions and genome features of microbialite-associated communities along a depth gradient

Aurélien Saghaï[1], **Yvan Zivanovic**[2], **David Moreira**[1], **Karim Benzerara**[3], **Paola Bertolino**[1], **Marie Ragon**[1], **Rosaluz Tavera**[4], **Ana Isabel López-Archilla**[5], and **Purificación López-García**[1,*]

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, Université Paris-Saclay, AgroParisTech, Orsay, France

[2]Institut de Biologie Intégrative de la Cellule, CNRS, Université Paris-Sud Orsay, Université Paris-Saclay, France

[3]Institut de Minéralogie et de Physique des Matériaux et de Cosmochimie, CNRS, Muséum National d'Histoire Naturelle, Université Pierre et Marie Curie, Sorbonne Universités, Paris, France

[4]Departamento de Ecología y Recursos Naturales, Universidad Nacional Autónoma de México, DF Mexico, Mexico

[5]Departamento de Ecología, Universidad Autónoma de Madrid, Madrid, Spain

Summary

Modern microbialites are often used as analogs of Precambrian stromatolites; therefore, studying the metabolic interplay within their associated microbial communities can help formulating hypotheses on their formation and long-term preservation within the fossil record. We performed a comparative metagenomic analysis of microbialite samples collected at two sites and along a depth gradient in Lake Alchichica (Mexico). The community structure inferred from single-copy gene family identification and long-contig (>10 kb) assignation, consistently with previous rRNA gene surveys, showed a wide prokaryotic diversity dominated by Alphaproteobacteria, Gammaproteobacteria, Cyanobacteria and Bacteroidetes, while eukaryotes were largely dominated by green algae or diatoms. Functional analyses based on RefSeq, COG and SEED assignations revealed the importance of housekeeping functions, with an overrepresentation of genes involved in carbohydrate metabolism, as compared to other metabolic capacities. The search for genes diagnostic of specific metabolic functions revealed the important involvement of Alphaproteobacteria in anoxygenic photosynthesis and sulfide oxidation, and Cyanobacteria in oxygenic photosynthesis and nitrogen fixation. Surprisingly, sulfate reduction appeared negligible. Comparative analyses suggested functional similarities among various microbial mat and

[*]For correspondence. Purificación López-García, Ecologie Systématique Evolution, CNRS UMR 8079, Université Paris-Sud, bâtiment 360, 91400 Orsay, France. puri.lopez@u-psud.fr; Tel. +33 169 154 991; Fax +33 169 154 697.

The authors declare no conflict of interest.

Europe PMC Funders Author Manuscripts

microbialite metagenomes as compared with soil or oceans, but showed differences in microbial processes among microbialite types linked to local environmental conditions.

Keywords

metagenomics; stromatolite; carbonate precipitation; anoxygenic photosynthesis; Chloroflexi; Cyanobacteria

Europe PMC Funders Author Manuscripts

Introduction

Microbialites are organosedimentary, most often carbonate-rich, structures formed under the influence of phylogenetically and functionally diverse microbial communities (Burne and Moore, 1987; Riding, 2000; Dupraz and Visscher, 2005). Their fabrics can be laminated (stromatolites) or clotted (thrombolites) (Dupraz *et al.*, 2009). Fossil stromatolites represent the most ancient macroscopic evidence of life on Earth (Tice and Lowe, 2004; Allwood *et al.*, 2006) and were abundant and widespread during the Proterozoic (0.5-2.5 Ga; Riding, 2006) before abruptly declining in the last 600 Ma (Awramik, 1971). Though relatively rare, they are nowadays found in diverse environments, including marine sites at the Bahamas (Dravis, 1983), Shark Bay (Australia; Logan, 1961) and South Africa (Smith and Uken, 2003) but also continental settings such as crater lakes (Kempe *et al.*, 1991; Kaźmierczak *et al.*, 2011), freshwater systems in karst-dominated areas (Laval *et al.*, 2000; Gischler *et al.*, 2008, Santos *et al.*, 2010), hypersaline lakes (Spadafora *et al.*, 2010; Farías *et al.*, 2013; Schneider *et al.*, 2013) and desert pools and streams (Breitbart *et al.*, 2009). Modern microbialites are valuable models to gain insights into the ecology and the biogeochemistry of their Precambrian analogs (Dupraz *et al.*, 2009). This particularly includes non-marine microbialites, since isotopic, microstructure and mineralogical evidence suggest that many ancient microbialites available in the fossil record developed in lacustrine environments (Laval *et al.*, 2000; Kazmierczak and Kempe, 2006; Bolhar and Van Kranendonk, 2007; Stüeken *et al.*, 2015).

Biologically influenced in situ precipitation of minerals, notably carbonates, has been shown as a major process in the formation of microbialites (Awramik and Riding, 1988; Reid *et al.*, 2000; Dupraz and Visscher, 2005; Bowlin *et al.*, 2012). $CaCO_3$ precipitation is favored by i) the increase of supersaturation of the local environment with $CaCO_3$ phases, which results from the increase of pH and/or the activities of Ca^{2+} and Mg^{2+}, or ii) the decrease of the energy barrier of nucleation provided by the presence of nucleation centers. While nucleation is sometimes suggested to be triggered by exopolymeric substances (EPS), produced in abundance by various bacterial phyla (Decho *et al.*, 2005; Benzerara *et al.*, 2006; Braissant *et al.*, 2007, 2009; Obst *et al.*, 2009), the achievement of high carbonate saturation indices can be due to several microbial metabolisms typically found in microbialites (Dupraz *et al.*, 2009). For instance, oxygenic and anoxygenic photosynthesis (Dupraz and Visscher, 2005; Bundeleva *et al.*, 2012), sulfate reduction (Visscher *et al.*, 2000; Gallagher *et al.*, 2012) and, eventually, anaerobic methane oxidation coupled to sulfate reduction (Michaelis *et al.*, 2002), induce a local pH increase that promotes carbonate precipitation. Conversely, microbial metabolisms such as aerobic respiration, sulfide

oxidation or fermentation tend to acidify the local environment, promoting carbonate dissolution (Dupraz and Visscher, 2005; Dupraz *et al.*, 2009). Therefore, microbialite formation in lacustrine systems ultimately relies upon the balance of these various metabolisms.

Microbialite-associated microbial communities are typically complex, as reveal molecular surveys based on small subunit (SSU) rRNA gene metabarcoding. Many of these studies have mainly focused on the prokaryotic component; e.g. in marine stromatolites from Shark Bay, Australia (Logan, 1961; Burns *et al.*, 2004; Papineau *et al.*, 2005; Leuko *et al.*, 2007; Allen *et al.*, 2009; Wong *et al.*, 2015; Suosaari *et al.*, 2016) and Highborne Cay, Bahamas (Baumgartner *et al.*, 2009; Foster *et al.*, 2009; Myshrall *et al.*, 2010; Khodadad and Foster, 2012; Mobberley *et al.*, 2012), but also in non-marine microbialites, such as those of Cuatro Ciénegas (Mexico, Falcón et al., 2007; Breitbart et al., 2009; Nitti et al., 2012), Lake Van (Turkey, López-García et al., 2005), Lake Alchichica (Mexico, Couradeau et al., 2011; Saghaï et al., 2015), and several other freshwater systems (Santos *et al.*, 2010; Russell *et al.*, 2014) or hypersaline lakes (Farías *et al.*, 2013, 2014; Schneider *et al.*, 2013). In all these systems, bacteria are highly diverse and clearly dominate microbial communities, archaea being scarce (Mobberley *et al.*, 2013; Ruvindy *et al.*, 2016; Saghaï *et al.*, 2015; White III *et al.*, 2015). Comparatively, microbial eukaryotes are less-well studied. Nonetheless, metabarcoding approaches suggest that they are moderately abundant and diverse in microbialites from Cuatro Ciénegas (Breitbart *et al.*, 2009), lake Alchichica (Couradeau *et al.*, 2011; Saghaï *et al.*, 2015) and several marine stromatolites (Al-Qassab *et al.*, 2002; Perkins *et al.*, 2012; Bernhard *et al.*, 2013; Mobberley *et al.*, 2013; Edgcomb *et al.*, 2014). Metagenomics can, via the direct sequencing of environmental DNA, inform about the metabolic potential of natural communities (Tringe and Rubin, 2005; Dinsdale *et al.*, 2008). In addition, it also provides information about resident viruses, so far seldom studied in microbialites (Desnues *et al.*, 2008). To date, only a handful of microbialite community metagenomes are available, including those of Clinton Creek (White III *et al.*, 2015), Cuatro Ciénegas (Breitbart *et al.*, 2009), Highborne Cay (Khodadad and Foster, 2012; Mobberley *et al.*, 2013, 2015), Little Darby Island (The Bahamas; Casaburi *et al.*, 2016), Pavilion Lake (White III *et al.*, 2016) and Shark Bay (Ruvindy *et al.*, 2016). These studies tend to highlight the importance of genes involved in EPS production and degradation, as well as in energy metabolism (e.g., photosynthesis or sulfatereduction) likely involved in microbialite formation. Yet, assessing potentially shared metabolic patterns associated with microbialite formation will require the comparison of an increasing number of microbialite metagenomes from a wide array of biotopes and spatial scales.

Lake Alchichica, an alkaline crater lake at pH~9 located at high altitude (2,300 m above sea level), harbors actively growing microbialites, which are mostly composed of hydromagnesite [$Mg_5(CO_3)_4(OH)_2.4(H_2O)$] and aragonite ($CaCO_3$) (Kaźmierczak *et al.*, 2011). Previous surveys of prokaryotic and eukaryotic diversity (Couradeau *et al.*, 2011; Saghaï *et al.*, 2015) revealed a wide diversity of largely dominant bacteria (85-95% total rRNA genes), whereas eukaryotes were fairly diverse, displaying moderate abundance (5-15%), and archaea exhibited poor diversity and extremely low (0.02-0.3%) relative abundances (Saghaï *et al.*, 2015). Cyanobacteria performing a new form of intracellular biomineralization were isolated from these microbialites (Couradeau *et al.*, 2012). In

parallel, extracellular biomineralization by members of the order Pleurocapsales seems particularly important. Pleurocapsales significantly increase in abundance with depth and are associated with the specific precipitation of aragonite, as shown by a combination of fine-scale (Couradeau *et al.*, 2013; Gérard *et al.*, 2013) and statistical analyses (Saghaï *et al.*, 2015). Here, in an effort to explore local spatial variation but also to initiate large-scale metagenomic comparisons among different microbialite systems, we analyze the functional potential of metagenomes obtained from two different sites in Lake Alchichica and, at one site, along a depth gradient (1, 5, 10 and 15 m-deep; Saghaï *et al.*, 2015), and we compare it with that of other microbialite and microbial mat metagenomes. Our results highlight that Alchichica microbialites harbor highly diverse microbial communities in which anoxygenic photosynthesis linked to sulfide oxidation, oxygenic photosynthesis and nitrogen fixation are important.

Results and Discussion

General features of samples and metagenomes

We generated metagenomic data by direct paired-end Ilumina sequencing of DNA extracted from microbialite fragments collected on the Western shore (AL-W; 0.5-1 m depth) and along a depth gradient on the Northern shore (1 to 15 m depth; AL-N series) of Lake Alchichica. The AL-W site was chosen based on the observation of visible seepage activity (bubbling) and the presence of two macroscopic morphotypes, brownish columnar microbialites and cauliflower whitish hydromagnesite-dominated microbialites (Kaźmierczak *et al.*, 2011). AL-W microbialite fragments belonged to the latter, but their surface was less mineralized and had more gelatinous aspect than that of microbialites sampled on the Northern shore, which were mineralized up to the surface, representing the most conspicuous microbialites visible in the lake (Fig.1). Three technical replicates were obtained for AL-W, which were highly congruent in terms of community structure based on rDNAs (Saghaï *et al.*, 2015). Therefore, the three AL-W metagenomes were merged into a single dataset to have a higher coverage providing better assembly and annotation.

The Illumina sequencing yielded between 97 and 306 million paired-end reads per sample (Table 1). The five metagenomes had an estimated coverage of 100% according to Nonpareil curves (Supplementary Fig.S1), these values being consistent with the high proportion of long contigs that we were able to assemble (see below). The reads were subsequently assembled into contigs using stringent criteria to facilitate gene prediction. N50 values ranged from 1,379 to 1,650 bp, indicating that, in each metagenome, half of the entire assembly length was contained in contigs equal to, or larger than, these values (Table 1). Typical prokaryotic genes are on the order of 1 kb in size, therefore being in the range of most of the assembled contigs. As expected, the total contig length was higher in the pooled AL-W metagenomes than in the AL-N series. However, despite the higher number of reads in AL-W, only 28% of them were assembled into contigs, contrasting with the 50-55% assembled reads in the AL-N metagenomes. Three non-mutually exclusive biological explanations might account for this difference. First, AL-W could harbor higher phylogenetic diversity. Although this does not appear to be the case in terms of large taxa (Saghaï *et al.*, 2015; see below), the internal diversity within phylogenetic lineages might be

higher. Second, eukaryotes, having longer genomes, could be more relatively abundant in AL-W. Indeed, based on the presence of SSU rRNA genes, eukaryotes appear to be on average twice as abundant (10-15% of total rRNA genes) as they are in AL-N samples (3-8%) (Saghaï *et al.*, 2015). Finally, eukaryotic genomes in AL-W (largely dominated by diatoms) might be bigger than those of AL-N samples (largely dominated by green algae) (Saghaï *et al.*, 2015). At any rate, the relatively low ratio of assembled reads and the relatively low N50 value are in agreement with highly diverse microbial communities.

As the average GC content of metagenomic reads appears to correlate with their environmental origin (Foerstner *et al.*, 2005), we explored the GC content distribution in Alchichica metagenomes. The GC distribution was bimodal, exhibiting two peaks at 40% and 65-70% GC in most of the samples. Intriguingly, there was a clear shift of the GC content with depth. Shallower samples were enriched in (AL-W, AL-N-1 and AL-N-5), or almost exclusively contained (AL-N-1), high GC sequences whereas the two deepest samples (AL-N-10, AL-N-15) displayed increasingly low-GC content (Supplementary Fig.S2A). This may be partly due to the increasing proportions of Pleurocapsales with depth (Saghaï *et al.* 2015), since contigs affiliating to Pleurocapsales (see below) show 38% GC content (average/median values). The GC content distribution in Alchichica metagenomes diverged from that of other microbialites and mat metagenomes. Marine microbialites display average low-GC sequences, whereas Yellowstone microbial mats exhibit high-GC sequences, both types of distribution being unimodal; only Guerrero Negro microbial mats exhibit a bimodal distribution, more akin to that of most Alchichica metagenomes (Supplementary Fig.S2B).

Community structure based on protein-coding marker genes

To better assess the prokaryotic community composition from metagenomic data, we compared the structure derived from small subunit (SSU) rRNA genes mined in the Alchichica metagenomes with that derived from protein-coding marker genes currently used for prokaryotic taxonomic profiling in metagenomic data (von Mering *et al.*, 2007; Sunagawa *et al.*, 2013) (Fig.2A). Thus, we searched for 40 single-copy gene families universally distributed in prokaryotic genomes (Creevey *et al.*, 2011) in the five Alchichica metagenomic datasets (Supplementary Table S1). Unfortunately, this strategy could not be applied to the eukaryotic component because of the paucity of phylogenetically diverse protist genomes; the updated COG database concentrating exclusively on prokaryotes (Galperin *et al.*, 2014). Between 7,333 and 18,203 predicted genes were affiliated to one of the 40 prokaryotic conserved COG families (Table 1). In general, there was good agreement between the relative proportions of the most abundant taxa identified by SSU rRNAs and single-copy genes (Fig.2A). Overall, Bray-Curtis distances calculated from the relative proportions of high-rank taxonomic groups ranged from 0.07 (AL-W) to 0.24 (AL-N10; average 0.22) in all samples but AL-N-15 (distance = 0.44). Hence, relatively similar results can be retrieved from the two types of markers when high-rank taxonomic groups are considered. Nonetheless, for some taxa and samples, SSU rRNAs were overrepresented as compared to single-copy genes, suggesting that these taxa contain multiple rRNA operons. This is the case for Gammaproteobacteria and Cyanobacteria in the AL-N series, especially in the deepest samples. By contrast, Alphaproteobacteria, Bacteroidetes or Planctomycetes

show similar proportions as estimated by the two types of markers, suggesting that they contain mostly one single rRNA operon.

Similarly to marine microbialites (Ruvindy *et al.*, 2016), the core members of Alchichica microbialites belonged to Proteobacteria (mainly Gamma and Alpha subdivisions), Bacteroidetes and Cyanobacteria (Fig.2A). However, although the same major prokaryotic groups were present, clear differences were observed between AL-W and AL-N samples, attesting to the different local conditions in the two areas of the lake (Saghaï *et al.*, 2015). Alphaproteobacteria represented around 30% of the sequences in the AL-W metagenome, and remained remarkably constant (ca. 20%) along the AL-N depth gradient, whereas Gammaproteobacteria, less represented in AL-W (˜7%), increased with depth from ca. 10 to 20% in the AL-N samples. Bacteroidetes were 2-3 times more abundant in AL-W (˜25%) than in AL-N metagenomes, a trend also observed in the Verrucomicrobia and Betaproteobacteria. Actinobacteria and Firmicutes, on the contrary, were more abundant in AL-N microbialites, their proportions increasing in the deepest sample. Other moderately abundant groups in AL-N, such as Chloroflexi and Nitrospirae, remained constant along the depth gradient. Consistent with other studies (Khodadad and Foster, 2012; Ruvindy *et al.*, 2016; Saghaï *et al.*, 2015; White III *et al.*, 2015), the presence of Archaea was negligible (0.1-0.9% of the single-copy genes). Archaeal sequences consisted mostly of Euryarchaeota, except in AL-N-5 where Thaumarchaeota dominated.

Phylogenetic affiliation of long contigs (>10 kb)

Although our metagenomic comparisons were based on all assembled contigs (see below), we explored in more detail contigs larger than 10 kb. This size is sufficiently long to reliably attribute these genome fragments to specific taxa, especially in the case of prokaryotes. In eukaryotes, however, the presence of long non-coding regions may result in a lower number of genes over a given length.

Since, in principle, the probability to assemble long contigs depends on the relative abundance of the corresponding organisms, the affiliation of long contigs might reflect the general diversity of the metagenomes in terms of high-rank taxa. At the same time, because only highly similar reads corresponding to conserved and/or abundant genomic fragments can be assembled, long contigs thus represent a source of information on (i) the intra-genomic variability of microbial lineages within communities and (ii) the level of «genomic novelty» present in the metagenomes. We conservatively assembled from 4,180 to 6,767 of such contigs >10 kb per metagenome, for a total size of 109 to 157 Mb (11-17% of total contig length; Supplementary Table S2). Some lineages were more abundantly retrieved in SSU rRNA and conserved single-copy genes than in long contig proportions, such as the Deltaproteobacteria and Planctomycetales, suggesting that these lineages cover a wide diversity and/or show high genomic variability. Conversely, some lineages exhibited a higher proportion of long contigs than expected, such as the Chloroflexi (all depths) or *Deinococcus-Thermus* (in AL-N-1 and AL-N-5). However, despite these differences, the phylogenetic distribution of prokaryotic contigs was relatively similar to the distribution observed with SSU rRNAs and single-copy genes (Fig.2A). We then sorted these long contigs in two classes, depending on the proportion of genes attributed to the most abundant

taxon: 20-50% and >50% (Supplementary Fig.S3 and Supplementary Table S2). Unsurprisingly, the same four dominant groups (Alphaproteobacteria, Gammaproteobacteria, Cyanobacteria, Bacteroidetes) in rRNA and single-copy gene-based diversity were overrepresented in the two categories of contigs, with up to 80-90% of the total sequence length in the five long-contig sets (Supplementary Fig.S3A). This offers the possibility to confidently assemble large genome fragments for abundant organisms; for instance, we were able to assemble 1,176 contigs (28,419,775 bp) where ≥50% predicted genes affiliated to Pleurocapsales, a group playing a key role in extra-cellular calcification in Alchichica microbialites (Couradeau *et al.*, 2013; Gérard *et al.*, 2013; Saghaï *et al.*, 2015). However, other lineages were comparatively more represented in the 20-50% gene-attribution class, notably Actinobacteria, Chloroflexi and, at lesser extent, Deltaproteobacteria and Planctomycetes (Supplementary Fig.S3B). This suggests that these genomes are divergent and lack closely related members in genome databases.

In the case of eukaryotes, long contigs represented a minor fraction in term of sequence length (Supplementary Fig.S3 and Supplementary Table S2). Nonetheless, and consistently with rRNA gene distribution (Saghaï *et al.*, 2015), they corresponded mostly to Chlorophyta (green algae) in the AL-N series and mostly to Bacillariophyta (diatoms) in AL-W (Fig.2B). We also assembled contigs from Fungi (Ascomycota), Haptophyta and Metazoa (Arthropoda). Finally, some contigs were affiliated to bacteriophages (Caudovirales; 352,658 bp in total); representatives of Archaea were absent.

Functional genes and metabolic potential

The distribution of functional genes from the five metagenomes in COG categories and SEED classes was statistically homogeneous according to the Kruskal-Wallis test ($H = 0.03$ and $P = 0.99$ in both cases; Fig.3, Supplementary Table S3). Interestingly, this functional homogeneity across metagenomes appeared to be due to phylogenetically distinct but functionally redundant microbial taxa, as previously shown at the class level in Cyanobacteria (Saghaï et al., 2015) and at the OTU level among both Bacteria and Eukaryotes (Couradeau et al., 2011). As expected, housekeeping genes, involved in DNA, amino acid and protein metabolism were among the most abundant. Functional categories more directly connected to microbialite formation, such as carbohydrate metabolism and cell wall biogenesis, were also relatively abundant. Carbohydrate metabolism concerns the uptake and utilization of carbon, including the metabolism of various saccharides (*e.g.* maltose, mannose, trehalose and xylose) potentially involved in the synthesis of EPS. They contribute to confer a negative charge to EPS, increasing their affinity to cations (e.g. Ca^{2+}, Mg^{2+}; Pereira *et al.*, 2009). The exploration of the carbohydrates subsystem revealed that genes associated with trehalose synthesis were dominated by Cyanobacteria, Alphaproteobacteria and Gammaproteobacteria (Supplementary Fig.S4). Within those, genes with the highest sequence similarity to sequenced genes/genomes were mainly related to Cyanobacteria (Chroococcales and Nostocales) and Gammaproteobacteria. Genes involved in maltose/maltodextrin utilization and mannose metabolism were particularly associated to various Cyanobacteria, Bacteroidetes, Alphaproteobacteria and Gammaproteobacteria whereas genes involved in xylose utilization corresponded predominantly to members of the Alphaproteobacteria (Rhodobacterales; Supplementary

Fig.S4). Genes involved in cell wall and envelope synthesis were dominated by peptidoglycan and lipopolysaccharide synthesis genes in all five metagenomes (Fig.4). Genes related to rhamnose synthesis, a common hydrophobic component of bacterial EPS (Pereira *et al.*, 2009), were relatively abundant, something previously observed in Cuatro Ciénegas (Breitbart *et al.*, 2009) and Highborne Cay (Mobberley *et al.*, 2015) metagenomes, and mostly affiliated to Cyanobacteria (Chroococcales), Alphaproteobacteria (Rhodobacterales) and various Gammaproteobacteria (data not shown). Interestingly, only a few functional genes involved in EPS production affiliated to the cyanobacterial order of Oscillatoriales, despite the fact that they were relatively abundant in shallower Alchichica samples (Couradeau *et al.*, 2011; Saghaï *et al.*, 2015). This is not surprising given that these cyanobacteria, lacking sheaths or being only slightly sheathed, essentially attach to the surface of microbialites rather than closely associate with mineral precipitates (Waterbury 2006).

Genes involved in the production of energy represented a significant proportion of the functional hits in the five metagenomes (Fig.3) and covered the typical functional groups found in microbialites (Dupraz *et al.*, 2009). Using the RefSeq annotations, we specifically searched for genes that are usually taken as diagnostic for particular metabolic pathways and identified the main taxa involved in those processes (Fig.5). Oxygenic photosynthesis obviously was an important process. Although dominated by cyanobacteria, with over 60% (AL-W) and 70-75% (AL-N samples) of photosystem I and II (*psa* and *psb*) genes identified, eukaryotic genes were also abundant. However, in agreement with observed differences in microbial community composition, diatom genes represented ca. 40% of oxygenic photosynthesis-related genes in AL-W, while green algal genes were more relatively abundant in AL-N samples (20-25%). Interestingly, photosystem genes affiliated to potentially calcifying Haptophyta (*Emiliana huxleyi; n=42, average identity = 0.92*) were identified in AL-N samples and appeared to increase with depth (Fig.5), although these organisms were not identified at the SSU rRNA gene level.

Genes involved in anoxygenic photosynthesis were also very abundant, confirming that this process is very important in Alchichica microbialites (Saghaï *et al.*, 2015). The phylogenetic affinity of genes involved in bacteriochlorophyll synthesis and the photosynthetic reaction centers (*puf* genes) revealed that, in all samples, the anoxygenic photosynthesis pathway was very largely dominated by members of the Alphaproteobacteria (70-80%; mainly Rhodobacterales), along with Chloroflexi and Gammaproteobacteria (Chromatiales). Because H_2S is a frequent electron donor for anoxygenic photosynthesis, we searched for *sox* genes, involved in sulfur oxidation. Not only the relative proportion of *sox* genes was comparable to that of genes involved in anoxygenic photosynthesis (Fig.5) but there was a clear correlation ($r = 0.992$, *p-value* = 1.7e-8) between anoxygenic photosynthesis and sulfur oxidation genes in Gammaproteobacteria and Alphaproteobacteria (mostly Rhodobacterales). This strongly suggests that anoxygenic photosynthesis is coupled to H_2S oxidation in these two clades.

Since, together with oxygenic and anoxygenic photosynthesis, sulfate reduction is one of the metabolisms invoked as potentially favoring carbonate precipitation in microbialites (Dupraz and Visscher, 2005), we further explored genes involved in the sulfur cycle. However, genes

typically involved in sulfate reduction, notably *dsrAB* encoding the dissimilatory sulfite reductase, were found in surprisingly low abundance (Fig.5). Most of them were detected in the AL-N-1 sample (mainly Desulfobacterales), where sulfate reducing Deltaproteobacteria were slightly more abundant (Saghaï *et al.*, 2015; Fig.2), although some sequences were also detected among Firmicutes. Genes encoding the dissimilatory adenosine-5′-phospho-sulphate reductase (*aprAB* genes) are markers for both sulfate-reducing and sulfur-oxidizing prokaryotes. This explains the presence of numerous hits related to Betaproteobacteria and Gammaproteobacteria, which have no known sulfate reducer members, and reinforce the observation based on *sox* genes that many members of those groups oxidize sulfur. The low abundance of *dsrAB* genes is intriguing, but seems a trend in other microbialites. For instance, they were absent in Shark Bay metagenomes (Ruvindy *et al.*, 2016). Given the conspicuous but recurrent presence of Deltaproteobacteria in all Alchichica microbialites (Fig.2), *dsrAB* low frequencies might suggest either that these Deltaproteobacteria do not reduce sulfate or that they use more divergent proteins to carry out such function.

Nitrogen fixation, as inferred from the presence of *nif* genes, was also present in all microbialite samples (Fig.5). More abundant in surface microbialites (AL-W and AL-N-1), their phylogenetic distribution was generally similar in the different samples, *nif* genes being mostly affiliated to Cyanobacteria (up to 85%) and to Alphaproteobacteria, Gammaproteobacteria and Betaproteobacteria. This is consistent with previous *nifH* surveys in Alchichica microbialites (Beltrán *et al.*, 2012). AL-N-1 displayed a somewhat different profile, with an equivalent proportion of deltaproteobacterial and cyanobacterial genes (Fig. 5). Within cyanobacteria, genes ascribing to Nostocales and Pleurocapsales had comparable proportions. Genes involved in denitrification (reductases for nitrate, nitrite, nitric oxide and nitrous oxide) were also abundant. Primarily affiliated to Alphaproteobacteria (Rhodobacterales, Rhizobiales) and Gammaproteobacteria, they were also recovered from various members of Betaproteobacteria, Deltaproteobacteria, Planctomycetes, Deinoccocus-Thermus and Cyanobacteria (Chroococcales, Oscillatoriales). The dissimilatory reduction of nitrate to N_2 can favor the dissolution of calcium carbonate (Visscher and Stolz, 2005), such that this activity, together with aerobic respiration and fermentation, can counteract metabolisms conducive to carbonate precipitation (oxygenic and anoxygenic photosynthesis or sulfate reduction). We also detected a few ammonia oxidation genes with high identities to Thaumarchaeota (Nitrosopumilales) sequences, but only in AL-N-5.

We also identified genes encoding alkaline phosphatases (*pho* genes), typically involved in dissolved organic phosphorus utilization, in the five Alchichica metagenomes (from 264 in AL-N-10 to 540 in AL-W). They mostly belonged to Alphaproteobacteria; Betaproteobacteria, Gammaproteobacteria, some cyanobacterial members (Chroococcales, Nostocales) and heterotrophic lineages such as Bacteroidetes and Planctomycetes were also represented (data not shown). Consistent with microbial diversity surveys, genes specifically associated with methanogenesis were absent.

Comparison with other metagenomes

To explore the functional differences between several microbialite types and microbial mats forming under diverse environmental conditions, we compared the functional profiles of

Alchichica metagenomes with other publicly available metagenomes. The datasets included stromatolites from The Bahamas, an oncolite (layered spherical microbialite) from a river in Cuatro Ciénegas, hypersaline mats from Guerrero Negro and hot-spring mats from Yellowstone National Park (Supplementary Table S4). The five metagenomes from Alchichica microbialites clustered together in correspondence analyses based on the relative abundance of COG categories, being well separated from other microbialite and microbial mat metagenomes along the first axis (Fig.6A). Likewise, the two stromatolites from Highborne Cay clustered together, as did the two samples from Yellowstone. The metagenome from Shark Bay clearly discriminated from the rest along the second axis, while the Guerrero Negro mat occupied an intermediate position between the Alchichica microbialites and Yellowstone mats, Rio Mesquites and Highborne Cay metagenomes. This reflects differences in microbial community compositions and suggests differences in microbial processes among microbialite types but also between microbialites and non-(massively) calcifying microbial mats, depending on the environmental conditions in which they develop. Nonetheless, when we added datasets from soil and deep-sea samples to our analysis, all microbialites and mat samples loosely clustered together away from soil and deep sea metagenomes (Fig.6B). They were also closer to the deep sea samples along the first and most explanatory axis (up to 60% of variance), which possibly relates to the fact that they develop in aqueous environments. Within the cluster of mats and microbialite metagenomes, Octopus and Mushroom mats clustered closer to Alchichica microbialites, which might be explained by the fact that they form in slightly alkaline springs (pH 8-8.5), closer to the Alchichica values (pH 9). The proximity between Guerrero Negro mats, Highborne Cay and Shark Bay stromatolites in this correspondence analysis might perhaps be due to the saline environment in which they develop, although the Rio Mesquites (a freshwater river) sample placed closer to both Highborne Cay and Guerrero Negro samples than to Alchichica microbialites (Fig.6B). This suggests that additional environmental parameters also influence the structure and function of these communities.

Although metagenomes from Alchichica cluster with metagenomes from other microbialites and microbial mats, it has to be noted that this resemblance is at a very general functional level. However, functional homogeneity can be due to phylogenetically distinct communities comprising functionally redundant lineages. Furthermore, organisms belonging to the same taxa might exhibit very different metabolic properties when it comes to much more specific functional categories or to regulatory aspects. To get deeper insights in these issues, the future comparison of genomes (or large genome fragments, as those generated in this work) from microbialite-associated microorganisms will be extremely useful.

Concluding remarks

Comparative metagenomics of microbialites from the alkaline Lake Alchichica confirm previous diversity studies carried out independently (Couradeau *et al.*, 2011) and reveal the functional potential of these microbial communities. Our study provides qualitative information about the nature of genes and metabolic functions present in these communities, collected at two different sites (AL-W and AL-N) and, at one site, along a depth gradient (AL-N, 1 to 15 m depth). However, it also provides useful quantitative information for comparison. Thus, although we only produced replicate metagenomes for AL-W samples,

they were remarkably similar (Saghaï *et al.*, 2015). In addition, AL-N samples were collected at the same place and depths as those from a previous study (Couradeau *et al.*, 2011), showing very congruent patterns of microbial diversity and offering inter-annual validation for the respective community structures. Taken globally, Alchichica metagenomes highlight the importance of photosynthesis, including anoxygenic photosynthesis, in these microbialites. The latter likely depends on H_2S as electron donor, as reveal the high relative abundance of sulfur oxidation-related genes and the co-occurrence of *sox* genes with genes involved in anoxygenic photosynthesis in the same bacterial groups (Fig.5). While oxygenic photosynthesis is mostly carried out by cyanobacteria, there is also an important contribution of diatoms (AL-W samples, affected by subaquatic seeping) and green algae (AL-N samples), indicating a eukaryotic contribution to carbon fixation in these structures. Nitrogen fixation and denitrification are also highly represented pathways. By contrast, sulfate reduction, despite the importance of sulfide oxidation and the presence of potentially sulfate-reducing deltaproteobacteria in these communities, does not seem to be a profuse metabolism in these microbialites. Together with photosynthesis, genes involved in carbohydrate metabolism are also very abundant, underlying the active role of heterotrophic bacteria in degrading EPS produced by photosynthesizers, notably cyanobacteria. Despite some differences, our comparative metagenomic study highlights common functional trends between microbial mats and microbialites at large scale. This may not be surprising considering that microbialites are calcifying microbial mats. Therefore, the same type of major metabolic functions related to stratified photosynthesis along redox gradients and the associated heterotrophic functions, notably carbohydrate metabolism, are represented. We have been able to assemble long contigs (>10 kb) for many different microorganisms associated with these microbialites and the analysis of their gene affiliation allows them to attribute them to particular lineages, which parallels the community structure derived from mining SSU rRNA genes. Together with cultivation attempts, the reconstruction and analysis of scaffold genomes from these metagenomes offers a promising way to characterize lineage-specific metabolisms and to further explore interspecific interactions within these highly complex ecosystems.

Experimental procedures

Sample collection, DNA purification and sequencing

Microbialite fragments were collected in winter (January 2012), when the lake was not stratified (Macek *et al.*, 2009), at two different sites of the lake: the Western shore at ca. 0.5–1 m depth (AL-W; 19° 25' 0.13”N; 97° 24' 41.07”W) and the Northern shore at 1, 5, 10, and 15 m depth (AL-N samples; 19° 25' 12.49”N; 97° 24' 12.35”W) (Fig.1 ; Saghaï *et al.*, 2015). Equivalent volumes of microbialite fragments were ground and carbonates partially dissolved prior to DNA extraction using Power Biofilm™ DNA Isolation Kit and further cleaning using the Power Clean™ ProDNA Clean-up Kit (MoBio, Carlsbad, CA, USA) as previously described (Saghaï *et al.*, 2015). DNA yield ranged from 0.8 μg (deepest AL-N samples) to 22 μg (AL-W). Illumina paired-end libraries were prepared without any amplification step for each AL-N sample and, in triplicate, for the AL-W sample (technical replicates), as described (Saghaï *et al.*, 2015). Sequencing of the seven metagenomic libraries was done in two independent runs (AL-N and AL-W samples) using Illumina

HiSeq2000 v3 (2x100 bp paired-end reads) by Beckman Coulter Genomics (Danver, MA, USA). The total number of paired-end reads ranged from 78 to 131 millions, i.e., 7.8-13.1 Gbp per library and orientation (forward and reverse) (Saghaï *et al.*, 2015). Raw sequencing data is available at NCBI Bioproject PRJNA315555. The level of coverage of the community achieved by each metagenomic dataset was estimated and projected using Nonpareil version 2.4 with default parameters (Rodriguez-R and Konstantinidis, 2014), after preprocessing the reads with SolexaQA version 3.1.5 and a minimum Phred quality score of 20 (Cox *et al.*, 2010).

Assembly and annotation

For each metagenomic dataset, forward and reverse reads were assembled into contigs using MEGAHIT (version 0.3.3; Li *et al.*, 2015) with default parameters but a minimum length of 200 bp for the assembled contigs and a carefully chosen kmer size (we also used MetaVelvet (Namiki *et al.*, 2012) but median and longest contig sizes obtained were poorer; data not shown). Since the length of the starting kmer size is essential for assembly quality, we tested a range of starting values (15-121 bp) and chose a kmer size of 27 (kmer list: 27, 47, 67, 87, 107, 123) for the AL-N libraries. As the three AL-W replicate libraries originated from the same sample, we merged the three paired-end read datasets to improve contig quality and size and, hence, gene prediction and potential genome scaffolding (kmer size=57, kmer list: 57, 77, 97, 117, 123). Gene prediction was performed on the newly assembled contigs using Prodigal (version 2.60, metagenomic mode; Hyatt *et al.*, 2012). For a more accurate functional annotation, we used the amino acid sequence of predicted genes. The annotation was done using the blastp command implemented in DIAMOND (version 0.7.9, with a maximum e-value of 10^{-5}; Buchfink *et al.*, 2015) with the non-redundant protein database (RefSeq nr release 63; January 12, 2014), the recently updated COG database (Galperin et al., 2014) and the SEED database (release of September 14, 2011, ftp://ftp.theseed.org/subsystems/; Overbeek et al., 2005). . For subsequent analyses, we retained only the best hit to represent each annotated gene, with a minimum amino acid identity of 50% over at least 80% of the query length. Various statistics regarding contigs assembly and annotation are provided in Table 1. COG hits were further used to describe the phylogenetic structure of the communities by mining for a set of 40 single copy gene families universally distributed in prokaryotic genomes (Supplementary Table S1; Creevey *et al.*, 2011). In a complementary approach, we mined the RefSeq annotations for genes diagnostic of the following metabolisms: anoxygenic photosynthesis (bacteriochlorophyll and *puf* genes), photosynthesis (*psa* and *psb* genes), sulfate reduction (*aprAB* and *dsrAB* genes), sulfur oxidation (*sox* genes), nitrogen fixation (*nif* genes), denitrification (reductases for nitrate, nitrite, nitric oxide and nitrous oxide), ammonia oxidation (*amo* genes), methanogenesis (formylmethanofuran dehydrogenases) and phosphorus utilization (alkaline phosphatase genes).

Selection and phylogenetic affiliation of the longest contigs

For a more in-depth analysis linking qualitative functional information to particular taxa, we conservatively selected long contigs from each metagenome that i) had at least 10 kb with a minimum of 10 genes (considering that the average prokaryotic gene size is around 1 kb; Xu *et al.*, 2006) and ii) contained genes with high similarity to RefSeq (minimum best hit

identity ≥50% over at least 80% of the gene length) to establish the phylogenetic affiliation of these long contigs. Contigs were attributed to a given taxon (i.e. at the phylum level) when more than 20% of their genes had phylogenetic affinity (using the above criteria) to that taxon. Groups of contigs having, respectively, 20-50% or ≥50% of their genes attributed to a same taxon were analyzed separately. Contigs with less than 20% genes affiliated to the same taxon were discarded and their taxonomic affiliation considered as uncertain. The relative proportion of taxa based on long contigs was determined using the total sequence length affiliated to each taxon rather than the number of contigs.

Statistical analyses

Statistical analyses were conducted with the R software (R Development Core Team, 2013). The comparison of the taxonomic distribution inferred from protein-coding marker genes with results obtained by mining for 16S rRNA genes in the same metagenomic datasets (Saghaï *et al.*, 2015) was done using Bray-Curtis dissimilarity distances. They were calculated on frequencies of high-rank bacterial and archaeal taxa using the 'Vegan' R package (version 2.0-10; Oksanen *et al.*, 2012) implemented in R (http://cran.r-project.org) with no prior transformation of the data. The H statistic of the Kruskal-Wallis test was computed, using the relative proportions of either functional categories (COG and SEED) or high-rank taxa obtained for the five metagenomes, to test whether they were statistically similar (R function kruskal.test). The Pearson correlation coefficient, and its associated p-value, corresponding to the relative proportion of *sox* and anoxygenic photosynthesis-related genes was calculated using the cor.test function implemented in R. The heatmap showing the relative abundance of the genes involved in Cell Wall and Capsule SEED subsystem was generated using the 'Pheatmap' package (version 1.0.7). The functional annotations of the five Alchichica metagenomes were also compared with several publicly available metagenome datasets. To do so, we recovered on the MG-RAST server (Meyer *et al.*, 2008) data on the abundance of COG categories from several metagenomes (Supplementary Table S4), using as criteria a maximum e-value cut-off of 10^{-5} and a minimum identity cutoff of 50%. They included microbialites from Cuatro Ciénegas, Highborne Cay and Shark Bay, hypersaline mats from Guerrero Negro, hot spring mats from Yellowstone National Park and forest and hot desert soil samples. Metagenomes of two deep-sea planktonic communities (fraction 0.2-5 μm) were also incorporated in the dataset as potential outliers. Correspondence analyses were performed using the 'Ade4' package (version 1.6-2) based on frequencies of the COG categories (calculated relatively to the total number of COG hits). To maximize high-level functional differences, we removed from the dataset two COG categories: Function unknown and General functional prediction only.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Eleonor Cortés for help and good company during the field trip and work at the UNAM lab and to Philippe Deschamps for technical advice. This research was funded by the European Research Council Grants *ProtistWorld* (PI P.L.G., AdG Agreement no. 322669) and *CALCYAN* (PI K.B., CoG Agreement no. 307110)

under the European Union's Seventh Framework Program and the RTP Génomique environnementale of the CNRS (project MetaStrom, PI D.M.).

References

Al-Qassab S, Lee WJ, Murray S, Geoffrey A, Simpson B, Patterson DJ. Flagellates from Stromatolites and Surrounding Sediments in Shark Bay , Western Australia. Acta Protozool. 2002; 41:91–144.

Allen MA, Goh F, Burns BP, Neilan BA. Bacterial, archaeal and eukaryotic diversity of smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. Geobiology. 2009; 7:82–96. [PubMed: 19200148]

Allwood AC, Walter MR, Kamber BS, Marshall CP, Burch IW. Stromatolite reef from the Early Archaean era of Australia. Nature. 2006; 441:714–8. [PubMed: 16760969]

Awramik SM. Precambrian Columnar Stromatolite Diversity: Reflection of Metazoan Appearance. Science. 1971; 174:825–27. [PubMed: 17759393]

Awramik SM, Riding R. Role of algal eukaryotes in subtidal columnar stromatolite formation. Proc Natl Acad Sci USA. 1988; 85:1327–29. [PubMed: 16593910]

Baumgartner LK, Spear JR, Buckley DH, Pace NR, Reid RP, Dupraz C, et al. Microbial diversity in modern marine stromatolites, Highborne Cay, Bahamas. Environ Microbiol. 2009; 11:2710–19. [PubMed: 19601956]

Beltrán Y, Centeno CM, García-Oliva F, Legendre P, Falcón LI. N2 fixation rates and associated diversity (nifH) of microbialite and mat-forming consortia from different aquatic environments in Mexico. Aquat Microb Ecol. 2012; 67:15–24.

Benzerara K, Menguy N, López-García P, Yoon T-H, Kazmierczak J, Tyliszczak T, et al. Nanoscale detection of organic signatures in carbonate microbialites. Proc Natl Acad Sci USA. 2006; 103:9440–45. [PubMed: 16772379]

Bernhard JM, Edgcomb VP, Visscher PT, McIntyre-Wressnig A, Summons RE, Bouxsein ML, et al. Insights into foraminiferal influences on microfabrics of microbialites at Highborne Cay, Bahamas. Proc Natl Acad Sci USA. 2013; 110:9830–4. [PubMed: 23716649]

Bhaya D, Grossman AR, Steunou A-S, Khuri N, Cohan FM, Hamamura N, et al. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J. 2007; 1:703–713. [PubMed: 18059494]

Bolhar R, Van Kranendonk MJ. A non-marine depositional setting for the northern Fortescue Group, Pilbara Craton, inferred from trace element geochemistry of stromatolitic carbonates. Precambrian Res. 2007; 155:229–250.

Bosak T, Greene SE, Newman DK. A likely role for anoxygenic photosynthetic microbes in the formation of ancient stromatolites. Geobiology. 2007; 5:119–126. [PubMed: 20890383]

Bowlin EM, Klaus JS, Foster JS, Andres MS, Custals L, Reid RP. Environmental controls on microbial community cycling in modern marine stromatolites. Sediment Geol. 2012; 263–264:45–55.

Braissant O, Decho aW, Dupraz C, Glunk C, Przekop KM, Visscher PT. Exopolymeric substances of sulfate-reducing bacteria: Interactions with calcium at alkaline pH and implication for formation of carbonate minerals. Geobiology. 2007; 5:401–411.

Braissant O, Decho AW, Przekop KM, Gallagher KL, Glunk C, Dupraz C, et al. Characteristics and turnover of exopolymeric substances in a hypersaline microbial mat. FEMS Microbiol Ecol. 2009; 67:293–307. [PubMed: 19049495]

Breitbart M, Hoare A, Nitti A, Siefert J, Haynes M, Dinsdale E, et al. Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. Environ Microbiol. 2009; 11:16–34. [PubMed: 18764874]

Buchfink B, Xie C, Huson DH. Fast and Sensitive Protein Alignment using DIAMOND. Nat Methods. 2015; 12:59–60. [PubMed: 25402007]

Bundeleva IA, Shirokova LS, Bénézeth P, Pokrovsky OS, Kompantseva EI, Balor S. Calcium carbonate precipitation by anoxygenic phototrophic bacteria. Chem Geol. 2012; 291:116–131.

Burne RV, Moore LS. Microbialites: Organosedimentary Deposits of Benthic Microbial Communities. Palaios. 1987; 2:241–254.

Europe PMC Funders Author Manuscripts

Burns BP, Goh F, Allen M, Neilan BA. Microbial diversity of extant stromatolites in the hypersaline marine environment of Shark Bay, Australia. Environ Microbiol. 2004; 6:1096–1101. [PubMed: 15344935]

Casaburi G, Duscher A, Reid R, Foster JS. Characterization of the stromatolite microbiome from Little Darby Island, The Bahamas using predictive and whole shotgun metagenomic analysis. Environ Microbiol. 2016; doi: 10.1111/1462-2920.13094

Centeno CM, Legendre P, Beltrán Y, Alcántara-Hernández RJ, Lidström UE, Ashby MN, et al. Microbialite genetic diversity and composition relate to environmental variables. FEMS Microbiol Ecol. 2012; 82:724–35. [PubMed: 22775797]

Couradeau E, Benzerara K, Gérard E, Estève I, Moreira D, Tavera R, et al. Cyanobacterial calcification in modern microbialites at the submicrometer scale. Biogeosciences. 2013; 10:5255–5266.

Couradeau E, Benzerara K, Gérard E, Moreira D, Bernard S, Brown GE, et al. An early-branching microbialite cyanobacterium forms intracellular carbonates. Science. 2012; 336:459–62. [PubMed: 22539718]

Couradeau E, Benzerara K, Moreira D, Gérard E, Kaźmierczak J, Tavera R, et al. Prokaryotic and eukaryotic community structure in field and cultured microbialites from the alkaline Lake Alchichica (Mexico). PLoS One. 2011; 6:e28767. [PubMed: 22194908]

Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics. 2010; 11:485. [PubMed: 20875133]

Creevey CJ, Doerks T, Fitzpatrick DA, Raes J, Bork P. Universally distributed single-copy genes indicate a constant rate of horizontal transfer. PLoS One. 2011; 6:e22099. [PubMed: 21850220]

Decho AW, Visscher PT, Reid RP. Production and cycling of natural microbial exopolymers (EPS) within a marine stromatolite. Palaeogeogr Palaeoclimatol Palaeoecol. 2005; 219:71–86.

Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M, et al. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature. 2008; 452:340–3. [PubMed: 18311127]

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, et al. Functional metagenomic profiling of nine biomes. Nature. 2008; 452:629–632. [PubMed: 18337718]

Dravis JJ. Hardened subtidal stromatolites, Bahamas. Science. 1983; 219:385–6. [PubMed: 17815316]

Dupraz C, Reid RP, Braissant O, Decho AW, Norman RS, Visscher PT. Processes of carbonate precipitation in modern microbial mats. Earth-Science Rev. 2009; 96:141–162.

Dupraz C, Visscher PT. Microbial lithification in marine stromatolites and hypersaline mats. Trends Microbiol. 2005; 13:429–38. [PubMed: 16087339]

Edgcomb V, Bernhard J. Heterotrophic Protists in Hypersaline Microbial Mats and Deep Hypersaline Basin Water Columns. Life. 2013; 3:346–62. [PubMed: 25369746]

Edgcomb VP, Bernhard JM, Summons RE, Orsi W, Beaudoin D, Visscher PT. Active eukaryotes in microbialites from Highborne Cay, Bahamas, and Hamelin Pool (Shark Bay), Australia. ISME J. 2014; 8:418–29. [PubMed: 23924782]

Falcón LI, Cerritos R, Eguiarte LE, Souza V. Nitrogen fixation in microbial mat and stromatolite communities from Cuatro Ciénegas, Mexico. Microb Ecol. 2007; 54:363–73. [PubMed: 17450393]

Farías ME, Contreras M, Rasuk MC, Kurth D, Flores MR, Poiré DG, et al. Characterization of bacterial diversity associated with microbial mats, gypsum evaporites and carbonate microbialites in thalassic wetlands: Tebenquiche and La Brava, Salar de Atacama, Chile. Extremophiles. 2014; 18:311–29. [PubMed: 24442191]

Farías ME, Rascovan N, Toneatti DM, Albarracín VH, Flores MR, Poiré DG, et al. The discovery of stromatolites developing at 3570 m above sea level in a high-altitude volcanic lake Socompa, Argentinean Andes. PLoS One. 2013; 8:e53497. [PubMed: 23308236]

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci USA. 2012; 109:21390–5. [PubMed: 23236140]

Foerstner KU, von Mering C, Hooper SD, Bork P. Environments shape the nucleotide composition of genomes. EMBO Rep. 2005; 6:1208–1213. [PubMed: 16200051]

Foster JS, Green SJ, Ahrendt SR, Golubic S, Reid RP, Hetherington KL, et al. Molecular and morphological characterization of cyanobacterial diversity in the stromatolites of Highborne Cay, Bahamas. ISME J. 2009; 3:573–87. [PubMed: 19148145]

Gallagher KL, Kading TJ, Braissant O, Dupraz C, Visscher PT. Inside the alkalinity engine: The role of electron donors in the organomineralization potential of sulfate-reducing bacteria. Geobiology. 2012; 10:518–530. [PubMed: 22925453]

Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2014; 43:D261–9. [PubMed: 25428365]

Gérard E, Ménez B, Couradeau E, Moreira D, Benzerara K, Tavera R, et al. Specific carbonate-microbe interactions in the modern microbialites of Lake Alchichica (Mexico). ISME J. 2013; 7:1997–2009. [PubMed: 23804151]

Gischler E, Gibson MA, Oschmann W. Giant Holocene Freshwater Microbialites, Laguna Bacalar, Quintana Roo, Mexico. Sedimentology. 2008; 55:1293–1309.

Hyatt D, Locascio PF, Hauser LJ, Uberbacher EC. Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics. 2012; 28:2223–2230. [PubMed: 22796954]

Kazmierczak J, Kempe S. Genuine modern analogues of Precambrian stromatolites from caldera lakes of Niuafo'ou Island, Tonga. Naturwissenschaften. 2006; 93:119–26. [PubMed: 16365738]

Kaźmierczak J, Kempe S, Kremer B, López-García P, Moreira D, Tavera R. Hydrochemistry and microbialites of the alkaline crater lake Alchichica, Mexico. Facies. 2011; 57:543–570.

Kempe S, Kazmierczak J, Landmann G, Konuk T, Reimer A, Lipp A. Largest known microbialites discovered in Lake Van, Turkey. Nature. 1991; 349:605–8.

Khodadad CLM, Foster JS. Metagenomic and metabolic profiling of nonlithifying and lithifying stromatolitic mats of Highborne Cay, The Bahamas. PLoS One. 2012; 7:e38229. [PubMed: 22662280]

Kunin V, Raes J, Harris JK, Spear JR, Walker JJ, Ivanova N, et al. Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. Mol Syst Biol. 2008; 4:198. [PubMed: 18523433]

Laval B, Cady SL, Pollack JC, McKay CP, Bird JS, Grotzinger JP, et al. Modern freshwater microbialite analogues for ancient dendritic reef structures. Nature. 2000; 407:626–9. [PubMed: 11034210]

Leuko S, Goh F, Allen MA, Burns BP, Walter MR, Neilan BA. Analysis of intergenic spacer region length polymorphisms to investigate the halophilic archaeal diversity of stromatolites and microbial mats. Extremophiles. 2007; 11:203–10. [PubMed: 17082971]

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015; 31:1674–1676. [PubMed: 25609793]

Logan BW. Cryptozoon and associate stromatolites from the recent, Shark Bay, Western Australia. J Geol. 1961; 69:517–533.

López-García P, Kazmierczak J, Benzerara K, Kempe S, Guyot F, Moreira D. Bacterial diversity and carbonate precipitation in the giant microbialites from the highly alkaline Lake Van, Turkey. Extremophiles. 2005; 9:263–74. [PubMed: 15959626]

Macek M, Alcocer J, Lugo Vázquez A, Martínez-Pérez ME, Peralta Soriano L, Vilaclara Fatjó G. Long term picoplankton dynamics in a warm-monomictic, tropical high altitude lake. J Limnol. 2009; 68:183–192.

Martín-Cuadrado A-B, López-García P, Alba J-C, Moreira D, Monticelli L, Strittmatter A, et al. Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. PLoS One. 2007; 2:e914. [PubMed: 17878949]

Martin-Cuadrado A-B, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martínez E, Henn MR, et al. Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. ISME J. 2008; 2:865–86. [PubMed: 18463691]

von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, et al. Quantitative phylogenetic assessment of microbial communities in diverse environments. Science. 2007; 315:1126–1130. [PubMed: 17272687]

Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M, et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008; 9:386. [PubMed: 18803844]

Michaelis W, Seifert R, Nauhaus K, Treude T, Thiel V, Blumenberg M, et al. Microbial reefs in the Black Sea fueled by anaerobic oxidation of methane. Science. 2002; 297:1013–1015. [PubMed: 12169733]

Mobberley JM, Khodadad CLM, Foster JS. Metabolic potential of lithifying cyanobacteria-dominated thrombolitic mats. Photosynth Res. 2013; 118:125–140. [PubMed: 23868401]

Mobberley JM, Khodadad CLM, Visscher PT, Reid RP, Hagan P, Foster JS. Inner workings of thrombolites: spatial gradients of metabolic activity as revealed by metatranscriptome profiling. Sci Rep. 2015; 5:12601. [PubMed: 26213359]

Mobberley JM, Ortega MC, Foster JS. Comparative microbial diversity analyses of modern marine thrombolitic mats by barcoded pyrosequencing. Environ Microbiol. 2012; 14:82–100. [PubMed: 21658172]

Myshrall KL, Mobberley JM, Green SJ, Visscher PT, Havemann SA, Reid RP, et al. Biogeochemical cycling and microbial diversity in the thrombolitic microbialites of Highborne Cay, Bahamas. Geobiology. 2010; 8:337–54. [PubMed: 20491947]

Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res. 2012; 40:e155. [PubMed: 22821567]

Nitti A, Daniels CA, Siefert J, Souza V, Hollander D, Breitbart M. Spatially Resolved Genomic, Stable Isotopic, and Lipid Analyses of a Modern Freshwater Microbialite from Cuatro Ciénegas, Mexico. Astrobiology. 2012; 12:685–698. [PubMed: 22882001]

Obst M, Dynes JJ, Lawrence JR, Swerhone GDW, Benzerara K, Karunakaran C, et al. Precipitation of amorphous CaCO3 (aragonite-like) by cyanobacteria: A STXM study of the influence of EPS on the nucleation process. Geochim Cosmochim Acta. 2009; 73:4180–4198.

Oksanen, AJ., Blanchet, FG., Kindt, R., Legendre, P., Minchin, PR., Hara, RBO., et al. Vegan: Community Ecology Package. 2012. R Package version 1.17-9. Available at http://CRAN.R–project.org/package=veg.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 2005; 33:5691–5702. [PubMed: 16214803]

Papineau D, Walker JJ, Mojzsis SJ, Pace NR. Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. Appl Environ Microbiol. 2005; 71:4822–32. [PubMed: 16085880]

Pereira S, Zille A, Micheletti E, Moradas-Ferreira P, De Philippis R, Tamagnini P. Complexity of cyanobacterial exopolysaccharides: Composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. FEMS Microbiol Rev. 2009; 33:917–941. [PubMed: 19453747]

Perkins RG, Mouget JL, Kromkamp JC, Stolz J, Pamela Reid R. Modern stromatolite phototrophic communities: A comparative study of procaryote and eucaryote phototrophs using variable chlorophyll fluorescence. FEMS Microbiol Ecol. 2012; 82:584–596. [PubMed: 22671029]

Pinto AJ, Raskin L. PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. PLoS One. 2012; 7:e43093. [PubMed: 22905208]

Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One. 2010; 5:e9490. [PubMed: 20224823]

R Development Core Team. A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2013.

Reid R, Visscher P, Decho A, Stolz J. The role of microbes in accretion, lamination and early lithification of modern marine stromatolites. Nature. 2000; 406:989–992. [PubMed: 10984051]

Riding R. Microbial carbonate abundance compared with fluctuations in metazoan diversity over geological time. Sediment Geol. 2006; 185:229–238.

Riding R. Microbial carbonates: The geological record of calcified bacterial-algal mats and biofilms. Sedimentology. 2000; 47:179–214.

Rodriguez-R LM, Konstantinidis KT. Nonpareil: A redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics. 2014; 30:629–635. [PubMed: 24123672]

Russell JA, Brady AL, Cardman Z, Slater GF, Lim DSS, Biddle JF. Prokaryote populations of extant microbialites along a depth gradient in Pavilion Lake, British Columbia, Canada. Geobiology. 2014; 12:250–264. [PubMed: 24636451]

Ruvindy R, White RA III, Neilan BA, Burns BP. Unravelling core microbial metabolisms in the hypersaline microbial mats of Shark Bay using high-throughput metagenomics. ISME J. 2016; 10:183–196. [PubMed: 26023869]

Saghaï A, Zivanovic Y, Zeyen N, Moreira D, Benzerara K, Deschamps P, et al. Metagenome-based diversity analyses suggest a significant contribution of non-cyanobacterial lineages to carbonate precipitation in modern microbialites. Front Microbiol. 2015; 6:797. [PubMed: 26300865]

Santos F, Peña A, Nogales B, Soria-Soria E, Del Cura MAG, González-Martín JA, et al. Bacterial diversity in dry modern freshwater stromatolites from Ruidera Pools Natural Park, Spain. Syst Appl Microbiol. 2010; 33:209–21. [PubMed: 20409657]

Schneider D, Arp G, Reimer A, Reitner J, Daniel R. Phylogenetic analysis of a microbialite-forming microbial mat from a hypersaline lake of the Kiritimati atoll, Central Pacific. PLoS One. 2013; 8:e66662. [PubMed: 23762495]

Smith AM, Uken R. Living marine stromatolites at Kei River mouth. S Afr J Sci. 2003; 99:200.

Spadafora A, Perri E, Mckenzie JA, Vasconcelos C. Microbial biomineralization processes forming modern Ca:Mg carbonate stromatolites. Sedimentology. 2010; 57:27–40.

Stüeken EE, Buick R, Schauer AJ. Nitrogen isotope evidence for alkaline lakes on late Archean continents. Earth Planet Sci Lett. 2015; 411:1–10.

Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger SA, Kultima JR, et al. Metagenomic species profiling using universal phylogenetic marker genes. Nat Methods. 2013; 10:1196–99. [PubMed: 24141494]

Suosaari EP, Reid RP, Stolz JF, Casaburi G, Foster JS, Hagen P, et al. New multi-scale perspectives on the stromatolites of Shark Bay, Western Australia. Sci Rep. 2016; 6:20557. [PubMed: 26838605]

Tice MM, Lowe DR. Photosynthetic microbial mats in the 3,416-Myr-old ocean. Nature. 2004; 431:549–52. [PubMed: 15457255]

Tringe SG, Rubin EM. Metagenomics: DNA sequencing of environmental samples. Nat Rev Genet. 2005; 6:805–814. [PubMed: 16304596]

Valdespino-Castillo PM, Alcántara-Hernández RJ, Alcocer J, Merino-Ibarra M, Macek M, Falcón LI. Alkaline phosphatases in microbialites and bacterioplankton from Alchichica soda lake, Mexico. FEMS Microbiol Ecol. 2014; 90:504–519. [PubMed: 25112496]

Visscher PT, Reid PR, Bebout BM. Microscale observations of sulfate reduction: Correlation of microbial activity with lithified micritic laminae in modern marine stromatolites. Geology. 2000; 28:919–922.

Visscher PT, Stolz JF. Microbial mats as bioreactors: populations, processes, and products. Palaeogeogr Palaeoclimatol Palaeoecol. 2005; 219:87–100.

Wade BD, Garcia-Pichel F. Evaluation of DNA Extraction Methods for Molecular Analyses of Microbial Communities in Modern Calcareous Microbialites. Geomicrobiol J. 2003; 20:549–561.

Waterbury, JB. The Cyanobacteria – Isolation, Purification and Identification. The Prokaryotes. 3rd edition. Dworkin, M.Falkow, S.Rosenberg, E.Schleifer, KH., Stackbrandt, E., editors. Vol. 4. Springer-Verlag; New York: 2006. p. 1053-1073.

White RA III, Chan AM, Gavelis GS, et al. Metagenomic analysis suggests modern freshwater microbialites harbor a distinct core microbial community. Front Microbiol. 2016; 6:1531. [PubMed: 26903951]

White RA III, Power IM, Dipple GM, Southam G, Suttle CA. Metagenomic analysis reveals that modern microbialites and polar microbial mats have similar taxonomic and functional potential. Front Microbiol. 2015; 6:966. [PubMed: 26441900]

Wong HL, Smith D-L, Visscher PT, Burns BP. Niche differentiation of bacterial communities at a millimeter scale in Shark Bay microbial mats. Sci Rep. 2015; 5:15607. [PubMed: 26499760]

Xu L, Chen H, Hu X, Zhang R, Zhang Z, Luo ZW. Average Gene Length Is Highly Conserved in Prokaryotes and Eukaryotes and Diverges Only Between the Two Kingdoms. Mol Biol Evol. 2006; 23:1107–1108. [PubMed: 16611645]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Originality-Significance Statement

Studying the metabolic potential of microbial communities associated to modern microbialites can help formulating hypotheses on how they form and can be preserved in the fossil record. However, the paucity of metagenomic data available from this type of structures limits potential generalizations applicable to the past. In this work, we analyzed several metagenomes generated from microbialite fragments collected from the Lake Alchichica (Mexico) harboring conspicuous microbialite structures. Our work shows the importance of a series of metabolic activities including, notably, anoxygenic photosynthesis. The comparison of Alchichica metagenomes with other metagenomes from marine and freshwater microbialites, microbial mats and outgroup environments reveals some common trends uniting microbialite and microbial mat communities.

Europe PMC Funders Author Manuscripts

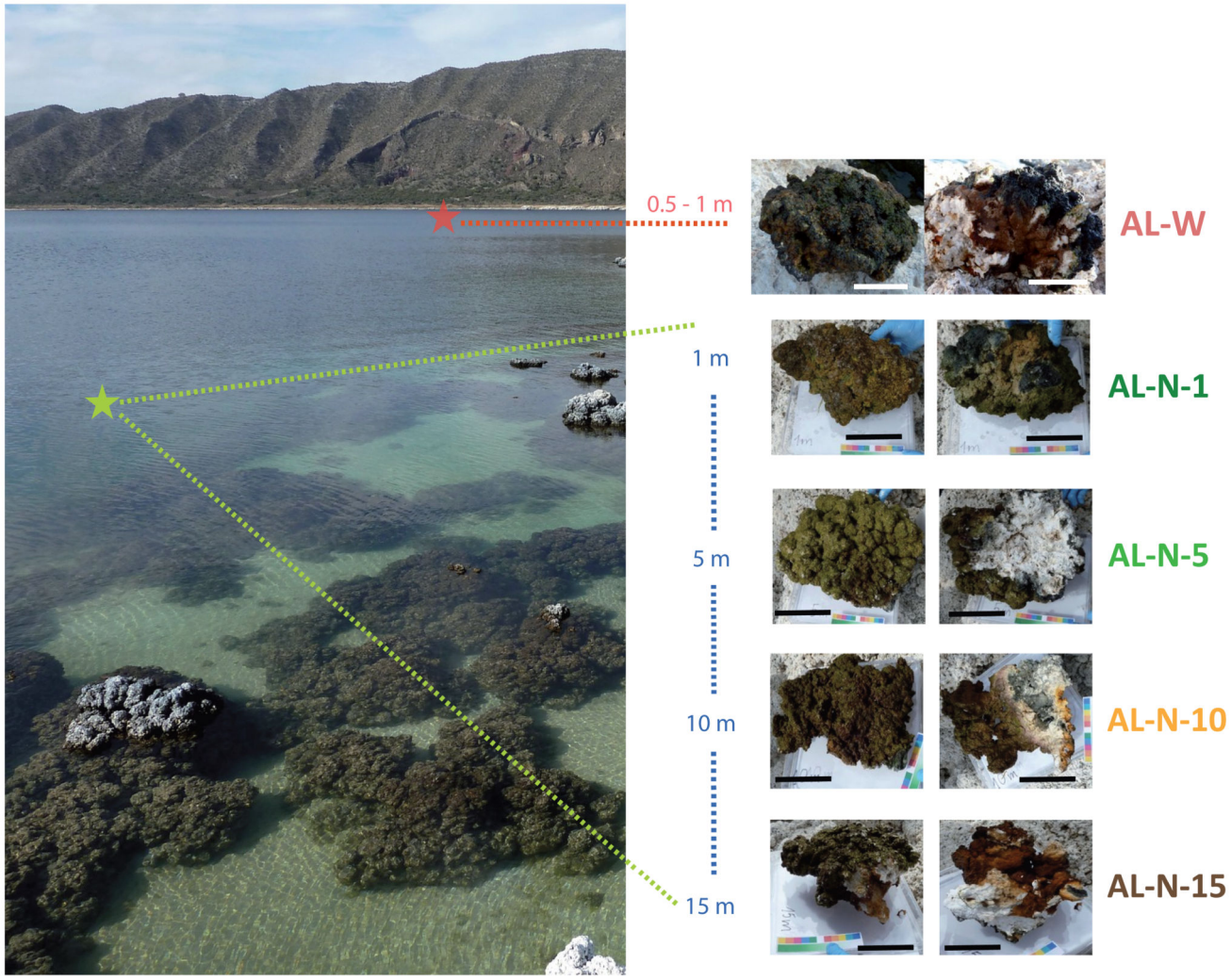

Fig. 1.

Lake Alchichica sampling sites and microbialite fragments collected. The two sampling sites AL-N and AL-W are indicated on the lake with stars. Photographs taken on site of the different samples are shown on the right. For the AL-N samples, the correspondence with the depth gradient is schematically indicated in blue. Upper and downside views of microbialite fragments are shown on the left and on the right part of the panels, respectively. The scale bar corresponds to 10 cm.

Europe PMC Funders Author Manuscripts

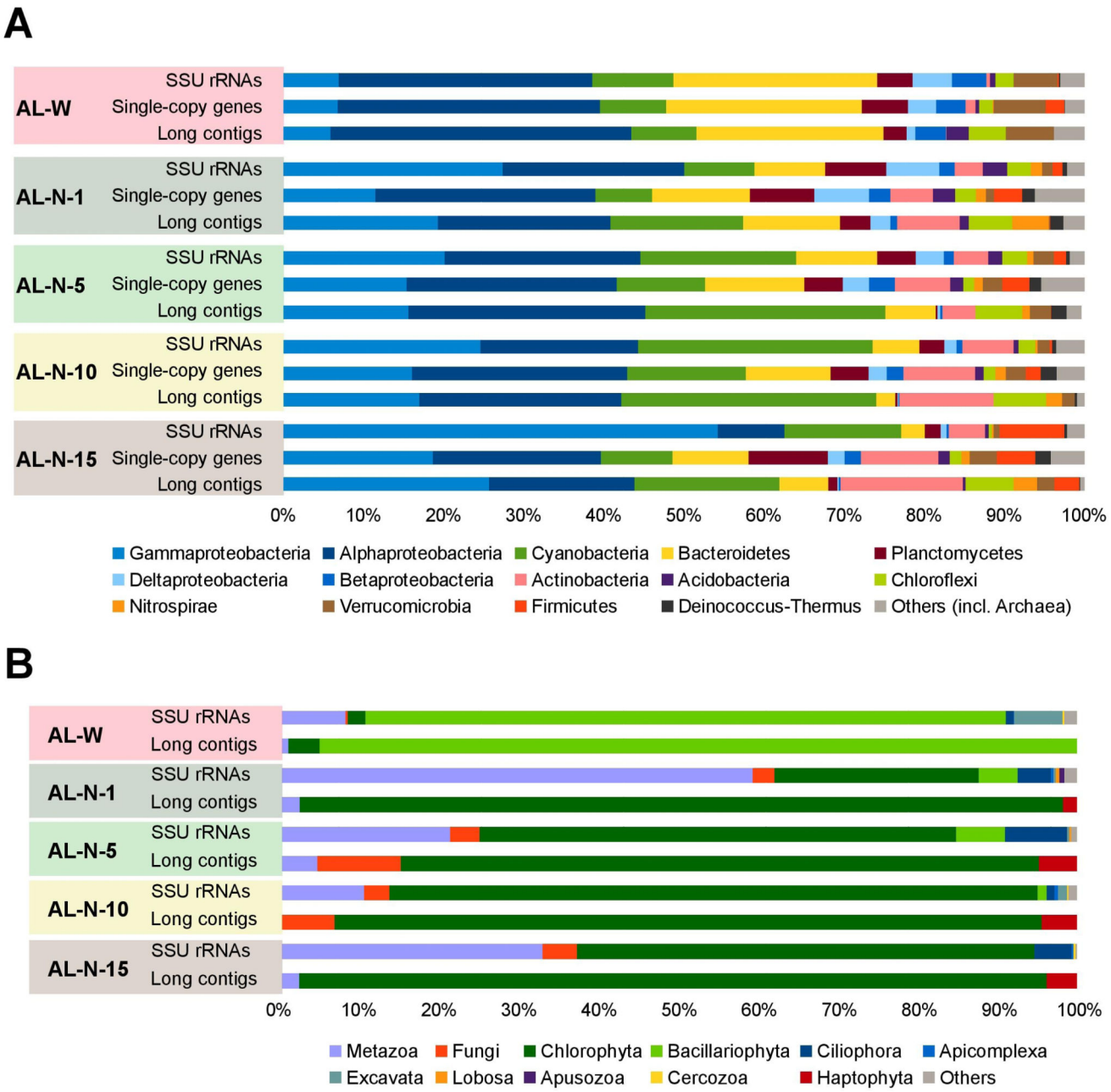

Fig. 2.

Relative proportions of SSU rRNAs genes, 40 single copy protein-encoding genes universally distributed in prokaryotic genomes and sequence lengths in taxon-attributed long contigs (>10 kb) in Alchichica microbialite metagenomes. A, prokaryotic sequences; B, eukaryotic sequences.

Europe PMC Funders Author Manuscripts

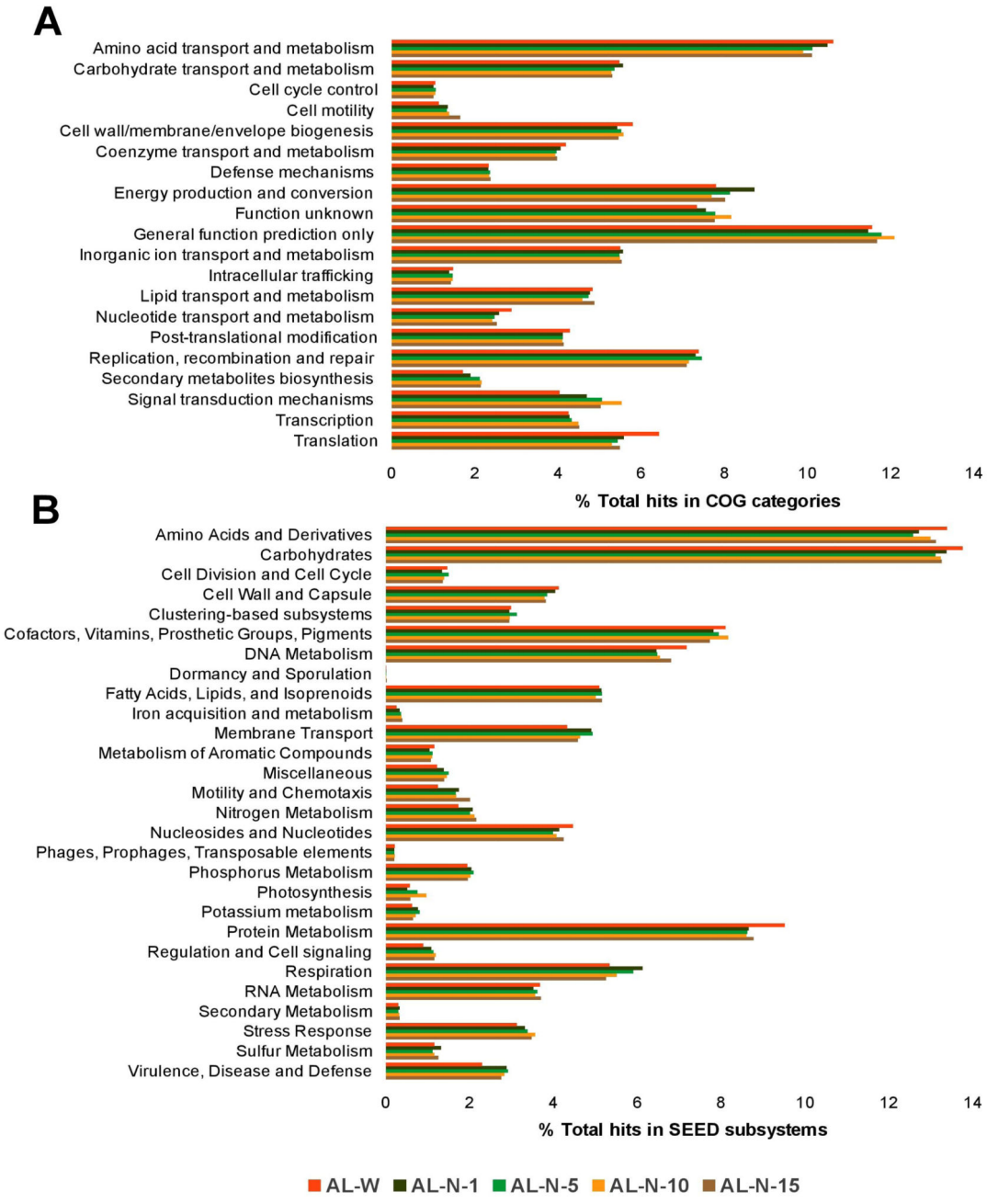

Fig. 3.

Histograms showing the functional characteristics of Alchichica metagenomes based on (A) COG and (B) SEED annotations of the predicted genes. Bars represent the percentage of hits in each functional category normalized to the total number of COG or SEED hits in each metagenome.

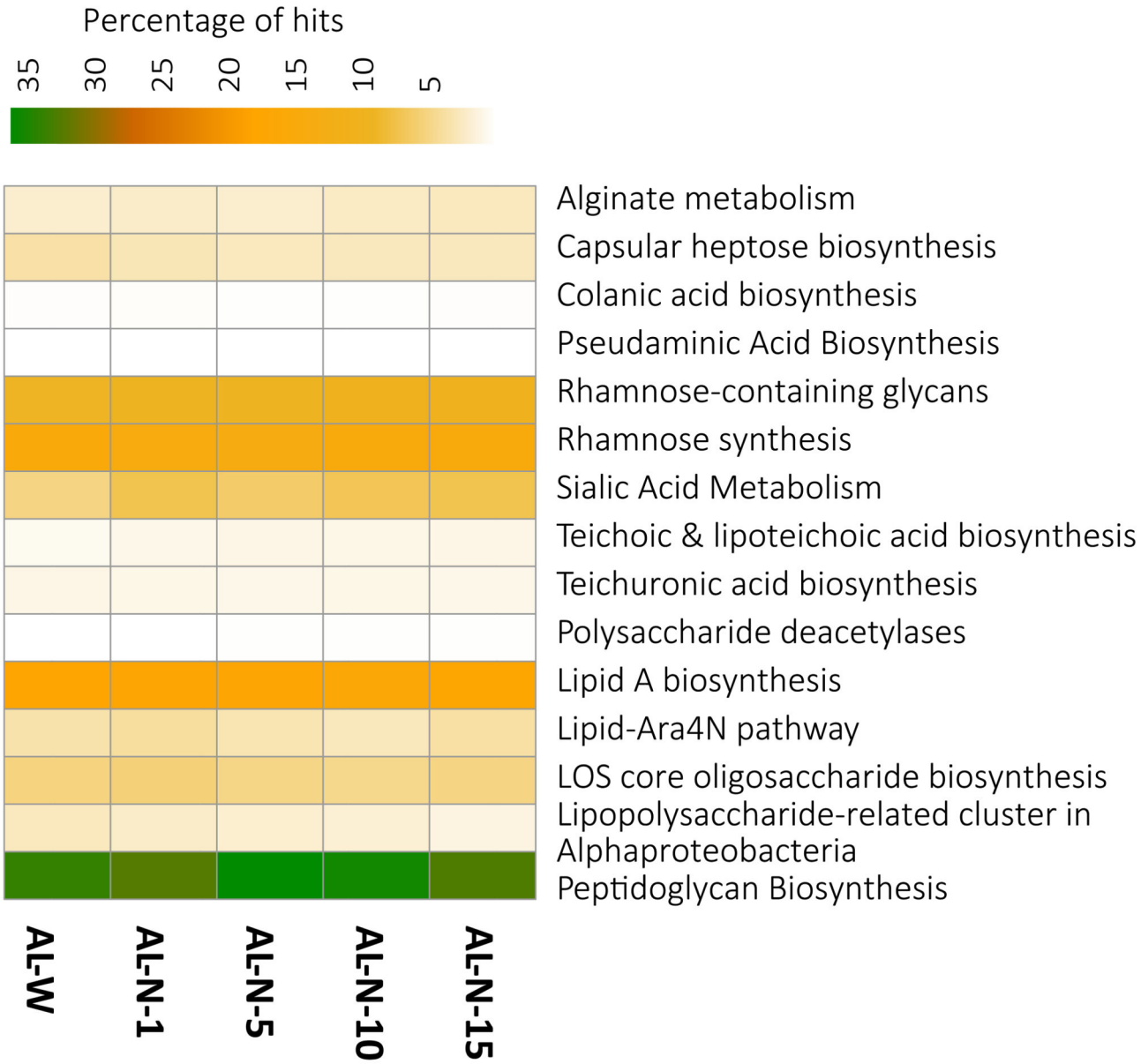

Fig. 4.

Heatmap showing the relative abundance of the genes involved in cell wall and capsule SEED subsystem. Proportions were calculated on the total number of hits associated to the cell wall and capsule SEED subsystem in each metagenome.

Europe PMC Funders Author Manuscripts

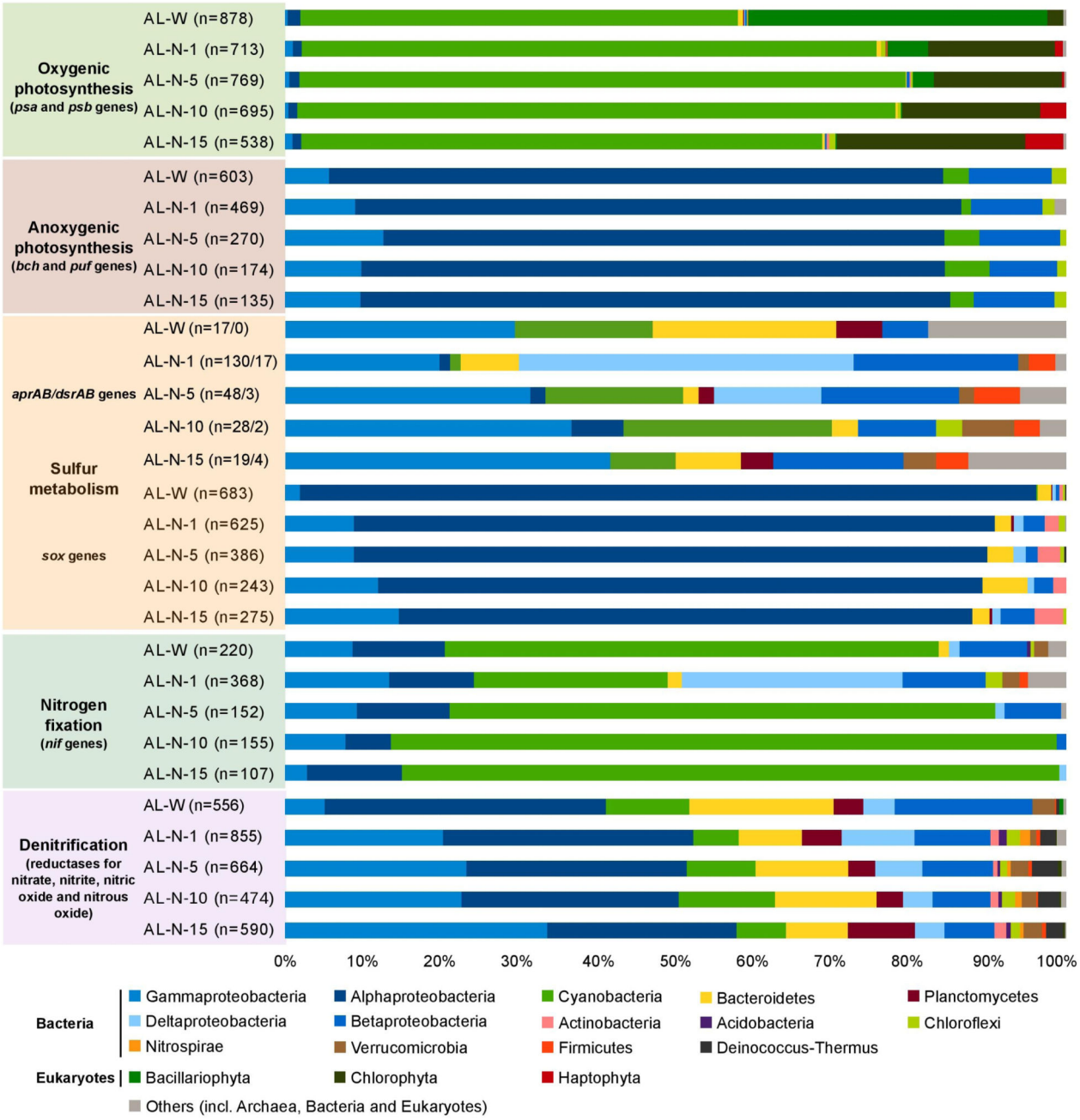

Fig. 5.

Histograms showing the relative proportions of taxa involved in various keys microbialite metabolisms (Dupraz *et al.*, 2009). Data were obtained by mining the RefSeq annotations for genes diagnostic of these metabolisms (see Material & Methods).

Europe PMC Funders Author Manuscripts

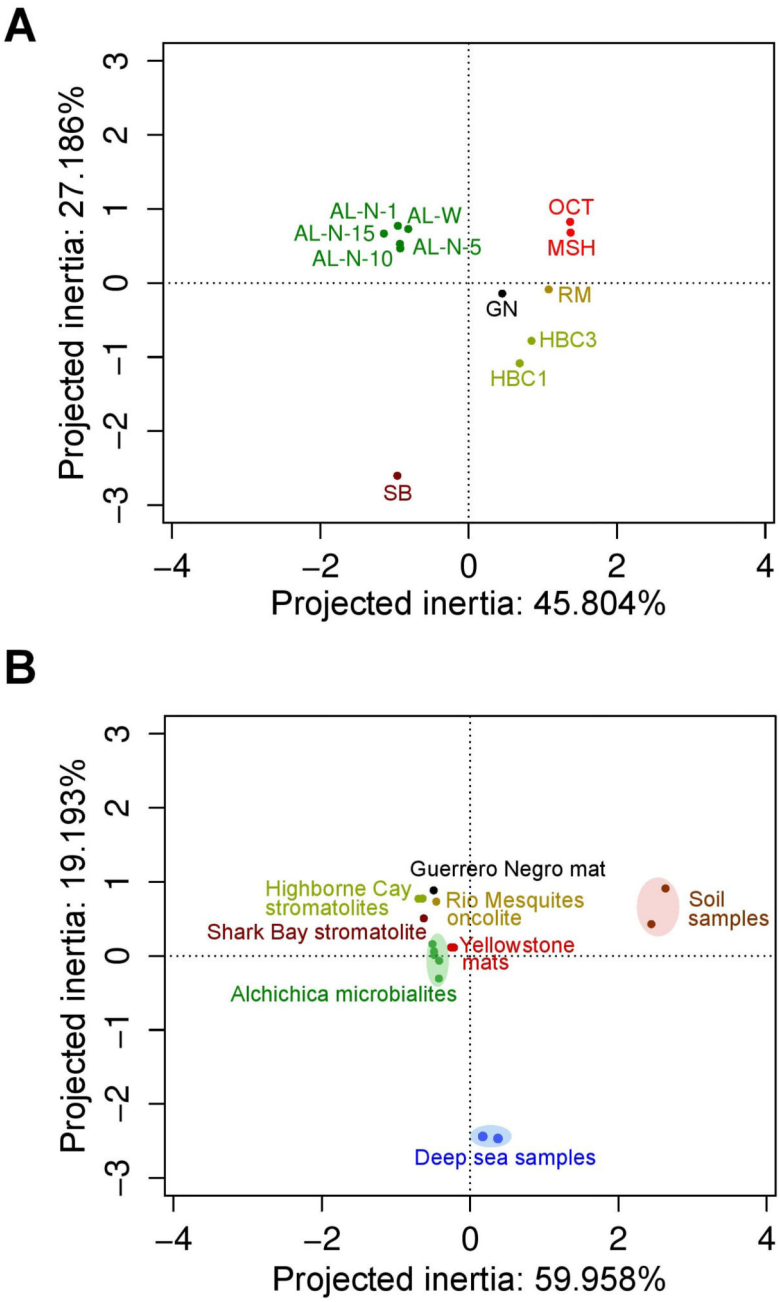

Fig. 6.

Correspondence analysis based on COG categories relative abundances. The five Alchichica metagenomes (AL) were compared to various microbial mat and microbialite metagenomes. (A) Were included microbialites from Highborne Cay (HBC1 and HBC3), Rio Mesquites (RM) and Shark Bay (SB), hypersaline mats from Guerrero Negro (GN) and hot spring mats from Yellowstone National Park (OCT, MSH). (B) Two soil and deep sea planktonic samples were added to the analysis. See Supplementary Table S2 for details on these

metagenomes. To maximize high-level functional differences we removed the «Function unknown» and «General Functional Prediction only» COG categories from the datasets.

Europe PMC Funders Author Manuscripts

Table 1

General features of the Alchichica microbialite metagenomes. AL-W corresponds to the merging of sequences from three replicate Illumina paired-end libraries.

	AL-W	AL-N-1	AL-N-5	AL-N-10	AL-N-15
No. of paired-end reads[1]	306,787,264	130,153,392	110,282,579	97,849,068	131,562,915
No. contigs	1,255,565	1,167,342	737,466	547,418	744,917
Reads assembled into contigs (%)	28	55	52	52	50
Cumulative contig length (bp)	1,405,565,865	1,225,479,094	779,631,887	627,878,408	809,071,698
Smallest / longest contig (bp)	200 / 457,307	200 / 405,581	200 / 496,458	200 / 509,452	200 / 509,451
Mean / median contig size (bp)	1,119 / 622	1,050 / 576	1,057 / 566	1,147 / 592	1,086 / 571
N50 (bp)	1,495	1,379	1,397	1,650	1,486
No. predicted genes	2,142,895	1,998,747	1,263,406	978,667	1,250,850
No. RefSeq hits	1,035,878	1,027,026	773,361	539,137	642,769
No. SEED hits	414,317	451,851	297,923	201,637	242,315
No. COG hits	924,425	942,147	725,454	485,842	569,026
No. single-copy genes[2]	18,203	15,593	11,305	7,333	9,056
No. bacterial 16S rRNAgenes[3]	45,144	14,266	12,431	12,710	32,078
No. 18S rRNA genes[3]	7,253	717	927	1,374	1,876

[1] Raw number of total pairs of reads produced (it implies the same number of forward and reverse sequences)

[2] Total number of occurrences of the 40 gene families found in single copy in prokaryotic genomes

[3] Data from Saghaï *et al.* (2015)