Distribution, abundance and ecogenomics of the cosmopolitan Gemmatimonadota group PAUC43f

Borja Aldeguer-Riquelme
University of Alicante
Josefa Antón (anton@ua.es)
University of Alicante
Fernando Santos
University of Alicante

Research Article

Keywords: Gemmatimonadota, PAUC43f, cosmopolitan, marine, ecology

Posted Date: May 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1666545/v1

Abstract

Background

The phylum *Gemmatimonadota* comprises mainly uncultured microorganisms and presents a high level of diversity, similar to that of *Proteobacteria* or *Actinobacteria*. This diversity reflects the phylum's wide range of functional capabilities, which allows its members to inhabit different environments such as soils, freshwater lakes, marine sediments, sponges or corals. Some studies have focused on the *Gemmatimonadota* present in soils or freshwater lakes, however, to date, the information about this group in the marine environment is very scarce. One of the most frequently retrieved *Gemmatimonadota* from marine environments is PAUC43f, an uncultured group that has been detected only by 16S rRNA gene sequence analyses, with not a single isolate or metagenome-assembled genome (MAG) affiliated with this group.

Results

Here, we carried out a broad study of the distribution, abundance, ecotaxonomy and potential metabolic capabilities of PAUC43f by mining information from 16S rRNA gene amplicon datasets, *Gemmatimonadota* MAGs, and fluorescent *in situ* hybridization (FISH). We detected PAUC43f in 4,965 amplicon datasets mainly from marine sediments, sponges, corals and saline soils and lakes around the world, which indicated this group is cosmopolitan and adapted to saline environments. PAUC43f represented a substantial fraction of the microbial community, mainly in sediments, sponges and soils where, in some cases, it could account for up to 34.3% of the 16S rRNA gene sequences. The potential metabolic capabilities inferred from the 39 PAUC43f MAGs suggested a facultatively aerobic and chemoorganotrophic metabolism, although some members could also oxidise hydrogen. Furthermore, PAUC43f could play a key environmental role as a N_2O consumer as well as a supplier of cysteine, serine and thiamine. Finally, the design and test of a FISH probe provided experimental evidence of the presence, activity, and morphology of PAUC43f in marine sediments. We propose its renaming as "*Candidatus* Palaucibacterales", in reference to the Republic of Palau where the first sequence of this group was recovered, and its classification as a new order within the *Gemmatimonadot*a phylum and *Gemmatimonadetes* class.

Conclusions

Our findings highlight the global distribution and the potentially relevant ecological roles of PAUC43f in different worldwide distributed environments, mainly as N_2O scavengers as well as vitamin and amino acid suppliers.

Background

Over the last three decades, the development of culture-independent techniques has allowed the study of many microbial taxa that had remained hidden due to culture limitations. Among these taxa, the phylum *Gemmatimonadota* (also known as *Gemmatimonadetes*) was discovered in 2001 by two independent studies which used 16S rRNA gene clone libraries to explore the microbial diversity of a reactor sludge and coastal marine sediments [1, 2]. Formerly designated as "candidate division BD" (or KS-B), this phylum was renamed in 2003 when the strain T-27^T, which belongs to this group, was isolated from a wastewater treatment plant and named as *Gemmatimonas aurantiaca* [3]. Since then, only five more species have been isolated: *G. phototrophica* [4], *G. groenlandica* [5], *Roseisolibacter agri* [6] and *Gemmatirosa kalamazoonesis* [7], included in the *Gemmatimonadetes* class; and *Longimicrobium terrae*, in the *Longimicrobia* class [8]. Although only two classes contain isolated species, the phylum *Gemmatimonadota* comprises up to seven classes according to SILVA r138 [9]. In addition to *Gemmatimonadetes* and *Longimicrobia*, there are the BD2-11 and S0134 terrestrial groups, the PAUC43f marine benthic group, and the MD2902-B12 and AKAU4049 classes, all of them lacking cultured representatives. In fact, approximately 86% of all 16S rRNA gene sequences of *Gemmatimonadota* deposited in the SILVA database have been retrieved from uncultured members of the phylum. Consequently, the available information about the ecology and physiology of this phylum is still scarce.

Previous studies based on 16S rRNA gene sequences have highlighted *Gemmatimonadota* as a cosmopolitan phylum, as diverse as *Actinobacteria* or *Proteobacteria* [10], which could reflect a broad physiological potential that allows this group to colonize a high variety of environments. The *Gemmatimonadota* are present in many types of soils, constituting one of the eight most abundant phyla and accounting for between 0.2% and 6.5% of total 16S rRNA gene sequences [11, 12]. Recently, Bay and co-workers suggested the metabolic potential of soil *Gemmatimonadota* to oxidize CH_4 and reduce N_2O, both potent greenhouse gases [13]. Indeed, *in vitro* experiments in other members of this phylum have confirmed its ability to reduce N_2O [14, 15]. On the other hand, *Gemmatimonadota* are also present in the water column and sediments of freshwater lakes [16–19]. A recent study in Czechia and Switzerland freshwater lakes estimated that *Gemmatimonadota* could represent up to 1% of the water column microbial community, with higher relative abundances in the hypolimnion than in the epilimnion [17]. In these environments, chemoorganotrophic and photoheterotrophic *Gemmatimonadota* are present, as revealed by cultures [4, 5] and metagenomics [16, 17]. *Gemmatimonadota* have also been found in marine environments, such as seawater [20, 21], marine sediments [22–25] and sponges [10, 26, 27]. Due to this ubiquity in marine environments, Hanada & Sekiguchi, 2014 suggested that *Gemmatimonadota* could play an important role, still unexplored, in the oceans [10].

PAUC43f is one of the most frequently detected classes of *Gemmatimonadota* in marine environments [10]. Although the first 16S rRNA gene sequence assigned to this class was discovered 20 years ago [28, 29], and it is the third largest class of *Gemmatimonadota* in the SILVA database, very little is known about its ecology and physiology. Furthermore, to date, PAUC43f members have only been detected through 16S rRNA gene sequences and there is not a single isolate or metagenome-assembled genome (MAG) affiliated with this group. There is some information that suggests PAUC43f is a salt-adapted group, present in marine sediments, hydrothermal vents, sponges and corals [23, 30–35] and also in ephemeral saline lake sediments [36, 37].

In this work, we aimed to fill the gap of information about the distribution, abundance, physiology, and ecological role of the *Gemmatimonadota* PAUC43f class. For this purpose, we retrieved all the PAUC43f 16S rRNA gene sequences from SILVA r138 and performed an extensive search for PAUC43f in 189,104 publicly available 16S rRNA gene amplicon datasets from the Sequence Read Archive (SRA). Furthermore, several databases were screened for PAUC43f MAGs that had been previously overlooked. Our results confirmed the widespread distribution of PAUC43f in salt-related environments (brackish to hypersaline, with salinities between 0.1-18.5%), with the highest abundances in sponges, marine sediments and soil samples. Based on 16S rRNA gene sequences, 16 genera were defined and linked to different ecological niches. The metabolic potential of PAUC43f, based on MAGs, indicated that these bacteria are chemoorganotrophs able to perform both aerobic and anaerobic respiration. Since some members of PAUC43f can reduce N_2O, a potent greenhouse gas, these bacteria may be helpful for mitigating the harmful effects of this gas. Also, the potential capability to synthetize cysteine, serine and vitamin B1 (thiamine) was found in almost all PAUC43f MAGs, suggesting that they could play an important role by supplying the community with these compounds.

Material And Methods

PAUC43f 16S rRNA gene analyses

A dataset was built with 3,686 complete and partial 16S rRNA gene sequences classified as “PAUC43f marine benthic group” (i) retrieved from SILVA r138 database [9] and (ii) 16S rRNA gene sequences from marine invertebrates (corals and sponges from the Mediterranean Sea [38]) and sediments from the Mar Menor lagoon (SE, Spain; Aldeguer-Riquelme et al., submitted). Sequences were aligned using SINA [39], in the ARB software [40], and introduced by parsimony into the SILVA 16S rRNA tree to check their taxonomy. Only sequences clustering within the PAUC43f group were kept for further analyses. To avoid redundancy, sequences were clustered with cd-hit-est [41] at 97% of identity, a threshold commonly used for species delineation [42], and 90% of coverage. As a result, 384 groups were generated, and the longest sequence of each group was selected as the representative for subsequent analyses. The global distribution of PAUC43f was drawn in R with the ggplot2 [43] and tidyverse [44] packages, based on the information provided by the sequence metadata. The 179 sequences with available geographic coordinates were included in the map.

The presence of PAUC43f in different environments was estimated using the IMNGS software [45]. The abovementioned 384 representative sequences were searched in a total of 189,104 16S rRNA gene amplicon datasets, available in the SRA repository, from 16 different environments (air, coral, estuary, fish, freshwater, human gut, human not gut, hydrothermal, hypersaline, marine sediment, marine sediment mat, sea water, oyster, skin, soil and sponge) using a 97% identity cut-off. To obtain a more precise value of PAUC43f relative abundances, estimated by the percentage of total 16S rRNA gene sequences, those SRA datasets where PAUC43f was detected by IMNGS (4965 datasets) were downloaded, BLASTN queried against the 384 sequences and only those hits above 97% identity and 70% coverage were considered. Therefore, we only estimated the abundance of each detected species (at 97% identity, as mentioned), although probably there are still more undetected, which implies that the calculated PAUC43f abundances were likely underestimates of their true abundances.

For precise taxonomic studies, the 66 sequences longer than 800 bp (from the 384 representative sequences) were analysed in the ARB software. SINA was used to align the sequences and, to exclude highly variable positions, a base frequency filter was applied prior to the tree construction with both neighbor-joining (Jukes-Cantor correction) and maximum likelihood (PHYML) algorithms (1,000 bootstraps). The tree was constructed with those sequences longer than 1,200 bp and then, sequences between 800 and 1200 bp were added by parsimony. BD2-11, MD2902-B12 and *Gemmatimonadetes* sequences were used as outgroups. A cluster representing a genus was defined when at least two sequences were monophyletic in both neighbor-joining and maximum likelihood trees [46, 47] and their identities were above 94.5%, the threshold for genus delineation [48]. Finally, iTOL was employed to draw the tree [49]. The environmental frequency and abundance of each genus was estimated as explained above.

Metagenome-assembled genome analyses

Metagenome-assembled genomes (MAGs) belonging to PAUC43f were searched for in the GTDB release 95 [50] and GEM databases [51] as well as in other public sources (https://data.ace.uq.edu.au/public/sponge_mags/; [26, 27]) and Mar Menor sediment metagenomes sequenced in our lab. Two different but complementary strategies were followed to retrieve PAUC43f MAGs. First, 16S rRNA gene sequences were extracted from all available *Gemmatimonadota* MAGs and classified in the online SILVA ACT service (https://www.arb-silva.de/aligner/). Then, MAGs carrying a 16S rRNA gene sequence of PAUC43f were classified using the whole genome classifier tool GTDB-tk [50]. The second strategy relied on comparison, based on AAI and phylogenomic trees, of all available *Gemmatimonadota* MAGs with those classified as PAUC43f by 16S rRNA gene sequences. AAI was estimated using the aai.rb script from the enveomics toolbox [52] while phylogenomic trees were constructed with PhyloPhlAn [53].

MAGs identified as PAUC43f were manually curated as previously proposed [54]. Completeness and contamination were estimated using CheckM [55]. Metabolic reconstruction was carried out using the annotation provided by KAAS KEGG [56], Interproscan (using Pfam, CDD, SMART and TIGRFAM databases) [57–61] and BLASTP against the NCBI non-redundant database [62]. Secondary metabolite biosynthetic gene clusters (BGCs) were identified by antiSMASH [63] with the “strict” detection level.

Fluorescence in situ hybridization (FISH)

To get experimental information (presence, activity, morphology, size) about this group, PCR primers and FISH probes were designed using DECIPHER [64] and PrimerQuest Design Tool (IDT, https://eu.idtdna.com/PrimerQuest/Home/Index). Since Mar Menor sediment samples (Murcia, Spain; 37°45'N 0°47'W), where PAUC43f had been previously detected, were readily accessible to our lab, we designed these primers and probe with the 16S rRNA gene

sequences of PAUC43f retrieved from these sediments (Aldeguer-Riquelme *et al*., submitted). *In silico* quality control was performed using OligoAnalyzer Tool (IDT, https://eu.idtdna.com/pages/tools/oligoanalyzer), searching for secondary structures and dimerization, while probe specificity was checked with TestProbe against SILVA database [9]. As a result, probe PAUC43f_826 (5'- AGGGTCAATCCTCCCAACACCTAGTAC-3'), which covered 32.7% of the PAUC43f sequences from SILVA, was selected as the best candidate. To test the probe, sediment samples from the Mar Menor lagoon were collected in the summer of 2021 and fixed with 4% formaldehyde at 4°C for 4 hours. The presence of PAUC43f in these samples was confirmed by PCR with specific primers for this group (272F: 5'-GTAAGTCGGGTGTGAAATTC-3'; 393R: 5'-TTCCCGATATCTACGCATTC-3') which covered the 11.2% of the PAUC43f sequences in SILVA. The hybridization was carried out on a filter as previously described [65] and the probe was optimized using a range of formamide concentrations (10, 20, 30, 40, 50 and 60%). Briefly, hybridization was done at 46°C for 4 hours followed by two washing steps at 48°C for 15 min. Then, filters were stained with DAPI (1 mg/mL), washed with milli-Q water, dehydrated with absolute ethanol (1 min each step) and finally visualized in the Zeiss LSM800 confocal laser scanning microscope.

Results And Discussion

Ecological distribution

PAUC43f 16S rRNA gene sequences were detected in several marine environments such as sediments, corals, sponges, oysters, estuaries, sea water and hydrothermal vents, in addition to samples from hypersaline lake sediments, soil and marine sediment mats (Fig. 1A). It was remarkable that, from the 179 sequences that were included in the map, 89 were recovered from marine sediment samples. Regarding the geographical distribution, PAUC43f was detected around the world in almost every latitude and longitude, and in both shallow and deep environments.

To get more insights into the PAUC43f ecological distribution, its relative abundance (% of 16S reads from the total) was estimated for each environment (Fig. 1B). From the total of 189,104 16S rRNA gene amplicon datasets analysed, PAUC43f was detected in 4,965 of them (Suppl. Table 1). The highest mean relative abundances were in sponges, marine sediments and soils (Fig. 1B) while the lowest values were found in sea water and hydrothermal samples. It is remarkable that PAUC43f could reach extremely high relative abundances as observed, for example, in an arid saline soil from China [66] and in petroleum-impacted sediments from a saline lake [67], where it represented 34.3% and 19.3% of the total community, respectively.

Since PAUC43f reached its highest relative abundances in sponges, sediments and soil samples, its distribution in these environments was explored more deeply. PAUC43f was detected in at least 30 different sponge species, being most frequently found in *Coscinoderma matthewsi*, where it accounted for up to 5.4% of the 16S rRNA gene sequences, *Xestospongia* spp., *Rhopaloeides odorabile* and *Suberites* spp. Regarding marine sediments, no clear pattern of distribution related to latitude or the temperature was observed; PAUC43f was as abundant in cold as in warm waters (Suppl. Figure 1A-1B), although the highest abundances were found in middle latitudes. No pattern related to the water column depth above the sediment was found, and it was remarkable that PAUC43f was detected even in a sediment with an overlaying water column of 5813 m, which suggests it can resist strong pressure (Suppl. Figure 1C). On the other hand, sediment depth seemed to be important since PAUC43f abundances were highest at the surface and decreased with depth (Suppl. Figure 1D). For soils, the highest abundances were found in middle latitudes in the northern hemisphere, although it is important to note that this hemisphere presents a higher proportion of land than the southern hemisphere (Suppl. Figure 2A). As for sediments, the abundance of PAUC43f in soils was also related to soil depth, with higher abundances at the surface (Suppl. Figure 2B).

Ecotaxonomy

The 16S rRNA-based phylogenetic tree revealed 16 different genera supported by both neighbour-joining and PHYML algorithms (Fig. 2), which included 62% of the total tree sequences. All these genera, and indeed almost all the sequences included in the tree (except AB305477.1.916), belonged to the same order and the same family, based on previously proposed thresholds for these taxonomic ranks (order: 82.0%, family: 86.5%; Yarza et al., 2014).

To analyse the ecological distribution of these genera, their frequencies and abundances in each environment were calculated. As shown in the Fig. 3A, the frequency of each genus differed across environments: some genera, such as 1, 3, 4, 6 and 9, displayed a wide environmental distribution, while others, such as genera 10, 11, 12 and 13, were limited to a few environments and samples. All genera were detected in corals, sea water, sediment and soil whereas only a few genera were found in fish, hydrothermal vents, hypersaline lake sediments and marine sediment mats. Regarding relative abundances (Fig. 3B), these genera are included within the rare biosphere in many environments (< 0.1%; [68]); however, in certain environments, some genera had moderate to high relative abundances (> 0.1%). For example, genus 6 was remarkably abundant in hypersaline lake sediments, marine sediments, and soil samples while genus 16 was clearly host-associated, with abundances above 0.1% in coral and sponge samples. Further, genera 7 and 9 had abundances above 0.1% in marine sediment and hydrothermal samples, and genus 10 had high abundances in hypersaline lake sediments and soils. These observations suggest that each genus might be adapted to specific environments. This finding implies that at least some genera are genuine members of microbiomes of corals, sponges, marine sediments, hypersaline lake sediments and soils. However, since their abundances and frequencies in fish, sediment mats and oysters are low, they are most likely detected in these types of samples as a passively transported bacteria.

Phylogenomic and metabolic analyses

Searching genome/MAG databases (GEM & GTDB r95) and recent publications [26, 27] led to the identification of 24 MAGs: 6 from the GEM database, 0 from GTDB, and 18 from the publications. An additional 15 MAGs were recovered from metagenomes sequenced in our lab from Mar Menor sediments (Aldeguer-Riquelme et al., in prep). All 39 MAGs (Table 1) met the criteria to be considered of good quality by having completeness above 80% and contamination below 5% [54, 69]. In addition, 14 MAGs also carried 16S rRNA genes (Table 1). The MAG sizes ranged from 1.80 to 4.07 Mb with a

minimum GC content of 65.4% and a maximum of 71.7%. Regarding their origins, the MAGs were obtained from sponge, marine sediment and soil metagenomes (18, 18 and 3 MAGs respectively, Table 1; Suppl. Table 2). A relationship between MAG origin and size, independent of the completeness, was observed, with the smallest genomes found in marine sediment MAGs and the highest in sponge MAGs (Suppl. Figure 3). On the other hand, in terms of abundance, most of these MAGs had abundances above 0.1% (0.05% – 12.52%, Table 1) in their original metagenomes and thus belonged to the abundant biosphere.

A phylogenomic tree inferred using all *Gemmatimonadota* genomes and MAGs available in the GTDB supported the monophyletic origin of PAUC43f within this phylum (Fig. 4A). MAGs recovered from sponges clustered in a PAUC43f sub-branch different from those from sediments and soils. A similar result was obtained when the AAI was calculated among them (Fig. 4B). Thus, PAUC43f MAGs clustered according to their origin, which agrees with the results of 16S rRNA gene analyses (Figs. 2 and 3). Indeed, 16S rRNA gene sequences retrieved from MAGs were also classified and showed that some genera were associated with specific environments, as showed above (Fig. 2). These results strongly supported the specialization of these MAG lineages on specific ecological niches.

With respect to the taxonomic range of MAGs, the phylogenomic tree and AAI values (Fig. 4B) indicated that the 39 MAGs represented 14 different species (AAI $\geq$ 95%; [69, 70]), 8 of which were recovered at least twice from different metagenomes. The sponge MAGs belonged to 8 different species of the same genus while the 6 species from soils and sediments represented 4 different genera (AAI $\leq$ 65%; [69]). Hereinafter, the analyses will focus on these 14 species.

