Metagenomes of Mediterranean Coastal Lagoons

Rohit Ghai[1], Claudia Mella Hernandez[1], Antonio Picazo[2], Carolina Megumi Mizuno[1], Karolina Ininbergs[3], Beatriz Díez[4], Ruben Valas[5], Christopher L. DuPont[5], Katherine D. McMahon[6], Antonio Camacho[2] & Francisco Rodriguez-Valera[1]

[1]Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, Alicante, Spain, [2]Cavanilles Institute of Biodiversity and Evolutionary Biology – University of Valencia E-46100 Burjassot, Spain, [3]Department of Botany, Stockholm University, Stockholm, Sweden, [4]Depto. Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Alameda 340, Santiago, Chile, [5]J Craig Venter Institute, San Diego, CA, USA, [6]Departments of Civil and Environmental Engineering and Bacteriology, University of Wisconsin Madison, Madison, USA.

Coastal lagoons, both hypersaline and freshwater, are common, but still understudied ecosystems. We describe, for the first time, using high throughput sequencing, the extant microbiota of two large and representative Mediterranean coastal lagoons, the hypersaline Mar Menor, and the freshwater Albufera de Valencia, both located on the south eastern coast of Spain. We show there are considerable differences in the microbiota of both lagoons, in comparison to other marine and freshwater habitats. Importantly, a novel uncultured sulfur oxidizing Alphaproteobacteria was found to dominate bacterioplankton in the hypersaline Mar Menor. Also, in the latter prokaryotic cyanobacteria were almost exclusively comprised by *Synechococcus* and no *Prochlorococcus* was found. Remarkably, the microbial community in the freshwaters of the hypertrophic Albufera was completely in contrast to known freshwater systems, in that there was a near absence of well known and cosmopolitan groups of ultramicrobacteria namely Low GC Actinobacteria and the LD12 lineage of Alphaproteobacteria.

C oastal lagoons are shallow water bodies separated from the sea by a barrier, connected at least intermittently to the sea by one or more restricted inlets, and usually oriented parallel to the shore. The formation of the barrier is crucial, as it allows lagoon waters to acquire significantly different characteristics compared to the nearby seawater. Mediterranean coastal lagoons commonly are not affected by significant tidal influences as tides in the Mediterranean Sea are very low. This avoids the diel inputs of seawater that are common in oceanic salt marshes. Because of the relatively increased isolation from the sea and their location within a hydrological catchment, these lagoons also become more susceptible to changes in salinity, dissolved oxygen, nutrient content, largely owing to the increased effect of evaporation in a restricted area, leading to increased salinity, and deposition of various salts (e.g. calcium carbonate), as well as to the strongest influence of the surrounding land. Because of the high population density in coastal Mediterranean areas, these lagoons are usually impacted by agricultural, mining, tourism and general developmental activities leading to the lagoon becoming a common sink for a wide variety of waste material[1]. These differences are also reflected in the organisms that inhabit these lagoons, which may largely contrast with those of the nearby marine environment. Such lagoons are very common environments in flat areas along the Mediterranean coasts, and may range from small to very large size.

Albufera de Valencia (from its Arabic name البحيرة *al-buhayra*, "the little sea"), located a few Km south from the city of Valencia, Spain (39°19′54″N, 0°21′8″W) is a shallow (1 m on average) coastal lagoon which nowadays holds freshwater. It was originally a marine harbour that got progressively separated from the sea by a sand strip growing from North to South, due to the dominant marine currents and the deposition of river sediments, later becoming a large brackish coastal lagoon nearly 300 km² in size since the Roman times up till the 18th century[1]. However, in the second half of the 19th century, nearly 60% of the lake was filled up to reclaim land for the cultivation of rice, and this decline continued, and today the lake size has shrunk to ~23 km², nearly 15 times less than its original size[2]. During the course of this regression, freshwater inflow to the lagoon increased, owing to increased rice cultivation, and the development of irrigation within the catchment (917 km²) croplands and today it is surrounded by ~223 km² of rice fields that largely determine its hydrological functioning. Moreover, the

outgoing connection to the sea became controlled by hydraulic gates and inflow of sea water was totally stopped, leading to the complete conversion of a once brackish ecosystem into a freshwater one. Because of the increased human activities in its densely populated surroundings, Albufera collapsed as a macrophyte-dominated lagoon and turned into a highly hypertrophic ecosystem with very dense phytoplankton populations primarily dominated by cyanobacteria[3].

The Mar Menor (also meaning "Little Sea" in Spanish) another huge coastal lagoon of nearly 135 km[2] in surface and somewhat triangular in shape, is among the largest lagoons in the Mediterranean. Located in the region of Murcia (Spain, 37°43′08″N, 00°47′14″W), it is separated from the sea by an extremely thin and nearly 20 km long strip of land (called La Manga), and is hypersaline (∼5%). Mar Menor receives water from a number of sources, mainly small streams flowing into the lagoon (usually seasonal), run-off from mining activities, wastewater treatment plants overflows, agricultural land runoff and from urban development and tourist activities[4]. This area of the Spanish coast is also among the most threatened by the rise in global sea levels[5]. Both Albufera and Mar Menor have received, and still partly receive substantial amounts of wastewaters that have led to a high trophic status of both waterbodies, though strongest inputs were historically received by Albufera, and led this lagoon to a extreme hypertrophic status and maintain a very high internal load of nutrients in its sediments in addition to external inputs[1] that support extensive algal growth. By studying these two systems, we cover the most representative ecological types of Mediterranean coastal lagoons, namely freshwater lakes dominated by continental hydrological processes like Albufera, and hypersaline lagoons maintaining wider connections with the sea but having higher salinity because of evaporative processes, like Mar Menor. Good examples of saline coastal lagoons are those located at the Languedoc-Roussillon region, in France, where a series of coastal lagoons mostly connected with the sea are widespread through the Mediterranean coast. Contrastingly, Albufera de Valencia probably represents the best example of a freshwater coastal lagoon in the Mediterranean.

The marine environment has now been the focus of a large number of metagenomic studies and they have already succeeded in providing us with a partial view of the organisms in this habitat[6–11]. The contribution of marine picoplanktonic cyanobacteria to global oxygen levels[12], the superabundance of Candidatus Pelagibacter species[13], vertical patterns in microbial diversity[10] and several other studies have served to illustrate the diversity and the of the marine microbial world. Moreover, salinity has been shown to be an extremely important factor in distribution of microbes[14], much more than temperature and pH. Saline environments have been well studied, using several different methodologies, 16S rRNA genes, cultures and metagenomics approaches[15–22]. However, most saline environments examined so far have been primarily solar salterns, which are usually controlled environments, shielded from extraneous inputs (e.g. urban waste). Coastal lagoons, on the other hand, are more susceptible to vagaries of natural and human origin. Thus, in spite of the widespread nature and importance of hypersaline coastal lagoons, we know very little about the microbial composition of these ecosystems. The same also occurs for freshwater coastal lagoons, whose microbiota, compared to that of the open sea, has been scarcely studied, even less using modern metagenomic approaches. A lack of information is evident for the composition of the microbiota of Mediterranean coastal lagoons, but new generation sequencing methodology promises to offer a view of the microbial diversity and to unmask the environmental characteristics that would act as selective factors in determining the microbiota of the two main types of such ecosystems. This will also allow us to compare our metagenomic description of the microbial component of the plankton, with a series of still relatively scarce metacommunity data from other aquatic ecosystems, both saline and freshwater.

On the other hand, freshwater ecosystems, like the ubiquitous marine SAR11 lineage in oceans, also have their own characteristic abundant microbes i.e. Low GC Actinobacteria[23–25], Betaproteobacteria (e.g. Polynucleobacter[26–29]) and the LD12 clade of Alphaprotebacteria[30.31] (related to marine SAR11). These are usually threatened ecosystems, facing ever increasing pressure due to continued human activities. Indeed, freshwater ecosystems (rivers and lakes together) comprise only 0.266% of all freshwater on earth[32]. In this small percentage of available freshwater, several factors influence microbial diversity in lakes e.g. trophic status[33,34], pH[25,35], landscape[36] and water retention time[35]. While lagoon salinity is indeed a major driver of microbial diversity[37], another important physical characteristic that is of importance is depth of the lagoons, as shallow water bodies have better light penetration, faster nutrient recycling and higher primary productivity. The average depth of Albufera is quite shallow (∼1 m), while Mar Menor is comparatively deeper (∼5 m). Though there has been much work on lakes using 16S rRNA and culture dependent approaches, recently, culture independent approaches have also begun to shed light on these systems[8,24,38,39]. However, eutrophic freshwater lagoons have also not been studied using high throughput metagenomic sequencing, (indeed there are no metagenomic datasets from a eutrophic freshwater system yet) and given the biases in culture-dependent and culture-independent approaches[7,40] it is important to study these habitats using less biased methods. In addition, both these lagoons, Albufera and Mar Menor, are eutrophic systems[1,4], especially the former. And both kinds, hypersaline and freshwater coastal lagoons are particularly widespread along the Mediterranean coast and indeed all over the world. We have very little information on the microbes of these habitats, and very few studies have been done with Mediterranean coastal lagoons in particular. Some studies have focused on cell counts e.g. for picocyanobacteria, in several coastal lagoons of the Adriatic[41], and some others have been all 16S rRNA surveys, e.g for lagoons on the French Atlantic[42] and Mediterranean coasts[43], and also for the large lagoon of Venice[44,45], which is a particularly productive environment, connected to the Adriatic Sea.

As part of the Global Ocean Sampling expedition, we have used high-throughput metagenomic sequencing to investigate the microbial diversity of the Albufera and Mar Menor lagoons and compare them with other closely related aquatic environments. To compare with the hypersaline Mar Menor metagenome, we have chosen three metagenomic datasets from saline environments i.e. the Mediterranean Deep Chlorophyll Maximum (referred to as DCM), a marine metagenome[7], of relevance for this case because the Mediterranean Sea is the primary water input into Mar Menor, and a metagenome from a very shallow hypersaline lagoon (salinity 6%, depth 0.3 m), called Punta Cormoran (referred to as PC6) in Galapagos Islands[8], as it is the only other hypersaline lagoon dataset available, and also because it is a relatively pristine environment. In addition, a 19% salinity dataset from a solar saltern (referred to as SS19) is also included as an extremely hypersaline environment[17]. For comparison with the freshwater Albufera dataset, we have included three metagenomes, two from lentic and another from a lotic habitat. The lentic datasets include Lake Lanier[46] (Georgia,US; forms the primary drinking water supply for the Atlanta metropolitan area)and Lake Gatun[8](located in the middle of the Panama Canal). The lotic metagenomic dataset is from the pristine upper water column of the Amazon River[24]. We compare the Mar Menor and Albufera metagenomes to these metagenomes in order unmask ecological factors that could be related to the microbial composition of their respective communities.

Results

Physico-chemical characteristics. The locations where the samples were taken are shown in Supplementary Figure S1. Some physico-chemical and biological properties of the samples are described in

Supplementary Table S1. Salinity, as represented by electrical conductivity, is much higher in Mar Menor than in the Mediterranean Sea. Albufera waters, however, appeared as highly mineralized freshwaters, showing a certain influence of the sea, with values of 2.8 mS cm^{-1}, as compared to freshwaters from the area, which commonly show conductivities of around 1 mS cm^{-1}. Even though Albufera is well separated from the sea, open inlets controlled by hydraulic gates sometimes allow some connection, and aquifers providing water to the lagoon are slightly influenced by marine waters. This demonstrates that we have chosen the two sides of the main environmental condition in determining the ecology of coastal lakes, this is, salinity, with both a highly saline and a freshwater lagoon. Similar to well mineralized waters, both samples were mildly alkaline (pH was 8.4 for Mar Menor and 7.69 for Albufera), but alkalinity (and bicarbonate concentrations) in Albufera was lower compared to surrounding freshwater systems. The lower alkalinity in Albufera is mainly due to the high rates of planktonic primary production of such hypertrophic system that uses large amounts of inorganic carbon, thus decreasing the alkaline reserve mainly formed by bicarbonate. Saline content of Albufera, though much lower than that of Mar Menor, is quite balanced in anions between bicarbonate, chloride and sulphate, whereas that of Mar Menor is much higher, and mostly due to chloride. These data indicate the differences in relative importance of continental and marine inputs in these two systems.

Total nitrogen (TN) and total phosphorus (TP) concentrations, taken together with chlorophyll concentrations, better reflect the extent and effects of eutrophication on both lagoons, as they show the amount of nutrients that are incorporated into biomass, mainly phytoplankton, in the form of particulate nutrients. Both TN and TP were around two and a half times higher in Albufera than in Mar Menor, showing that, in addition to salinity, these two systems also maintain a large difference in another quite important environmental feature, namely, the trophic status.

Chlorophyll-a concentration reveals even higher differences than TN and TP, as the chlorophyll levels of Mar Menor (3.94 μg/l) are actually very similar to that of the DCM of the Mediterranean[7] (3.4 μg/l), while Albufera displayed levels corresponding to extremely hypertrophic conditions (271.31 μg/l). Following OECD criteria[47], all these values categorized Albufera as a hypertrophic system, whereas those of Mar Menor correspond to a mesotrophic system but with a strong trend towards eutrophication as indicated by concentrations of soluble nutrients. Remarkably, when considering both nitrogen and phosphorus, these nutrients are mostly included within the particulate fraction in the Albufera, with comparatively low amounts in the soluble forms of phosphorus (soluble reactive phosphorus, mainly orthophosphate) and, even lower, of nitrogen (ammonia), compared to overall amounts that are mainly owed to the biomass of phytoplankton, as shown by Chl-a concentrations. Because of its long residence time during most of the year, Albufera acts as a bioreactor that converts most of the incoming nutrients in phytoplankton biomass, most of which is later retained in the sediments and represents a strong internal load that further supports hypertrophic conditions. Moreover, most of this phytoplankton biomass is composed by cyanobacteria, as shown by the dominance of taxa-specific carotenoids (Supplementary Table S1) from these phytoplanktonic organisms, such as zeaxanthin, as was further confirmed by microscopic and molecular analyses. Contrastingly, most nitrogen and phosphorus in Mar Menor was detected as soluble forms, with relatively low levels of phytoplankton biomass that are still comparable to productive areas of the sea, such as the DCM, but much poorer compared to Albufera. Ammonium is, in contrast to Albufera, the main form of nitrogen in the waters of Mar Menor. The very high planktonic biomass in Albufera quickly assimilates available nutrients, especially those which are limiting, and ammonium is the preferred form of nitrogen to be assimilated by organisms as it

has the same redox status than organic nitrogen. In Mar Menor, however, the high availability of soluble (biologically available) forms of nitrogen and phosphorus compared to the low chlorophyll levels indicates the occurrence of recent peaks of nutrient inputs into this lagoon, occurring briefly before the sampling, that have not yet had the time to be converted into biomass. Massive occasional nutrient inputs are a common feature of this lagoon and are associated to time-restricted discharges of wastewaters[48] or increased agricultural runoff linked to heavy rains. These inputs commonly cause algal blooms that are associated with such nutrient dynamics[4]. Recent modelling estimated that, only accounting from agriculture sources associated to irrigations procedures, more than 2000 tonnes of nitrogen and around 60 tonnes of phosphorus enter per year in the Mar Menor, which, together with other sources, such as urban wastewaters, explain the high levels of soluble nitrogen found in this lagoon. Additionally to this modeling, previous empirical evidence of the high amounts of nutrients received by Mar Menor was given by Velasco et al[49], who during a hydrological cycle measured nutrient inputs as high as 2010 tones of inorganic nitrogen and 178 tonnes of soluble reactive (biologically available) phosphorus in a year. Thus, our measurements of dissolved inorganic nitrogen, even if chlorophyll concentrations are not so high, reveal a relatively high (mesotrophic to eutrophic) trophic status of Mar Menor compared to the coastal waters of the nearby Mediterranean Sea, though much lower than that of Albufera, where nutrients are likely quickly bioconverted into phytoplankton biomass.

Phytoplankton diversity and abundance. In contrast to the very different abundance of phytoplankton (quantified as Chl-a concentration), both systems showed similar densities of heterotrophic bacterioplankton (in the range of 4–5 10^6 cells per ml), higher than those commonly found in surface waters of the Mediterranean Sea[50,51]. However the abundance of phototrophic picoplankton, mainly unicellular *Synechococcus*-like cyanobacterial cells, was almost twenty times higher in La Albufera than in Mar Menor. These autotrophic picoplankton (APP) cells are similar to those of surface waters, phycocyanin-rich cells mostly lacking phycoerythrin[52]. However, although APP abundance is much higher in Albufera, they represented up to 9.4 % of phytoplankton biomass (biovolume) in Mar Menor. This contribution was 3.3 % in Albufera, where filamentous cyanobacteria, diatoms and chlorophytes accounted for most of the biomass (Supplementary Figure S2). The relatively high diversity of phytoplankton in Albufera (Figure 1, Supplementary Figure S3) revealed by our sampling is a relative novelty in this lake within the last years associated with sewage diversion[53] compared to the previous decades, when filamentous cyanobacteria, like *Planktothrix agardhii, Pseudanabaena galeata* and *Geitlerinema* sp. widely dominated the community[54]. This relatively high diversity related to increased relevance of chlorophytes and diatoms compared to cyanobacteria is also shown by taxa-specific pigments. In addition to the high concentrations of the cyanobacterial-specific carotenoid zeaxanthin, high concentrations of the diatom-marker carotenoid fucoxanthin were also found (Supplementary Table S1). The high contribution of chlorophytes in terms of total phytoplankton biomass, mostly due to the presence of very big colonial species of *Pediastrum* (*P. boryanum* and *P. duplex*), which at the time of sampling accounted for 46.6% of total phytoplanktonic biovolume (Figure 1; Supplementary Figure S2) but only for 1.3 % of phytoplankton individuals, is likely the reason that chlorophyte-specific carotenoids are not so abundant. Sewage diversion, together with increased flushing during some periods associated to rice cultivation, sometimes promotes clear water phases in late winter and spring, as it occurred in 2010, when sampling was performed and the more evident clear water phase has been reported for the last four decades. Contrastingly, Dinoflagellates
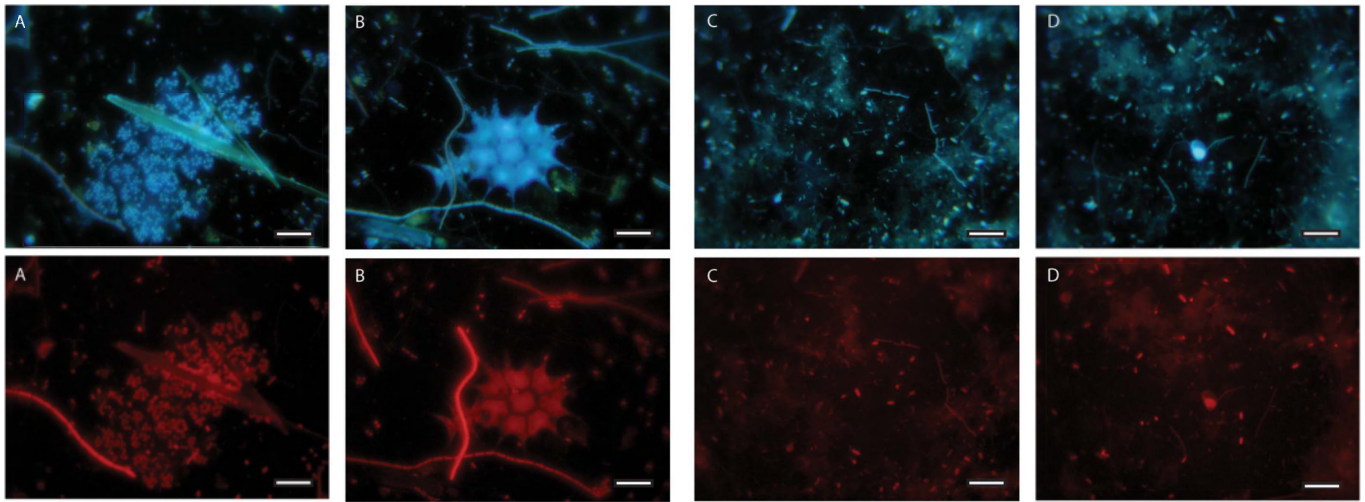
Figure 1 | Pairs of microphotographs, DAPI stain (blue, up) and photosynthetic pigment autofluorescence (red, down) of samples from Albufera (A and B) and Mar Menor (C and D) showing different microorganisms. A) a colony of unicellular picocyanobacteria B) several filamentous cyanobacteria and coenobia of the chlorophytes *Pediastrum* sp. and *Scenedesmus* sp. C) Different morphologies of heterotrophic bacterioplankton (cells not showing red autofluorescence in lower pictures) and autotrophic picocyanobacteria (cells showing red autofluorescence in lower pictures). D) Heterotrophic bacterioplankton and autotrophic picocyanobacteria with a eukaryotic nanoflagellate. White bar corresponds to 10 µm in all pictures.

dominated by far phytoplankton in Mar Menor, both in terms of total phytoplankton biomass and number of cells (excluding APP for the later count), with also relevant contributions of diatoms and unicellular picocyanobacteria (Figure 1; Supplementary Figure S2). These are also reflected in the abundance of the taxa-specific carotenoids (Supplementary Table S2), which, although at much lower concentrations than those of Albufera, also shows the relative importance of the dominant phytoplankton groups. Neither Albufera nor Mar Menor hold planktonic anoxygenic phototrophic bacteria, as revealed by the absence of bacteriochlorophylls.

GC Content. We obtained nearly equal amount of sequence data from each one three different filter sizes for each dataset (0.1, 0.8 and 3.0 µm, See Supplementary Table S2). The sequence data from the three filters of Mar Menor shows some differences in GC content (Supplementary Figure S4), with the 3 µm filter showing a high GC peak, likely because of the increased number of eukaryotic sequences captured in this filter. A comparison of the GC content of the two smaller filter sizes (0.1 and 0.8µm) with other available marine and hypersaline metagenomes (DCM, PC6 and SS19) is shown in Supplementary Figure S4. The Mar Menor metagenome shows a single distinct peak at ~50%, similar to the marine metagenome in being unimodal, but of very different GC% and a broader GC range, and also distinct from the other hypersaline datasets which have clear bimodal GC distributions (both PC6 and SS19). All the hypersaline metagenomes do have at least a single peak at around 50% GC. The figure indicates that across a range of salinities, (from 3.5% to 19%) a diverse range of GC content may be found. Moreover, Mar Menor GC distribution appears to be quite different from PC6, although both habitats have nearly identical salinity (however, as no other physical-chemical data is available for PC6 dataset apart from salinity so the factors relating to these differences cannot be adequately discussed). There does not appear to be an abundance of very high GC organisms (~70% GC as in PC6) in Mar Menor (Supplementary Figure S4). On the other hand, sequences from all three filters of Albufera tended to show a GC profile skewed towards high GC content (Supplementary Figure S4). Comparison to three other freshwater datasets (Lake Gatun in Panama, Lake Lanier in Atlanta,US[46], and the River Amazon[24]) (Supplementary Figure S4), does not show any kind of clear pattern, apart from a low GC peak (~45–50%) in all datasets except Albufera. So in this initial examination, the GC% profiles of both Mar Menor and Albufera appear quite different from other metagenomic datasets, and this already is an indication of the different communities in these ecosystems compared to other related available datasets.

Community Structure. Among prokaryotes, the results of classification of the 16S rRNA sequences and all reads comparison to the NR database indicated almost exclusively the presence of Bacteria (Supplementary Figures S5, S6 and S7; Supplementary Tables S3, S4, S5 and S6). No archaeal 16S rRNA reads were detected in the Mar Menor dataset, and an extremely low number (<1%, n=322) of all metagenomic reads could be assigned to Archaea in Albufera. This extremely low fraction of reads from Albufera was assigned primarily to Euryarchaeota. This is indeed a little unusual, as Archaea are usually at least minor components of most systems (with exceptions, e.g. solar saltern crystallizer ponds), typically in the range of 5–10%[24] but in Mar Menor we have barely detectable levels of archaeal sequences.

Phages. Among the most abundant organisms recruiting the maximum number of reads from the Mar Menor metagenome was a viral genome, that of *Roseobacter phage SIO1*. (see Supplementary Table S3). Roseophages are lytic podoviruses of Roseobacters, first isolated for *Roseobacter SIO67*, an aerobic, heterotrophic alphaproteobacterium. The currently sequenced Roseophages have been isolated from California near-shore locations. Comparative genome analysis of Roseophages has revealed largely conserved genomes, with three distinct pockets of variability (thyX gene, phosphage metabolism genes and structural genes like the tail-fiber protein). However, our sequence data indicates the presence of a population of organisms belonging to the order Rhodobacterales (see 16S rRNA section above). The average %identity of the metagenomic hits mapping to the Roseobacter genome was ~40%, i.e. rather low, so the dominant phage might be an abundant podovirus, similar to Roseophages, but its host specificity is as yet uncertain, as the host itself is as yet undescribed. In comparison to the nearly 11% reads in Mar Menor metagenome being assigned to phages, only ~3.6% reads could be assigned to phages in Albufera. Even then, a phage genome, *Prochlorococcus* phage P-SSM2 appeared as a genome that recruited several hits in Albufera (Supplementary Table S4). P-SSM2 is a myovirus, that is specific for cross-infections between *Prochlorococcus* strains[55]. However, there is no *Prochlorococcus* population in
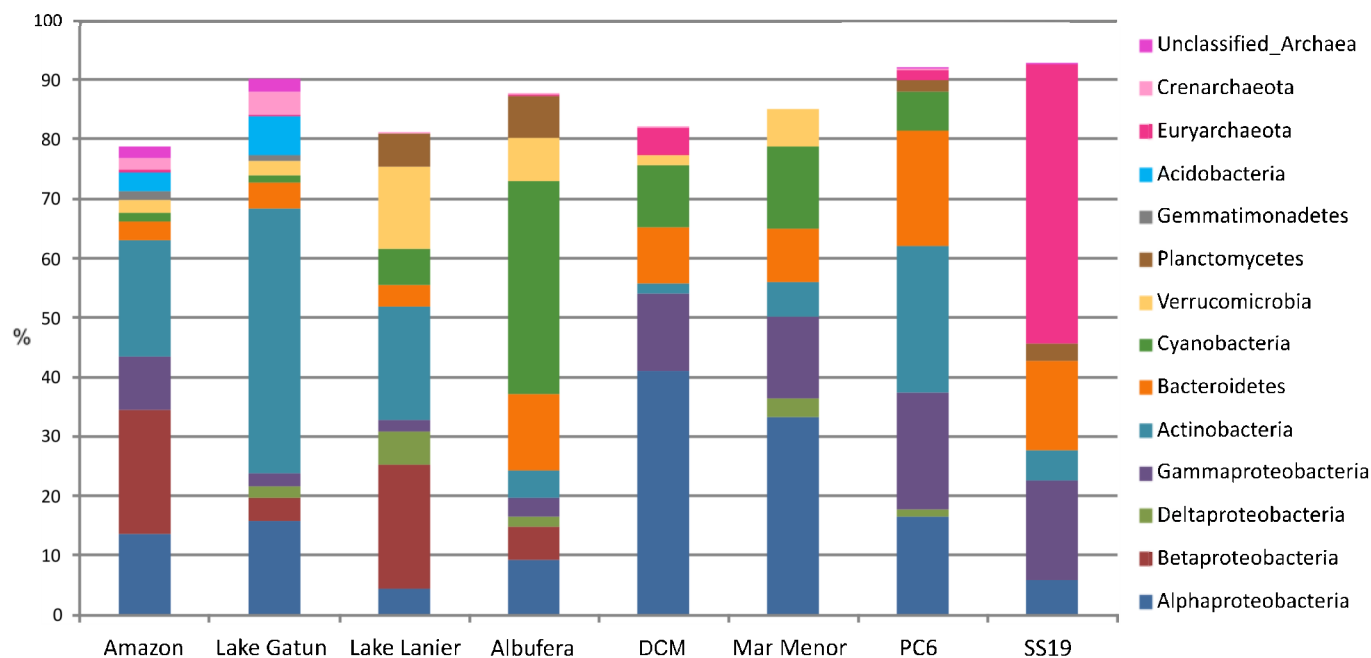
Figure 2 | Cross comparison of comparative distribution 16S rRNA sequences from selected abundant high level bacterial taxa from Mar Menor and Albufera metagenomes to several freshwater and saline metagenomes. Results from all filters have been combined.

Albufera, so these reads likely belong to an abundant myovirus, which might be infecting the abundant *Synechococcus* or even *Cyanobium*.

Alphaproteobacteria. Alphaproteobacteria form a large part (~33%) of the community in Mar Menor (Figure 2), similar to the DCM. The marine metagenome of the DCM is dominated by *Candidatus* Pelagibacter (belonging to the SAR11 cluster). In Mar Menor as well, the majority of alphaproteobacterial 16S sequences could be ascribed to the SAR11 cluster, and nearly half (43.6%, n=69) of all 16S reads to which we could assign a tentative genus could be affiliated to *Candidatus* Pelagibacter (Supplementary Table S5). Moreover, *alphaproteobacterium HIMB114*, also a member of the SAR11 cluster (a marine microbe, isolated from Hawaii) was among the organisms that recruited the maximum number of reads from the metagenome (Supplementary Table S3). These results point towards the abundance of a SAR11 representative in Mar Menor. In addition to these organisms (both belonging to the order Rickettsiales), a number of hits were classified into the order Rhodobacterales (~30% of all alphaproteobacterial reads), that are known to comprise, among others, abundant microbes (e.g. *Marivita, Cetrimonas, Roseisalinus, Roseovarius* were identified). Only a small number of reads were classified into the order Rhizobiales (~10% of all alphaproteobacterial reads).

Contrastingly, in the similarly hypersaline lagoon of Punta Cormoran, which has a similar high percentage of alphaproteobacterial reads, the SAR11 clade does not appear to have any abundant representatives, with only a very small minority of reads assigned to *Candidatus* Pelagibacter. the major taxa belonged to the order Rhodobacterales (e.g. *Dinoroseobacter, Roseovarius, Loktanella* etc) and to a lesser extent, Rhizobiales (rhizobacteria) (e.g *Parvibaculum, Mesorhizobium* etc.)[17].

So it appears, firstly, that an as yet unknown but abundant SAR11 cluster representative inhabits the hypersaline lagoon of Mar Menor supported by the recruitment plots of both *Candidatus* Pelagibacter and Alphaproteobacteria HIMB114 (Figure 3). Secondly, the Alphaproteobacteria inhabiting two hypersaline lagoons of similar salinity, are substantially different. Punta Cormoran, the pristine lagoon appears to have a thriving *Roseobacter*-community compared

to Mar Menor that has both SAR11 representatives and *Roseobacter* species.

In contrast to Mar Menor, in the freshwater Albufera, the Alphaproteobacteria are in a minority (~9% of all reads). This is surprising as they are usually detectable across a range of freshwater bodies[31]. Based on 16S rRNA phylogenetic analyses, freshwater Alphaproteobacteria have been divided into a number of different lineages, called alfI, alfII, alfIII, alfIV, alfV (LD12 sister group to SAR11), alfVI and alfVII[31,56]. In general, freshwater Alphaproteobacteria are not a well studied group, and we have very little information regarding their ecology, functional roles or genomic characteristics. However, the freshwater datasets chosen here do show clearly that there appears to be a wide variation in the abundance and occurrence of the freshwater alphaproteobacterial lineages (Supplementary Figure S8), particularly the complete absence of the LD12 clade in Albufera, which could mean that LD12 distribution might be affected by nutrient status, as supported by our results.

Cyanobacteria. Cyanobacteria form a sizeable percentage of the marine microbial community, especially the deep chlorophyll maximum[7] have been shown to progressively decrease in numbers with increasing salinity[17]. Here also, we can clearly see (Figure 2) that the number of cyanobacterial sequences shows a decline from the marine, to 5% salinity and finally nearly absent at 19% salinity. The total percentage of cyanobacteria identifiable in the Mar Menor dataset is similar to that of the marine metagenome of the DCM, and nearly twice that of Punta Cormoran. The top organisms identified as cyanobacterial were only *Synechococcus* strains (e.g. WH7803, WH7805), which have been identified before in hypersaline habitats, both by 16S rRNA cloning studies[57] and by metagenomic analyses[17]. Comparisons of the metagenomic reads against *Synechococcus* genomes show a very high level of fragment recruitment (Figure 3), indicating close relatedness between the free-living cyanobacteria in Mar Menor to the already sequenced strains. However, in comparison to the DCM, where the cyanobacterial population comprises both *Prochlorococcus* and *Synechococcus*, it appears that among free-living unicellular picocyanobacteria, *Synechococcus* alone contributes to the primary productivity of this system, where it accounts for ~10% of phytoplankton biomass (Figure 2), and the
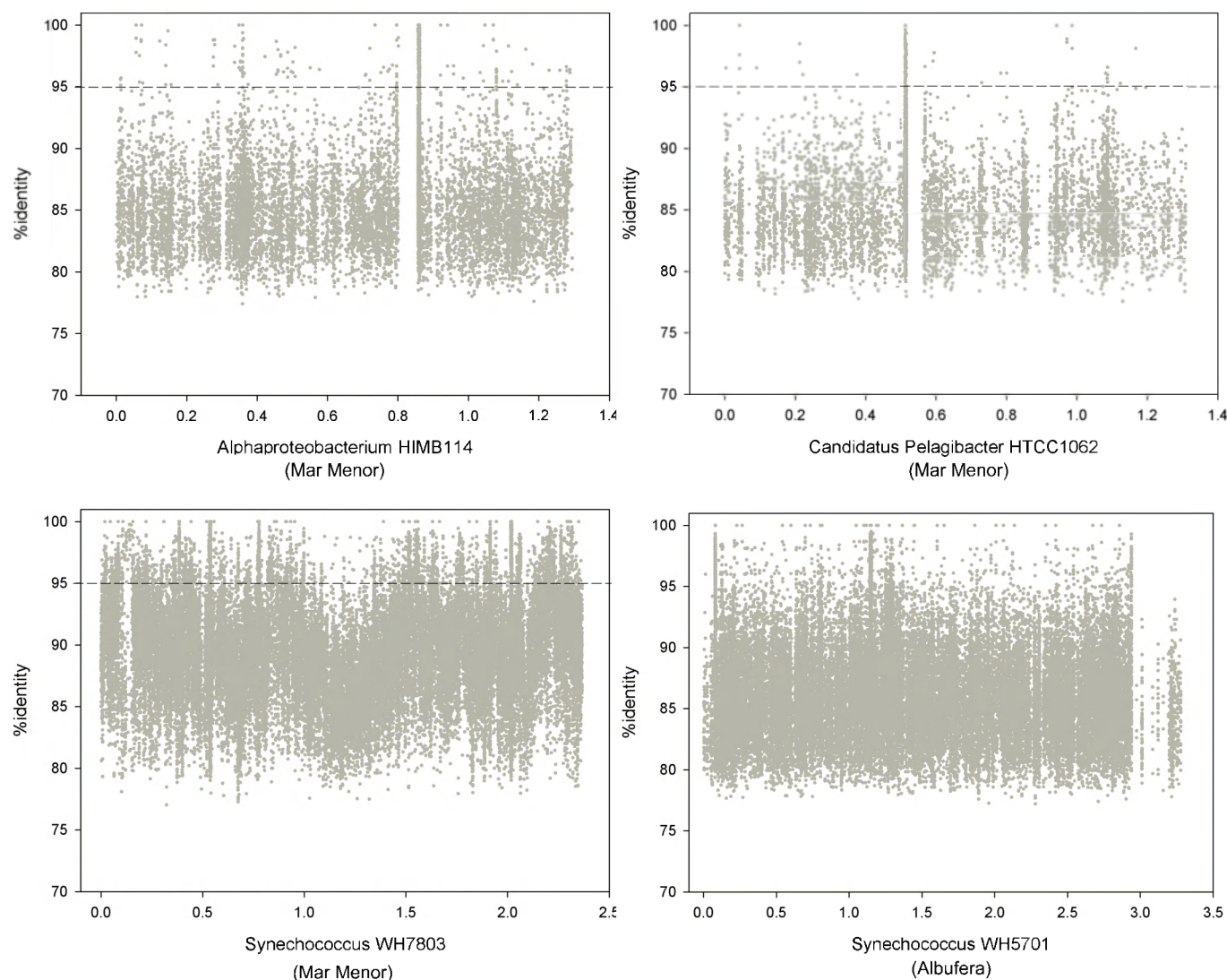
Figure 3 | Fragment recruitment plots of selected organisms versus the Mar Menor and the Albufera metagenomes. The comparisons were done using BLASTN, and a minimum length of 50 bp and an evalue of 1e-5 was considered a hit. The X-axis is scaled in Mb and the Y-axis shows the %identity.

range of *Prochlorococcus* does not extend into the high salinity waters of Mar Menor. Indeed, at higher taxonomic levels, there appears to be very little difference between Mar Menor and the DCM, e.g. similar levels of Alphaproteobacteria, Cyanobacteria, Gammaproteobacteria, Bacteroidetes. Differences appear to emerge at the organismal levels, e.g. absence of *Prochlorococcus* in Mar Menor, a heterogenous alphaproteobacterial population etc.

The most striking characteristic of the Lake Albufera is clearly the high-abundance of cyanobacterial sequences (comprising nearly 35% of all 16S rRNA sequences, and nearly 23% of all metagenomic reads, see Figure 2 and Supplementary Figure S7). Albufera exhibits a highly hypertrophic status, which makes a difference with other freshwater bodies previously studied, like the Amazon river, Lake Gatun and Lake Lanier (or even from the other saline/hypersaline datasets), which do not display such cyanobacterial abundances. Both the Amazon and Lake Gatun show only a very small percentage of cyanobacteria (<2%), while Lake Lanier appears to have a little more (~6%). In comparison to Mar Menor, where we were able to identify mainly *Synechococcus*, the diversity of cyanobacteria in Albufera is clearly higher, with a number of different, and abundant genera, e.g. *Synechococcus, Cyanobium, Pseudanabaena, Merismopedia*, all of which have been previously isolated from freshwaters and mostly detected in this lake[3].

Microscopic counts using the Utermöhl sedimentation technique on inverted microscope (Supplementary Figure S2) are useful for distinguishing morphologically different species of a certain size, mainly ranging from nanoplankton to bigger planktonic microorganisms, including filamentous cyanobacteria and eukaryotic algae. This does not apply for picocyanobacteria, such as *Cyanobium* and *Synechococcus*, which jointly accounted for up to 48 % of the 16S rRNA sequences assigned to a genus in samples from Albufera (Supplementary Table S6). Larger cyanobacteria, like colonial forms of genus *Merismopedia* or filamentous species, like *Pseudanabaena*, were detectable both from sequencing techniques (Supplementary Table S5) and by microscopy (Supplementary Figure S3), showing a partial agreement in both methods.

Even though the measured chlorophyll a levels in Albufera are far higher (271.31 µg/l) than Mar Menor (3.94 µg/l), the difference in the percentage of cyanobacteria (by 16S rRNA analysis) is not proportionately larger (~35% in Albufera and ~12% in Mar Menor). This is likely due to the presence of an enormous diversity and abundance of eukaryotic photosynthetic algae in Albufera, that are not very well detected by sequencing due to much larger eukaryotic genome size but are identified clearly under the microscope (Figure 1, Supplementary Figure S3). Similarly to Mar Menor, we found no evidence for presence of *Prochlorococcus* in the metagenomic data
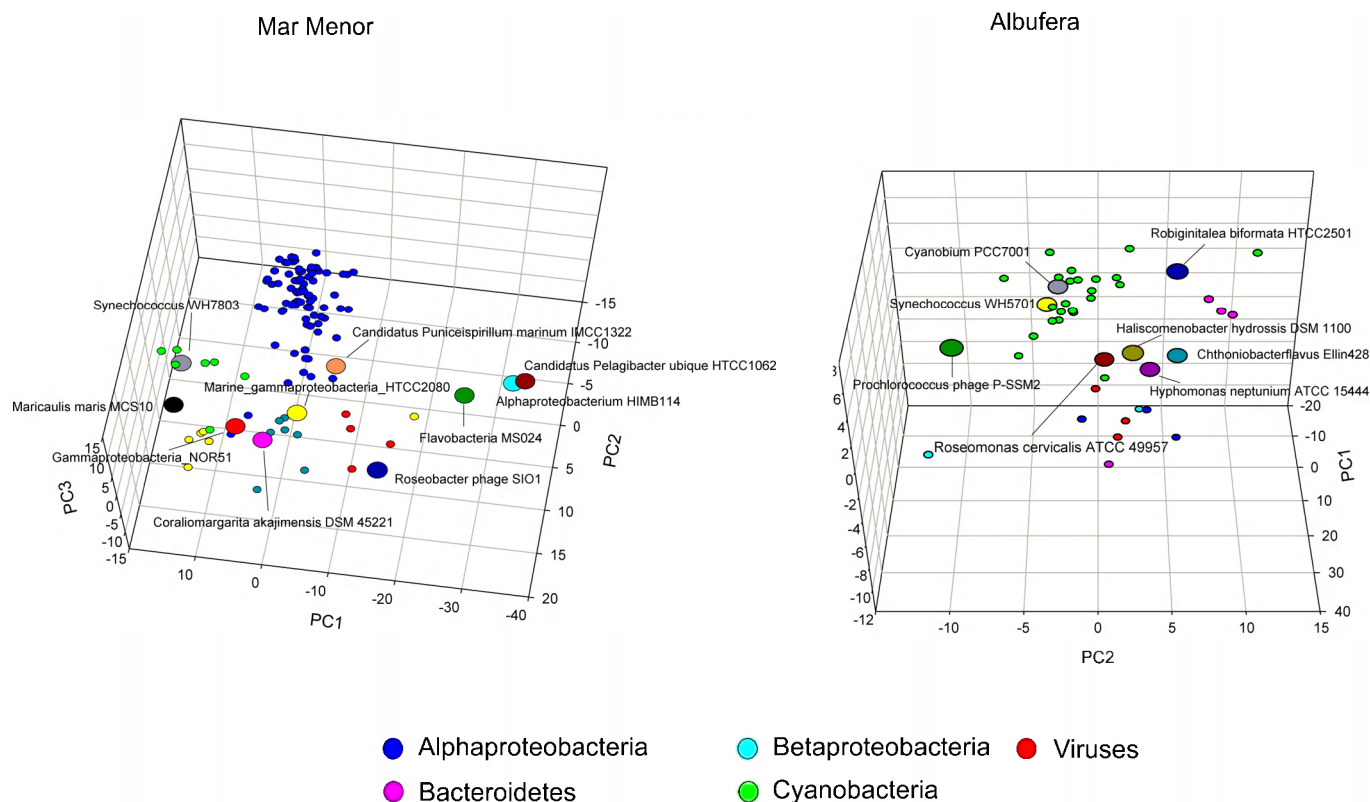
Figure 4 | PCA of tetranucleotide frequencies of assembled contigs from Mar Menor and Albufera. Only those contigs longer than 2 kb that had a consistent phylogenetic profile are shown (see methods).

from Albufera although *Prochlorococcus*-like populations have been reported in freshwater systems before[58], and a study on Yellowstone Lake has also detected *Prochlorococcus* ecotypes in freshwater[59]. Also, even though we detected small amounts of chlorophyll b (Supplementary Table S1), and *Prochlorococcus* cells have characteristic divinyl derivatives of chlorophyll a and b, this may most likely be attributed to chlorophytes that also have these pigments and were identified as very abundant by microscopic counts (Supplementary Figure S2). Another indication of the relative homogeneity of the cyanobacterial populations in Mar Menor, compared to Albufera, is also visible in the GC% profile of the cyanobacterial reads (Supplementary Figure S9), i.e. a single peak at ~62%, while in Albufera, there are two distinct peaks (one at ~55% and the other at ~62%).

Verrucomicrobia. This group of microbes is another point of difference between the pristine Punta Cormoran and Mar Menor. Verrucomicrobiae are widely distributed and have been isolated from a number of different habitats, e.g. soils, lakes, marine sediments, hot springs, and even in man-made ecosystems like acid-rock drainage and municipal solid-waste landfill leachates[60]. They are recognized as an increasingly significant group of soil bacteria and according to several estimates may comprise up to 10% of total bacteria in soil[60]. In Mar Menor we find that a *Coraliomargarita akajimensis* (isolated from seawater[61]) related microbe is quite abundant (Supplementary Table S3). Another abundant organism (by 16S rRNA) was *Haloferula*, which lacks a sequenced genome. However, *Haloferula* species have been isolated from marine environments[62] so it is likely these are close relatives. Moreover, it is clear from Figure 2, that the Verrucomicrobia are abundant in Albufera as well. However, here instead of *Coraliomargarita* or *Haloferula* (which appear to be more salt-tolerant), there is a *Chthionibacter*[63] (which was isolated from soil) related Verrucomicrobia.

Actinobacteria. Actinobacteria have been primarily thought of as soil bacteria. This can be attributed to the ease of cultivation of this group, which have been referred to as high GC microbes. However, several studies, using different approaches (16S rRNA, FISH, and metagenomics) have shown now that Actinobacteria are very common and abundant members of freshwater communities[24,30,56,64,65], and many are not even high GC[17,23,24]. The abundance of Actinobacteria varies greatly across the datasets (Figure 2). In the Albufera metagenome we were able to identify as corresponding to this group only 5-6% of reads (by 16S, and all reads). This is an extremely low percentage relative to the other datasets, (e.g. Amazon ~20%, Lake Gatun ~40%, and Lake Lanier ~20%). This reduced relevance of Actinobacteria is indeed striking.

Some saline datasets also show an abundant actinobacterial presence, e.g. ~24% of all 16S rRNA reads in the hypersaline lagoon Punta Cormoran are actinobacterial. This is in sharp contrast to the very low numbers in the DCM (~2%) or Mar Menor and SS19 (~5% each). In addition, most of the actinobacterial reads from the Mar Menor metagenome were high GC (Supplementary Figure S10) while those from Albufera showed three clear GC% peaks, indicating that in spite of the low number of actinobacterial reads, there might be at least three different clades of Actinobacteria present here.

We examined the 16S rRNA actinobacteril reads from all these datasets in the framework of a well-defined taxonomy (Figure 5), which the freshwater taxa have been classified into seven lineages (~10–15% identical in 16S rRNA to each other)[31]. Each lineage is subclassified into clades (>=95% identity to at least one member), and clades into tribes (>=97% identity to at least one member). The results of this classification show the variation in abundance of these lineages across all the datasets. However, apart from these differences, it is very difficult to arrive at more conclusions as there is not even a single sequenced representative yet from low GC Actinobacteria.
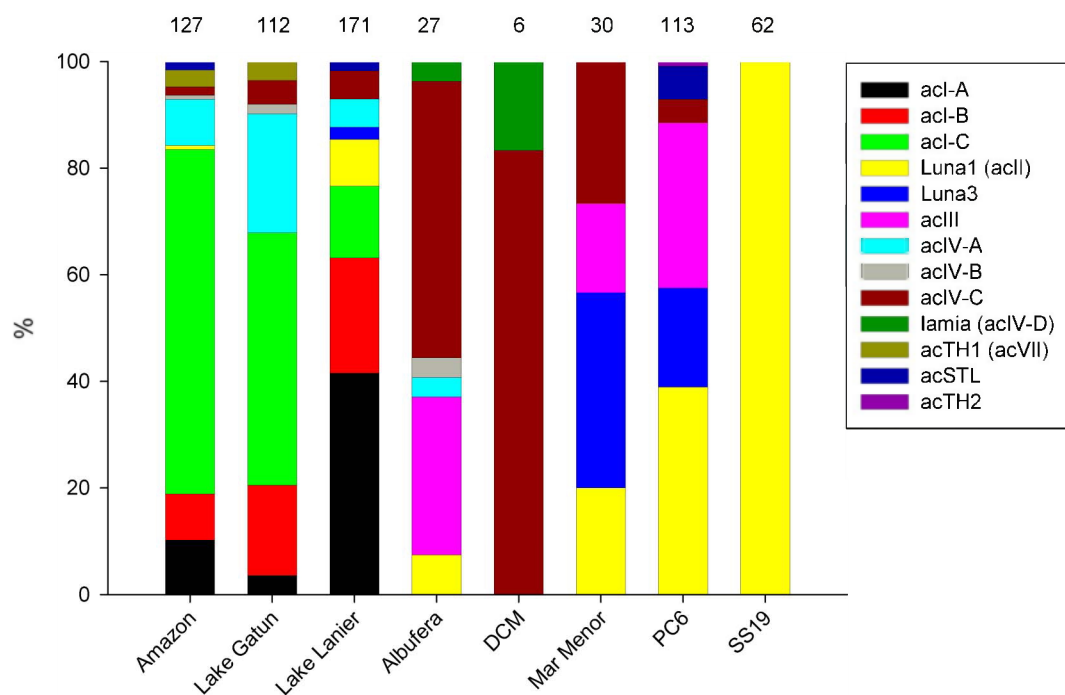
Figure 5 | Classification of actinobacterial 16S Reads from Albufera, Mar Menor and several other metagenomic datasets into known lineages of freshwater Actinobacteria. The numbers above the bars indicate the total number of actinobacterial 16S sequences detected in each dataset.

Both Mar Menor and Albufera contain very similar percentage of actinobacterial reads. However, they differ in the type of their resident Actinobacteria. The majority of reads in Mar Menor could be affiliated to the Luna1, Luna3, acIII and acIV lineages. Albufera also has acIII and acIV, Luna3 lineage is absent, and several others are present in small numbers. The Amazon River and Lake Gatun show very similar populations, with acI-C clade being the most abundant. Lake Lanier is quite different from both of these and has acI-A as the dominant clade. But Albufera is drastically different from any of the other freshwater datasets, with the acI lineage completely absent. Instead, acIII and acIV are nearly equally dominant. The saline samples also show a different trend. Very few reads are detected in the DCM so these might not be very reliable, but Mar Menor and PC6 actually appear quite similar in their actinobacterial load, apart from the extra presence of the acSTL lineage in PC6. However, one of the most striking results is the total dominance of the Luna1 lineage (previously called acII), in the SS19 dataset. It is also present in significant amounts in Mar Menor and Punta Cormoran. Broadly however, the data clearly show the separation between the various lineages on grounds of salinity. For example, the acI lineage, without doubt among the most abundant freshwater lineages, is restricted to freshwater alone, and is not available in saline habitats. However, none of these lineages have a sequenced representative yet, so we cannot speculate further on the nature of these differences.

Betaproteobacteria. Betaproteobacteria are among the most dominant taxa in freshwater systems. This has been shown by several approaches (16S rRNA, FISH, metagenomics)[24,64,66]. In simple abundance levels, in comparison to Albufera(~6%) with other freshwater datasets, only Lake Gatun has a similar abundance levels of betaproteobacteria(~3%), while the Amazon and Lake Lanier appear to have very high levels (nearly 20%) (Figure 2). Although betaproteobacteria are detectable in Albufera, the most prominent, nearly universally available, and arguably the best studied freshwater beta-

proteobacteria, *Polynucleobacter*, was conspicuous by its absence. Moreover, only a handful of betaproteobacterial 16S rRNA sequences could be affiliated to known genera (*Methylibium*-3 and *Thauera*-1) while the others were all unclassified betaproteobacteria. Comparisons from using all reads did suggest that nearly 6% of all sequences in Albufera were betaproteobacterial (Supplementary Figure S7). The ubiquity and absence of *Polynucleobacter* across a wide variety of lakes of different characteristics (altitude, pH, water chemistry, landscape position, trophic status etc) has been discussed extensively[28] before and it has been suggested that high levels of dissolved organic carbon are negatively correlated to the abundance of this microbe. In Albufera, we detected very high values of dissolved organic matter (data not shown), and it is likely that this factor is important in the absence of this ubiquitous microbe in this habitat.

Freshwater betaproteobacteria are broadly divided into seven lineages (betI to betVII) based on 16S rRNA phylogenetic analyses[31]. We classified the betaproteobacterial reads from several datasets into these lineages (Supplementary Figure S11). Both the Amazon and Lake Lanier both showed a wide variety of lineages, and with nearly equal amounts of betI lineage. Indeed, betI lineage does appear to be nearly universal across all freshwater datasets. This lineage does have some cultured representatives (e.g. *Limnohabitans*[67,68]). However, some lineages of betaproteobacteria found in both are different, e.g. betIII (order Burkholderiales) is dominant in Lake Lanier and betIV (order Methylophilales) in the Amazon.

The betII lineage, to which *Polynucleobacter* belongs, is seen only in two datasets (Amazon and Lake Lanier). Albufera also contained sequences belonging to the betVI lineage (~50%). In Albufera, the betaproteobacterial sequences appear nearly evenly divided between the betI and the betVI lineages, with a small amount of betIV sequences. However, the betI and the betVI lineages appear to be widely distributed in all freshwater datasets. But apart from this, there does not appear to be any kind of simple commonality regarding the distribution of the lineages within the datasets studied, with each dataset having its own characteristic features. More

metagenomic datasets complemented with environmental data will be required to elucidate more clearly the various reasons for distribution of these lineages.

We also examined in more detail the distribution of *Polynucleobacter* specifically in all the metagenomic datasets compared in this study. In the betII lineage there are four different "tribes" namely PnecA, B C and D, named after *Polynucleobacter*[31]. The four tribes refer to different *Polynucleobacter* species i.e. PnecA, B, C D refer to *P. rarus*, *P. acidiphobus*, *P. necessarius* and *P. cosmopolitanus* respectively. Only the Amazon and the Lake Lanier datasets showed evidence of presence of *Polynucleobacter*. However, both are quite enriched in *Polynucleobacter* (betII lineage). More specifically, all four tribes PnecA, B, C and D were identified in the Amazon dataset, while only PnecB was identified in the Lake Lanier dataset. We could not identify any Pnec 16S sequences in any of the other datasets.

It is clear however, that betaproteobacteria are not numerically abundant in saline waters, e.g. in Mar Menor, only about ~1% of all the reads could be assigned to betaproteobacteria (Supplementary Figure S7). They are at similarly low levels in the Deep Chlorophyll Maximum, Punta Cormoran and SS19 datasets as well. This is in concordance with similar results regarding the low abundance of betaproteobacteria in marine metagenomic datasets that have been obtained before[8].

Eukaryotes. From the collected metagenomic data it is possible to identify eukaryotic sequences ~12% Mar Menor, ~2% Albufera from comparison to the complete NR database. Indeed, the number of eukaryotic reads increased progressively with increasing filter size (Supplementary Figure S5). The total number of 18S sequences identified in Mar Menor and Albufera were 28 (~5% of total SSUs) and 22 (~5% of total SSUs) respectively. The main eukaryote identified in Mar Menor was *Alexandrium* (~18%, n=5) a marine armored dinoflagellate that produces neurotoxins that cause paralytic shellfish poisoning. *Alexandrium* is well known in coastal lagoons in the Mediterranean[69] and has both autotrophic and heterotrophic species. *Alexandrium* blooms are harmful and are famously referred to as red tides. The toxins it produces can have adverse effects when consumed by humans, usually in the form of contaminated seafood (shellfish, fish etc)[70]. Moreover, these blooms are common in coastal habitats, and affect marine trophic structure, increase mortality of marine fish, birds and mammals and disrupt recreational activities[70]. Dinoflagellate blooms are usually correlated with increased levels of reduced nitrogen sources, particularly ammonia and urea (at least for *Alexandrium*)[71]. Photosynthetic dinoflagellates can supplement photosynthetic growth by organic sources and the increase in the levels of inorganic nutrients (particularly nitrogen and phosphorus)[72], coupled by their ability to produced paralyzing toxins make them strong competitors in eutrophic systems, affecting multicellular and unicellular life alike[73]. However, toxin production by *Alexandrium* is inconsistent, and not all species are toxic. Additionally to *Alexandrium*, other dinoflagellates were also detected (e.g. *Gymnodinium*, *Protoceratium*).

Another abundant organism present by 18S rRNA in Mar Menor was *Chrysochromulina* (n=4), which is a haptophyte from the class Prymnesiophyceae. Haptophytes (e.g *Chrysochromulina*, *Phaeocystis*, *Prymnesium*), are all bloom forming organisms. The particular feature of haptophytes is the presence of a haptonema, a flagella-like (though only superficially), retractile, coiled protuberance, performing several functions (e.g. sensory responses, prey capture)[74]. *Chrysochromulina* is also photosynthetic, and (like some *Alexandrium* species), can supplement photosynthetic growth by mixotrophic feeding. Indeed, some *Chrysochromulina* species are actually euryhaline as well, with a much higher level of optimum salinity for growth[75] than marine levels.

In a microscopic examination and enumeration of the planktonic species, we were able to identify a number of abundant diatoms (e.g.

Cyclotella, *Entomoneis*, *Nitszchia*). *Cyclotella* was identified by its 18S rRNA sequence in the metagenomic data as well. It is a well known abundant centric diatom. Some *Cyclotella* species are known to be associated with high nutrient concentrations, particularly phosphorus, and thus are actually associated with polluted, eutrophic waters[76,77]. However, the most abundant organism by far identified by microscopy was a dinophyte *Gyrodinium*. In contrast with diatoms, whose main sequences (*Cyclotella* sp.) corresponded to taxa already identified by microscopic observations, molecular identifications of dinoflagellates (dinophytes) did not coincided with microscopic determinations, which demonstrates that taxonomy of this group is yet far to be elucidated, even though in our microscopic determinations we only considered autotrophic or mixotrophic species which hold chloroplasts.

Apart from protists, crustacean copepods, that are among the most important group of marine invertebrates[78], particularly for the carbon flux in the food web of the oceans[79], were identified as well in Mar Menor (e.g. *Paracyclopina*, *Oithona*, *Diarthrodes*). These can be considered the zooplanktonic community in the lagoon. *Parcyclopina* species can be found in brackish waters but are tolerant to high salinities as well[80].

The planktonic community of protists and zooplankton in Albufera was clearly different from the dinoflagellate and copepod dominated Mar Menor. The 18S rRNA sequences from Albufera could be assigned primarily to diatoms (e.g. *Cymbella*, *Nistzchia*, *Sellaphora*), that also matched with microscopic determinations (Supplementary Figure S2), or ciliates (e.g. *Halteria*, *Strombidium*). While microscopic observations confirmed the presence of *Nitszchia* and *Cyclotella*, the vast majority of organisms in the sample were identified as the filamentous or colonial nonoplanktonic cyanobacteria (prokaryotes), primarily *Merismopedia*, which form a dense layer of loosely arranged cells in a somewhat planar (rectangular or square) topology, sometimes enclosed by a mucilaginous matrix. *Merismopedia* is commonly found floating in freshwater, several species are planktonic, and can also be found in somewhat halophilic habitats (e.g. coastal areas) or even in thermal springs. They are actually distributed all over the world[81]. In addition to the abundant cyanobacterial (prokaryotic) taxa, several type of chlorophytes (eg. species of *Pediastrum*, which accounted for a big portion of the phytoplankton biovolume, *Coenochloris*, *Chlamydomonas*, *Tetraedron*, *Scenedesmus*, etc) were detected. The photosynthetic organisms in Albufera clearly dwarfed those available in Mar Menor, both in sheer numbers and also in diversity.

Rhodopsins. We identified 52 rhodopsin sequences in the Mar Menor dataset and 34 in Albufera. In Mar Menor, though Firmicutes represented less than 1% of the classified sequences, 10 sequences (nearly 20%) of all rhodopsin sequences appeared related to firmicute rhodopsins (*Exiguobacterium* sp.). Nearly all other sequences in Mar Menor were related to proteobacterial rhodopsins (primarily a collection of Alphaproteobacteria and Gammaproteobacteria). In Albufera, the phylogenetic distribution of rhodopsins appeared more diverse , with the majority affiliated to Proteobacteria (11 sequences) and Planctomycetes (10 sequences). In addition, actinorhodopsins (7) and firmicute rhodopsins (4) were also found.

Metagenomic assembly. Assembly of the metagenomes resulted in a total of 104 contigs from Mar Menor and 35 contigs from Albufera (See methods for details). Nearly one-third of all contigs (77%, n=80) assembled from Mar Menor were primarily alphaproteobacterial (average GC 51.7%, average length 3.1 kb, total length 250 kb). The only other significant sized fraction of assembled contigs could be assigned to viruses (12%, n=8, total length= 34 kb). A small number of actinobacterial contigs (n=6, average length 2.8 kb, total length 17 kb) could also be assembled. Five of these contigs were high GC (57 to 60%) while the last contig, (size
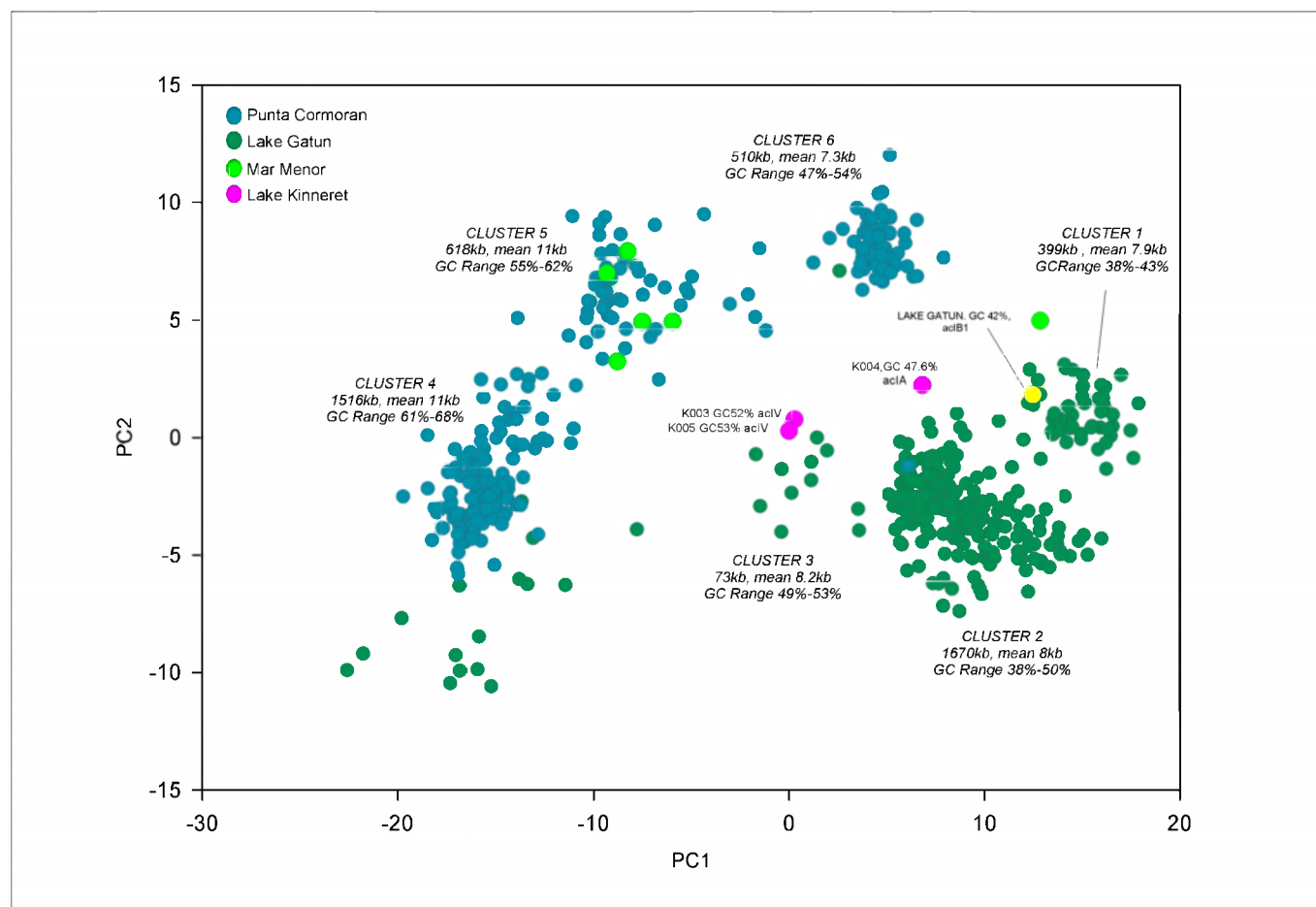
Figure 6 | PCA of tetranucleotide frequencies of actinobacterial contigs from Mar Menor. For reference, actinobacterial contigs from Lake Gatun, Punta Cormoran and three fosmids from Lake Kinneret fosmids are also included. Six clusters of contigs are indicated, and the total sequence,mean contig length, and GC% range, and likely phylogenetic affiliation of the cluster are also shown.

4.3 kb) had a much lower GC content (43%) and contained, among some hypothetical genes, the genes coding for the alpha and beta subunits of ribonucleotide reductase, that are crucial for conversion of ribonucleotides to deoxyribonucleotides.

We performed a principal component analysis on the tetranucleotide frequencies of the assembled contigs (see methods) (Figure 4). In this analysis it is possible observe, besides the actinobacterial cluster formed by the 5 (high GC) of the 6 Actinobacterial contigs (shown in yellow), 4 other clusters corresponding to cyanobacteria, gammaproteobacteria, alphaproteobacteria and viral contigs. However, the largest cluster is formed by the alphaproteobacterial contigs. But this cluster has no proximity to the reference genomes of two organisms of the SAR11 cluster (found by recruitment) namely, *Candidatus* Pelagibacter (GC=29.7%) of the SAR11 cluster and Alphaproteobacterium HIMB114, but instead is closer to *Candidatus* Puniceispirillum marinum(GC=48.9%), which is a member of the SAR116 cluster. A total of 320 genes were predicted in the 80 alphaproteobacterial contigs, and of these 120 genes gave a best hit to Rhizobiales (mean similarity 72.38%), while 108 genes gave best blast hits to Rhodobacterales (mean similarity 73%). It did not appear to be related to Rickettsiales. So even though it appears that there are at least two unidentified microbes in Mar Menor, by 16S rRNA analysis, one related to Rickettsiales, SAR11 and the other to Rhodobacterales, we did not assemble any reads from the SAR11 related microbe, but from the other.

In one of these assembled alphaproteobacterial contigs, we identified a nearly complete cluster of Sox genes that provide the necessary apparatus for performing sulfur oxidation. This cluster has been demonstrated to operate in photo- and chemo-

trophic Alphaproteobacteria that oxidize thiosulfate to sulfate without inorganic sulfur globule formation as free intermediate, and was first described in the alphaproteobacterium *Paracoccus pantotrophus*, a facultative lithoautotrophic organism that grows with thiosulfate (and other electron donors e.g. molecular hydrogen) as an energy source[82]. The cluster of *P. pantotrophus* coding for sulfur-oxidizing proteins comprises at least two transcriptional units with 15 genes. Seven genes, soxXYZABCD, code for proteins essential for constituting a periplasmic system for sulfur oxidation in vitro and are induced by thiosulfate. The SoxY gene has a C-terminal invariant binding site motif (VKVTIGGCGG), that binds different oxidation states of sulfur[83]. The exact motif was present in the assembled SoxY gene in the assembled contig, providing more confidence in the function assignment and assembly. Although several pathways of thiosulfate oxidation are known, two main pathways exist, the difference between them being related directly to the presence or absence of the SoxCD genes[84]. In the presence of SoxCD proteins, thiosulfate is converted to two sulfate molecules, and this is the pathway in *P. pantotrophus*, while in the absence of SoxCD, only a single sulfate is produced, the other sulfur atom being deposited in the form of inorganic globules (e.g. *Beggiatoa*). In the case of the assembled contigs, it clearly possesses a SoxC gene, while the SoxD part is likely not assembled. So it appears that the organism to which this cluster belongs is able to fully oxidize sulfur to two sulfates and does not deposit any sulfur granules either intra or extracellularly.

Comparison of the assembled Sox genes cluster with the sox gene cluster of *Roseobacter* sp. MED193 and *Aurantimonas*

10

manganoxydans SI85-9A1 (Supplementary Figure S12) showed nearly complete synteny between the genomic regions and the assembled contig. This suggests that the organism to whom these contigs belong is a novel sulfur-oxidising Alphaproteobacteria, likely adapted to a higher salinity. This is interesting because the close relatives of this microbe, e.g. *Candidatus* Pelagibacter, Alphaproteobacterium HIMB114 and *Candidatus* Puniceispirillum do not have the Sox cluster in their genomes and are likely incapable of sulfur oxidation.

Some other contigs that were assembled from the data for Mar Menor could be assigned to cyanobacteria (Figure 4). These contigs appeared closely related to *Synechococcus* species. However, the assembly from Albufera represented a much more diverse set of contigs, with several contigs assembled, primarily from cyanobacteria (66%) but also from other taxa (Viruses 9%, Bacteroidetes 11% and Betaproteobacteria 6%).

A more focused analysis of the assembled actinobacterial contigs from Mar Menor, in the context of actinobacterial contigs from other metagenomic datasets is shown in Figure 6. We collected actinobacterial contigs from Lake Gatun and Punta Cormoran (see methods), and also three fully sequenced actinobacterial fosmids from Lake Kinneret[85]. We could identify at least six distinct clusters, each representing a dominant lineage of freshwater actinobacteria. Because of the presence of 16S sequences in the contigs, at least three of the clusters can be assigned a tentative name, i.e. two sub groups of acI, acIA and acIB1, and a lineage acIV. In comparison to the Lake Gatun, the contigs from Punta Cormoran, appear to have higher GC content (see Clusters 4, 5 and 6 in Figure 6). Also, out of the 6 assembled contigs from Mar Menor, five cluster very clearly with Punta Cormoran contigs in Cluster 5 (GC% 55–62). Only a single contig, a low GC contig, clusters with the acIB1 cluster (Cluster 1). So it does appear that there is a minor low GC actinobacterial population in Mar Menor (also seen in the GC% profile of Mar Menor Actinobacterial Reads, Supplementary Figure S10). No rRNA sequences were detected either in the contigs from Punta Cormoran, or from Mar Menor contigs so assignment of names to clusters 4, 5 or 6 was not possible. Also, since a number of different actinobacterial clades are nearly equally abundant (Figure 5), it is not possible to associate with any degree of confidence these contigs with the known actinobacterial lineages.

Discussion

In this work, we have compared the relative relevance of the different groups of microorganisms in two coastal lagoons, one freshwater and another hypersaline, with other related aquatic systems, in the framework of the main environmental features characterizing them. The analysis of the metagenomic and metacommunity data of these two hypertrophic lagoons has revealed interesting general patterns. We have discovered, using assembly of the metagenomic data, a novel, as yet uncultured, sulfur oxidizing alphaproteobacterium, that is abundant in the hypersaline Mar Menor. We also found evidence of the presence of only *Synechococcus* as the abundant cyanobacteria and the complete absence of *Prochlorococcus*, which is abundant in the parent Mediterranean water body from which Mar Menor waters are derived. Also, even the freshwaters of Albufera, though abundant in cyanobacteria, did not show any indication of presence of *Prochlorococcus*. Microscopy and sequence data of the phytoplankton revealed differences in the two lagoons, Mar Menor dominated by dinoflagellates and Albufera by chlorophytes, while diatoms were observed in both. The main distinctive characteristic of Albufera is its highly hypertrophic status. It contained a considerably different microbiota than less nutrient rich freshwaters. Importantly, canonical freshwater microbial groups like low GC Actinobacteria, LD12 lineage of Alphaproteobacteria and even the cosmopolitan betaproteobacteria, *Polynucleobacter*, are all conspicuously absent. That cyanobacteria are a major component of hypertrophic waters (like

Albufera) is not new, however, the absence of the other major freshwater microbes certainly is a significant departure from other freshwater systems. Many of the groups that are absent in Albufera such as the low GC Actinobacteria or the LD12 lineage are very small sized bacteria, with a high surface to volume ratio, which might be of advantage in low nutrient situations. However, we speculate that this competitive edge is likely to be lost in a hypertrophic situation like the one in Albufera, where fast growing cyanobacteria might become dominant.

Methods

Sampling. Samples were collected from Albufera de Valencia on May 12, 2010 (39°19′54″N, 0°21′8″W) and on May 7, 2010 from Mar Menor (37°43′08″N, 00°47′14″W) as part of the J. Craig Venter Institute European Sampling Expedition. Approximately 20 L (Albufera de Valencia) and 40 L (Mar Menor) were sequentially filtered using three different filter sizes (0.1 μm, 0.8 μm and 3 μm). Filters were stored at −80°C in protective buffer (10 mL sterile filtered sampling water, 10 mL RNAlater, 200 μl 100x TE buffer, 400 μl 0.5 M EGTA and 400 μl 0.5 M EDTA) until DNA extraction. Then filters were thawed on ice and then treated with 1 mg/ml lysozyme and 0.2 mg/ml proteinase K (final concentrations). Nucleic acids were extracted with phenol/chloroform/isoamyl alcohol and chloroform/isoamyl alcohol and DNA integrity was checked by agarose gel electrophoresis. Samples were sequenced using the Roche 454 GS-FLX system, titanium chemistry. The raw data for all samples has been deposited in the Camera Database (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis, http://camera.calit2.net) and is publicly available.

Analytical methods: water chemistry, pigment analyses by HPLC, and microscopic counts. Using replicated water samples obtained simultaneously to those used for metagenomic analyses, a series of physical, chemical, and biological determinations were done. Electrical conductivity was measured with a WTW LF-191 conductivity meter, and pH with a Orion electrode ion Analyzer EA920 (Orion Research). Soluble reactive phosphorus was determined by the phosphomolibdic ascorbic acid method, nitrate was analysed after reduction in a cadmium column by the Griess method, both from *in situ* filtered (through GF/F glass-fiber filters) samples. Alkalinity was measured by titration with HCl and chloride was determined by the argentimetric method. Analyses were performed following Standard Methods for Water Analyses. Ammonium was also measured on filtered water using the modified indophenol blue method. Unfiltered samples for total phosphorus and total nitrogen determinations were digested through a double alkaline-acid persulfatic digestion. Once extracted, total phosphorus was determined as SRP after pH-neutralization. Total nitrogen was also determined on the digested samples following Bachmann and Canfield[86]. Carbon forms (CO_3, CO_3H and CO_2) were calculated from pH and alkalinity measurements following Rodier[87]. Chromophoric dissolved organic matter (CDOM) we quantify by means of the excitation-emission matrix (EEM) method[88] using a F-7000 Hitachi fluorescence spectrophotometer.

Phytoplankton abundance from Lugol-fixed samples was determined with an Olympus IX50 inverted microscope by using the Utermöhl sedimentation method[89]. Algae were identified according to several described taxonomic keys[90–93]. Samples for photosynthetic pigments determination were collected onto GF/F filters and extracted in the dark, overnight at −20°C, with 100 % acetone with several sonication times, samples were injected into a Waters HPLC system with a Waters 996 photodiode Array Detector. The system uses two columns (Spherisorb S5 ODS2) in series and running at 35°C in a methanol/ammonium acetate/acetone gradient following Pinckney et al[94] and Van Heukelen et al[95]. Peaks were identified according to their absorption spectra and concentration was calculated using commercial standards (DHI, Denmark).

For the cytometric identification and quantification of the bacterioplankton and APP cells, a Coulter Cytomics FC500 flow cytometer equipped with an argon laser (488 excitation), a red emitting diode (635 excitation), and five filters for fluorescent emission (FL1-FL5), was used. Bacterioplankton abundance was determined with argon laser by green fluorescence (Sybr Green I) using FL1 detector (525 nm). APP abundance was determined combining argon laser and red diode by red fluorescence (Chlorophyll-*a* and phycobiliproteins autofluorescence) using FL4 detector (675 nm).

Community Structure. 16S Ribosomal RNA genes were identified by comparing the datasets against the RDP database[37]. All reads that matched an rRNA sequence with an alignment length of more than 100 bases and an e-value of 0.001 against the database were extracted. The best hit was used to assign to a high taxonomic level. When possible, the sequences were further assigned to genus if they shared ≥95% rRNA sequence identity with a known species. Moreover, the 16S sequences were also run through the Metaxa program[96] to cross-check identified genera. Additionally, the entire datasets were compared to the NCBI NR database (using BLASTX, 1e-5) and analysed using the MEGAN software[97]. Classification of 16S rRNA reads into specific reference taxonomies was performed using mothur[98].

Assembly of the metagenomic reads (only reads >100 bp) was performed using a stringent criteria of overlap of at least 80 bp of the read and 99% identity and at most a single gap in the alignment (using Geneious Pro 5.4). Assembled contigs that were less

than 2 kb in length, and those with less than three predicted genes were discarded. We retained only those contigs that gave consistent hits to only a single high level taxon (e.g. Alphaproteobacteria, Euryarchaeota, Bacteroidetes, Actinobacteria). The strict assembly requirements combined with a taxonomic uniformity condition imposed on the assembled sequence resulted in a total of 88 contigs that were more than 5 kb in length and had a consistent phylogenetic profile and were hence more likely to originate from a single organism.

1. Vicente, E. & Miracle, M. A. The coastal lagoon Albufera de Valencia - An ecosystem under stress. *Limnetica* **8**, 87–100 (1992).
2. Soria, J. M. Past, present and future of la Albufera of Valencia Natural Park. *Limnetica* **25**, 135–142 (2006).
3. Romo, S., García-Murcia, A., Villena, M. J., Sánchez, V. & Ballester, A. Albufera de Valencia and implications for its ecology, management, and recovery. *Limnetica* **27**, 11–28 (2008).
4. Perez-Ruzafa, A., Fernandez, A. & Marcos, C. Spatial and temporal variations of hydrological conditions, nutrients and chlorophyll a in a Mediterranean coastal lagoon (Mar Menor, Spain). *Hydrobiologia* **550**, 11–27 (2005).
5. Cendrero-Uceda, A., Sanchez-Arcilla Conejo, A. & Zazo-Cardena, C. A preliminary general assessment of the impacts in Spain due to the effects of climate change impacts on coastal areas. (2005).
6. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
7. Ghai, R. *et al.* Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *The ISME Journal* **4**, 1154–1166 (2010).
8. Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**, e77 (2007).
9. Quaiser, A., Zivanovic, Y., Moreira, D. & Lopez-Garcia, P. Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J* **5**, 285–304, doi:ismej2010113 [pii]10.1038/ismej.2010.113 (2011).
10. DeLong, E. F. *et al.* Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496–503, doi:311/5760/496 [pii]10.1126/science.1120250 (2006).
11. Feingersch, R. *et al.* Microbial community genomics in eastern Mediterranean Sea surface waters. *ISME J* **4**, 78–87, doi:ismej200992 [pii]10.1038/ismej.2009.92 (2010).
12. Falkowski, P. G. The ocean's invisible forest. *Sci Am* **287**, 54–61 (2002).
13. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
14. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **104**, 11436–11440 (2007).
15. Benlloch, S. *et al.* Archaeal Biodiversity in Crystallizer Ponds from a Solar Saltern: Culture versus PCR. *Microb Ecol* **41**, 12–19 (2001).
16. Benlloch, S. *et al.* Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ Microbiol* **4**, 349–360 (2002).
17. Ghai, R. *et al.* New Abundant Microbial Groups in Aquatic Hypersaline Environments. *Scientific Reports* **1**, doi:doi:10.1038/srep00135 (2011).
18. Narasingarao, P. *et al.* De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J* (2011).
19. Oren, A. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems* **4**, 2 (2008).
20. Bodaker, I. *et al.* Comparative community genomics in the Dead Sea: an increasingly extreme environment. *ISME J* **4**, 399–407 (2010).
21. Baati, H., Guermazi, S., Gharsallah, N., Sghir, A. & Ammar, E. Novel prokaryotic diversity in sediments of Tunisian multipond solar saltern. *Res Microbiol* **161**, 573–582, doi:S0923-2508(10)00131-2 [pii]10.1016/j.resmic.2010.05.009 (2010).
22. Trigui, H. *et al.* Characterization of heterotrophic prokaryote subgroups in the Sfax coastal solar salterns by combining flow cytometry cell sorting and phylogenetic analysis. *Extremophiles* **15**, 347–358, doi:10.1007/s00792-011-0364-5 (2011).
23. Ghai, R., McMahon, K. D. & Rodriguez-Valera, F. Breaking a paradigm: cosmopolitan and abundant freshwater actinobacteria are low GC. *Environ Microbiol Reports* **4**(1), 29–35 (2012).
24. Ghai, R. *et al.* Metagenomics of the water column in the pristine upper course of the Amazon river. *PLoS One* **6**, e23785, doi:10.1371/journal.pone.0023785PONE-D-11-06938 [pii] (2011).
25. Newton, R. J., Jones, S. E., Helmus, M. R. & McMahon, K. D. Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Appl Environ Microbiol* **73**, 7169–7176 (2007).
26. Hahn, M. W. Isolation of strains belonging to the cosmopolitan Polynucleobacter necessarius cluster from freshwater habitats located in three climatic zones. *Appl Environ Microbiol* **69**, 5248–5254 (2003).
27. Hahn, M. W. *et al.* Polynucleobacter cosmopolitanus sp. nov., free-living planktonic bacteria inhabiting freshwater lakes and rivers. *Int J Syst Evol Microbiol* **60**, 166–173 (2010).
28. Jezberova, J. *et al.* Ubiquity of Polynucleobacter necessarius ssp. asymbioticus in lentic freshwater habitats of a heterogeneous 2000 km area. *Environ Microbiol* **12**, 658–669 (2010).
29. Wu, Q. L. & Hahn, M. W. Differences in structure and dynamics of Polynucleobacter communities in a temperate and a subtropical lake, revealed at three phylogenetic levels. *FEMS Microbiol Ecol* **57**, 67–79 (2006).
30. Zwart, G., Crump, B. C., Agterveld, M. P. K. V., Hagen, F. & Han, S. K. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology* **28**, 141–155 (2002).
31. Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**, 14–49 (2011).
32. Shiklomanov, I. & Rodda, J. (2003).
33. Jones, S. E., Newton, R. J. & McMahon, K. D. Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. *Environmental Microbiology* **11**, 2463–2472 (2009).
34. Yannarell, A. C., Kent, A. D., Lauster, G. H., Kratz, T. K. & Triplett, E. W. Temporal patterns in bacterial communities in three temperate lakes of different trophic status. *Microb Ecol* **46**, 391–405 (2003).
35. Lindstrom, E. S., Kamst-Van Agterveld, M. P. & Zwart, G. Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Appl Environ Microbiol* **71**, 8201–8206 (2005).
36. Yannarell, A. C. & Triplett, E. W. Geographic and environmental sources of variation in lake bacterial community composition. *Appl Environ Microbiol* **71**, 227–239 (2005).
37. Tamames, J., Abellan, J. J., Pignatelli, M., Camacho, A. & Moya, A. Environmental distribution of prokaryotic taxa. *BMC Microbiol* **10**, 85, doi:1471-2180-10-85 [pii]10.1186/1471-2180-10-85 (2010).
38. Pope, P. B. & Patel, B. K. Metagenomic analysis of a freshwater toxic cyanobacteria bloom. *FEMS Microbiol Ecol* **64**, 9–27 (2008).
39. Debroas, D. *et al.* Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget--France). *Environ Microbiol* **11**, 2412–2424 (2009).
40. Feingersch, R. & Beja, O. Bias in assessments of marine SAR11 biodiversity in environmental fosmid and BAC libraries? *ISME J* **3**, 1117–1119, doi:ismej200980 [pii]10.1038/ismej.2009.80 (2009).
41. Paoli, A. *et al.* Picocyanobacteria in Adriatic transitional environments. *Estuarine, Coastal and Shelf Science* **75**, 13–20, doi:10.1016/j.ecss.2007.02.026 (2007).
42. Wit, R. Microbial diversity in the Bassin d'Arcachon coastal lagoon (SW France). *Hydrobiologia* **611**, 5–15, doi:10.1007/s10750-008-9461-6 (2008).
43. Benlloch, S., Rodriguez-Valera, F. & Martinez-Murcia, A. J. Bacterial diversity in two coastal lagoons deduced from 16S rDNA PCR amplification and partial sequencing. *FEMS Microbiol Ecol* **18**, 267–280 (1995).
44. Simonato, F., Gómez-Pereira, P. R., Fuchs, B. M. & Amann, R. Bacterioplankton diversity and community composition in the Southern Lagoon of Venice. *Systematic and Applied Microbiology* **33**, 128–138, doi:10.1016/j.syapm.2009.12.006 (2010).
45. Celussi, M., Pugnetti, A. & Del Negro, P. Structural dynamics of bacterioplankton assemblages in the Lagoon of Venice. *Estuarine, Coastal and Shelf Science* **84**, 154–160, doi:10.1016/j.ecss.2009.05.028 (2009).
46. Oh, S. *et al.* Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl Environ Microbiol* **77**, 6000–6011, doi:AEM.00107-11 [pii]10.1128/AEM.00107-11 (2011).
47. O.E.C.D. *Organisation for Economic Co-operation and Development.* (1982).
48. García-Pintado, J., Martínez-Mena, M., Barberá, G. G., Albaladejo, J. & Castillo, V. M. Anthropogenic nutrient sources and loads from a Mediterranean catchment into a coastal lagoon: Mar Menor, Spain. *Science of The Total Environment* **373**, 220–239, doi:10.1016/j.scitotenv.2006.10.046 (2007).
49. Velasco, J. *et al.* Nutrient and particulate inputs into the Mar Menor Lagoon (SE Spain) from an intensive agricultural watershed. *Water, Air, and Soil Pollution* **176**, 37–56 (2006).
50. Robarts, R. D., Zohary, T., Waiserl, M. J. & Yacobi, Y. Z. Bacterial abundance, biomass, and production in relation to phytoplankton biomass in the Levantine Basin of the southeastern Mediterranean Sea. *Mar Ecol Prog Ser* **137**, 273–281 (1996).
51. Vaqué, D., Bloughl, H. A. & Duarte, C. M. Dynamics of ciliate abundance, biomass and community composition in an oligotrophic coastal environment (NW Mediterranean). *Aquat Microb Ecol* **12**, 71–83 (1997).
52. Camacho, A., Miracle, M. R. & Vicente, E. Which factors determine the abundance and distribution of picocyanobacteria in inland waters? A comparison among different types of lakes and ponds. *Arch. Hydrobiol.* **157**, 321–338 (2003).
53. Villena, M. J. & Romo, S. Phytoplankton changes in a shallow Mediterranean lake (Albufera of Valencia, Spain) after sewage diversion. *Hydrobiologia* **506–509**, 281–287 (2003).
54. Romo, S. & Miracle, M. A. Population dynamics and ecology of subdominant phytoplankton species in a shallow hypertrophic lake (Albufera of Valencia, Spain). *Hydrobiologia* **273**, 37–56 (1994).
55. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations. *PLoS Biology* **3**, e144, doi:10.1371/journal.pbio.0030144 (2005).
56. Glockner, F. O. *et al.* Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. *Appl Environ Microbiol* **66**, 5053–5065 (2000).

57. Clementino, M. M. *et al.* Prokaryotic diversity in one of the largest hypersaline coastal lagoons in the world. *Extremophiles* 12, 595–604, doi:10.1007/s00792-008-0162-x (2008).

58. Corzo, A., Jiminez-Gomez, F., Gordillo, F. J. L., Garcia-Ruiz, R. & Niell, F. X. Synechococcus and Prochlorococcus-like populations detected by flow cytometry in a eutrophic reservoir in summer. *Journal of Plankton Research* 21, 1575–1581 (1999).

59. Clingenpeel, S. *et al.* Yellowstone Lake: high-energy geochemistry and rich bacterial diversity. *Environ Microbiol* 13, 2172–2185, doi:10.1111/j.1462-2920.2011.02466.x (2011).

60. Wagner, M. & Horn, M. The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* 17, 241–249, doi:S0958-1669(06)00064-4 [pii]10.1016/j.copbio.2006.05.005 (2006).

61. Yoon, J. *et al.* Coraliomargarita akajimensis gen. nov., sp. nov., a novel member of the phylum 'Verrucomicrobia' isolated from seawater in Japan. *Int J Syst Evol Microbiol* 57, 959–963, doi:57/5/959 [pii]10.1099/ijs.0.64755-0 (2007).

62. Yoon, J. *et al.* Haloferula rosea gen. nov., sp. nov., Haloferula harenae sp. nov., Haloferula phyci sp. nov., Haloferula helveola sp. nov. and Haloferula sargassicola sp. nov., five marine representatives of the family Verrucomicrobiaceae within the phylum 'Verrucomicrobia'. *Int J Syst Evol Microbiol* 58, 2491–2500, doi:58/11/2491 [pii]10.1099/ijs.0.2008/000711-0 (2008).

63. Sangwan, P., Chen, X., Hugenholtz, P. & Janssen, P. H. Chthoniobacter flavus gen. nov., sp. nov., the first pure-culture representative of subdivision two, Spartobacteria classis nov., of the phylum Verrucomicrobia. *Appl Environ Microbiol* 70, 5875–5881, doi:70/10/5875 [pii]10.1128/AEM.70.10.5875-5881.2004 (2004).

64. Hiorns, W. D., Methe, B. A., Nierzwicki-Bauer, S. A. & Zehr, J. P. Bacterial diversity in Adirondack mountain lakes as revealed by 16S rRNA gene sequences. *Appl Environ Microbiol* 63, 2957–2960 (1997).

65. Denisova, L. Y., Belkova, N. L., Tulokhonov, I. I. & Zaichikov, E. Diversity of bacteria at various depths in the southern part of lake Baikal as detected by 16S rRNA sequencing. *Mikrobiologiia* 68, 547–556 (1999).

66. Buck, U., Grossart, H. P., Amann, R. & Pernthaler, J. Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake. *Environ Microbiol* 11, 1854–1865,doi:EMI1910 [pii]10.1111/j.1462-2920.2009.01910.x (2009).

67. Hahn, M. W. *et al.* Limnohabitans curvus gen. nov., sp. nov., a planktonic bacterium isolated from a freshwater lake. *Int J Syst Evol Microbiol* 60, 1358–1365, doi:ijs.0.013292-0 [pii]10.1099/ijs.0.013292-0 (2010).

68. Hahn, M. W., Kasalicky, V., Jezbera, J., Brandt, U. & Simek, K. Limnohabitans australis sp. nov., isolated from a freshwater pond, and emended description of the genus Limnohabitans. *Int J Syst Evol Microbiol* 60, 2946–2950, doi:ijs.0.022384-0 [pii]10.1099/ijs.0.022384-0 (2010).

69. Penna, A. *et al.* Phylogenetic relationships among the Mediterranean Alexandrium (Dinophyceae) species based on sequences of 5.8 S gene and Internal Transcript Spacers of the rRNA operon. *European Journal of Phycology* 43, 163–178, doi:10.1080/09670260701783730 (2008).

70. Anderson, D. M. *et al.* The globally distributed genus Alexandrium: Multifaceted roles in marine ecosystems and impacts on human health. *Harmful Algae* 14, 10–35, doi:10.1016/j.hal.2011.10.012 (2012).

71. Gilbert, P. M. *et al.* Harmful algal blooms in the Chesapeake and coastal bays of Maryland, USA: comparison of 1997, 1998, and 1999 events. *Estuaries* 24, 875–883 (2011).

72. Leong, S. C. Y. & Taguchi, S. Detecting the bloom-forming dinoflagellate Alexandrium tamarense using the absorption signature. *Hydrobiologia* 568, 299–308, doi:10.1007/s10750-006-0202-4 (2006).

73. Tillmann, U., John, U. & Cembella, A. On the allelochemical potency of the marine dinoflagellate Alexandrium ostenfeldii against heterotrophic and autotrophic protists. *Journal of Plankton Research* 29, 527–543, doi:10.1093/plankt/fbm034 (2007).

74. Kawachi, M. & Inouye, I. Functional roles of the haptonema and the spine scales in the feeding process of Chrysochromulina spinifera (Fournier) Pienaar et Norris. *Phycologia* 34, 193–200 (1995).

75. Edvardsen, N. & Paasche, E. in *Physiological Ecology of Harmful Algal Blooms* (eds D. M. Anderson, A. D. Cembella & G. M. Hallegraeff) 193–208 (Springer, 1998).

76. Sgro, G. V. *et al.* A diatom quality index from a diatom-based total phosphorus inference model. *Environmental Bioindicators* 2, 15–34 (2007).

77. Weckstrom, K. & Juggins, S. Coastal diatom-environment relationships from the Gulf of Finland, Baltic Sea. *Journal of Phycology* 42, 21–35. (2006).

78. Humes, A. G. How many copepods? *Hydrobiologia* 292/293, 1–7 (1994).

79. Legendre, L. & Rivkin, R. B. Fluxes of carbon in the upper ocean: regulation by food-web control nodes. *Mar Ecol Prog Ser* 242, 95–109 (2002).

80. Lee, K. W., Park, H. G., Lee, S. M. & Kang, H. K. Effects of diets on the growth of the brackish water cyclopoid copepod Paracyclopina nana Smirnov. *Aquaculture* 256, 346–353 (2006).

81. Komarek, J. & Anagnostidis, K. in *Süsswasserflora von Mitteleuropa* (eds H. Ettl, G. Gartner, H. Heynig & D. Mollenhauer) (Gustav Fischer, 1998).

82. Friedrich, C. G. *et al.* Novel Genes Coding for Lithotrophic Sulfur Oxidation of Paracoccus pantotrophus GB17. *J Bacteriol* 182, 4677–4687 (2000).

83. Quentmeier, A. & Friedrich, C. G. The cysteine residue of the SoxY protein as the active site of protein-bound sulfur oxidation of Paracoccus pantotrophus GB17. *FEBS Lett* 503, 168–172, doi:S0014-5793(01)02727-2 [pii] (2001).

84. Friedrich, C. G., Bardischewsky, F., Rother, D., Quentmeier, A. & Fischer, J. Prokaryotic sulfur oxidation. *Current Opinion in Microbiology* 8, 253–259, doi:10.1016/j.mib.2005.04.005 (2005).

85. Philosof, A., Sabehi, G. & Beja, O. Comparative analyses of actinobacterial genomic fragments from Lake Kinneret. *Environ Microbiol* 11, 3189–3200, doi:EMI2024 [pii]10.1111/j.1462-2920.2009.02024.x (2009).

86. Bachmann, R. W. & Canfield, D. E. Use of an alternative method for monitoring total nitrogen concentration in Florida Lakes. *Hydrobiologia* 323, 1–8 (1996).

87. Rodier, J. *L'analyse de l'eau (9ème édition)*. 9 edn, 1526 (Berdas, 2009).

88. Stedmond, C. A. Resolving the variability in dissolved organic mat- ter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol Oceanogr* 50, 686–697 (2005).

89. Utermöhl, H. Zur vervollkommung der quantitativen phytoplankton-methodik. *Mittenlungen Internationale Vereinigung für Theoretical und Angewandte Limnologie* 9, 1–38 (1958).

90. Anagnostidis, K. & Komárek, J. Modern approach to the classification system of cyanophytes. 3-Oscillatoriales. *Archiv für Hydrobiologie* Supplement 80, Algological Studies 50–53, 327–472 (1988).

91. Anton, A. & Duthie, H. C. Use of cluster analysis in the systematics of the algal genus Cryptomonas. *Canadian Journal of Botany* 59, 992–1002 (1981).

92. Huber-Pestalozzi, G. *Das Phytoplankton des Süsswassers. Systematik und Biologie. Euglenophyceen. Teil 4.*, 1135 (E. Schweizerbart'sche Verlag, 1955).

93. Komárek, J. & Fott, B. *Das Phytoplankton des Süsswassers. Systematik und Biologie. Chlorophyceae (Grünalgen). Ordung: Chlorococcales. Teil 7.* 1044 (E. Schweizerbart'sche Verlag, 1983).

94. Pinckney, J. L., Millie, D. F., Howe, K. E., Paerl, H. W. & Hurley, J. P. Flow scintillation counting of 14C-labeled microalgal photosynthetic pigments. *Journal of Plankton Research* 18, 1867–1880 (1996).

95. Van Heukelem, L., Lewitus, A. J., Kana, T. M. & Craft, N. E. Improved separations of phytoplankton pigments using temperature-controlled high performance liquid chromatography. *Marine Ecology Progress Series* 114, 303–313 (1994).

96. Bengtsson, J. *et al.* Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit (12S/16S/18S) sequences of archaea, bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Antonie Van Leeuwenhoek* 100, 471–475, doi:10.1007/s10482-011-9598-6 (2011).

97. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res* 17, 377–386 (2007).

98. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75, 7537–7541, doi:AEM.01541-09 [pii]10.1128/AEM.01541-09 (2009).

Acknowledgements

Author contributions

FRV conceived the work. Sample collection was done by CLD, KI, BD, FRV and AC. RG, CM, AP, CMM, RV, CLD, KM and AC analysed the data. RG, AC and FRV wrote the manuscript. All authors read and approved the final manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

How to cite this article: Ghai, R. *et al.* Metagenomes of Mediterranean Coastal Lagoons. *Sci. Rep.* 2, 490; DOI:10.1038/srep00490 (2012).