Comparative Genomics Highlights Symbiotic Capacities and High Metabolic Flexibility of the Marine Genus *Pseudovibrio*

Dennis Versluis[1], Bart Nijsse[1,2], Mohd Azrul Naim[1], Jasper J. Koehorst[2], Jutta Wiese[3], Johannes F. Imhoff[3], Peter J. Schaap[2], Mark W.J. van Passel[1,4], Hauke Smidt[1], and Detmer Sipkema[1,*]

[1]Laboratory of Microbiology, Wageningen University & Research, The Netherlands

[2]Laboratory of Systems and Synthetic Biology, Wageningen University & Research, The Netherlands

[3]Marine Microbiology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

[4]National Institute for Public Health and the Environment, Bilthoven, The Netherlands

*Corresponding author: E-mail: detmer.sipkema@wur.nl.

Accepted: December 18, 2017

Data deposition: This project has been deposited at the European Nucleotide Archive under the accession PRJEB20602.

Abstract

Pseudovibrio is a marine bacterial genus members of which are predominantly isolated from sessile marine animals, and particularly sponges. It has been hypothesized that *Pseudovibrio* spp. form mutualistic relationships with their hosts. Here, we studied *Pseudovibrio* phylogeny and genetic adaptations that may play a role in host colonization by comparative genomics of 31 *Pseudovibrio* strains, including 25 sponge isolates. All genomes were highly similar in terms of encoded core metabolic pathways, albeit with substantial differences in overall gene content. Based on gene composition, *Pseudovibrio* spp. clustered by geographic region, indicating geographic speciation. Furthermore, the fact that isolates from the Mediterranean Sea clustered by sponge species suggested host-specific adaptation or colonization. Genome analyses suggest that *Pseudovibrio hongkongensis* UST20140214-015B^T is only distantly related to other *Pseudovibrio* spp., thereby challenging its status as typical *Pseudovibrio* member. All *Pseudovibrio* genomes were found to encode numerous proteins with SEL1 and tetratricopeptide repeats, which have been suggested to play a role in host colonization. For evasion of the host immune system, *Pseudovibrio* spp. may depend on type III, IV, and VI secretion systems that can inject effector molecules into eukaryotic cells. Furthermore, *Pseudovibrio* genomes carry on average seven secondary metabolite biosynthesis clusters, reinforcing the role of *Pseudovibrio* spp. as potential producers of novel bioactive compounds. Tropodithietic acid, bacteriocin, and terpene biosynthesis clusters were highly conserved within the genus, suggesting an essential role in survival, for example through growth inhibition of bacterial competitors. Taken together, these results support the hypothesis that *Pseudovibrio* spp. have mutualistic relations with sponges.

Key words: symbiosis, phylogeny, secondary metabolites, domainome, sponge, microbiota.

Introduction

Pseudovibrio is a genus of Gram-negative, heterotrophic, facultative anaerobic, marine α-proteobacteria. The genus was proposed by Shieh et al (2004), and presently six species have been described, including *Pseudovibrio denitrificans* (Shieh et al. 2004), *Pseudovibrio ascidiaceicola* (Fukunaga et al. 2006), *Pseudovibrio japonicus* (Hosoya and Yokota 2007), *Pseudovibrio axinellae* (O'Halloran et al. 2013), *Pseudovibrio hongkongensis* (Xu et al. 2015), and *Pseudovibrio stylochi* (Zhang et al. 2016). To date, members have been isolated from seawater (Hosoya and Yokota 2007), ascidians (Fukunaga et al. 2006), a flatworm (Zhang et al. 2016), a sea cucumber (Zhang et al. 2013), tunicates (Sertan-de Guzman et al. 2007; Riesenfeld et al. 2008), corals (Chen et al. 2012), and sponges (Santos-Gandelman et al. 2013; Versluis et al. 2017). Members of the genus *Pseudovibrio* are characterized as motile, rod-shaped marine bacteria. For energy production, they are capable of oxygen-dependent respiration and/or nitrate- and nitrite-dependent (anaerobic) respiration as well as fermentation. An analysis of the

genomes of *Pseudovibrio* sp. FO-BEG1 and *Pseudovibrio* sp. JE062 showed that *Pseudovibrio* spp. are metabolically versatile, that is, they are capable of utilizing a wide range of organic and inorganic compounds to meet their carbon, nitrogen, phosphorous, and energy requirements (Bondarev et al. 2013).

The relationship between *Pseudovibrio* and sponges is particularly interesting because these bacteria have consistently been isolated from different sponge species (Lafi et al. 2005; Muscholl-Silberhorn et al. 2008; Menezes et al. 2010), while never simultaneously being isolated from nor detected in surrounding seawater (Webster and Hill 2001; Enticknap et al. 2006). Although it should be noted that the relative abundance of *Pseudovibrio* spp. in the sponge microbiota is generally low considering that their presence is rarely detected by cultivation-independent assays (Enticknap et al. 2006), these findings indicate that *Pseudovibrio* spp. are enriched in sponges, reinforcing the current hypothesis that they are symbionts of sponges. Further support for this hypothesis comes from the fact that *Pseudovibrio* spp. were reported to be vertically transmitted through sponge larvae of *Mycale laxissima* (Enticknap et al. 2006). Furthermore, *Pseudovibrio* spp. were found in reduced relative abundance in the cultured bacterial community of diseased specimen of *Rhopaloeides odorabile* (Webster and Hill 2001; Webster et al. 2002) and *Callyspongia* aff. *biru* (Sweet et al. 2015). This suggests that *Pseudovibrio* spp. may benefit sponge health.

The symbiotic role of *Pseudovibrio* in the sponge host is still unclear but it could include functions in nutrient uptake (Webster and Hill 2001), denitrification (Shieh et al. 2004; Fiore et al. 2010; Han et al. 2013), or host defence (Penesyan et al. 2011; Graca et al. 2013). In addition, genomic data indicated several mechanisms for establishing and maintaining symbiosis, for example by interactions of *Pseudovibrio* with the host immune system or cytoskeleton (Bondarev et al. 2013; Alex and Antunes 2015; Romano et al. 2016). *Pseudovibrio* spp. have received particular interest as sources of clinically relevant antimicrobials (Santos et al. 2010; O'Halloran et al. 2013), with multiple studies reporting that members of this genus are highly bioactive (Flemer et al. 2012; Graca et al. 2013). So far, *Pseudovibrio*-derived secondary metabolites that have been identified are phenazine (Schneemann et al. 2011), tropodithietic acid (TDA) (Penesyan et al. 2011; Harrington et al. 2014) and pseudovibrocin (Vizcaino 2011). Pathways of polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS), and/or their hybrids, are particularly important in the production of secondary metabolites by *Pseudovibrio* spp., albeit not for TDA biosynthesis (Kennedy et al. 2009; O'Halloran et al. 2011; Crowley et al. 2014).

In this study, we compared the genomes of 31 *Pseudovibrio* isolates, including 25 isolates obtained from sponges, 1 isolate retrieved from a bryozoan, and 5 type strains, by employing SAPP, the Semantic Annotation Pipeline with Provenance (Koehorst et al. 2016a, 2016b). Twenty-eight of these *Pseudovibrio* isolates were genome sequenced in this study, namely 22 isolates from sponges, 1 isolate from a bryozoan, and the 5 type strains. We resolved the phylogeny, and studied metabolic and secondary metabolite biosynthesis (SMB) pathways. In addition, we predicted features of antibiotic resistance and host–symbiont interactions, providing additional insights into the nature of the association between *Pseudovibrio* spp. and marine sponges.

Materials and Methods

Sample Collection and Data Deposition of All Analyzed Genomes

The type strains *P. ascidiaceicola* DSM 16392^T, *P. axinellae* DSM 24994^T, *P. denitrificans* JCM 12308^T, and *P. japonicus* NCIMB 14279^T were obtained from the respective culture collections. Sponge-associated *Pseudovibrio* strains were isolated from the sponges *Corticium candelabrum* ($n = 2$), *Petrosia ficiformis* ($n = 10$), and *Aplysina aerophoba* ($n = 7$) as described previously (Versluis et al. 2017). *Pseudovibrio* sp. AB108, *Pseudovibrio* sp. AB111, *Pseudovibrio* sp. AB113, and *Pseudovibrio* sp. BC118 were isolated from the sponges *Ircinia* sp., *Chondrilla nucula*, *Acanthella acuta*, and the bryozoan *Cellepora pumicosa*, respectively, collected in the Limski Channel in Croatia (45°7′54.56″N, 13°39′13.02″E) as described previously (Thiel and Imhoff 2003; Heindl et al. 2010). The type strain *P. hongkongensis* UST20140214-015B^T was genome sequenced but not included in growth experiments. Publicly available genome sequences of *Pseudovibrio* sp. FO-BEG1 (GenBank accession numbers: CP003147 and CP003148), *Pseudovibrio* sp. JE062 (ABXL00000000), *Pseudovibrio* sp. POLY-S9 (LCWX00000000), *Stappia stellulata* DSM 5886^T (AUIM00000000), *Nesiotobacter exalbescens* LA33B^T (AUGS00000000), and *Labrenzia alexandrii* DFL-11^T (ACCU00000000) were downloaded and reannotated applying the methods particular to this study. The assembled genomes and the corresponding GenBank files of the *Pseudovibrio* strains sequenced as part of this study were deposited under European Nucleotide Archive study accession PRJEB20602.

DNA Isolation and Sequencing

All *Pseudovibrio* strains were inoculated from glycerol stocks onto marine agar 2216 (Difco, Detroit, USA). Subsequently, a single colony was grown at 20 °C in marine broth 2216 (Difco), and DNA was isolated using the MasterPure™ DNA Purification Kit (Epicentre, Madison, USA). Genome sequencing was done using the Illumina MiSEQ (paired end, 2×300 bp reads, 500 bp average insert size) at GATC Biotech (Konstanz, Germany). As an exception, DNA from

the liquid culture of *P. hongkongensis* UST20140214-015B^T was isolated as described previously (Ausubel et al. 2002). In addition, the genome of this strain was sequenced on the Illumina Hiseq 2000 (paired end, 2 × 101 bp reads, 500 bp average insert size) (Shanghai Majorbio Bio-pharm Technology Co., Ltd, Shanghai).

Antibiotic Resistance

Resistance profiles were determined by inoculation of pregrown liquid cultures onto marine agar 2216 containing 1 of the following 17 (combinations of) antibiotics: (1) 20 μg/ml polymyxin B, (2) 20 μg/ml daptomycin, (3) 50 μg/ml vancomycin, (4) 50 μg/ml penicillin, (5) 20 μg/ml erythromycin, (6) 20 μg/ml ciprofloxacin, (7) 100 μg/ml cefotaxime, (8) 50 μg/ml tetracycline, (9) 50 μg/ml chloramphenicol, (10) 95 μg/ml sulfamethoxazole and 5 μg/ml trimethoprim, (11) 20 μg/ml lincomycin, (12) 100 μg/ml kanamycin, (13) 20 μg/ml rifampicin, (14) 20 μg/ml linezolid, (15) 50 μg/ml ampicillin, (16) 20 μg/ml imipenem, and (17) 50 μg/ml D-cycloserine. Antibiotic resistance was evaluated 2 days postinoculation and compared with growth on media without antibiotics as a reference. We defined three levels of antibiotic resistance: (1) "resistant"; growth of the bacteria was identical to their growth on media without antibiotics, (2) "intermediately resistant"; growth of the bacteria was slower than growth on media without antibiotics, and (3) "susceptible"; no growth.

Genome Assembly and Quality Control

The read quality and presence of Illumina Truseq adapter sequences was investigated with FASTQC (Andrews 2010), and Illumina Truseq adapter sequences were removed with Cutadapt 1.8.1 using default settings (Martin 2011). All genomes were assembled with the A5-miseq assembler (Coil et al. 2015) version 20150522 using default settings with as exception the *P. hongkongensis* UST20140214-015B^T genome that was assembled with Velvet 1.2.10 (Zerbino and Birney 2008). Subsequently, Pilon 1.13 (Walker et al. 2014) was used to improve the assemblies based on the assembled scaffolds and error-corrected reads obtained as output from the A5 assembler. Assemblies were investigated for contamination by BLASTn v2.3.0 (Altschul et al. 1990) using the NCBI nr/nt database (June 2015) as a reference, and using default settings. Overall, seven <3 kb contigs were removed that were assigned to *Clostridia* and *Bacilli* spp. Sequences from *Enterobacteria* phage phiX174 were removed as well as DNA of this phage is routinely included as a positive control during sequencing. In order to determine the coverage, A5 error-corrected reads were aligned with the assemblies by Bowtie2 2.2.5 (Langmead and Salzberg 2012) using default settings. The resulting sequence alignment map (SAM) file was converted by Samtools 0.1.19 (Li et al. 2009) into a binary alignment map (BAM) file, after which the coverage per base was calculated using Bedtools 2.25.0 (Quinlan and Hall 2010).

Genome Annotation

To obtain a manageable and homogeneous annotation with provenance, the genomes of all *Pseudovibrio* strains were de novo annotated using a defined set of algorithms and an in-house pipeline for annotation and data storage (Koehorst et al. 2016a). In addition, the genomes of *L. alexandrii* DFL-11^T, *S. stellulata* DSM 5886^T, and *N. exalbescens* DSM 16456^T were de novo annotated and added in order to serve as a reference. The complete genome sequences were used to detect rRNA genes by RNAmmer 1.2 (Lagesen et al. 2007), and to detect tRNA and tmRNA genes by ARAGORN 1.2.36 (Lagesen et al. 2007). Prodigal 2.6.2 (Hyatt et al. 2010) was applied to predict genes, and subsequently the InterPro Database 5.4-47.0 (Mitchell et al. 2015) was used to classify the genes and to predict domains and sites. At runtime, the following databases were selected: TIGRFAM, ProDom, SMART, PROSITE, PfamA, PRINTS, SUPERFAMILY, Coils, and Gene3d. PRIAM (Claudel-Renard et al. 2003), SignalP 4.0 (Petersen et al. 2011), and TMHMM (Krogh et al. 2001) were used to assign Enzyme Commission numbers (ECs), to detect signal peptides, and to detect transmembrane protein topology, respectively. Predictions and provenance were converted into the RDF data model with a self-defined ontology (van Dam et al. 2015). Results are stored as Turtle files and analyzed as described (Koehorst et al. 2016b).

Phylogeny, Functional Diversity, and Environmental Distribution

A phylogenetic tree was made in ARB (Ludwig et al. 2004) with the 16S rRNA gene sequences from the genomes. When multiple 16S rRNA genes were found in a genome that were >99% identical, a single gene was randomly picked to be included in the tree. The three closest non-*Pseudovibrio* relatives in the Silva guide tree (release 115) (Yilmaz et al. 2014) from which the genomes were publicly available were used as outgroup. Gene alignments were manually curated, and a Maximum Likelihood tree was constructed with 1,000 iterations of rapid bootstrapping. Average nucleotide identity (ANI) between the different genomes based on BLAST results (ANIb) was calculated using JSpecies v1.2.1 with default parameters (Richter and Rossello-Mora 2009). The dissimilarity in terms of protein diversity was analyzed and visualized in a distance tree. Therefore, to identify homologous proteins shared between the genomes, an all-against-all BLASTp was performed, after which OrthAgogue v1.0.3 (Ekseth et al. 2014) and MCL 14-137 (van Dongen 2000) were used to identify orthologous groups of proteins (OGPs), where the OGPs included proteins with both orthologous and inparalogous relations. A presence/absence matrix was created using all types of OGPs. With R version 3.2.2. (R development Core Team 2010), a NxN Jaccard distance

matrix was created based on the presence/absence of the OGPs that are present in greater than two genomes. This distance matrix was used for complete linkage hierarchical clustering. To assess the uncertainty in the hierarchical cluster analysis, we used the R package pvclust 2.0-0 (Suzuki and Shimodaira 2006) with 10,000 bootstraps. Canonical (constrained) correspondence analysis as implemented in Canoco 5 (Šmilauer and Lepš 2014) was used to investigate which variables (i.e., membership of the *Pseudovibrio* genus, geographic origin, and sponge of origin) could explain the variation in OGP composition among the genomes. The OGP compositional table contained information about the presence/absence of the OGPs in the genomes. The presence of metabolic pathways was predicted using OGPs that had EC numbers assigned, and using the MetaCyc database (Caspi et al. 2016) as a reference. The presence/absence data of OGPs that are present in at least three genomes were used as input for SIMPER from the package PRIMER 6 v6.1.9 (PRIMER-E Ltd, Plymouth, UK) (Clarke and Gorley 2006) in order to calculate the contribution of each OGP to the observed dissimilarity between genomes grouped by (environmental) variables. No cut-off was used for low contributions. DIAMOND v.0.7.12 (Buchfink et al. 2015) was used to assign all protein sequences to Clusters of Orthologous Groups (COGs) with the 2014 edition of the COGs database as a reference (Galperin et al. 2015). The environmental distribution of *Pseudovibrio* spp. was investigated using the Integrated Microbial Next Generation Sequencing (IMNGS) platform (Lagkouvardos et al. 2016). Full-length 16S rRNA gene sequences of the type strains were used as query, and the similarity threshold was set at 99% nucleotide identity.

Annotation of Symbiosis-Related Genes

The standalone version of Interproscan 5.17-6.0 (Jones et al. 2014; Mitchell et al. 2015) was used to detect the following potentially symbiosis-related gene products: ankyrin-repeat proteins (ARPs; IPR020683 and IPR002110), proteins with SEL1 repeats (IPR006597), tetratricopeptide repeat proteins (TRPs; IPR011990, IPR019734, IPR013105, IPR001440, and IPR011717), invasion protein B (IPR010642), proteins with the TadE-like domain (IPR012495), and LuxR (IPR000792 and IPR005143). Proteins with SEL1 repeats and tetratricopeptide repeats are occasionally wrongly classified due to their high similarity. Therefore, we used TPRpred (Karpenahalli et al. 2007) to correctly categorize proteins with these repeats using an *e*-value of $1.0E^{-7}$ as cut-off. The presence of Type III secretion systems (T3SSs), Type IV secretion systems (T4SSs), and Type VI secretion systems (T6SSs) was investigated in the genomes with MultiGeneBlast 1.1.14 (Medema et al. 2013) using default settings with as query previously identified gene clusters that were described by Romano et al. (2016), that is, FO-BEG1_3657-03696 (T3SS), PsW64_00918-00938 (T4SS), FO-BEG1_01844-01855 (T6SS-I), and FO-BEG1_02827-02846 (T6SS-II).

Detection of Antibiotic Resistance Genes and SMB Clusters

The presence of the *tdaABCDEFH* genes and the genes *paaI*, *paaJ*, *paaK*, *cysI*, and *malY* in the genomes was assessed to predict the capacity for TDA biosynthesis (Geng and Belas 2010; Harrington et al. 2014), whereas the presence of the *tdaR1*, *tdaR2*, and *tdaR3* genes (Wilson et al. 2016) was assessed to predict TDA resistance. Therefore, BLASTp using the sequences of UniProtKB accessions G8PKJ3_PSEUV (TdaA), G8PKJ1_PSEUV (TdaB), G8PKJ0_PSEUV (TdaC), G8PKI9_PSEUV (TdaD), G8PKI8_PSEUV (TdaE), G8PKH6_PSEUV (TdaF), G8PGC0_PSEUV (TdaH), I7DVM6_PHAI2 (tdaR1), I7EHC6_PHAI2 (tdaR2), and I7ES62_PHAI2 (tdaR3) was performed against the genome-encoded proteins. Furthermore, the presence of the genes *paaI*, *paaJ*, *paaK*, *cysI*, and *malY* was analyzed by investigating if genome-encoded proteins were assigned the EC numbers 3.1.2. (PaaI), 2.3.1.174 (PaaJ), 6.2.1.30 (PaaK), 1.8.1.2 (CysI), and 4.4.1.8. (MalY). The presence of gene clusters involved in curli fiber formation was investigated by BLASTp against the genome-encoded proteins using sequences of UniProtKB accessions G8PUY1_PSEUV (curlin associated repeat protein), G8PUY2_PSEUV (curlin associated repeat protein), G8PUY3_PSEUV (curlin associated repeat protein), G8PUY4_PSEUV (CsgG), and G8PUY5_PSEUV (CsgF). AntiSMASH version 3.0.4 was used to identify SMB clusters (Weber et al. 2015). Similarity among SMB clusters from the same type was investigated with MultiGeneBlast by calculating the number cluster-genes that shared $\geq$30% nucleotide identity (Medema et al. 2013). Antibiotic resistance functions were predicted with HMMER 3.1b2 (http://hmmer.org/) using the pHMMs of the Resfams database (core) v1.2 (Gibson et al. 2015). In addition, BLASTp of amino acid sequences against the CARD database (McArthur et al. 2013) was performed.

Results and Discussion

Metadata and Genome Characteristics

The genomes of 31 *Pseudovibrio* strains were analyzed (table 1) including 5 type strains that are currently characterized within the genus. The genomes of 28 *Pseudovibrio* strains are new and were sequenced as part of this study, whereas the genomes of strains *Pseudovibrio* sp. FO-BEG1, *Pseudovibrio* sp. JE062, and *Pseudovibrio* sp. POLY-S9 were obtained from previous studies (Bondarev et al. 2013; Alex and Antunes 2015). Twenty-five strains were isolated from sponges, four strains were isolated from other sessile marine animals, and two strains were isolated from seawater (table 1). The genome sizes of all *Pseudovibrio* strains, except for *P. hongkongensis* UST20140214-015B^T, were large (6 Mb $\pm$ 0.6 [s.d.]). The genome of *P. hongkongensis* was only 3.75 Mb and correspondingly, the genome also contained fewer unique OGPs (2,878 in total). On average,

Table 1

Metadata, Genome Assembly, and Annotation Statistics of the 31 *Pseudovibrio* Strains and the Three Close Relatives (Outgroup Members) Analyzed in This Study

Genome Metadata			Assembly Statistics					Genome Annotation Statistics		
Strain	Isolation Source	Geographic Location	Number of Scaffolds	Genome Size (Mb)	N50 (Mb)	GC content (%)	Coverage (X)	Number of ORFs	Number of OGP-assigned proteins	Number of unique OGPs
**P. hongkongensis* UST20140214-015B^T	Flatworm *Stylochus* sp.	Yung Shue O, Hong Kong	39	3.75	0.34	51.7	131	3,514	2,939	2,878
**P. ascidiaceicola* DSM 16392^T	Ascidian *P. proliferus*	Beach of Boso peninsula, Japan	40	5.88	0.42	50.3	178	5,365	5,112	4,992
**P. axinellae* DSM 24994^T	Sponge *A. dissimilis*	A marine lake, Ireland	78	5.20	0.39	50.7	230	4,786	4,122	3,982
**P. denitrificans* JCM 12308^T	Seawater	Nanwan Bay, Taiwan	27	6.12	0.64	50.7	199	5,428	5,123	5,033
**P. japonicus* NCIMB 14279^T	Seawater	Coast of Boso peninsula, Japan	17	4.97	0.65	50.9	245	4,554	3,945	3,870
**Pseudovibrio* sp. 2G02	Sponge *P. ficiformis*	Coast of Blanes, Spain	25	5.87	0.41	50.2	246	5,369	5,204	5,108
**Pseudovibrio* sp. 2H03	Sponge *P. ficiformis*	Coast of Blanes, Spain	39	6.00	0.47	50.0	233	5,524	5,349	5,210
**Pseudovibrio* sp. 2H05	Sponge *P. ficiformis*	Coast of Blanes, Spain	28	5.94	0.42	50.1	137	5,389	5,234	5,138
**Pseudovibrio* sp. 2H06	Sponge *P. ficiformis*	Coast of Blanes, Spain	21	6.09	0.57	50.7	221	5,474	5,328	5,200
**Pseudovibrio* sp. 4B07	Sponge *P. ficiformis*	Coast of Blanes, Spain	52	5.13	0.38	50.4	226	4,692	4,281	4,164
**Pseudovibrio* sp. 4H03	Sponge *P. ficiformis*	Coast of Blanes, Spain	26	5.79	0.88	50.3	212	5,295	5,196	5,078
**Pseudovibrio* sp. 4H10	Sponge *P. ficiformis*	Coast of Blanes, Spain	14	5.78	0.86	50.8	250	5,218	5,116	5,029
**Pseudovibrio* sp. 8G08	Sponge *P. ficiformis*	Coast of Blanes, Spain	36	5.94	0.48	50.3	165	5,361	5,240	5,144
**Pseudovibrio* sp. 8H04	Sponge *P. ficiformis*	Coast of Blanes, Spain	18	5.68	0.50	48.2	295	5,174	4,653	4,512
**Pseudovibrio* sp. 8H06	Sponge *P. ficiformis*	Coast of Blanes, Spain	40	6.06	0.40	49.7	264	5,554	5,342	5,207
**Pseudovibrio* sp. 1F12	Sponge *A. aerophoba*	Gulf of Roses, Spain	28	5.84	0.49	50.9	253	5,289	5,272	5,291
**Pseudovibrio* sp. 1D03	Sponge *A. aerophoba*	Gulf of Roses, Spain	22	5.77	0.50	50.3	711	5,239	5,126	5,051
**Pseudovibrio* sp. 3C10	Sponge *A. aerophoba*	Gulf of Roses, Spain	36	6.05	0.85	49.9	316	5,539	5,188	5,082
**Pseudovibrio* sp. 7F08	Sponge *A. aerophoba*	Gulf of Roses, Spain	27	5.78	0.48	51.3	124	5,243	5,127	5,021
**Pseudovibrio* sp. 7G12	Sponge *A. aerophoba*	Gulf of Roses, Spain	29	5.85	0.45	51.0	138	5,288	5,271	5,192
**Pseudovibrio* sp. 7H02	Sponge *A. aerophoba*	Gulf of Roses, Spain	28	5.84	0.52	50.9	119	5,288	5,272	5,192
**Pseudovibrio* sp. 7H04	Sponge *A. aerophoba*	Gulf of Roses, Spain	23	5.59	0.58	50.4	119	5,041	4,884	4,815
**Pseudovibrio* sp. 7A08	Sponge *C. candelabrum*	Coast of Blanes, Spain	22	5.76	0.59	50.8	126	5,176	5,116	5,014
**Pseudovibrio* sp. 2GO5	Sponge *C. candelabrum*	Coast of Blanes, Spain	20	5.76	0.59	50.8	189	5,182	5,117	5,020
**Pseudovibrio* sp. AB108	Sponge *Ircinia* sp.	Limski Channel, Croatia	49	5.91	0.44	44.6	200	5,369	5,202	5,098
**Pseudovibrio* sp. AB111	Sponge *C. nucula*	Limski Channel, Croatia	36	5.92	0.42	49.8	228	5,423	5,211	5,074
**Pseudovibrio* sp. AB113	Sponge *A. acuta*	Limski Channel, Croatia	13	5.37	0.88	51.0	191	4,886	4,796	4,727
**Pseudovibrio* sp. BC118	Bryozoan *C. pumicosa*	Limski Channel, Croatia	23	5.81	0.52	51.3	237	5,304	5,169	5,070
Pseudovibrio sp. FO-BEG1	A scleractinian coral	Coast of Florida, USA	1	5.48	5.48	52.4	n/a	4,926	4,842	4,673
Pseudovibrio sp. JE062	Sponge *M. laxissima*	Key Largo, Florida, USA	19	5.73	0.51	52.1	n/a	5,138	4,928	4,842
Pseudovibrio sp. POLY-S9	Sponge *P. penicillus*	Coast of Portugal	271	6.60	0.12	51.0	n/a	6,331	5,530	5,308
S. stellulata DSM 5886^T	Marine sediment		15	4.62	0.69	63.9	n/a	4,212	2,879	2,810
N. exalbescens LA33B^T	Water in a hypersaline lake	Laysan Atoll, Hawaiian Islands	40	4.15	0.29	54.6	n/a	3,771	3,094	3,022
L. alexandrii DFL-11^T	Culture of the dinoflagellate *A. lusitanicum*	Helgoland, Germany	6	5.50	5.30	57.4	n/a	5,128	3,276	3,154

NOTE.—An asterisk indicates that the genome was sequenced as part of this study.

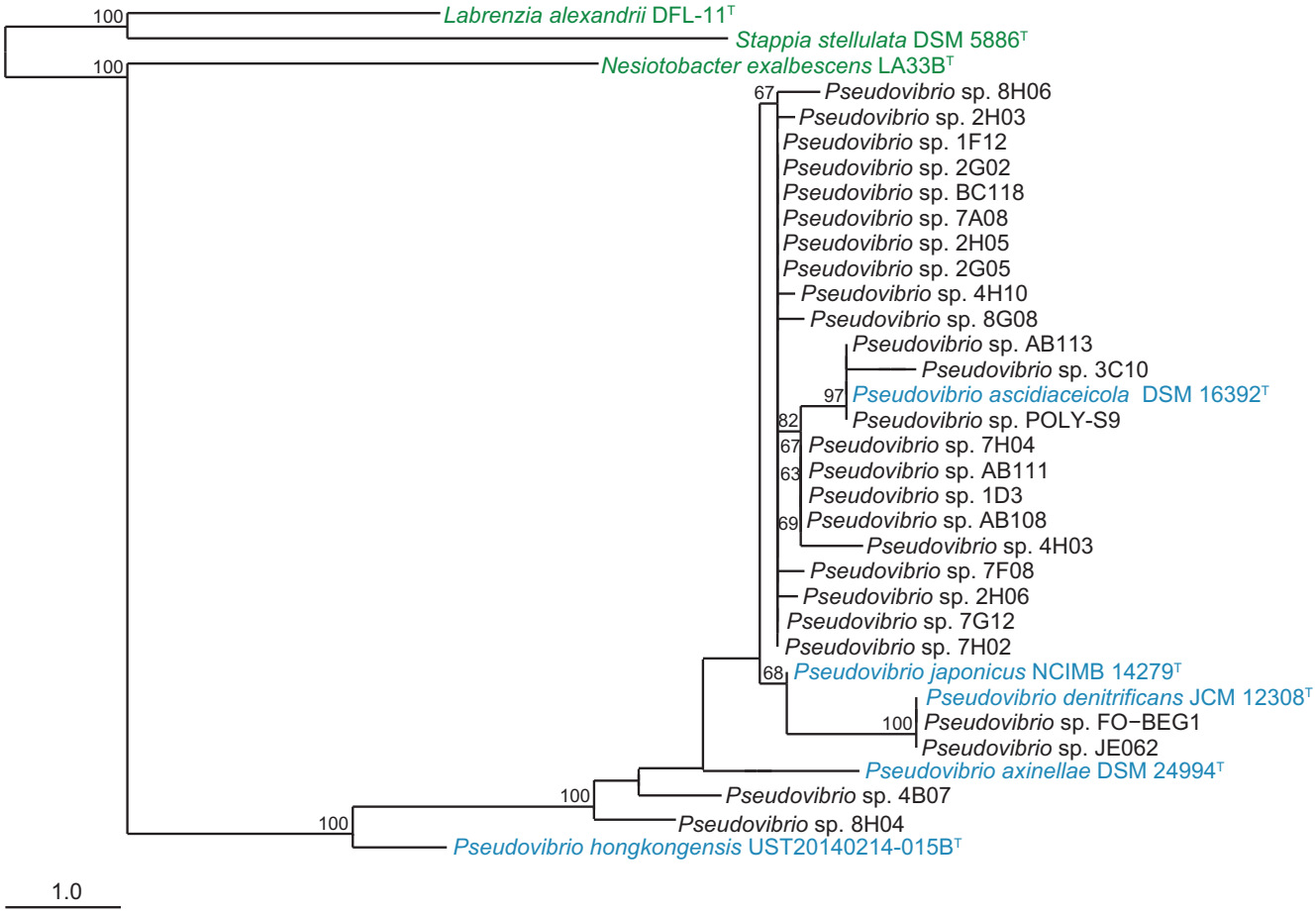

Fig. 1.—Phylogenetic tree based on 16S rRNA gene sequences from the 31 *Pseudovibrio* spp. and 3 closely related species that were included in the comparative genomics analysis. *Pseudovibrio* type strains are in blue and the closely related species are in green. The tree was constructed in ARB by Maximum likelihood analysis using 1,000 iterations of bootstrapping. Bootstrap values <50 are not shown. The horizontal bar indicates the number of substitutions per site.

Pseudovibrio genomes, excluding *P. hongkongensis* UST20140214-015B^T, contained 4,857 ± 508 (s.d.) unique OGPs (table 1, supplementary table S1, Supplementary Material online). The GC-content of the *Pseudovibrio* strains ranged from 44.6% for *Pseudovibrio* sp. AB108 to 52.4% for *Pseudovibrio* sp. FO-BEG1, whereas the GC-content of the three close relatives was 63.9% for *S. stellulata* DSM 5886^T isolated from marine sediment (Ruger and Hofle 1992), 54.6% for *N. exalbescens* LA33B^T isolated from water in a hypersaline lake (Donachie et al. 2006), and 57.4% for *L. alexandrii* DFL-11^T isolated from a culture of the dinoflagellate *Alexandrium lusitanicum* (Fiebig et al. 2013).

Phylogeny and Taxonomy

Phylogeny of the strains was assessed based on 16S rRNA gene sequences retrieved from the respective genomes as well as the presence of OGPs. Phylogenetic analysis based on 16S rRNA gene sequences placed all *Pseudovibrio* type strains (DSM 16392^T, DSM 24994^T, JCM 12308^T, NCIMB 14279^T, and UST20140214-015B^T) in separate branches of the corresponding tree (fig. 1). Most *Pseudovibrio* strains (23 of 31) could be classified at the species level to *P. ascidiaceicola* based on the phylogenetic tree and 16S rRNA gene identity values (supplementary table S2, Supplementary Material online). Further phylogenomic analysis of these strains based on the presence of OGPs placed *P. hongkongensis* UST20140214-015B^T closer to the non-*Pseudovibrio* outgroup members when compared with the 16S rRNA gene-based phylogenetic analysis (fig. 2). In addition, the ANI between the *P. hongkongensis* type strain and all other *Pseudovibrio* spp. strains was <75%, which was comparable to ANI values observed for the three outgroup species that all belong to different genera (supplementary table S3, Supplementary Material online). As such, the data presented here suggest that *P. hongkongensis* UST20140214-015B^T should not be considered a member of the *Pseudovibrio* genus and therefore might need to be placed in a separate

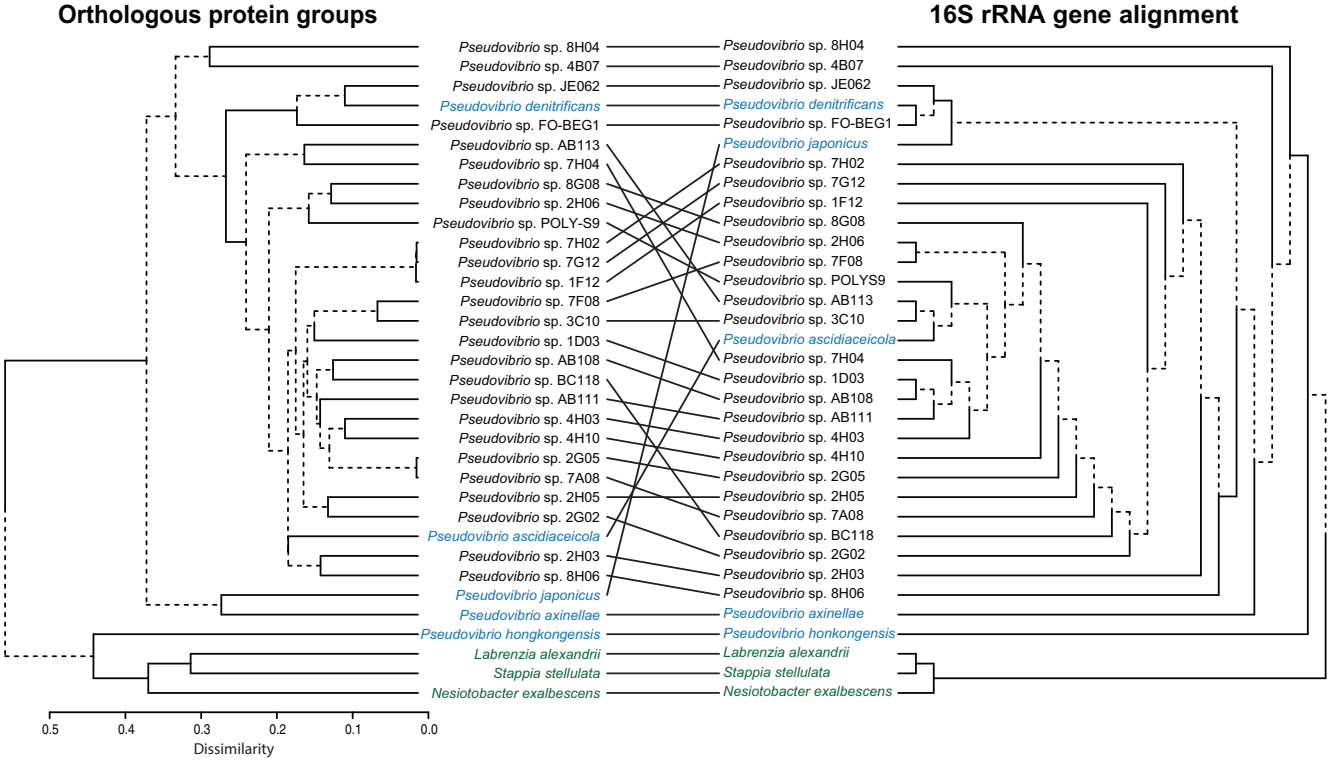

Fig. 2.—Functional similarity and phylogenetic distance of *Pseudovibrio* spp.. Comparison of two phylogenetic trees made by hierarchical linkage clustering based (1) on presence/absence of OGPs that are present in greater than two genomes and (2) the 16S rRNA gene alignment. *Pseudovibrio* type strains are depicted in blue and the outgroup members are depicted in green. Hierarchical clustering was performed using 10,000 iterations of bootstrapping. Dashed branches can be swapped.

genus. For this reason, it was treated as an outgroup member in all subsequent analyses. Furthermore, *Pseudovibrio* sp. 8H04 and *Pseudovibrio* sp. 4B07 were found in isolated branches in both trees (fig. 2), and 16S rRNA gene identity values did not allow species classification as both these strains shared between 97.6% and 98.6% nucleotide identity with the four most closely related type strains: *P. ascidiaceicola* DSM 16392^T, *P. denitrificans* JCM 12308^T, *P. japonicus* NCIMB 14279^T, and *P. axinellae* DSM 24994^T. These results, and the fact that the ANI values between these two strains and the *Pseudovibrio* spp. type strains are rather divergent (Richter and Rossello-Mora 2009) (*Pseudovibrio* strain 4B07 and strain 8H04 share, respectively, 94.3% and 97.1% ANI with the most similar type strain), indicate that they likely represent two distinct novel species within the *Pseudovibrio* genus. These strains would not have been recognized as novel species based on 16S rRNA gene similarity alone, and hence, 16S rRNA gene identity values underrepresent the genomic diversity within *Pseudovibrio*. This is also in line with the observation that *P. japonicus* NCIMB 14279^T is closely related to *P. denitrificans* JCM 12308^T based on the 16S rRNA gene tree (99% sequence identity) even though the two strains were highly dissimilar based on the OGP-based tree and ANI (81.4%). ANI scores showed strong r^2 correlation values of >0.93 with DNA–DNA hybridization (DDH) values and therefore can confidently assess species delineation (Konstantinidis and Tiedje 2005; Goris et al. 2007). The 70% DDH cut-off value for species delineation corresponds to 94–95% ANI (Richter and Rossello-Mora 2009). It should be noted, however, that genome-to-genome sequence comparison methods such as ANI are not (yet) allowed to be used as substitute for DDH in the characterization of a novel taxon (Auch et al. 2010; Tindall et al. 2010; Kim et al. 2014).

Clustering by Metadata and Group-Differences

To identify variables that best explain the variation in the distribution of OGPs in the genomes, we performed a canonical correspondence analysis (CCA). Strains that are members of the outgroup (including *P. hongkongensis* UST20140214-015B^T) clustered separately from members of the *Pseudovibrio* genus ($P = 0.002$) (fig. 3*A*). Furthermore, we found clustering by geographic origin (Mediterranean, Caribbean, Pacific, or Atlantic, $P = 0.002$) (fig. 3*B*). The fact that geographic origin can partially explain the variation in OGPs suggests that in different geographic regions adaptation has occurred to local environmental conditions (e.g., temperature, salinity, local flora and fauna, or nutrient availability). This is underlined by the fact that clusters in two of four geographic regions included different

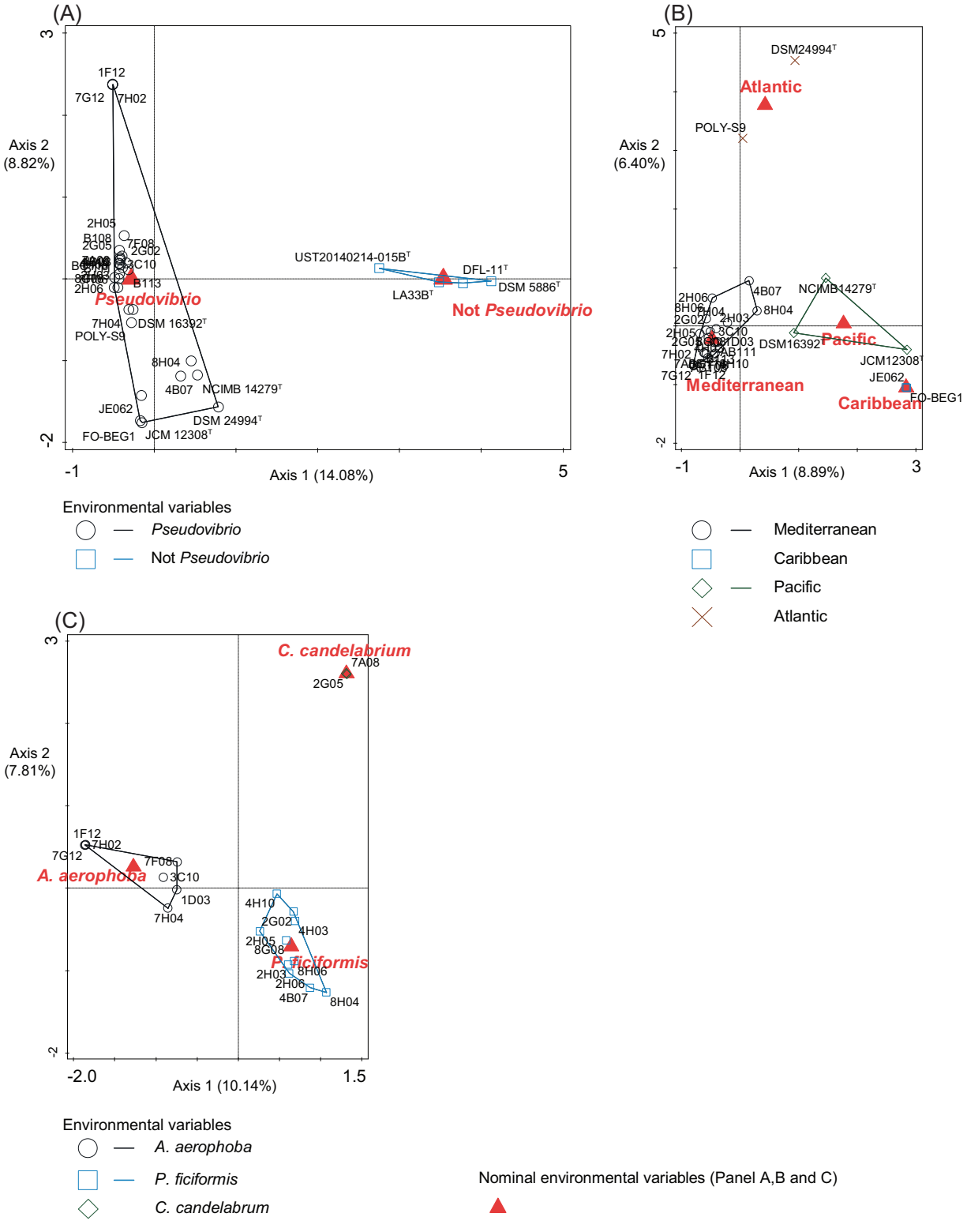

Fig. 3.—Clustering of strains by genus membership, geographic origin and sponge of origin. CCA was used to investigate which environmental variables could explain the variation in orthologous protein group composition among the genomes. All clustering shown in this figure is statistically significant ($P < 0.01$). Panel (*A*) shows results based on grouping of the genomes according to membership of the *Pseudovibrio* genus. In panel (*B*), the *Pseudovibrio* genomes are grouped according to their geographic origin. In panel (*C*), the *Pseudovibrio* strains that were isolated at the coast of Spain are grouped according to their sponge of origin.

Pseudovibrio species. Namely, the Atlantic Ocean cluster includes isolates belonging to *P. ascidiaceicola* (POLY-S9) and *P. axinellae* (DSM 24994^T), and the Pacific Ocean cluster includes *P. japonicus* NCIMB 14279^T, *P. denitrificans* JCM 12308^T, and *P. ascidiaceicola* DSM 16392^T. Most studies find that microbial composition and biogeography correlate in the aquatic environment, suggesting selective pressure of at least one environmental variable (Hanson et al. 2012). However, to what extent (trait) selection plays a role in the genetic variation among *Pseudovibrio* spp. by geography, as opposed to other suggested major processes such as drift, dispersal, and mutation, remains to be elucidated. This question can be further investigated by attempting to link variation in OGPs to environmental variables, and reinforces the need for proper recording of metadata for environmental samples and derived microbial isolates (DeLong 2009; Sun et al. 2011). We also found that *Pseudovibrio* strains isolated near the coast of Spain clustered by sponge of origin, that is, *A. aerophoba*, *C. candelabrum*, or *P. ficiformis* ($P = 0.006$) (fig. 3C). This might indicate that *Pseudovibrio* strains harbor functional genes that are associated with their ability to preferentially colonize certain sponge species. In turn, it might indicate that the *Pseudovibrio* strains selectively evolved in the sponge host after colonization. At least in some cases, sponge-associated bacteria are transmitted vertically (Sharp et al. 2007; Schmitt et al. 2008; Lee et al. 2009; Sipkema et al. 2015), and there is evidence that *Pseudovibrio* can be vertically transmitted via sponge larvae (Enticknap et al. 2006).

To identify differences in substrate utilization and product formation between the strains, the presence of metabolic pathways was investigated using the MetaCyc database as a reference. This database contains pathways and enzymes that are predominantly found in microorganisms and plants. We analyzed pathways that contained at least three enzymes with an EC number (supplementary fig. S1, Supplementary Material online) and pathways that contained at least four enzymes with an EC number (supplementary fig. S2, Supplementary Material online). In terms of metabolism, all members of the *Pseudovibrio* genus were highly similar, that is, across all genomes the same pathways were detected with differences between strains nearly always amounting to the presence and/or absence of single genes. *Pseudovibrio* spp. are metabolically highly versatile, which may enable them to survive in distinct habitats containing different substrates. Furthermore, *Pseudovibrio* sp. FO-BEG1 has been shown to sustain growth in ultra-oligotrophic seawater by simultaneously degrading different compounds (Schwedt et al. 2015). Although high metabolic versatility and large genome sizes are not considered typical for symbiotic bacteria (McCutcheon and Moran 2011; Dutta and Paul 2012), it could be that these characteristics benefit *Pseudovibrio* spp. in a variable and nutrient-rich environment such as in sponges. In contrast to members of the outgroups, all *Pseudovibrio* spp. were predicted to be able to synthesize Coenzyme F420. Coenzyme F420 is best known as an essential coenzyme of methanogenesis (Hendrickson and Leigh 2008). However, because *Pseudovibrio* spp. are incapable of methanogenesis, F420 might be involved in SMB (Mccormick and Morton 1982; Peschke et al. 1995) or other metabolic activities such as aerobic catabolism of 2,4,6-trinitrophenol or as an electron carrier to an F_{420}-dependent glucose-6-phosphate dehydrogenase (Purwantini et al. 1997; Ebert et al. 1999; Stover et al. 2000). In contrast to the three non-*Pseudovibrio* close relatives, all *Pseudovibrio* strains (including *P. hongkongensis* UST20140214-015B^T) were predicted to have the capacity to synthesize biotin (vitamin B_7). It has been suggested that sponges can benefit from the presence of bacteria that produce biotin, as they cannot synthesize this vitamin themselves (Webster and Thomas 2016).

High similarity between the *Pseudovibrio* spp. was also found with respect to the number of proteins assigned to COG functional categories (supplementary table S4, Supplementary Material online). The most pronounced difference between *Pseudovibrio* strains and the outliers was found for COG category "Mobilome: prophages, transposons," to which *Pseudovibrio* strains had on average twice the number of proteins assigned compared with the members of the outgroup.

Next, the differences regarding the presence of individual OGPs were investigated with SIMPER to investigate which functional capacities define our analyzed genomes when grouped according to geographic origin, sponge of origin, and membership of the *Pseudovibrio* genus (supplementary table S1, Supplementary Material online). Genes encoding enzymes that have urease activity (EC 3.5.1.5) were exclusively predicted in the outgroup members, the three *P. denitrificans* strains and *P. axinellae* DSM 24994^T. The low abundance of nitrogen may be growth limiting in the oceans (Antia et al. 1991), and regeneration of urea (i.e., by urease), which is a waste product of many animals, may alleviate this restraint. Following this hypothesis, bacteria in sessile marine animals could enhance their growth by using the urea excreted by their host as nitrogen source (Su et al. 2013). Alternatively, the bacteria could contribute to the nitrogen budget of the host by fixation of environmental urea (Wilkinson and Fay 1979). Given that the urease gene was detected in only a few (3 of 29) *Pseudovibrio* strains that reside in marine animals, we do not expect it to play an important role in the hosts–symbiont relationship, or this gene only plays a role in specific hosts. SIMPER analysis also showed that OGPs corresponding to enzymes involved in vitamin B_{12} (cobalamin) and vitamin B_1 (thiamin) biosynthesis were exclusively found in *Pseudovibrio* spp., but were absent in the outgroup members. Bacteria may be an essential contributor to the nutritional requirements of marine animals through production of these essential vitamins that the animals cannot synthesize themselves (Bondarev et al. 2013). However, we found that only a few enzymes of the multienzyme pathways

to produce vitamins B_1 and B_{12} were present in *Pseudovibrio* indicating that the vitamins cannot be de novo synthesized. Regarding vitamin B_1 biosynthesis, all *Pseudovibrio* strains lack a phosphotransferase (EC number: 2.7.4.7) and a thiazole tautomerase (5.3.99.10), both of which catalyse the final step to respectively produce pyrimidine and thiazole moieties that are required to form thiamine phosphate. Cobalamin cannot be synthesized de novo by *Pseudovibrio* spp. due to the absence of essential enzymes such as the adenosylcobinamide phosphate synthase (EC number: 6.3.1.10) and the adenosyl-cobric acid synthase (6.3.5.10). No further characteristic functional traits could be ascribed to *Pseudovibrio* genomes grouped by environmental variables because a large number of the idiosyncratic OGPs were hypothetical proteins or were proteins involved in routine processes.

Symbiosis

Eukaryotic-like Proteins, Virulence Factors, and Quorum Sensing

The interactions that are hypothesized to occur between *Pseudovibrio* spp. and the sponge host are still not well understood. Eukaryotic-like proteins (ELPs) were first predicted to play a role in symbiosis when they were discovered to be highly prevalent in sponge-associated bacteria (Liu et al. 2011; Siegl et al. 2011; Thomas et al. 2010; Fan et al. 2013). ARPs, which are ELPs, were predicted to interfere with phagocytosis of *Pseudovibrio* spp. by sponge host cells by retarding phagosome biogenesis, or by blocking fusion of the phagosome with the lysosome and its digestive enzymes (Nguyen et al. 2014). It has also been estimated that the average number of ARPs per sponge symbiont genome is between 10 and 40 when compared with an average of 2.5 ARPs in genomes of other symbiotic bacteria (Fan et al. 2012). Our analysis predicted ARPs in 23 of 25 sponge-associated *Pseudovibrio* strains, with an overall average of 1.57 ± 1.03 ARPs per genome (fig. 4 and supplementary table S5, Supplementary Material online), which does rank them among known symbiotic bacteria based on the number of ARPs. Proteins containing SEL1 repeats are also predominantly found in eukaryotes, and these proteins have been predicted to be symbiotic factors by which resident bacteria can interact with the host cells (Mittl and Schneider-Brachert 2007; Alex and Antunes 2015). Our analysis predicted on average 6.93 ± 1.51 proteins with SEL1 repeats in the *Pseudovibrio* genomes. In the genomes of the *Pseudovibrio* strains that were not associated with marine animals (i.e., those isolated from seawater), that is, *P. japonicus* NCIMB 14279^T and *P. denitrificans* JCM 12308^T, nine and eight proteins with SEL1 repeats were detected, respectively. ELPs with TRPs have been reported to participate in the delivery of virulence factors to the host cell (Cerveny et al. 2013). On average, we detected 34.27 ± 2.43 proteins with TRPs in the *Pseudovibrio* genomes. We also detected on average 6.27 ± 1.53 proteins with invasion associated locus B, which has been reported to be involved in host cell invasion (Coleman and Minnick 2003; Eicher and Dehio 2012). Furthermore, in each *Pseudovibrio* genome, except for *Pseudovibrio* sp. POLY-S9, one or two proteins were predicted that contain a TadE-like domain, suggesting that these might be involved in adherence of *Pseudovibrio* to host cells (Tomich et al. 2007). Based on membership of the above-mentioned symbiotic factors to OGPs, we can draw conclusions about their ancestry (supplementary table S5, Supplementary Material online). ARPs and proteins containing a TadE-like domain belonged to a few distinct OGPs that were each present in a subset of the *Pseudovibrio* strains, where no proteins containing a TadE-like domain and only one ARP were detected in the outgroup members. Therefore, we speculate that these proteins were either selectively maintained or acquired by *Pseudovibrio* spp.. A subset of the OGPs that encoded proteins with the SEL1 repeat (2/9), proteins with invasion associated locus B (2/10) and TRPs (15/59) were present in all genomes, including those of the outgroup members, indicating that these proteins were highly conserved. Subsequently, we investigated whether the LuxR/LuxI quorum sensing system was present in the *Pseudovibrio* genomes. The enzyme encoded by *lux*I catalyses the final step in the production of N-acyl homoserine lactone (AHL). AHL can bind to the regulatory protein LuxR, which together activate transcription of genes involved in AHL biosynthesis (auto-induction) as well as a wide array of other genes (Miller and Bassler 2001; Nasser and Reverchon 2007). AHL is secreted by the bacterium; therefore, as the population density increases, the concentration of AHL will increase, resulting in activation of gene expression across the bacterial community. In each *Pseudovibrio* genome, we detected more than nine copies of the *lux*R gene whereas in none of the genomes *lux*I was detected. An absence of *lux*I suggests that *Pseudovibrio* strains are dependent on quorum sensing molecules produced by neighbouring bacteria for expression of *lux*R/AHL-controlled genes, which has been previously hypothesized based on the genomes of *Pseudovibrio* sp. FO-BEG1 and *Pseudovibrio* sp. JE062 (Case et al. 2008; Bondarev et al. 2013). The genes regulated by quorum sensing may be involved in production of secondary metabolites (e.g., against microbial competitors or host pathogens) or biofilm formation (Dobretsov et al. 2009; Subramoni and Venturi 2009; Zan et al. 2012). In 29 of the 31 *Pseudovibrio* genomes analyzed here, we detected a cluster of genes putatively involved in biogenesis of proteinaceous extracellular fibres called curli. These extracellular curli are involved in biofilm formation, and as such may contribute to sponge colonization (Barnhart and Chapman 2006). Compared with the *Pseudovibrio* genomes, the genomes of the outgroup members encoded similar numbers of ELPs, LuxR regulators, proteins with invasion associated locus B and proteins containing the TadE-like domain. However, no outgroup member encoded proteins involved in curli fiber biosynthesis.

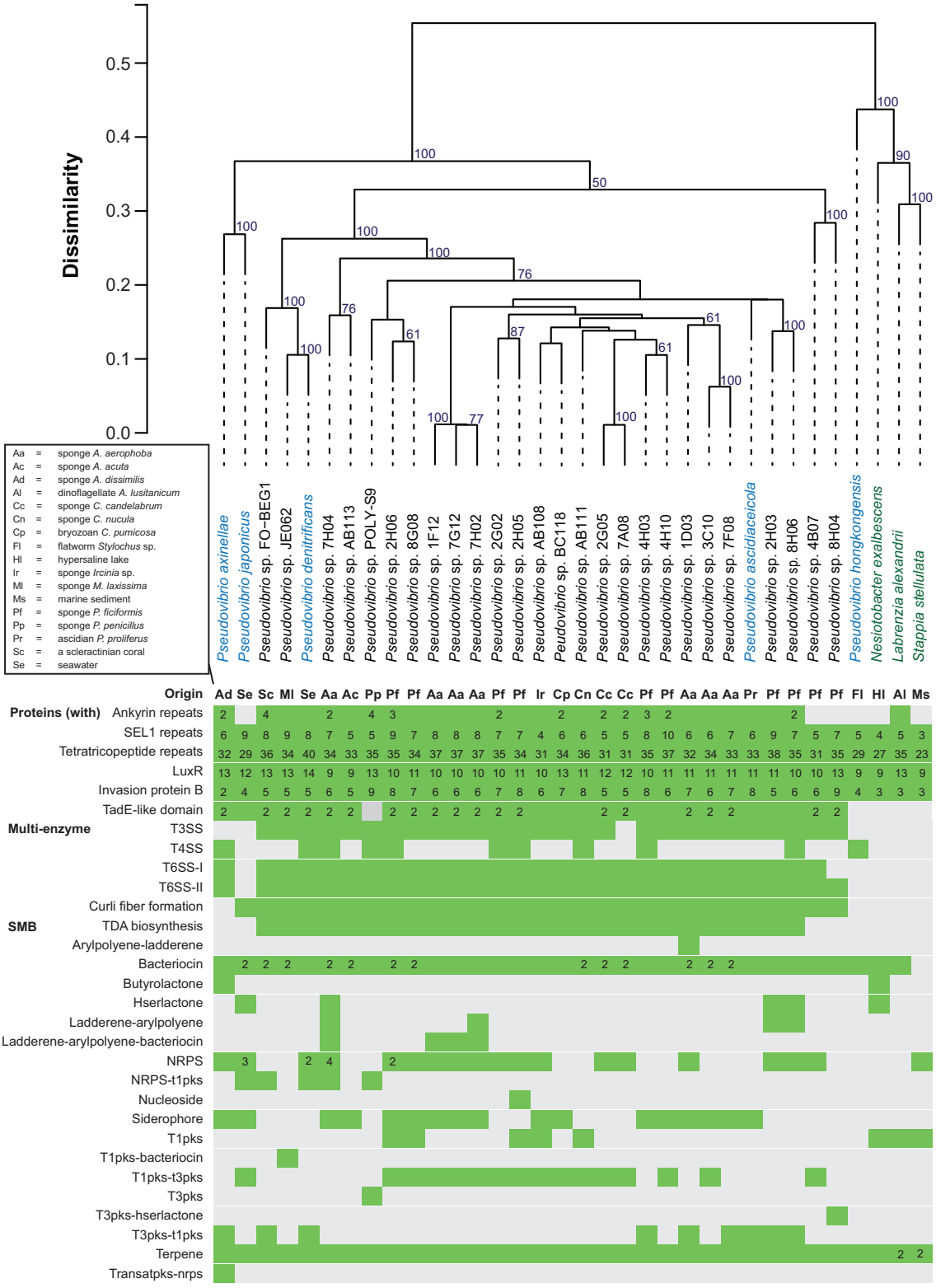

Fig. 4.—Symbioses-related proteins, secretions systems, and SMB clusters. Green squares indicate presence. Numerical values were included if functions were detected more than once. The tree on top was made by hierarchical linkage clustering based on the presence/absence of OGPs that are present in greater than two genomes. Type strains are in blue and the close relatives are in green. Hierarchical clustering was performed using 10,000 iterations of bootstrapping.

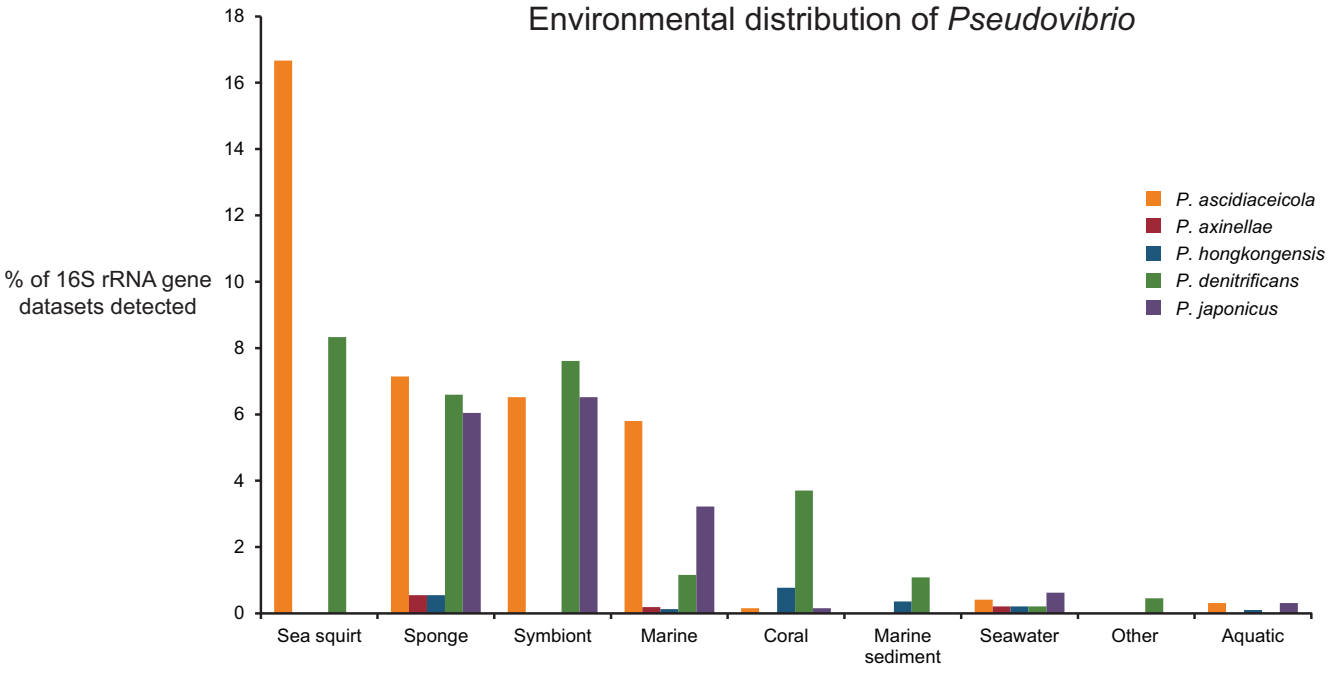

Fig. 5.—The presence of *Pseudovibrio* in 16S rRNA gene data sets from different niches. The percentage of 16S rRNA gene amplicon data sets in the SRA database in which the *Pseudovibrio* type strains were detected. The analysis was performed by the IMNGS platform, which divides the data sets in the SRA in 105 categories according to which biological niche was sampled. Full-length 16 S rRNA gene sequences of the type strains were used as query. The similarity threshold was set at 99% nucleotide identity. All environmental categories are shown in which at least one type strain was detected at > 0.1% relative abundance.

Secretion Systems

The distribution of T3SSs, T4SSs, and T6SSs was investigated (fig. 4) because these systems have been ascribed important roles in a range of symbiotic interactions. All three types of secretion systems allow bacteria to deliver effector molecules into a target cell via needle-like structures in the cellular membrane (Cornelis 2006; Coulthurst 2013; Low et al. 2014). Besides effector molecules, T4SSs can also translocate genetic material to the target cell. In all *Pseudovibrio* genomes, except the genome of *P. japonicus* NCIMB 14279^T, at least two of these secretions systems were detected. Interestingly, in the genome of *P. japonicus* NCIMB 14279^T, not a single of the aforementioned secretion systems was found indicating that the strain might lack equipment to form symbiotic relations. This notion is in line with the fact that *P. japonicus* NCIMB 14279^T was isolated from the open sea, and that members of this species were so far never isolated from marine animals. On the other hand, the notion is contradicted by *P. japonicus* spp. being detected more often in sponge 16S rRNA gene data sets when compared with seawater data sets (fig. 5). T6SS-I and/or T6SS-II clusters were detected in all *Pseudovibrio* genomes (except *P. japonicus* NCIMB 14279^T), and the genes in these clusters were all homologous and showed perfectly conserved synteny (supplementary table S6, Supplementary Material online). The fact that the clusters are conserved in *Pseudovibrio* strains from diverse geographical locations and species suggests that they are important for survival and propagation. T3SS clusters were limited to *Pseudovibrio* strains (27/28) belonging to the species *P. ascidiaceicola* and *P. denitrificans*, whereas T4SSs were detected in only 10 of all *Pseudovibrio* genomes. These results suggest *Pseudovibrio* spp. developed different strategies to carve a niche in a competitive environment. Romano et al. carried out a more detailed analysis of these clusters for *Pseudovibrio* sp. FO-BEG1 and *Pseudovibrio* sp. JE062, as well as for 10 other *Pseudovibrio* strains, by analyzing which effector molecules are putatively delivered into host cells (Romano et al. 2016). It was predicted that T3SSs were mainly used to reduce phagocytosis and to block the inflammatory response of the host, that T4SSs were mainly involved in DNA mobilization, and that T6SSs might be involved in evasion of the host immune response, or they could have an antagonistic role toward other bacteria. Only a single secretion system, namely the T4SS in *P. hongkongensis* UST20140214-015B^T, was detected in the outgroup members. The members of the outgroup (except *P. hongkongensis* UST20140214-015B^T) do not belong to genera that are implicated in forming symbiotic relations. In addition, none of the outgroup members, except *L. alexandrii* DFL-11^T that was washed from the dinoflagellate *A. lusitanicum* (Biebl et al. 2007), were isolated from a host. