Vol. 74: 107–119, 2015
doi: 10.3354/ame01734

AQUATIC MICROBIAL ECOLOGY
Aquat Microb Ecol

Published online February 12

Diversity of quorum sensing autoinducer synthases in the Global Ocean Sampling metagenomic database

Margot Doberva[1], Sophie Sanchez-Ferandin[2], Eve Toulza[3,4], Philippe Lebaron[1], Raphaël Lami[1,*]

[1]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Laboratoire de Biodiversité et Biotechnologie Marines (LBBM), Observatoire Océanologique, 66650 Banyuls/Mer, France

[2]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, 66650, Banyuls/Mer, France

[3]UPVD Université de Perpignan, UMR 5244, IHPE, 66860 Perpignan cedex 9, France

[4]CNRS, UMR 5244, 2EI, 66860 Perpignan cedex 9, France

ABSTRACT: Quorum sensing (QS) is a cell-to-cell signalling pathway that allows bacteria to synchronize their genetic expression. It is mediated by autoinducers (AI), including (1) acyl-homoserine lactones (AHLs or AI-1), produced by *Proteobacteria* using AinS, LuxI and HdtS synthase families and (2) furanosyl-diester-borate (FDB or AI-2), produced by a large range of phylogenetically diverse bacteria and synthetized by the LuxS family. Few data have been collected about the presence and importance of QS in marine waters using culture independent methods. In this study, we examined the presence and the diversity of AI-1 and AI-2 synthases in the Global Ocean Sampling (GOS), a large metagenomic database, covering 68 stations across 3 oceans. We built 4 reference protein databases with maximal phylogenetic coverage containing all known AI synthase sequences to retrieve AI synthases sequences from the GOS metagenomes. We retrieved 29 environmental sequences affiliated to LuxI (synthesizing AI-1), 653 related to HdtS (AI-1), 31 related to LuxS (AI-2) and only one for AinS (AI-1). AI synthases sequences were found in the 3 oceans covered by the GOS cruise and spanned a large phylogenetic diversity. These data revealed a large number of new marine AI sequences, suggesting that QS based on AI-1 diffusion is a widespread mechanism in the marine environment.

KEY WORDS: Quorum sensing · Acyl-homoserine lactone synthases · Global Ocean Sampling · Environmental metagenomes

INTRODUCTION

Quorum sensing (QS) is a cell-to-cell bacterial communication mechanism, allowing bacteria to sense their population density (Nealson 1977) and coordinate their gene expression levels (Bassler 1999, Fuqua & Greenberg 2002) and physiological activities (Miller & Bassler 2001). To perform QS-based communication, bacteria produce, secrete and accumulate small hormone-like molecules called autoinducers (AI) in the nearby environment (Fuqua et al. 1994). As cell density increases, the concentration of these AI increases. When a given threshold concentration is reached, targeted bacterial partners can respond to these compounds and modify their gene expression, and thus their behaviour and phenotype (Withers et al. 2001, Bassler 2002). It has been shown that QS systems regulate and synchronize activities like biofilm production (Parsek & Greenberg 2005, Dickschat 2010), nodulation (Cha et al. 1998, Loh et al. 2002), bioluminescence (Waters & Bassler 2005), virulence factor production (Smith & Iglewski 2003) and many others (Diggle et al. 2007). The coordination of bacterial community activities conveys an ecological advantage to the population (Case et al. 2008).

*Corresponding author: raphael.lami@obs-banyuls.fr

QS mechanisms are well described in medical or agronomic model bacterial strains (Cha et al. 1998, Loh et al. 2002) known for their pathogenicity (de Kievit & Iglewski 2000) or bioengineering potential (March & Bentley 2004, Brenner et al. 2008). By contrast, little attention has been paid to QS in the functioning of natural microbial communities, including in marine waters (Decho et al. 2010). This is because bacterial densities in seawater (10^5 to 10^6 cells ml^{-1}) are typically below the known thresholds that enable QS (Mohamed et al. 2008). However, at a micrometer scale many ecological niches (such as organic matter particles or microalgal blooms) harbour bacterial concentrations compatible with QS (Gram et al. 2002, Mohamed et al. 2008). In support of this hypothesis, the potential for QS in planktonic marine bacteria has been reported in many cultivated strains. These results were acquired either by directly detecting communication compounds or inferred by sequencing genes involved in QS in *Proteobacteria* (Gram et al. 2002, Wagner-Döbler et al. 2005).

The AI synthases are the key enzymes involved in AI production. It is now well established that the acylhomoserine lactone synthases (AHL or AI-1 synthases) are encoded by 3 groups of genes: *luxI*-like (Engebrecht & Silverman 1984), *ainS*-like (Gilson et al. 1995) and *hdtS*-like (Laue et al. 2000, Burton et al. 2005, Rivas et al. 2007). By contrast, *luxS*-like genes encode the (2*S*,4*S*)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran-borate synthase or *S*-THMF-borate synthase or furanosyl diester borate synthase or (2*R*,4*S*)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran synthase or *R*-THMF synthase, also known as the AI-2 synthases (Xavier & Bassler 2003, Miller et al. 2004). The AI-1 synthases have been identified in cultivated marine strains from the groups *Roseobacter* and *Vibrio* (Gram et al. 2002, Schaefer et al. 2002, Wagner-Döbler et al. 2005) and the AI-2 (Chen et al. 2002) have been observed in the *Vibrio* genus (Bassler et al. 1997, Bassler 1999). By contrast, very little attention has been paid to the diversity of AI synthases in whole marine microbial communities. As only 1% of marine bacterial communities in the water column are readily cultivable by standard methods (Amann et al. 1995), an overview of potential QS mechanisms in the marine environment remains largely incomplete.

The Global Ocean Sampling (GOS) database is a large collection of marine metagenomes from 68 sampling sites. A total of 6.3 billion bp were Sanger sequenced from microplankton samples (Rusch et al. 2007) from which the metaproteome (all coding sequences) was also predicted (Yooseph et al. 2007). The GOS datasets mainly comprise sequences from the free-living fraction (FL) of bacterioplankton (0.1 to 0.8 µm pore size filter). Additionally, for 8 stations the particle-attached fraction (PA) (0.8 to 3 µm) has been sequenced (GS000, GS001, GS048, GS108, GS110, GS112, GS117 and GS122). The GOS offers a large snapshot of the diversity and functional potential of marine microbial communities and has been successfully explored, for example, for the potential to utilize selenium (Zhang & Gladyshev 2008) or to metabolize various chemical forms of iron (Toulza et al. 2012).

In this study, we investigated the diversity patterns of QS AI synthases in the bacterial communities represented in the GOS metagenomes. We report the presence, the large diversity of both AI-1 and AI-2 synthase sequences and their distribution throughout the 3 oceans covered by the GOS expedition, and we highlight new sequences of AI synthases.

MATERIALS AND METHODS

Construction of reference databases for AI synthase proteins

We built a total of 4 databases, one for each family of AI synthases: AinS-like, LuxI-like, HdtS-like and LuxS-like. Following the approach described in Toulza et al. (2012), we first selected protein sequences that have been functionally characterized in previous publications. To improve taxonomic coverage, we added to our reference databases additional sequences annotated as AI synthase from the GenBank non-redundant (NR) database (www.ncbi.nlm.nih.gov/). Each database was aligned using Clustal W (Larkin et al. 2007) in Jalview (Waterhouse et al. 2009) to discard redundant and partial sequences. Our final AI databases comprised 569 LuxI, 30 AinS, 214 HdtS and 297 LuxS annotated protein sequences.

Screening for AI synthases in the GOS metagenome

The GOS expedition and its subsequent metagenome analysis have been described previously (Rusch et al. 2007, Yooseph et al. 2007). Briefly, 200 l of seawater was collected at 68 different sampling stations and filtered onto 4 different successive filters with a porosity of 20, 3, 0.8 and 0.1 µm. In this work, we discarded 10 stations where no AI sequences

were found (GS006, GS038-040, GS042-046, GS050). In our study, we also focused on both FL (0.1 to 0.8 µm) and PA (0.8 to 3 µm) fractions, which constitute the most complete sequence datasets in terms of sequencing effort. The GOS database contains sequences collected from various habitats within the Atlantic, Pacific and Indian Oceans. A total of 11 different habitat types were sampled, including open ocean (23 stations), coastal (22 stations), coral reefs (4 stations), estuary (2 stations) and 7 others (Rusch et al. 2007). Environmental variables were also retrieved from Venter et al. (2004) and Rusch et al. (2007).

The detection of AI-1 and AI-2 synthases from the GOS database (proteins predicted from reads) was conducted by sequence homology searches using the BLAST (Basic Local Alignment Search Tool) algorithm (Altschul et al. 1990) and our AI synthase databases as the query sequences. All our AI protein BLAST searches were performed using the CAMERA portal (http://camera.crbs.ucsd.edu/projects/details.php?id=CAM_PROJ_GOS) (Sun et al. 2011) using blastall default parameters (Seshadri et al. 2007) (including FL and PA fractions). A Reciprocal Best BLAST Hit (RBH) analysis (Moreno-Hagelsieb & Latimer 2008) was then performed on the sequences retrieved from the GOS dataset. Briefly, putative environmental AI synthases were searched against NR database and those GOS sequences with a RBH to the same annotated AI synthases were assigned to the gene and designated as most related to that taxon.

We also compared the distribution of AI environmental protein sequences in both the FL and PA fractions from the 8 GOS sampling stations for which these data are available. We pooled all the retrieved AI protein sequences from PA and FL fractions from all 8 sites, and clustered the environmental and annotated protein sequences using CD-HIT (Li & Godzik 2006, Fu et al. 2012). The clustering parameters were 40% sequence identity and a word size equal to 2. The differences in the number of environmental sequences affiliated to HdtS between the FL and PA fractions were tested for significance using a Wilcoxon test with R (www.r-project.org).

Normalization of AI counts in GOS metagenomes to single copy genes

The number of sequences matching to AI synthases in the GOS was normalized to the number of RpoB protein sequences present in the sample. That gene is present in only a single copy in bacterial genomes and is not prone to horizontal transfer and thus is also a good taxonomic marker (Case et al. 2007). Following the same procedure as described for the AI sequences, we built an RpoB protein reference databases with a large taxonomic coverage containing 1000 sequences. We finally reported the number of each family of AI environmental sequences related to our reference sequences relative to the number of RpoB protein sequences in the GOS dataset. This allowed normalization to the number of bacterial genomes in the dataset.

Phylogenetic analysis

Phylogenetic analyses were conducted for each targeted family of AI synthase. All environmental and annotated sequences were aligned using Clustal W (Larkin et al. 2007) in MEGA v.5.1 software (Tamura et al. 2011), and alignments adjusted manually. Phylogenetic trees were constructed using both Neighbour-Joining (NJ) and Maximum Likelihood (ML) methods for each kind of AI synthase protein sequence alignments. NJ analyses were conducted with gamma correction and 1000 bootstrap replicates. Protein model selection was performed from each protein dataset, and phylogenetic analyses were then conducted with the selected models (WAG model for LuxI, HdtS and LuxS) using gamma correction and 100 bootstrap replicates. As we obtained similar topologies from the 2 methods, only those obtained by NJ are presented here, but when useful, percentage of bootstrap supports obtained by ML searches are also mentioned.

RESULTS

Features of AI synthase protein reference databases

All our reference databases were built to have as broad a taxonomic coverage as possible, consisting of protein sequences assignable to AI synthases from the NR database at the time this study was conducted. The LuxI, LuxS and HdtS-like protein family databases had a large taxonomic coverage, which included the alpha-, beta-, gamma- and delta-lineages of *Proteobacteria*. The LuxS protein reference database also contained annotated sequences from *Spirochaetes*, *Bacteroidales* and the Gram-positive *Firmicutes* and *Actinobacteria*. The AinS family,

Table 1. Protein families and features of environmental sequences

Features	Protein families				
	AinS	LuxI	HdtS	LuxS	RpoB
Number of annotated sequences	30	569	214	297	1000
Conservation (%)	34	27	27	78	62
Average number of amino acids	380	200	250	170	1200
Taxonomic affiliation	*Vibrio-naceae*	*Proteo-bacteria*	*Alpha-*, *Beta-* and *Gamma-proteobacteria*	All bacterial phyla	All bacterial phyla
Number of Global Ocean Sampling (GOS) stations where sequences were found	1/58	9/58	57/58	3/58	58/58
Total number of environmental sequences retrieved in GOS	1	29	653	31	4293
Ratio: number of environmental sequences to number of RpoB copies (%)	0.02	0.7	14.8	0.7	100

limited to the *Vibrionaceae*, contained not only AinS but also LuxM and VanM synthases (AinS-like synthases). A total of 634 annotated protein sequences from the National Center for Biotechnology Information (NCBI) were affiliated to LuxI with a mean identity of 27% (number of conserved amino acids divided by total amino acids), 214 sequences to HdtS with a mean identity of 27%, 297 sequences to LuxS with a mean identity of 78% and 30 sequences to the AinS with a mean identity of 34% (Table 1). The mean length of these sequences was about 380 amino acids for AinS, 200 for LuxI, 250 for HdtS and 170 for LuxS.

Abundance and diversity of LuxI synthases in the GOS dataset

We retrieved 29 sequences related to LuxI AHL synthases in the GOS dataset. Those sequences were distributed over 10 stations of the GOS cruise (Table 1, Fig. 1) and appeared related to a large taxonomic diversity of AI-1 sequences in NCBI. These GOS sequences were aligned with annotated sequences and phylogenetic reconstruction was performed (Fig. 2). All affiliated environmental sequences arose from *Alphaproteobacteria* (Fig. 2). A total of 19 environmental sequences (65%) matched with sequences belonging to *Rhodobacterales* (*Rhodobacteraceae* family), 2 (7%) with sequences belonging to *Sphingomonadales* (*Sphingomonadaceae* family) and 3 (14%) with sequences belonging to *Rhizobiales* (*Bradyrhizobiaceae* and *Beijerinckiaceae* families) (Fig. 2). These affiliations were supported by strong NJ bootstrap percentages (>99 for *Rhodobacterales*; 100 for *Sphingomonadales*; 97 and 100 for *Rhizobiales*; Fig. 2) and also in some cases by strong ML bootstrap percentages. Although well identified as AI-1 synthases by RBH, a large fraction of environmental sequences (14%) could not be unambiguously related to any annotated sequences.

