General Patterns of Microbial Diversity Across Samples

Two key advantages result from direct, PCR-free, metagenomic-based diversity analyses. First, it is possible to access to 16S/18S rRNA genes for all organismal diversity simultaneously, including archaea, bacteria, and eukaryotes. Second, because PCR-induced quantitative differences are avoided, the number of rDNA sequences identified in metagenomes reflects the corresponding relative abundance of genes present in the environment. However, although copy number may vary across taxa, especially in the case of eukaryotes, and might not reflect relative cell abundance, quantitative PCR and mock community analyses suggest that rDNA copy number may be a good proxy for biomass (Godhe et al., 2008; Egge et al., 2013). Of course, cell lysis and DNA purification biases might exist, which we tried to minimize by pooling DNA from multiple DNA purification experiments (20–32 for each sample). In the case of Alchichica microbialites, the overall distribution of 16S/18S rRNA genes in the three domains of life confirmed previous ideas suggesting that bacteria dominate this kind of samples, with roughly 85–95% sequences, whereas archaea were excessively minor components (<1%), consistently with other studies (Goh et al., 2009; Couradeau et al., 2011). Eukaryotes varied from 2–8% in the AL-N samples to 10–15% in the AL-W samples (**Figure 3A**). These values are comparable to those found in metagenomic datasets from Cuatro Ciénegas (around 10% eukaryotic sequences; Breitbart et al., 2009) or the Bahamas (less than 13%; Khodadad and Foster, 2012).

We subsequently analyzed the diversity of 16S/18S rRNA gene sequences at a finer level of taxonomic resolution within the bacteria and the eukaryotic domains. Archaea encompassed only a few sequences in all the metagenomes (**Table 2**) and will not be commented further. Both Thaumarchaeota and Euryarchaeota sequences were detected in all samples, except AL-N-10 and AL-N-15 where only Thaumarchaeota were present (data not shown). This corroborated the presence of the two archaeal lineages previously detected (Couradeau et al., 2011). Most reads matched poorly characterized sequences, although at least some Euryarchaeota belonged to the Haloarchaea. Whether these archaea are truly associated to the microbialites or are planktonic archaea from the surrounding waters remains to be determined. By contrast, bacterial and eukaryotic sequences were much more abundant and diverse. The within-domain bacterial and eukaryotic diversity patterns observed for the three AL-W replicate samples were remarkably similar (**Figure 3**; Supplementary Figures S1 and S2). This strongly suggests that the patterns observed for the depth profile (AL-N samples) are also reliable even if replicates could not be obtained for these samples due to the much lower DNA yield.

Although we tried to eliminate little invertebrates feeding on the biofilm, larvae or eggs when collecting our samples, a significant portion of metazoan sequences were detected, ranging from almost 10 to 60% of the total 18S rDNAs depending on the metagenome (Supplementary Figure S1). These essentially affiliated to Arthropoda (mainly Acari), Annelids, and Nematoda. Even if some metazoans are truly living in close association with the microbialites within the more superficial, less mineralized zones of the biofilm, e.g., nematodes, and might play a control on the microbial community (Tarhan et al., 2013), many of the identified sequences corresponded to arthropods or other exogenous animals that skewed the relative abundance of the various taxa toward metazoan members. Therefore, we subsequently excluded metazoan sequences from our datasets to compare the relative distribution of the rest of eukaryotic taxa (**Figure 3B**). The eukaryotic diversity covered a large variety of high-rank taxa in all the samples, but was numerically dominated by photosynthetic taxa that contributed between 75 and 95% of the microbial eukaryotic sequences. However, whereas AL-W metagenomes were dominated by diatom sequences (almost 90% of microbial eukaryotic sequences), samples from the AL-N vertical profile were all dominated by green algal sequences (**Figure 3B**). This radical difference between the two sites does not appear related to the mineral composition of the substrate (**Table 1**) but may be related to the local hydrochemistry, which seems locally influenced by seepage activity and perhaps also by occasional spring discharge at the steepest part of the crater. Indeed, the chemical analysis of seepage water collected at this site in 2007 revealed much less saline (668 mg/l of total dissolved solids versus ca. 8,800 mg/l for the lake water) but far more Si-rich water. Si was on average 40 times more concentrated (694 μmol/l) than the rest of the water lake (Kazmierczak et al., 2011). This local Si input is likely determinant for the dominance of diatoms in these samples. The differences in the eukaryotic component of the surface biofilm may partly explain the distinct macroscopic aspect of the biofilm (**Figure 1**). In addition to diatoms and green algae, likely photosynthetic euglenids (Euglenozoa, Excavata) were present in the two sites, having a noticeable percentage (6–8%) in the AL-W site. The heterotrophic component of the microbial eukaryotic community was more varied and encompassed protist grazers (ciliates, amoebae, cercozoans, or apusozoans) and fungi. Some apicomplexan sequences were identified, which might correspond to parasites (**Figure 3B**). The analysis of singletons revealed the presence of some groups not detected in high abundance; for instance, haptophytes were identified in AL-N-1, AL-N-10, and AL-N-15 (Supplementary Figure S2B).

Metagenome sequences were largely dominated by bacteria, which were also highly diverse, encompassing numerous phyla (**Figure 3C**). Although in varying proportions, Cyanobacteria, Gammaproteobacteria and Alphaproteobacteria, and Bacteroidetes were generally the most abundant groups, collectively accounting for up to 70 or 80% of all bacterial sequences, depending on the metagenome. Gammaproteobacteria were nevertheless much less abundant in AL-W samples (<10% bacterial sequences) as compared to the AL-N samples, where they accounted for 20% up to more than 50%. Bacteroidetes, on the contrary, were more abundant in the AL-W samples, accounting for 20% of sequences. Cyanobacteria were less abundant in surface microbialites (around 10%), but increased their proportions to 15–30% in deeper samples. Other relatively abundant groups (5–10% sequences) comprised the Planctomycetes, Deltaproteobacteria, Chloroflexi, Verrucomicrobia and, in the AL-N samples, also the