Table 1
General characteristics of PAUC43f MAGs. aStrain heterogeneity. bMAG abundance is shown as percentage of recruited reads from the total metagenome reads.

BinID	Contigs	bp	%GC	16S (Genus)	Completness (%)	Contamination (%)	SH[a] (%)	Abundance[b] (%)	Origin	Reference
IRC1_bin_13	43	3603896	69.29	No	91.21	3.3	0	1.96	Sponge	Engelberts et al., 2020
IRC2_bin_12	180	3551308	69.26	No	93.41	3.3	0	0.94	Sponge	Engelberts et al., 2020
IRC3_bin_28	79	3616708	69.29	No	95.6	3.3	0	1.38	Sponge	Engelberts et al., 2020
IRC4_bin_13	52	3197555	68.94	No	93.41	3.3	0	1.48	Sponge	Engelberts et al., 2020
RHO1_bin_50	58	4078012	68.89	No	94.51	4.4	0	5.38	Sponge	Engelberts et al., 2020
RHO2_bin_49	54	3475617	69.01	Yes (16)	95.54	3.3	0	2.72	Sponge	Engelberts et al., 2020
RHO3_bin_38	43	3421550	69.05	No	91.14	3.3	0	6.59	Sponge	Engelberts et al., 2020
VXMQ01000000	219	3760526	69.11	No	95.54	3.3	0	0.70	Sponge	Engelberts et al., 2020
VXMR01000000	158	3163883	69.03	No	94.51	3.36	0	0.55	Sponge	Engelberts et al., 2020
VXOJ01000000	71	3403461	69.49	No	94.51	3.3	0	1.73	Sponge	Engelberts et al., 2020
VXQX01000000	96	3603938	69.05	No	94.51	3.3	0	12.52	Sponge	Engelberts et al., 2020
VXSM01000000	159	3454019	68.84	No	93.41	3.3	0	1.16	Sponge	Engelberts et al., 2020
VXWG01000000	114	3599158	69.44	No	95.6	3.3	0	2.06	Sponge	Engelberts et al., 2020
VXXU01000000	424	2999705	68.98	No	84.55	2.2	0	1.27	Sponge	Engelberts et al., 2020
VXYT01000000	102	3583507	69.04	No	95.6	3.3	0	1.99	Sponge	Engelberts et al., 2020
VYCV01000000	309	3082878	69.08	No	90.11	4.5	0	0.51	Sponge	Engelberts et al., 2020
VYDQ01000000	196	3551893	69.02	No	94.51	3.3	0	1.27	Sponge	Engelberts et al., 2020
VYFI01000000	341	3232310	68.91	No	87.3	2.2	0	0.66	Sponge	Engelberts et al., 2020
3300025550_6	399	2468409	67.08	No	80.49	3.85	0	0.23	Marine sediment	Tringe, unpublished
3300025554_5	315	2894191	67.11	Yes (2)	91.21	2.2	0	0.20	Marine sediment	Tringe, unpublished
3300026122_2	273	2964121	71.34	Yes	90.91	3.3	0	0.48	Soil	Zhou et al., 2022
3300026127_2	119	3341153	71.38	Yes	96.7	3.3	0	0.98	Soil	Zhou et al., 2022
3300026196_11	67	3192904	71.26	Yes	89.01	3.3	0	1.44	Soil	Zhou et al., 2022
3300027962_15	199	2956959	71.68	No	96.15	4.4	0	0.62	Marine sediment	Kimbrel, unpublished
Bin_M15_27	604	2793946	65.52	Yes (4)	90.11	0	0	1.64	Marine sediment	Aldeguer-Riquelme et al., unpublished

BinID	Contigs	bp	%GC	16S (Genus)	Completness (%)	Contamination (%)	SH[a] (%)	Abundance[b] (%)	Origin	Reference
Bin_S15_23	531	2769550	65.51	Yes (4)	90.56	0.55	0	1.09	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_M162_005	197	2934314	65.49	Yes (4)	95.3	3.3	0	0.34	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_M182_011	153	2696093	65.49	Yes (4)	94.51	4.4	25	0.19	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_M42_016	78	1799666	65.37	Yes (4)	80.22	1.1	0	0.07	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_M51_014	235	2719955	65.52	No	94.51	2.75	0	0.20	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_M92_019	219	2740149	65.42	Yes (4)	95.6	4.4	25	0.16	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S162_005	153	2826554	65.49	Yes (4)	96.7	4.4	25	0.37	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S162_007	280	2573992	66.45	No	93.96	3.6	16.67	0.25	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S182_010	257	2980185	65.47	No	96.7	3.3	0	0.26	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S212_14	88	2387045	65.52	Yes (4)	91.71	3.3	0	0.47	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S43_010	308	2799597	65.57	No	94.51	4.95	16.67	0.26	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S43_056	396	2710074	65.57	Yes (2)	82.99	3.3	0	0.05	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S53_006	260	2329960	65.44	No	93.46	2.75	33.33	0.68	Marine sediment	Aldeguer-Riquelme et al., unpublished
Bin_S91_007	166	2340983	65.52	No	92.49	3.3	0	0.28	Marine sediment	Aldeguer-Riquelme et al., unpublished

To shed light on the ecological role of PAUC43f, the potential metabolic capabilities of each species were explored (Fig. 5, Suppl. Table 3). In terms of cell wall structure, PAUC43f was a gram-negative bacterium which lacked the genes for the flagellar assembly (except species 10). Regarding central carbon metabolism, complete glycolysis and tricarboxylic acid cycle (TCA) pathways were found in almost all species, as were sugar transporters, which suggested that PAUC43f has a chemoorganotrophic metabolism. The lack of phosphoglucose isomerase in species 10, 11, 12, 13, and 14, and phosphoglycerate kinase and malate dehydrogenase in species 8 may have been due to MAG incompleteness since all other enzymes were present in these MAGs. Genes related to carbon fixation or photosynthetic metabolism were not found. However, species from sediments and soils presented 1c and 1f hydrogenases [71], which indicated they could potentially shift between chemoorganotrophic and chemolithotrophic metabolism. It is noteworthy that hydrogenotrophic metabolism in other *Gemmatimonadota* members has been recently demonstrated [72].

Members of PAUC43f seemed to be facultative aerobes since genes for complex IV cytochrome oxidase, which transfers electrons to oxygen, were detected in all MAGs, although most of them also encoded the genes for the nitrate, nitrite and/or nitrous oxide respiration. In addition, species found in sponges were able to respire thiosulfate, and species from sediments and soils could likely carry out alcoholic fermentation. The potential for PAUC43f lineages represented by MAGs from sediment and soils to reduce N_2O is in agreement with previous observations from other *Gemmatimonadota* members [14, 15, 73] and highlights its ecological relevance. N_2O is a potent greenhouse gas, which, due to human activities such as agricultural fertilization of combustion of fossil fuels [74], is increasing its atmospheric concentrations at a rate of 0.8 ppb per year [75], with some of the highest concentrations measured in coastal and estuarine waters [76, 77]. Thus, N_2O consumers, such as PAUC43f, could play a key role in mitigating the harmful effects of this gas.

Biosynthetic pathways for amino acids, B vitamins, and secondary metabolites, which are relevant molecules for microbial metabolism and physiology, were explored in PAUC43f. With respect to amino acid biosynthesis, all PAUC43f species encoded the complete pathways for synthesis of glutamate, alanine, aspartate, glycine, threonine, cysteine and leucine, while the pathways for the rest of amino acids were complete in only some species. It is noteworthy that species from sponges were able to synthesize more amino acids (16–18) than species from sediments and soils (10–13). The most extended putative auxotrophies were found for valine, isoleucine, lysine and histidine; each of these could be only synthetized by two species. Species from sponges were all auxotrophic for histidine, whereas species from sediments and soils were auxotrophic for valine, lysine, isoleucine, tyrosine, proline and phenylalanine. However, species from sponges could perhaps circumvent valine and isoleucine auxotrophies by encoding a branched-amino acid transporter and, in sediment and soil species, tyrosine might be acquired by cotransport with H^+. Furthermore, transporters for oligopeptide import were found in all species. A possible explanation for these auxotrophies might be due to each amino acid's frequency of use and biosynthetic cost. The most used amino acids in the PAUC43f proteome were alanine, leucine and glycine (Suppl. Figure 4) which could be produced by these species with a low metabolic cost (alanine and glycine; [78]). In contrast, histidine and lysine, whose biosynthesis is metabolically more expensive [78], were among the less frequent amino acids and most species were auxotrophic for them. On the other hand, PAUC43f could play an important ecological role providing cysteine and serine to the marine community since their biosynthetic pathway and efflux transporters were encoded in almost all species, and serine auxotrophy has been demonstrated for important marine bacteria such as *Pelagibacter ubique* [79].

In regards to the potential for vitamin B production (Suppl. Table 3), essential core biosynthetic genes for thiamine (vitamin B1) (*thiC*, *thiG* and *thiE*), which is a cofactor of several essential enzymes [80], were detected in all species except species 10 that lacked *thiC*. Given that B1 auxotrophy is frequent in both eukaryotic and prokaryotic marine communities [81–83], and is the second most common in marine environments [84], PAUC43f could be a key supplier of B1 to the marine communities. Regarding riboflavin (vitamin B2), which is a precursor of the coenzymes FAD and FMN [85], all PAUC43f MAGs encoded the complete biosynthetic operon for this vitamin. Niacin or vitamin B3 act as coenzymes in redox reactions and could also be synthetized by PAUC43f species from soils (species 9) and marine sediments (species 10, 11 and 14). On the other hand, the pathway for pantothenate (vitamin B5), a precursor of coenzyme A, was found complete only in species 9 while species 1, 3, 4, 5, 6, 7, 8, 10 and 11 lacked one gene and species 2, 12, 13 and 14 lacked several. Similarly, most genes involved in the biosynthesis of folate (vitamin B9), an important molecule in anabolic reactions, were found in PAUC43f MAGs (*folE*, *DHPS*, *folC*, *phoD*, *folB*, *PTPS*) but dihydrofolate reductase (*DHFR*) and 4-amino-4-deoxychorismate lyase were only present in species 9 and species 11, respectively. Thus, the presence of most genes involved in pantothenate and folate biosynthesis suggests that some PAUC43f species may be capable of synthesizing these vitamins, and that missing genes were likely not binned in the MAGs, however, this remains to be tested in future studies. Biosynthetic pathways for vitamins B6, B7 and B12 were not found, and the presence, in all species, of the *bioY* gene, which encodes a biotin (vitamin B7) transporter [86], and *btuF* and *btuB*, which are part of the cobalamin (vitamin B12) transporter [87], suggests PAUC43f imports these vitamins from the extracellular environment.

Secondary metabolites are usually involved in growth, development and defense [88], and they are interesting molecules for medicine due to their potential uses as antibiotic, anti-tumoral and cholesterol-lowering drugs. PAUC43f MAGs were analysed with antiSMASH [63], which revealed that sponge MAGs presented a higher number and diversity of BGC (4–9 BGC per MAG) than those from sediments and soils (1–2 BGCs per MAG) (Fig. 6, Suppl. Table 4). Species from sponges encoded type I polyketide synthase (T1PKS), ranthipeptide and ribosomally synthesised and post-translationally modified peptides (RiPPs). Most of them had low similarity to previously described BGCs so their products could not be predicted. However, some T1PKS were similar to those known to synthetize azinomycin B, a potent antibiotic with antitumor activity [89, 90], cyphomycin, an antifungal compound [91], and vazabitide A and funisamine, both with unknown biological properties [92, 93]. On the other hand, terpene and β-lactone producing bacteria are frequently found in marine sediments and sponge symbionts [88, 94], thus it was not unexpected to find them encoded by most PAUC43f species.

Antibiotic and heavy metal resistance genes were detected in PAUC43f MAGs. With respect to antibiotic resistance genes, PAUC43f MAGs encoded β-lactamases, tetracycline/H^+ antiporters and fosmidomycin and macrolide efflux pumps. With respect to heavy metal resistance genes, they were mainly present in sediment and soil species. Bacterioferritin, an iron storage protein which also protects the cell from the reactive Fe^{2+}, was present in species 11, 12, 13 and 14. Additionally, some species could be resistant to As^{3+} (species 9, 10, 11, 12, 13 and 14), Zn^{2+}, Pb^{2+} and Cd^{2+} (species 9, 11, 12, 13 and 14) by exporting these metals out of the cell.

Fluorescence in situ hybridization (FISH)

To visualize PAUC43f cells and obtain evidence of their presence and metabolically active state in the environment, FISH probes were designed and tested. *In silico* analyses indicated that the probe matched 32% of the PAUC43f sequences deposited in the SILVA database and sequences of genera 2, 3, 4, 5, 6, 7 and 8 (Fig. 2). Thus, the probe does not target the whole PAUC43f group but rather a set of closely related sequences, most of them associated with marine sediments and soils, the most important environments for the free-living PAUC43f. Since the probe matched with 16S rRNA gene

sequences within MAGs recovered from Mar Menor sediments, fluorescence *in situ* hybridization was performed with sediment samples from this lagoon. The best result was obtained with 40% formamide and, as shown in Fig. 7, PAUC43f cells displayed a small but wide rod morphology. Since more cells than the highlighted ones seemed to be hybridized with the PAUC43f_826 probe but not DAPI signal, we suspected DAPI may have been quenched by the probe fluorophore or by pigments present in the cells. With this assay, we provide experimental evidence of the presence and metabolically active state of PAUC43f in marine sediments.

Conclusions

In this study we explored the environmental distribution and relevance of PAUC43f, an overlooked cosmopolitan and salt-related group within the *Gemmatimonadota* phylum. We show that PAUC43f is present in a wide variety of samples and environmental conditions reaching, in some cases, very high abundances, indicating that the ecological role of this group needs to be further studied and considered. Indeed, Hanada & Seguchi, 2014 predicted an important, but still unexplored, environmental role for marine *Gemmatimonadota*. Our data shows that PAUC43f could be an important environmental ally to mitigate global warming effects due to its potential capability to reduce N_2O and its widespread presence in marine sediments, the most extensive area on Earth. Furthermore, this group could be key for supplying thiamine, cysteine and serine to the marine community. Together these findings highlight the ecological relevance of PAUC43f and support the prediction made by Hanada & Seguchi, 2014. We provide, for the first time, insights into the ecology, metabolic potential and ecosystem services of PAUC43f at the global scale. Based on phylogenomic analyses we propose its classification as a new order within the *Gemmatimonadot*a phylum and *Gemmatimonadetes* class with the name of "*Candidatus* Palaucibacterales", in reference to the Republic of Palau where the first sequence of this group was recovered.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The list of SRA 16S rRNA gene amplicon runs used in this study are available in the Suppl. Table 1. The MAG dataset recovered in this study and supporting the conclusions of this article are available in the NCBI repository, under BioProject PRJNA838580. MAGs recovered from public databases are available in:

https://data.ace.uq.edu.au/public/sponge_mags/

https://portal.nersc.gov/GEM/

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the EU-H2020 MetaFluidics project with grant agreement number 685474 (to J.A.). B. A.-R. is an ACIF fellow (Generalitat Valenciana).

Authors' contributions

B.A-R, F.S. and J.A. conceived and designed the study. B.A-R performed the analyses under the supervision and guidance of F.S. and J.A. All authors discussed, wrote, read and approved the manuscript.

Acknowledgements

We thank Ramon Rosselló-Móra for his help with 16S rRNA gene tree construction and uncultured taxa naming and Heather Maughan for her professional English editing and the critical reading of the manuscript.

References

1. Hugenholtz P, Tyson GW, Webb RI, Wagner AM, Blackall LL. Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria, with no known pure-culture representatives. Appl Environ Microbiol. 2001;67:411–9.

2. Madrid VM, Aller JY, Aller RC, Chistoserdov AY. High prokaryote diversity and analysis of community structure in mobile mud deposits off French Guiana: Identification of two new bacterial candidate divisions. FEMS Microbiol Ecol. 2001;37:197–209.
3. Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, Kamagata Y, et al. *Gemmatimonas aurantiaca* gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes* phyl. nov. Int J Syst Evol Microbiol. 2003;53:1155–63.
4. Zeng Y, Selyanin V, Lukeš M, Dean J, Kaftan D, Feng F, et al. Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium *Gemmatimonas phototrophica* sp. Nov., and emended descriptions of the genus *Gemmatimonas* and *Gemmatimonas aurantiaca*. Int J Syst Evol Microbiol. 2015;65 Pt_8:2410–9.
5. Zeng Y, Nupur Y, Wu N, Madsen AM, Chen X, Gardiner AT, et al. *Gemmatimonas groenlandica* sp. nov. is an aerobic anoxygenic phototroph in the phylum *Gemmatimonadetes*. Front Microbiol. 2021;11:606612.
6. Pascual J, Foesel BU, Geppert A, Huber KJ, Boedeker C, Luckner M, et al. *Roseisolibacter agri* gen. Nov., sp. nov., a novel slow-growing member of the under-represented phylum *gemmatimonadetes*. Int J Syst Evol Microbiol. 2018;68:1028–36.
7. DeBruyn JM, Fawaz MN, Peacock AD, Dunlap JR, Nixon LT, Cooper KE, et al. *Gemmatirosa kalamazoonesis* gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum *Gemmatimonadetes*. J Gen Appl Microbiol. 2013;59:305–12.
8. Pascual J, García-López M, Bills GF, Genilloud O. *Longimicrobium terrae* gen. nov., sp. nov., an oligotrophic bacterium of the under-represented phylum *Gemmatimonadetes* isolated through a system of miniaturized diffusion chambers. Int J Syst Evol Microbiol. 2016;66:1976–85.
9. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 2013;41 Database issue:D590-6.
10. Hanada S, Sekiguchi Y. The phylum *Gemmatimonadetes*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes. Berlin, Heidelberg: Springer; 2014.
11. Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol. 2006;72:1719.
12. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, et al. A global atlas of the dominant bacteria found in soil. Science (80-). 2018;359:320–5.
13. Bay SK, Dong X, Bradley JA, Leung PM, Grinter R, Jirapanjawat T, et al. Trace gas oxidizers are widespread and active members of soil microbial communities. Nat Microbiol. 2021;6:246–56.
14. Chee-Sanford J, Tian D, Sanford R. Consumption of N2O and other n-cycle intermediates by *Gemmatimonas aurantiaca* strain T-27. Microbiol. 2019;165:1345–54.
15. Park D, Kim H, Yoon S. Nitrous oxide reduction by an obligate aerobic bacterium, *Gemmatimonas aurantiaca* strain T-27. Appl Environ Microbiol. 2017;83:00502–17.
16. Fecskeová LK, Piwosz K, Hanusová M, Nedoma J, Znachor P, Koblížek M. Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. Sci Rep. 2019;9:18766.
17. Mujakić I, Andrei A-Ş, Shabarova T, Fecskeová LK, Salcher MM, Piwosz K, et al. Common presence of phototrophic *Gemmatimonadota* in temperate freshwater lakes. mSystems. 2021;6:e01241-20.
18. Song H, Li Z, Du B, Wang G, Ding Y. Bacterial communities in sediments of the shallow Lake Dongping in China. J Appl Microbiol. 2012;112:79–89.
19. Zhang J, Yang Y, Zhao L, Li Y, Xie S, Liu Y. Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. Appl Microbiol Biotechnol. 2015;99:3291–302.
20. Nunoura T, Takaki Y, Hirai M, Shimamura S, Makabe A, Koide O, et al. Hadal biosphere: Insight into the microbial ecosystem in the deepest ocean on Earth. Proc Natl Acad Sci U S A. 2015;112:E1230–6.
21. Nunoura T, Hirai M, Yoshida-Takashima Y, Nishizawa M, Kawagucci S, Yokokawa T, et al. Distribution and niche separation of planktonic microbial communities in the water columns from the surface to the hadal waters of the Japan Trench under the Eutrophic Ocean. Front Microbiol. 2016;7:1261.
22. Durbin AM, Teske A. Microbial diversity and stratification of South Pacific abyssal marine sediments. Environ Microbiol. 2011;13:3219–34.
23. Bergo NM, Bendia AG, Ferreira JCN, Murton BJ, Brandini FP, Pellizari VH. Microbial diversity of deep-sea ferromanganese crust field in the Rio Grande rise, southwestern Atlantic ocean. Microb Ecol. 2021;82:344–55.
24. Marcial Gomes NC, Borges LR, Paranhos R, Pinto FN, Mendonça-Hagler LCS, Smalla K. Exploring the diversity of bacterial communities in sediments of urban mangrove forests. FEMS Microbiol Ecol. 2008;66:96–109.
25. Schauer R, Bienhold C, Ramette A, Harder J. Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. ISME J. 2010;4:159–70.
26. Robbins SJ, Song W, Engelberts JP, Glasl B, Slaby BM, Boyd J, et al. A genomic view of the microbiome of coral reef demosponges. ISME J. 2021. https://doi.org/10.1038/s41396-020-00876-9.
27. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. Characterization of a sponge microbiome using an integrative genome-centric approach. ISME J. 2020;14:1100–10.
28. Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, et al. Molecular evidence for a uniform microbial community in sponges from different oceans. Appl Environ Microbiol. 2002;68:4431–40.

29. Webster NS, Wilson KJ, Blackall LL, Hill RT. Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile*. Appl Environ Microbiol. 2001;67:434–44.
30. Hardoim CCP, Ramaglia ACM, Lôbo-Hajdu G, Custódio MR. Community composition and functional prediction of prokaryotes associated with sympatric sponge species of southwestern Atlantic coast. Sci Rep. 2021;11:9576.
31. Griffiths SM, Antwis RE, Lenzi L, Lucaci A, Behringer DC, Butler MJ, et al. Host genetics and geography influence microbiome composition in the sponge *Ircinia campana*. J Anim Ecol. 2019;88:1684–95.
32. López-García P, Duperron S, Philippot P, Foriel J, Susini J, Moreira D. Bacterial diversity in hydrothermal sediment and epsilonproteobacterial dominance in experimental microcolonizers at the Mid-Atlantic Ridge. Environ Microbiol. 2003;5:961–76.
33. Radwan M, Hanora A, Zan J, Mohamed NM, Abo-Elmatty DM, Abou-El-Ela SH, et al. Bacterial community analyses of two red sea sponges. Mar Biotechnol. 2010;12:350–60.
34. Gerçe B, Schwartz T, Syldatk C, Hausmann R. Differences between bacterial communities associated with the surface or tissue of mediterranean sponge species. Microb Ecol. 2011;61:769–82.
35. Cerqueira T, Pinho D, Froufe H, Santos RS, Bettencourt R, Egas C. Sediment microbial diversity of three deep-sea hydrothermal vents southwest of the Azores. Microb Ecol. 2017;74:332–49.
36. Genderjahn S, Alawi M, Mangelsdorf K, Horn F, Wagner D. Desiccation-and saline-tolerant bacteria and archaea in kalahari pan sediments. Front Microbiol. 2018;9:2082.
37. Aerts JW, van Spanning RJM, Flahaut J, Molenaar D, Bland PA, Genge MJ, et al. Microbial communities in sediments from four mildly acidic ephemeral salt lakes in the yilgarn craton (Australia) – terrestrial analogs to ancient mars. Front Microbiol. 2019;10:779.
38. Rubio-Portillo E, Ramos-Esplá AA, Antón J. Shifts in marine invertebrate bacterial assemblages associated with tissue necrosis during a heat wave. Coral Reefs. 2021;40:395–404.
39. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.
40. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar A, et al. ARB: A software environment for sequence data. Nucleic Acids Res. 2004;32:1363–71.
41. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.
42. Stackebrandt E, Goebel BM. Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence snalysis in the present species definition in Bacteriology. Int J Syst Evol Microbiol. 1994;44:846–9.
43. Wickham H. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York; 2016.
44. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4:1686.
45. Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, et al. IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci Rep. 2016;6:33721.
46. Pohlner M, Dlugosch L, Wemheuer B, Mills H, Engelen B, Reese BK. The majority of active *Rhodobacteraceae* in marine sediments belong to uncultured genera: A molecular approach to link their distribution to environmental conditions. Front Microbiol. 2019;10:659.
47. Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A, et al. Phylogenetic and ecological analysis of novel marine stramenopiles. Appl Environ Microbiol. 2004;70:3528–34.
48. Yarza P, Yilmaz P, Pruesse E, Glockner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Micro. 2014;12:635–45.
49. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res. 2019;47:W256–9.
50. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2020;36:1925–7.
51. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2020. https://doi.org/10.1038/s41587-020-0718-6.
52. Rodríguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. Peer J Prepr. 2016;4:e1900v1.
53. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun. 2020;11:2500.
54. Ramos-Barbero MD, Martin-Cuadrado A-B, Viver T, Santos F, Martinez-Garcia M, Antón J. Recovering microbial genomes from metagenomes in hypersaline environments: The Good, the Bad and the Ugly. Syst Appl Microbiol. 2019;42:30–40.
55. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
56. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35:W182–5.

57. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
58. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.
59. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and genome properties in 2013. Nucleic Acids Res. 2013;41:D387–95.
60. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46:D493–6.
61. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43:D222–6.
62. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021;49:D10–7.
63. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, et al. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39:W339–46.
64. Wright ES, Safak Yilmaz L, Corcoran AM, Ökten HE, Noguera DR. Automated design of probes for rRNA-targeted fluorescence in situ hybridization reveals the advantages of using dual probes for accurate identification. Appl Environ Microbiol. 2014;80:5124–33.
65. Snaidr J, Amann R, Huber I, Ludwig W, Schleifer KH. Phylogenetic analysis and in situ identification of bacteria in activated sludge. Appl Environ Microbiol. 1997;63:2884–96.
66. Xie K, Deng Y, Zhang S, Zhang W, Liu J, Xie Y, et al. Prokaryotic community distribution along an ecological gradient of salinity in surface and subsurface saline soils. Sci Rep. 2017;7:13332.
67. Mustafa GA, Abd-Elgawad A, Ouf A, Siam R. The Egyptian Red Sea coastal microbiome: A study revealing differential microbial responses to diverse anthropogenic pollutants. Environ Pollut. 2016;214:892–902.
68. Pedrós-Alió C. The rare bacterial biosphere. Ann Rev Mar Sci. 2012;4:449–66.
69. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. ISME J. 2017;11:2399–406.
70. Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol. 2005;187:6258–64.
71. Søndergaard D, Pedersen CNS, Greening C. HydDB: A web tool for hydrogenase classification and analysis. Sci Rep. 2016;6:34212.
72. Islam ZF, Welsh C, Bayly K, Grinter R, Southam G, Gagen EJ, et al. A widely distributed hydrogenase oxidises atmospheric H2 during bacterial growth. ISME J. 2020;14:2649–58.
73. Jones CM, Spor A, Brennan FP, Breuil MC, Bru D, Lemanceau P, et al. Recently identified microbial guild mediates soil N2O sink capacity. Nat Clim Chang. 2014;4:801–5.
74. Reay DS, Davidson EA, Smith KA, Smith P, Melillo JM, Dentener F, et al. Global agriculture and nitrous oxide emissions. Nat Clim Chang 2012 26. 2012;2:410–6.
75. Hofmann DJ, Butler JH, Dlugokencky EJ, Elkins JW, Masarie K, Montzka SA, et al. The role of carbon dioxide in climate forcing from 1979 to 2004: Introduction of the Annual Greenhouse Gas Index. In: Tellus, Series B: Chemical and Physical Meteorology. 2006.
76. Arévalo-Martínez DL, Kock A, Löscher CR, Schmitz RA, Bange HW. Massive nitrous oxide emissions from the tropical South Pacific Ocean. Nat Geosci. 2015;8:530–3.
77. Barnes J, Upstill-Goddard RC. N2O seasonal distributions and air-sea exchange in UK estuaries: Implications for the tropospheric N2O source from European coastal waters. J Geophys Res Biogeosciences. 2011;116:G01006.
78. Akashi H, Gojobori T. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. Proc Natl Acad Sci U S A. 2002;99:3695–700.
79. Tripp HJ, Schwalbach MS, Meyer MM, Kitner JB, Breaker RR, Giovannoni SJ. Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. Environ Microbiol. 2009;11:230–8.
80. Jurgenson CT, Begley TP, Ealick SE. The structural and biochemical foundations of thiamin biosynthesis. Annu Rev Biochem. 2009;78:569–603.
81. Paerl RW, Bertrand EM, Allen AE, Palenik B, Azam F. Vitamin B1 ecophysiology of marine picoeukaryotic algae: Strain-specific differences and a new role for bacteria in vitamin cycling. Limnol Oceanogr. 2015;60:215–28.
82. Paerl RW, Sundh J, Tan D, Svenningsen SL, Hylander S, Pinhassi J, et al. Prevalent reliance of bacterioplankton on exogenous vitamin B1 and precursor availability. Proc Natl Acad Sci U S A. 2018;115:E10451.
83. Tang YZ, Koch F, Gobler CJ. Most harmful algal bloom species are vitamin B1 and B 12 auxotrophs. Proc Natl Acad Sci U S A. 2010;107:20756–61.
84. Sañudo-Wilhelmy SA, Gómez-Consarnau L, Suffridge C, Webb EA. The role of B vitamins in marine biogeochemistry. Ann Rev Mar Sci. 2014;6:339–67.
85. Averianova LA, Balabanova LA, Son OM, Podvolotskaya AB, Tekutyeva LA. Production of Vitamin B2 (Riboflavin) by microorganisms: An overview. Front Bioeng Biotechnol. 2020;8:570828.
86. Finkenwirth F, Kirsch F, Eitinger T. Solitary bio Y proteins mediate biotin transport into recombinant Escherichia coli. J Bacteriol. 2013;195:4105–11.
87. Van Bibber M, Bradbeer C, Clark N, Roth JR. A new class of cobalamin transport mutants (btuF) provides genetic evidence for a periplasmic binding protein in Salmonella typhimurium. J Bacteriol. 1999;181:5539–41.

88. Gozari M, Alborz M, El-Seedi HR, Jassbi AR. Chemistry, biosynthesis and biological activity of terpenoids and meroterpenoids in bacteria and fungi isolated from different marine habitats. Eur J Med Chem. 2021;210:112957.
89. Mari Ohtsuka SI, Irinoda K, Kukita K, Nagaoka K, Nakashima T. Azinomycins A And B, New Antitumor Antibiotics III. Antitumor Activity. J Antibiot (Tokyo). 1987;40:60–5.
90. Zhao Q, He Q, Ding W, Tang M, Kang Q, Yu Y, et al. Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. Chem Biol. 2008;15:693–705.
91. Chevrette MG, Carlson CM, Ortega HE, Thomas C, Ananiev GE, Barns KJ, et al. The antimicrobial potential of *Streptomyces* from insect microbiomes. Nat Commun. 2019;10:516.
92. Hasebe F, Matsuda K, Shiraishi T, Futamura Y, Nakano T, Tomita T, et al. Amino-group carrier-protein-mediated secondary metabolite biosynthesis in *Streptomyces*. Nat Chem Biol. 2016;12:967–72.
93. Covington BC, Spraggins JM, Ynigez-Gutierrez AE, Hylton ZB, Bachmann BO. Response of secondary metabolism of hypogean actinobacterial genera to chemical and biological stimuli. Appl Environ Microbiol. 2018;84:e01125-18.
94. Robinson SL, Christenson JK, Wackett LP. Biosynthesis and chemical diversity of β-lactone natural products. Nat Prod Rep. 2019;36:458–75.

Figures

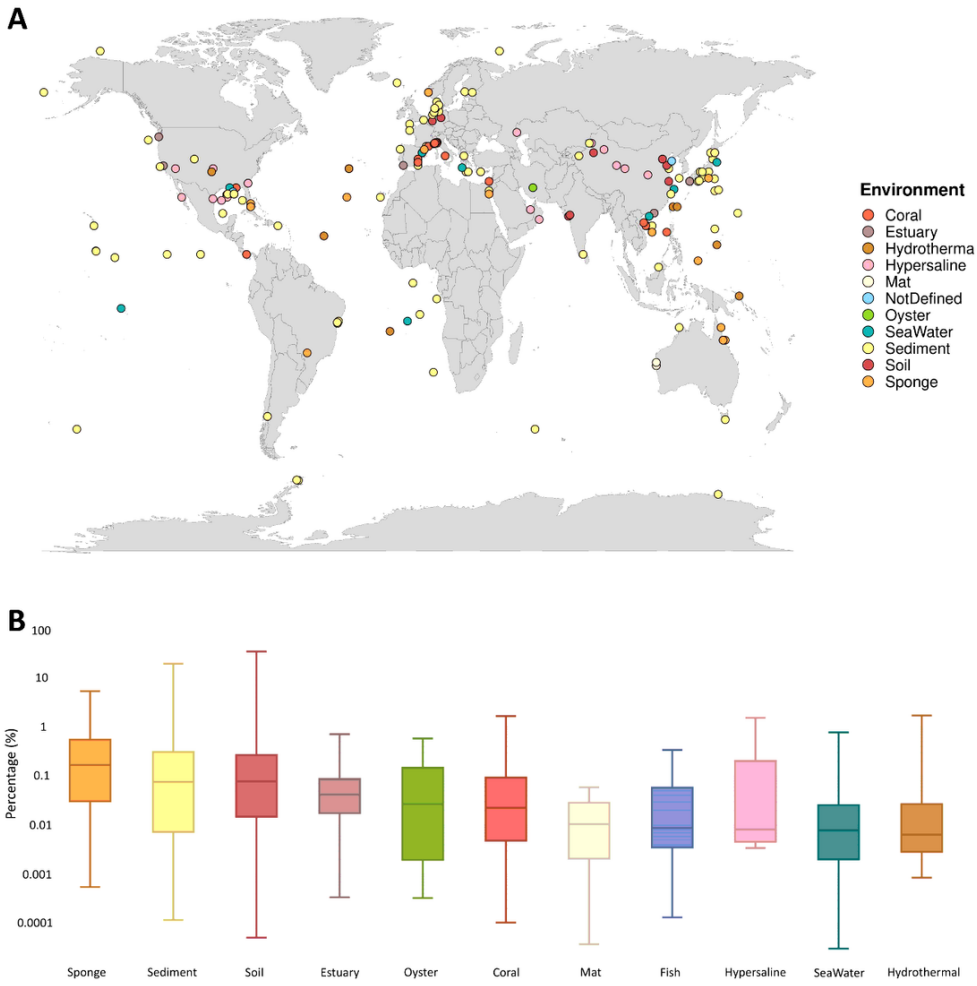

Figure 1

Ecological distribution and abundance of PAUC43f based on 16S rRNA gene sequences. A) Worldwide distribution and environments where PAUC43f have been detected. Colors correspond to the sampled environment as shown in B. B) Boxplot, in logarithmic scale, of PAU43f relative abundances in samples from different environments.

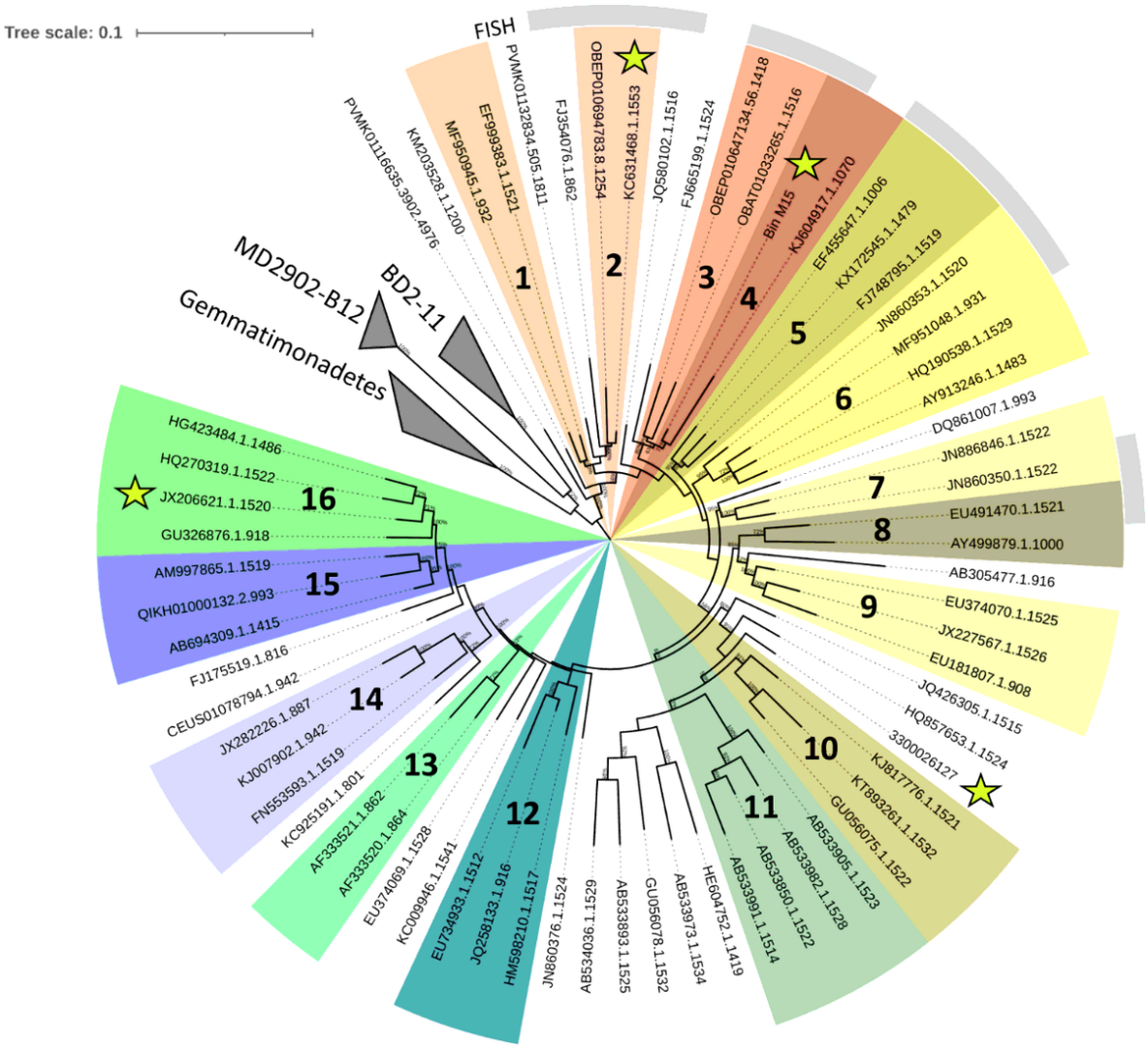

Figure 2

Maximum-likelihood tree based on 16S rRNA gene sequences longer than 800 bp (n = 66). Monophyletic clusters in both NJ and maximum-likelihood tree with identities above 94.5%, the threshold for delineating genera, are displayed with different colors and numbers. The 16S rRNA gene sequences from MAGs are marked with stars. The external grey circle indicates the sequences targeted by the FISH probe.

A

Environment \ Genus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Coral (80/2956)	7.5	3.8	6.3	2.5	2.5	3.8	1.3	10.0	15.0	1.3	1.3	1.3	7.5	27.5	27.5	40.0
Estuary (286/413)	82.2	19.6	82.2	22.7	16.8	18.2	0.7	40.2	2.4	1.0	0.0	18.2	0.0	43.4	3.8	16.4
Fish (26/1779)	11.5	0.0	26.9	69.2	0.0	0.0	0.0	7.7	11.5	0.0	0.0	0.0	0.0	3.8	0.0	11.5
Hydrothermal (13/285)	0.0	0.0	0.0	0.0	0.0	38.5	30.8	0.0	84.6	0.0	0.0	15.4	0.0	23.1	38.5	0.0
Hypersaline (4/57)	25.0	0.0	0.0	25.0	0.0	75.0	0.0	0.0	0.0	75.0	25.0	0.0	0.0	0.0	0.0	0.0
Mat (25/720)	0.0	60.0	60.0	28.0	4.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Oyster (37/319)	48.6	86.5	64.9	8.1	0.0	2.7	0.0	0.0	2.7	16.2	40.5	0.0	0.0	2.7	8.1	8.1
Sea water (289/17906)	35.6	28.0	29.1	16.3	27.0	6.2	3.5	5.9	13.1	2.8	1.7	3.5	6.2	16.6	20.8	8.0
Sediment (1002/4032)	23.3	32.7	40.3	13.3	16.1	58.4	43.3	14.3	53.6	1.5	1.3	33.7	0.9	36.2	32.5	5.7
Soil (883/32795)	43.4	33.1	16.4	39.0	21.3	62.4	10.3	7.9	12.2	17.8	5.7	4.9	0.8	7.8	3.2	3.4
Sponge (175/655)	6.9	1.1	6.3	5.1	1.1	1.1	0.0	3.4	5.7	0.0	0.0	1.1	25.1	8.6	5.1	90.3

B

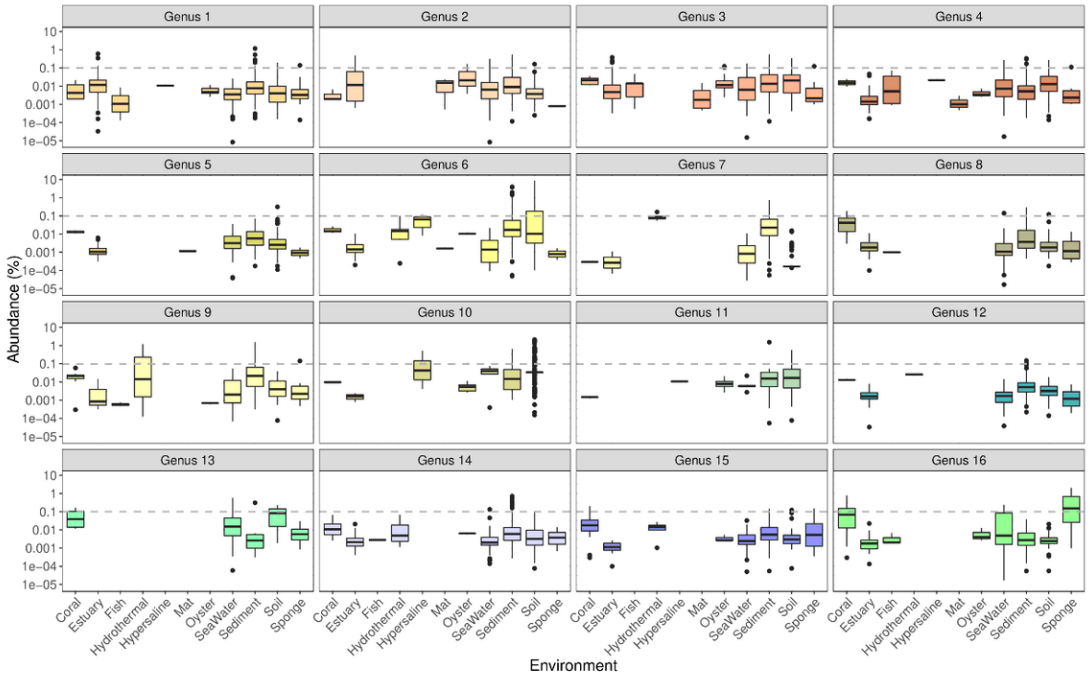

Figure 3

Ecological distribution of PAUC43f genera. A) Frequency of detection of each genus in each environment with respect to the total number of samples in which PAUC43f sequences could be detected. Highest values are displayed in green and lowest in white. Next to the environment name, in parentheses, are the numbers of samples with PAUC43f genera out of the total number of samples analysed for each environment. B) Relative abundance (%) of each genus in each environment. The horizontal dashed line indicates a relative abundance of 0.1%, a threshold for abundant and rare biospheres.

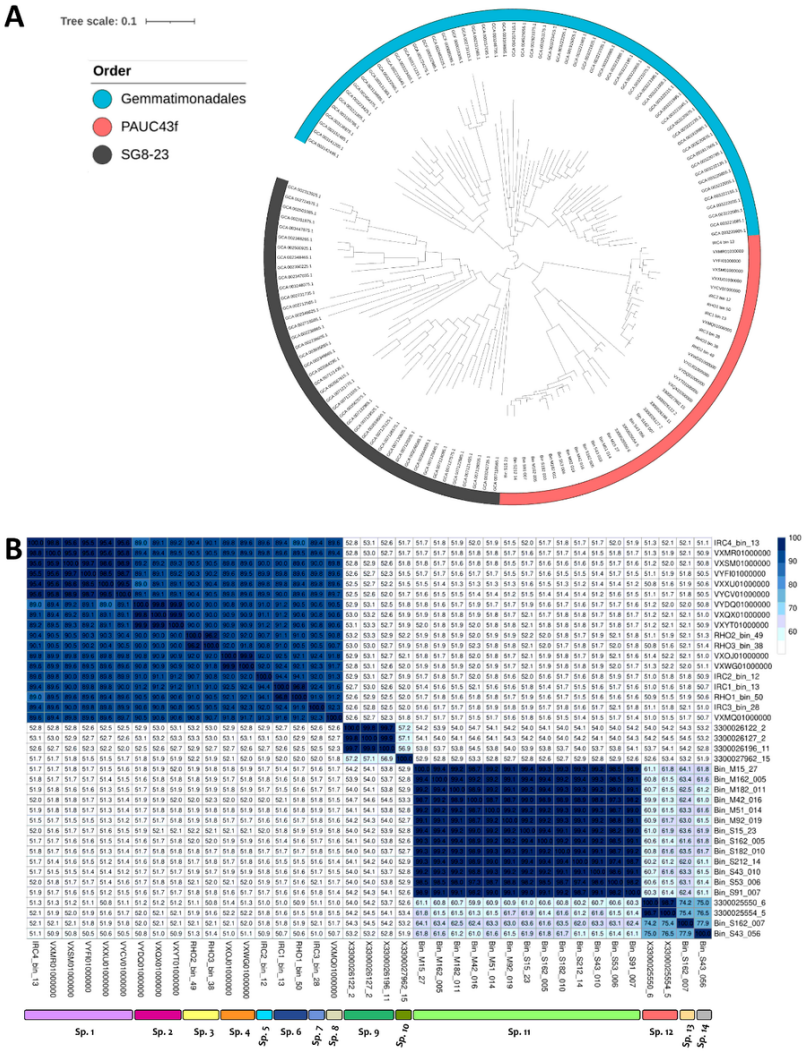

Figure 4

Taxonomic classification of PAUC43f MAGs. A) Phylogenomic tree with all *Gemmatimonadota* MAGs present in the GTDB database and PAUC43f MAGs. The external circle indicates the taxonomic classification at the order level. B) Heatmap based on average amino acid identity (AAI) values for the PAUC43f MAGs. Values above 95% of AAI, the threshold for species delimitation, are highlighted in dark blue.

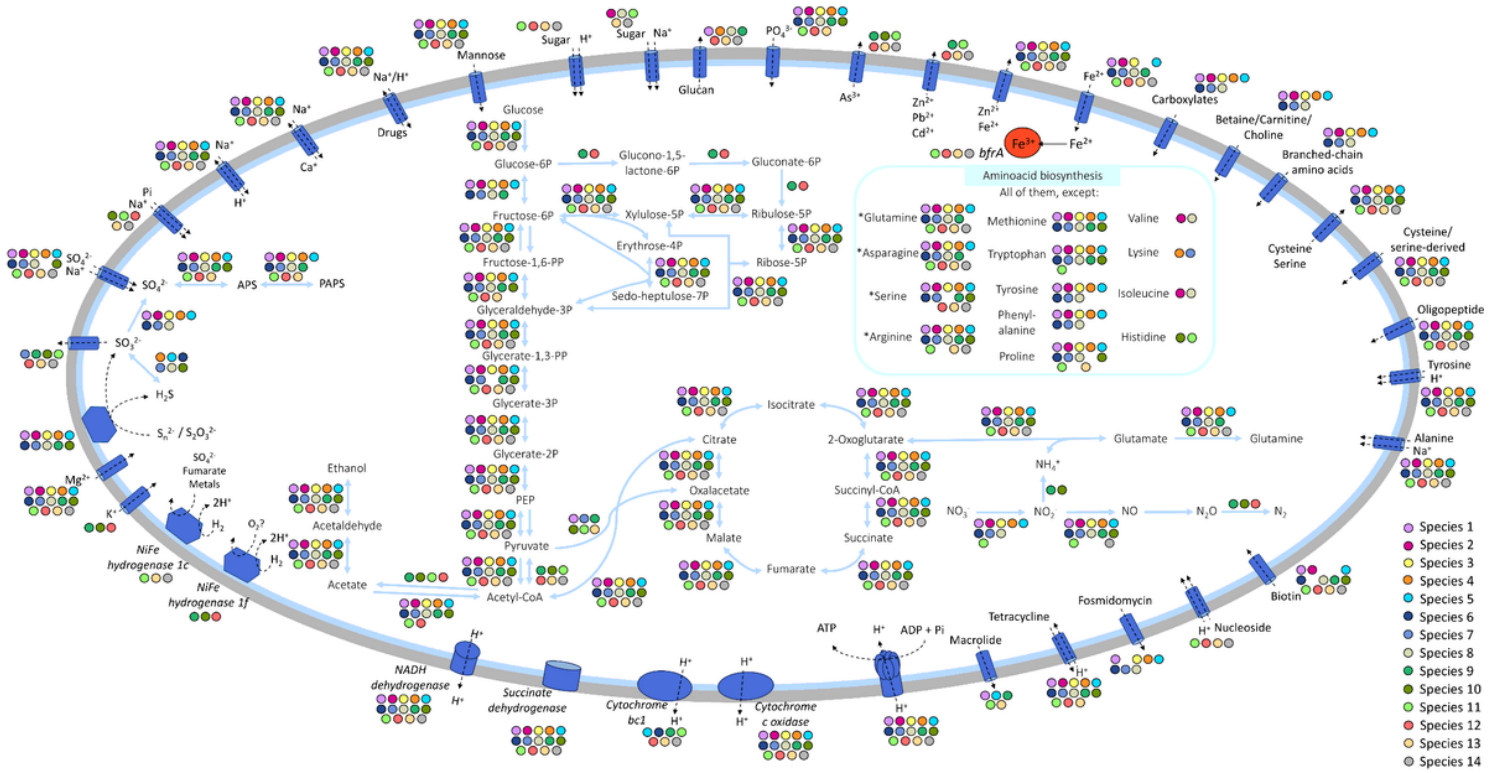

Figure 5

Predicted metabolic capabilities for PAUC43f species. Each species is represented by a colored dot (see legend) named in the same order as in Figure 4. The annotation of MAGs used to reconstruct the metabolism can be found in Supplementary Table 2.

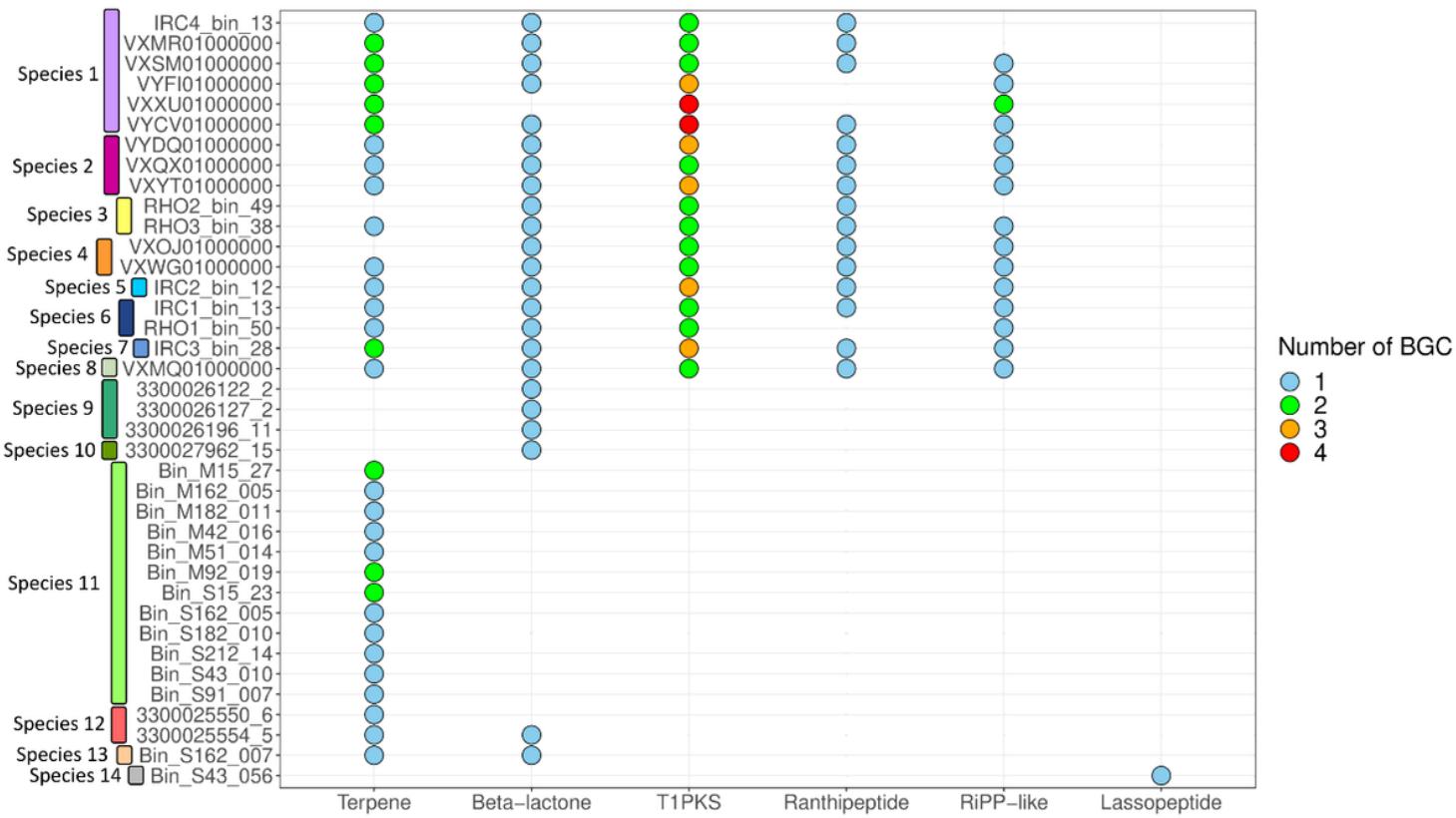

Figure 6

Secondary metabolite biosynthetic gene clusters (BGC) predicted by antiSMASH for each MAG. Colored dots indicate the number of each BGC per MAG (1: blue; 2: yellow; 3: orange; 4: red).

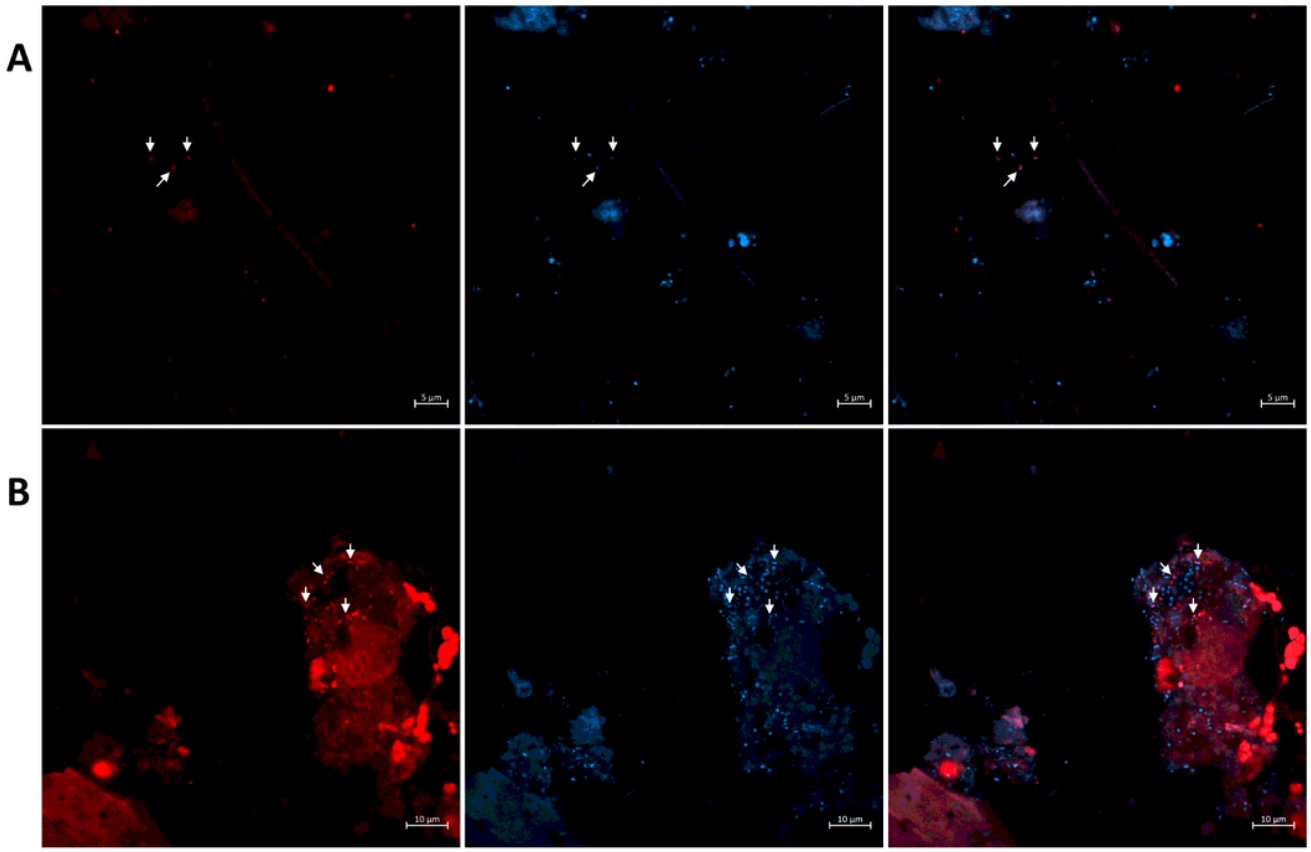

Figure 7

FISH of Mar Menor sediment samples with a PAUC43f specific probe. A and B show two different microscopic fields observed with two different color channels, red for probe and blue for DAPI. From left to right: PAUC43f_826, DAPI and merged channels. Arrows indicate cells displaying signals in both FISH and DAPI channels.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tiff
- FigureS2.tiff
- FigureS3.tiff
- FigureS4.tiff
- Suppl.Table1.xlsx
- Suppl.Table2.xlsx
- Suppl.Table3.xlsx
- Suppl.Table4.xlsx