LuxI sequences arising from PA fractions represented 10% of total retrieved LuxI sequences in the GOS dataset. These sequences were present at Stns GS048 and GS110. All these sequences appear enclosed in a larger phylogenetic group (defined with bootstraps values of 53/86 for NJ/ML, respectively; Fig. 2) comprising members of the *Sphingomonodaceae* family.

Abundance and diversity of HdtS synthases in the GOS dataset

A total of 653 environmental sequences related to HdtS were retrieved from the GOS metagenomes (Table 1). Those sequences were found in all oceans crossed by the GOS expedition and in a large number of stations (57 different stations from a total of 58 sampled by the expedition) (Fig. 1). Collectively, all of these sequences were related to *Alpha-*, *Beta-* and *Gammaproteobacteria*. More precisely, a large fraction (19% or 127 sequences) of these sequences were related to the SAR11 group (*Alphaproteobacteria*, with bootstrap supports of 100 and 96 for NJ and ML, respectively), including its cultivated representative member *Pelagibacter ubique* (indicated by 'SAR11' sequences;

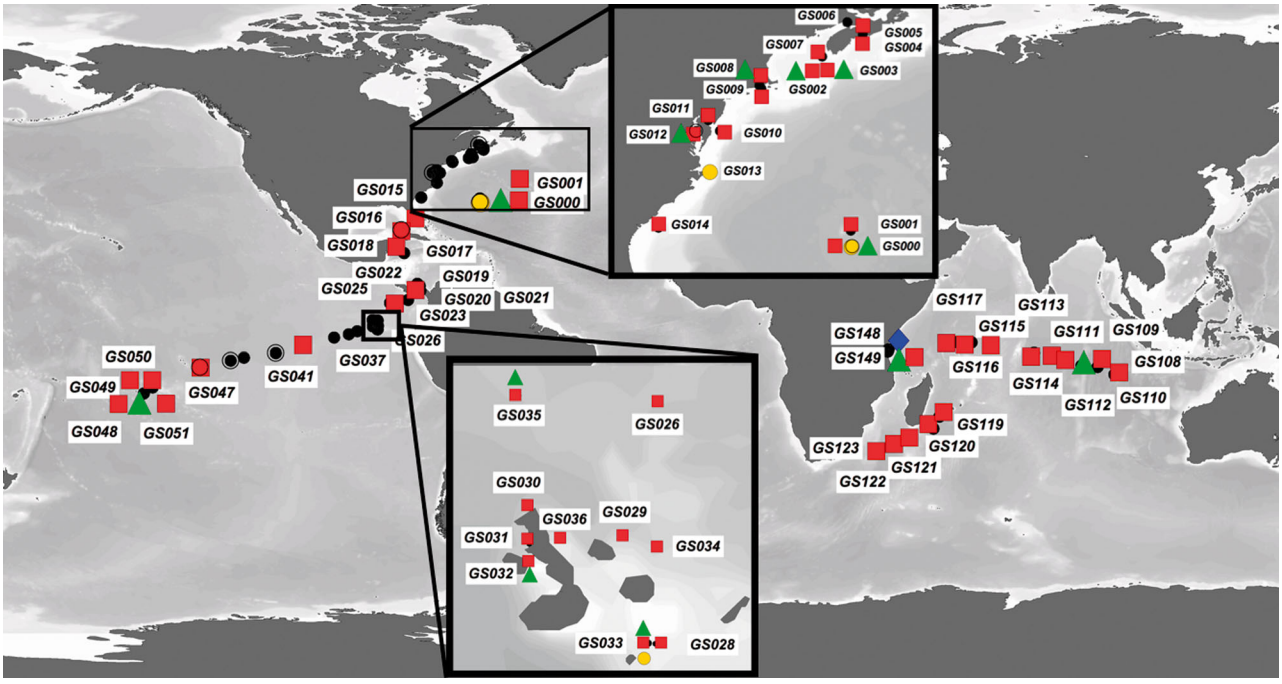

Fig. 1. Geographic distribution of AI-1 (LuxI, HdtS and AinS-like sequences) and AI-2 (LuxS sequences) retrieved from the Global Ocean Sampling (GOS) metagenomes. Black dots: GOS sampling stations where no AI sequences were found; blue diamonds: AinS affiliated sequences; green triangles: LuxI affiliated sequences; red squares: HdtS affiliated sequences; yellow dots: LuxS affiliated environmental sequences. To display the geographic distribution of the environmental sequences related to the 4 different protein families (AinS, LuxI, HdtS and LuxS) a geographic map was built using Ocean Data View software v.4 (ODV Schlitzer, R, Ocean Data View, http://odv.awi.de, 2013)

Fig. 3). A lower fraction (3% or 21 sequences) of these sequences was related to the SAR116 group (*Alphaproteobacteria*, with bootstrap supports of 95 and 67 for NJ and ML, respectively; Fig. 3). Interestingly, one sequence was strongly associated to a sequence from *Oceanicaulis* (*Rhodo-bacterales*), which belongs to the *Hyphomonadaceae* family (bootstrap supports of 100 and 83 for NJ and ML, respectively). Among *Gammaproteobacteria*, a large fraction of sequences were related to the SAR86 group (23% or 153 sequences) with high bootstrap supports both in NJ and ML methods (100 and 62, respectively; Fig. 3). One sequence was related to *Pseudomonadales*, and a large fraction (21% or 140 sequences) appeared more closely related to other orders within *Gammaproteobacteria* (*Oceanospirillales*, *Chromatiales*, *Acidithiobacillales* and *Triotrichales*). Lastly, a significant number of sequences appeared closely related to the *Betaproteobacteria* (represented by the *Burkholderiales* order, 19% or 126 sequences) and unclassified *Gammaproteobacteria* (5% or 32 sequences) with relatively strong bootstrap supports (63 and 57 for NJ and ML, respectively; Fig. 3).

A few environmental sequences (19%) were recovered from the 8 sampling stations covering both the FL and PA fractions. Overall, a total of 63 sequences were recovered from the PA fraction. Interestingly, the 2 sequences GS051 FL and GS117 PA were more closely related to (NJ bootstrap support of 98), and found within, the SAR11 *Alphaproteobacteria* group. Six sequences from the PA fraction of the GS110 sampling site were recovered together and closely related to the *Gammaproteobacteria*, *Francisella tularensis* (NJ bootstrap support of 97; Fig. 3). To test whether there was any difference in the taxonomic composition between the FL and the PA fractions, we clustered all HdtS sequences using a threshold of 40% sequence identity, and found a total of 14 clusters (Table 2). As an example, cluster 6 contains environmental sequences related to SAR86 clade in the 2 fractions of the GOS Stns GS000, GS110 and GS112 (Table 2). In the environmental sequences, of which 11% were from the FL fraction and 15% from the PA fraction, we did not find any significant difference between the cluster affiliations of environmental sequences distributed in the 2 metagenome fractions (Wilcoxon test, $p = 0.83$).

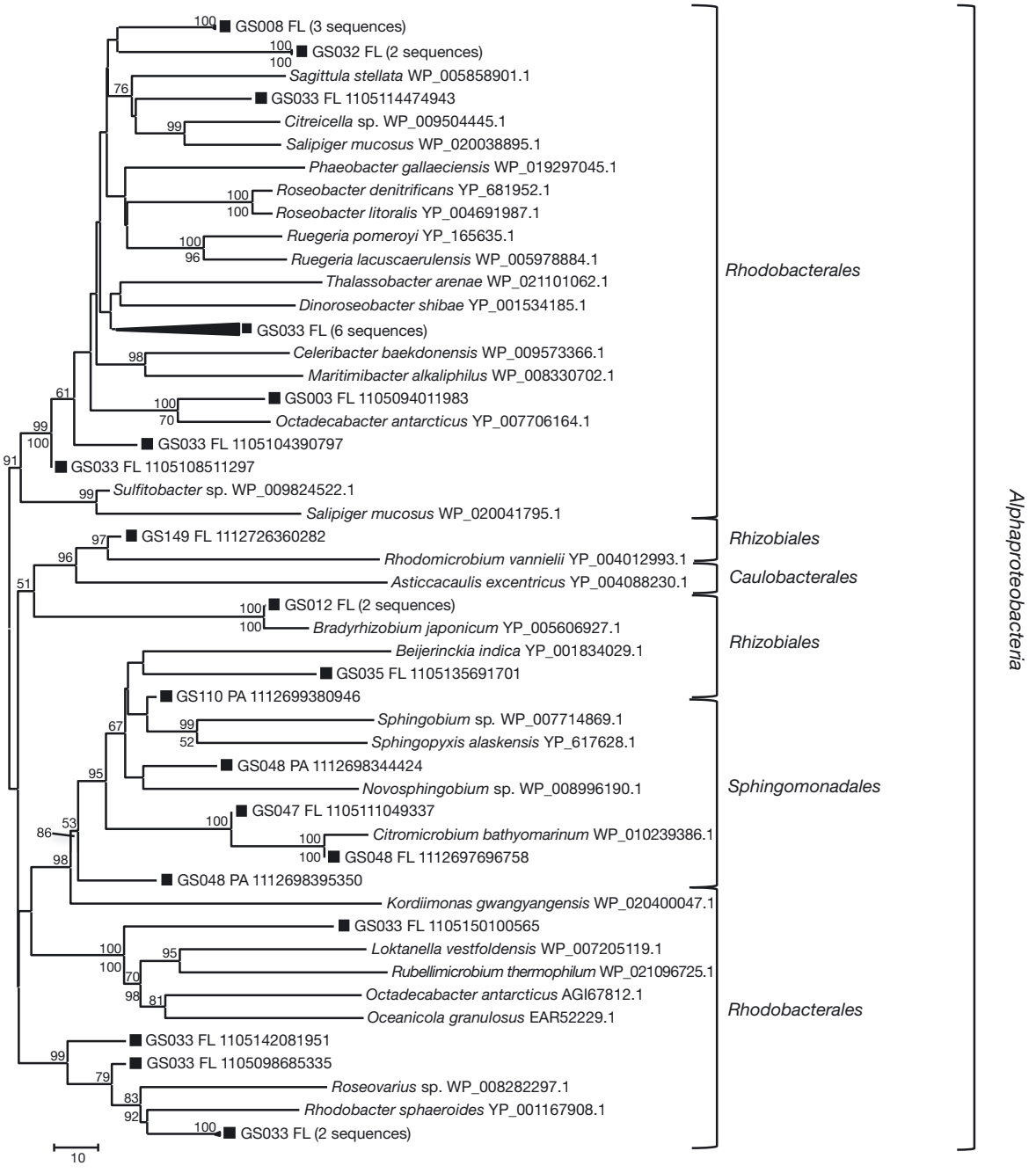

Fig. 2. Phylogenetic tree of annotated and environmental sequences predicted from Global Ocean Sampling (GOS) metaproteome LuxI protein sequences produced using the Neighbour-Joining method with 1000 bootstrap replicates. Only bootstrap support > 50 are shown. Values above branches are those obtained from NJ method; values under branches are those obtained from ML methods. Black squares: environmental sequences; FL: sequences retrieved from free-living fractions; PA: sequences retrieved from particle-attached fractions

Abundance and diversity of LuxS synthases in the GOS dataset

A total of 31 environmental sequences related to LuxS were detected in the GOS dataset at 3 sampling sites: GS000, GS013 and GS033 (Table 1, Fig. 1). All environmental sequences affiliated to LuxS were retrieved from the FL fractions. As previously noted, the LuxS family showed the widest taxonomic coverage compared to the other families. The environmental sequences found were related to *Gammaproteobacteria* (97% or 30 sequences, NJ and ML bootstraps support of 100 and 99%, respectively; Fig. 4) and to the Gram-positive bacteria from the

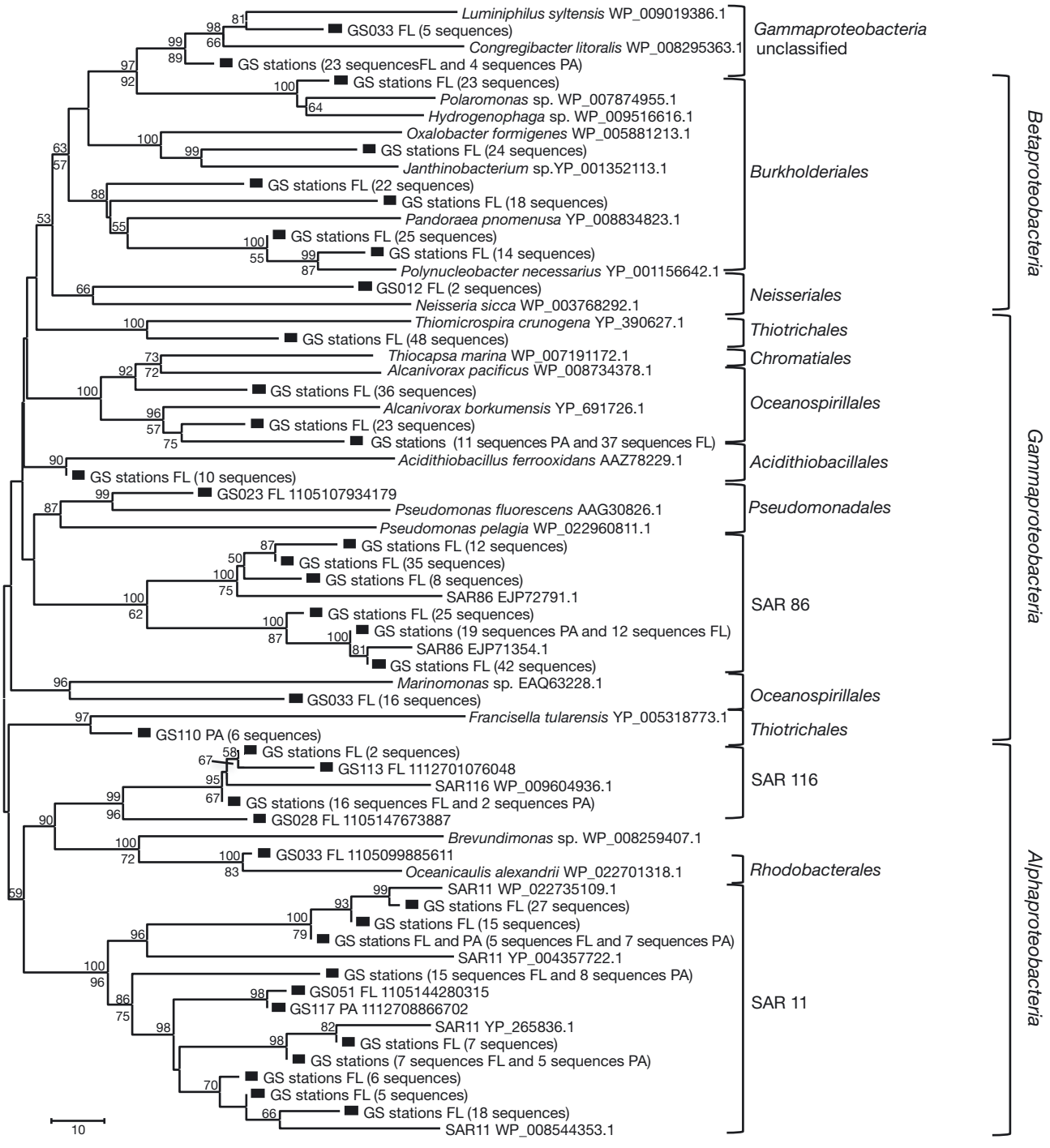

Fig. 3. Phylogenetic tree of annotated and environmental HdtS protein sequences produced using the Neighbour-Joining method with 1000 bootstrap replicates. Only bootstrap support > 50 are shown. Values above branches are those obtained from NJ method; values under branches are those obtained from ML methods. See Fig. 2 for definitions of symbols

Deinococcales (3% or 1 sequence, NJ and ML bootstraps support of 100 and 90%, respectively; Fig. 4). Among *Gammaproteobacteria*, 4 sequences (13%) group inside a clade found in both NJ and ML phylogenies, comprising sequences from *Alishewanella* and *Rheinheimera* species (ML bootstrap support of 92%; Fig. 4), but which were more closely related to *Reinheimera* (NJ and ML bootstraps support of 93 and 91%, respectively; Fig. 4). A large number of environmental sequences (65% or 20 sequences) seemed related to *Shewanella* species, although it was not well supported (except 3 GS000 FL sequences with *Shewanella frigidimarina*, ML bootstrap support of 60%).

Table 2. Comparison of HdtS phylogenetic cluster abundances (in % of total sequences) in the free-living (FL) and particle-attached (PA) fractions over 8 Global Ocean Sampling (GOS) stations. HdtS sequences were clustered with CD-HIT software at 40% identity. Statistical analysis on these data revealed no significant difference of HdtS diversity between the 2 fractions (see 'Materials and methods' and 'Results' for details)

Cluster no.	Phylogenetic affiliation of protein sequence clusters	Fraction (%) FL	PA
1	*Luminiphilus*/*Congregagibacter*	0	3.1
2	*Alcalinivorax*	7.1	8.7
3	*Oceanocaulis*/SAR116	3.1	1.6
4	*Roseobacter* clade	0.8	0
5	*Hahella*/*Methyloglobulus*	7.1	5.5
6	SAR11	8.7	9.5
7	SAR86	11.1	15
8		3.9	2.3
9		3.1	0.8
10		3.1	0.8
11	Non-affiliable environmental	0	1.6
12	sequences	0	0.8
13		0.8	0
14		0.8	0
Total		50	50

Diversity of AinS synthases in the GOS dataset

We retrieved only 1 sequence of AinS in the GOS dataset, located at Stn 148, in the Indian Ocean close to the Zanzibar coast in a fringing reef. The top high scoring BLAST pair was 'N-(3-hydroxybutanoyl)-L-homoserine lactone synthase luxM' (WP_002541906.1) (70% sequence coverage and 90% identity) from *Grimontia* sp., a *Vibrionaceae* family member.

Correlation with environmental variables

For the LuxI group, we found no correlation with the environmental data. For HdtS family there was a positive correlation with metagenome size, and negative correlation between chlorophyll (Spearman $\rho = -0.425$, $p = 0.002$), nitrate (Spearman $\rho = -0.553$, $p = 5.475 \times 10^{-5}$), silica concentration (Spearman $\rho = -0.573801$, $p = 2.481 \times 10^{-5}$), primary production (Spearman $\rho = -0.374$, $p = 0.01$) and the number of environmental sequences related to HdtS. A link with LuxS-related environmental sequences and salinity was also observed (Spearman $\rho = 0.341$, $p = 0.020$). We did not find any correlation with the other environmental variables (temperature, iron concentration, habitat type) (Table 3).

Table 3. Correlation tests between sequencing effort or environmental variables and AI synthase abundance. For each station, we first checked whether the number of environmental sequences was related to metagenome size with the Spearman coefficient. Then, we verified the existence of a significant relationship between environmental variables and the number of environmental sequences retrieved for each AI family, normalized by the number of potential bacterial genomes at each Global Ocean Sampling (GOS) station (assessed by RpoB, see 'Materials and methods' for details) using a Spearman correlation test (discrete quantitative variables, non-normal). For sequence–habitat correlations, a Kruskal-Wallis test was used (non-parametric, qualitative data against quantitative)

Variable	HdtS	LuxI	LuxS
Station metagenome size			
Observed stat	10795.42	29835.53	21963.30
p-value	4.425×10^{-8}	0.8070	0.0297
ρ	0.6501	0.0331	0.2882
Temperature			
Observed stat	25650.83	28923.05	29455.69
p-value	0.5881	0.7531	0.6497
ρ	0.0746	−0.0434	−0.0626
Salinity			
Observed stat	17155.63	17576.5	10683.97
p-value	0.7018	0.579	0.0203
ρ	−0.0580	−0.0840	0.3411
Chlorophyll			
Observed stat	27924.82	16952.69	17674.31
p-value	0.0023	0.3548	0.5018
ρ	−0.4247	0.1350	0.0967
Fe			
Observed stat	18117.4	21233.64	15623.73
p-value	0.7512	0.1238	0.5179
ρ	−0.0474	−0.2277	0.1239
NO_3			
Observed stat	26867.15	15388.56	14099.07
p-value	5.475×10^{-5}	0.4605	0.34
ρ	−0.5534	0.1102	0.1848
Si			
Observed stat	27220.46	12747.96	16131.44
p-value	2.481×10^{-5}	0.0741	0.6529
ρ	−0.5738	0.2629	0.0673
Primary production			
Observed stat	23759.31	13477.96	14815.8
p-value	0.0097	0.1359	0.3363
ρ	−0.3737	0.2207	0.1434
Habitat type			
χ^2	235.216	239.168	172.158
p-value	0.0737	0.0665	0.3061
df	15	15	15

DISCUSSION

Our study revealed the presence of AI-1 and 2 synthases, key enzymes in QS pathways, within the GOS predicted metaproteome. To date, QS genes have only been directly found in cultured marine

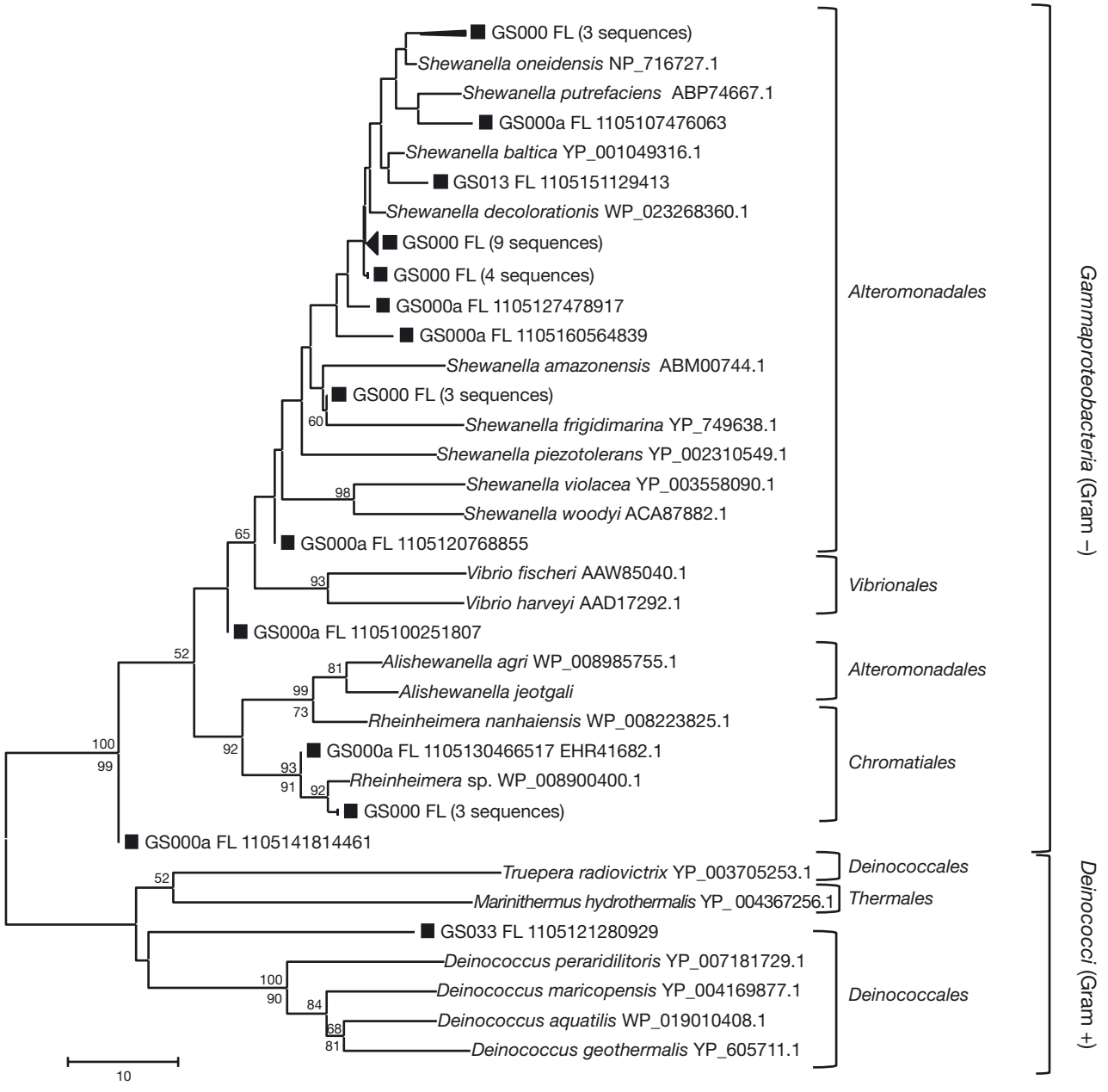

Fig. 4. Phylogenetic tree of annotated and environmental LuxS protein sequences produced using the Neighbour-Joining method with 1000 bootstrap replicates. Only bootstrap support > 50 are shown. Values above branches are those obtained from NJ method; values under branches are those obtained from ML methods. See Fig. 2 for definitions of symbols

bacteria (Engebrecht et al. 1983, Gilson et al. 1995, Gram et al. 2002, Wagner-Döbler et al. 2005) or indirectly inferred from AHL production and genome sequence annotation (Mohamed et al. 2008, Cuadrado-Silva et al. 2013, Doberva et al. 2014a,b). By contrast, no previous studies have examined the presence and diversity of those genes in the marine water column using culture-independent approaches.

We detected a large phylogenetic diversity of AI synthases in the GOS metagenomic libraries. However, it has not been possible to affiliate a large fraction of the retrieved sequences at a fine phylogenetic scale, as only a few marine cultivated strains harbour AI synthase genes that would allow a more accurate taxonomic association. At a higher phylogenetic level, the AI sequences retrieved from the GOS were affiliated to many different groups of bacteria. LuxI

environmental sequences were mainly related to *Rhodobacterales* (Cude & Buchan 2013), *Shingomonadales* and *Rhizobiales*, while HdtS sequences were mainly related to SAR11 and *Pseudomonadales*. These results must be very carefully interpreted, as it has been demonstrated that the phylogenetic signal in AI sequences may be affected by large horizontal transfer between bacterial groups (Gray & Garey 2001, Lerat & Moran 2004). However, our data clearly increase the currently known diversity of AI sequences. This high percentage of AI sequences without close relatives from known species in the GOS datasets clearly demonstrates that QS potential is present in a much wider diversity of marine bacteria than previously suspected from cultivable marine bacteria.

We may have underestimated the extent of AI diversity present in the GOS databases. Since our approach was based on sequence homology, we retained only the environmental sequences that were close to known and annotated sequences in databases to avoid false positives. Because of the high stringency of the search criteria used in this study (e-value 10^{-50} for first BLAST and 10^{-20} for RBH), it is possible that divergent AI environmental sequences were missed. This means that all AI sequences we can detect are dependent on the number and diversity of annotated sequences currently available in the NCBI NR database. The diversity of AI synthases is probably much greater than detectable by similarity searching. Despite this, the number of new AI sequences detected in this study was high, even though total diversity is still probably largely underestimated.

Another recurrent bias to assess when working with metagenomic datasets is the potential functional diversity of annotated proteins (Uchiyama & Miyazaki 2009, Prakash & Taylor 2012). Our work indicates that the majority of AI-1 synthases in the GOS dataset are related to the *hdtS* gene family. HdtS is a member of the lysophosphatidic acid acyltransferase family (Laue et al. 2000) and harbours a dual functionality: acylation of lysophosphatidic acid (Cullinane et al. 2005) and AHL synthesis (Laue et al. 2000, Churchill & Chen 2011). The production of AHL based on HdtS has been well demonstrated experimentally in *Pseudomonas fluorescens* (Laue et al. 2000), *Acidithiobacillus ferooxidans* (Rivas et al. 2007) and *Nitrosomonas europea* (Burton et al. 2005). When expressed in *Escherichia coli*, HdtS protein enabled the production of 3-OH-C14:1-AHL, C10-AHL and C6-AHL (Laue et al. 2000). One possible mechanism is that HdtS could transfer acyl chains from acyl-ACP or acyl-CoA to S-adenosylmethionine to generate AHLs (Cullinane et al. 2005). Although the experimental evidence accumulated from the strains mentioned above indicates a strong link with AHL production, the large diversity of HdtS found in the GOS metagenomic libraries highlights the need to further confirm a role of diverse HdtS-like proteins in QS.

Our results did not reveal any major difference in AI diversity for HdtS between the FL (0.1–0.8 µm) and PA (0.8–3 µm) fractions. It is commonly thought that bacteria in sea-water occur in low densities and thus AHLs molecules produced by a cell would quickly be diluted before reaching a receiving cell, hampering QS (Hmelo & Van Mooy 2009). Thus, a recurrent hypothesis is that QS in marine communities may have more chance to occur in particle-attached conditions where bacteria could be concentrated, such as on the surface or inside a particle of sinking organic matter (Mohamed et al. 2008, Hmelo et al. 2011). The absence of a significant difference (Wilcoxon test, $p = 0.83$) in HdtS diversity distribution between PA and FL fractions is not surprising as marine bacterial groups can present both FL and PA lifestyles (Ghiglione et al. 2007, Crespo et al. 2013). As the presence of QS genes is similar between PA and FL microenvironments, future studies should focus on detecting differential expression of these genes. Interestingly, recent work conducted on *Dinoroseobacter shibae* tends to support this hypothesis (Patzelt et al. 2013).

The LuxI and HdtS AI-1 synthase sequences retrieved in this study were detected in metagenomic libraries from the Atlantic, Pacific and Indian Oceans covered by the GOS. This result highlights that AI synthases were present in a wide range of marine environmental conditions, including open and coastal oceans, coral reefs and salt marshes. This observation supports the hypothesis that QS mechanisms may be widespread in many marine environments, and opens the wider question of the potential role of QS in marine prokaryotic species. Furthermore, the negative correlation between environmental sequences related to HdtS and environmental variables linked to coastal marine habitats suggests that HdtS was more prevalent in bacteria living in the open ocean.

Previous studies have revealed a wide phylogenetic distribution of the LuxS group among *Proteobacteria*, including the marine bacteria genera *Vibrio* (Bassler 1999) and *Shewanella* (Bodor et al. 2008), as well as within the *Firmicutes*, *Actinobacteria*, *Bacteroidales*, *Deinococcales* (*Deinococcus*)

and *Spirochaetes* (Schauder et al. 2001, Lerat & Moran 2004). The involvement of LuxS in QS remains a matter of debate (see Rezzonico & Duffy 2008 and Platt & Fuqua 2010 and references therein). However, many studies also report LuxS-based synthesized molecules as the initiators of a bacterial Esperanto (Winans 2002), i.e. allowing interspecies communication between bacteria. However, our results show a low abundance and a narrow taxonomic coverage of LuxS sequences retrieved from the GOS dataset. This may be due to the lifestyle of bacteria harbouring the AI-2 synthases that are frequently found in association with large organisms (Dworkin & Falkow 2006, Bodor et al. 2008) and thus may be undetectable with the protocols employed by the GOS cruise. Similarly, AinS AI synthases were very poorly represented in the GOS dataset. This result is perhaps not surprising as these AI synthases are known in only a few members of the *Vibrionaceae* (Milton et al. 2001), and therefore the low detection rate could be due to the small search database size. Interestingly, Rusch et al. (2007) also did not detect any member of the *Vibrionaceae* family in the GOS dataset based on 16S rRNA and *rpoB* gene analysis. This makes it more likely that the low number of AinS detected was due to the low sequence coverage of the GOS metagenomes being unable to detect rare members of the community such as *Vibrionaceae*. Again, these biases probably underestimated the real extent of marine AI diversity, which we have extended with this study.

Overall, our work has revealed a large number of new AI sequences. Our results show that diverse AI synthase genes are present in the marine environment and that many are related to presently uncultivated bacteria. It appears that AI synthases are diverse and present in all oceans crossed by the GOS expedition, supporting the hypothesis that QS cell-to-cell bacterial communication systems are widespread in marine waters. Thus, this study opens the door to the wider questions about the conditions under which expression of these genes occurs and the potential role of QS in marine bacteria and ecosystems.

Acknowledgements. We thank Olivier Naret for his help in reference database constitution. We thank Sheree Yau for ther advice on the revised manuscript version. We are grateful to the CNRS-UPMC ABiMS bioinformatics platform (http://abims.sb-roscoff.fr) for providing computational resources and support. This project was supported by an Emergence UPMC grant. The authors declare no conflict of interest.

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

Bassler BL (1999) How bacteria talk to each other: regulation of gene expression by quorum sensing. Curr Opin Microbiol 2:582–587

Bassler BL (2002) Small talk: cell-to-cell communication in bacteria. Cell 109:421–424

Bassler BL, Greenberg EP, Stevens AM (1997) Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*. J Bacteriol 179:4043–4045

Bodor A, Elxnat B, Thiel V, Schulz S, Wagner-Dobler I (2008) Potential for *luxS* related signalling in marine bacteria and production of autoinducer-2 in the genus *Shewanella*. BMC Microbiol 8:13

Brenner K, You L, Arnold FH (2008) Engineering microbial consortia: a new frontier in synthetic biology. Trends Biotechnol 26:483–489

Burton EO, Read HW, Pellitteri MC, Hickey WJ (2005) Identification of acyl-homoserine lactone signal molecules produced by *Nitrosomonas europaea* strain Schmidt. Appl Environ Microbiol 71:4906–4909

Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S (2007) Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 73:278–288

Case RJ, Labbate M, Kjelleberg S (2008) AHL-driven quorum-sensing circuits: their frequency and function among the *Proteobacteria*. ISME J 2:345–349

Cha C, Gao P, Chen YC, Shaw PD, Farrand SK (1998) Production of acyl-homoserine lactone quorum-sensing signals by gram-negative plant-associated bacteria. Mol Plant Microbe Interact 11:1119–1129

Chen X, Schauder S, Potier N, Van Dorsselaer A, Pelczer I, Bassler BL, Hughson FM (2002) Structural identification of a bacterial quorum-sensing signal containing boron. Nature 415:545–549

Churchill MEA, Chen L (2011) Structural basis of acyl-homoserine lactone-dependent signaling. Chem Rev 111:68–85

Crespo BG, Pommier T, Fernández-Gómez B, Pedrós-Alió C (2013) Taxonomic composition of the particle-attached and free-living bacterial assemblages in the Northwest Mediterranean Sea analyzed by pyrosequencing of the 16S rRNA. Microbiol Open 2:541–552

Cuadrado-Silva CT, Castellanos L, Arévalo-Ferro C, Osorno OE (2013) Detection of quorum sensing systems of bacteria isolated from fouled marine organisms. Biochem Syst Ecol 46:101–107

Cude WN, Buchan A (2013) Acyl-homoserine lactone-based quorum sensing in the Roseobacter clade: complex cell-to-cell communication controls multiple physiologies. Front Microbiol 4:336

Cullinane M, Baysse C, Morrissey JP, O'Gara F (2005) Identification of two lysophosphatidic acid acyltransferase genes with overlapping function in *Pseudomonas fluorescens*. Microbiology 151:3071–3080

de Kievit TR, Iglewski BH (2000) Bacterial quorum sensing in pathogenic relationships. Infect Immun 68:4839–4849

Decho AW, Norman RS, Visscher PT (2010) Quorum sensing in natural environments: emerging views from microbial

mats. Trends Microbiol 18:73–80
Dickschat JS (2010) Quorum sensing and bacterial biofilms. Nat Prod Rep 27:343–369
Diggle SP, Crusz SA, Cámara M (2007) Quorum sensing. Curr Biol 17:R907–R910
Doberva M, Sanchez-Ferandin S, Ferandin Y, Intertaglia L and others (2014a) Genome sequence of the sponge-associated *Ruegeria halocynthiae* strain MOLA R1/13b, a marine *Roseobacter* with two quorum-sensing-based communication systems. Genome Announc 2:e00993-14
Doberva M, Sanchez-Ferandin S, Ferandin Y, Intertaglia L, Joux F, Lebaron P, Lami R (2014b) Genome sequence of *Maribius* sp. strain MOLA 401, a Marine *Roseobacter* with a quorum-sensing cell-dependent physiology. Genome Announc 2:e00997-14
Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds) (2006) The Prokaryotes, 3rd edn, Vol 6: *Proteobacteria*: gamma subclass. Springer Science & Business Media, New York, NY
Engebrecht J, Silverman M (1984) Identification of genes and gene products necessary for bacterial bioluminescence. Proc Natl Acad Sci USA 81:4154–4158
Engebrecht J, Nealson K, Silverman M (1983) Bacterial bioluminescence: isolation and genetic analysis of functions from *Vibrio fischeri*. Cell 32:773–781
Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152
Fuqua C, Greenberg EP (2002) Listening in on bacteria: acyl-homoserine lactone signalling. Nat Rev Mol Cell Biol 3:685–695
Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J Bacteriol 176: 269–275
Ghiglione JF, Mevel G, Pujo-Pay M, Mousseau L, Lebaron P, Goutx M (2007) Diel and seasonal variations in abundance, activity, and community structure of particle-attached and free-living bacteria in NW Mediterranean Sea. Microb Ecol 54:217–231
Gilson L, Kuo A, Dunlap PV (1995) AinS and a new family of autoinducer synthesis proteins. J Bacteriol 177: 6946–6951
Gram L, Grossart HP, Schlingloff A, Kiørboe T (2002) Possible quorum sensing in marine snow bacteria: production of acylated homoserine lactones by *Roseobacter* strains isolated from marine snow. Appl Environ Microbiol 68: 4111–4116
Gray KM, Garey JR (2001) The evolution of bacterial LuxI and LuxR quorum sensing regulators. Microbiology 147: 2379–2387
Hmelo L, Van Mooy BAS (2009) Kinetic constraints on acylated homoserine lactone-based quorum sensing in marine environments. Aquat Microb Ecol 54:127–133
Hmelo LR, Mincer TJ, Van Mooy BAS (2011) Possible influence of bacterial quorum sensing on the hydrolysis of sinking particulate organic carbon in marine environments. Environ Microbiol Rep 3:682–688
Larkin MA, Blackshields G, Brown NP, Chenna R and others (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948
Laue BE, Jiang Y, Chhabra SR, Jacob S and others (2000) The biocontrol strain *Pseudomonas fluorescens* F113 produces the *Rhizobium small* bacteriocin, *N*-(3-hydroxy-7-*cis*-tetradecenoyl)homoserine lactone, via HdtS, a putative novel *N*-acylhomoserine lactone synthase. Microbiology 146:2469–2480
Lerat E, Moran NA (2004) The evolutionary history of quorum-sensing systems in bacteria. Mol Biol Evol 21: 903–913
Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659
Loh J, Pierson EA, Pierson LS III, Stacey G, Chatterjee A (2002) Quorum sensing in plant-associated bacteria. Curr Opin Plant Biol 5:285–290
March JC, Bentley WE (2004) Quorum sensing and bacterial cross-talk in biotechnology. Curr Opin Biotechnol 15: 495–502
Miller MB, Bassler BL (2001) Quorum sensing in bacteria. Annu Rev Microbiol 55:165–199
Miller ST, Xavier KB, Campagna SR, Taga ME, Semmelhack MF, Bassler BL, Hughson FM (2004) *Salmonella typhimurium* recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2. Mol Cell 15:677–687
Milton DL, Chalker VJ, Kirke D, Hardman A, Cámara M, Williams P (2001) The LuxM homologue VanM from *Vibrio anguillarum* directs the synthesis of *N*-(3-hydroxyhexanoyl)homoserine lactone and *N*-hexanoylhomoserine lactone. J Bacteriol 183:3537–3547
Mohamed NM, Cicirelli EM, Kan J, Chen F, Fuqua C, Hill RT (2008) Diversity and quorum-sensing signal production of *Proteobacteria* associated with marine sponges. Environ Microbiol 10:75–86
Moreno-Hagelsieb G, Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics 24:319–324
Nealson KH (1977) Autoinduction of bacterial luciferase. Arch Microbiol 112:73–79
Parsek MR, Greenberg EP (2005) Sociomicrobiology: the connections between quorum sensing and biofilms. Trends Microbiol 13:27–33
Patzelt D, Wang H, Buchholz I, Rohde M and others (2013) You are what you talk: quorum sensing induces individual morphologies and cell division modes in *Dinoroseobacter shibae*. ISME J 7:2274–2286
Platt TG, Fuqua C (2010) What's in a name? The semantics of quorum sensing. Trends Microbiol 18:383–387
Prakash T, Taylor TD (2012) Functional assignment of metagenomic data: challenges and applications. Brief Bioinform 13:711–727
Rezzonico F, Duffy B (2008) Lack of genomic evidence of AI-2 receptors suggests a non-quorum sensing role for luxS in most bacteria. BMC Microbiol 8:154
Rivas M, Seeger M, Jedlicki E, Holmes DS (2007) Second acyl homoserine lactone production system in the extreme acidophile *Acidithiobacillus ferrooxidans*. Appl Environ Microbiol 73:3225–3231
Rusch DB, Halpern AL, Sutton G, Heidelberg KB and others (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol 5:e77
Schaefer AL, Taylor TA, Beatty JT, Greenberg EP (2002) Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. J Bacteriol 184:6515–6521
Schauder S, Shokat K, Surette MG, Bassler BL (2001) The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. Mol Microbiol 41:463–476

Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007) CAMERA: a community resource for metagenomics. PLoS Biol 5:e75

Smith RS, Iglewski BH (2003) *P. aeruginosa* quorum-sensing systems and virulence. Curr Opin Microbiol 6:56–60

Sun S, Chen J, Li W, Altintas I and others (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. Nucleic Acids Res 39:D546–D551

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Toulza E, Tagliabue A, Blain S, Piganeau G (2012) Analysis of the Global Ocean Sampling (GOS) project for trends in iron uptake by surface ocean microbes. PLoS ONE 7: e30931

Uchiyama T, Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr Opin Biotechnol 20:616–622

Venter JC, Remington K, Heidelberg JF, Halpern AL and others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74

Wagner-Döbler I, Thiel V, Eberl L, Allgaier M and others (2005) Discovery of complex mixtures of novel long-chain quorum sensing signals in free-living and host-associated marine *Alphaproteobacteria*. ChemBioChem 6:2195–2206

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189–1191

Waters CM, Bassler BL (2005) QUORUM SENSING: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 21:319–346

Winans SC (2002) Bacterial Esperanto. Nat Struct Biol 9: 83–84

Withers H, Swift S, Williams P (2001) Quorum sensing as an integral component of gene regulatory networks in Gram-negative bacteria. Curr Opin Microbiol 4:186–193

Xavier KB, Bassler BL (2003) LuxS quorum sensing: more than just a numbers game. Curr Opin Microbiol 6: 191–197

Yooseph S, Sutton G, Rusch DB, Halpern AL and others (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. PLoS Biol 5:e16

Zhang Y, Gladyshev VN (2008) Trends in selenium utilization in marine microbial world revealed through the analysis of the Global Ocean Sampling (GOS) project. PLoS Genet 4:e1000095

Editorial responsibility: Craig Carlson, Santa Barbara, California, USA

Submitted: August 22, 2014; Accepted: November 22, 2014
Proofs received from author(s): January 30, 2015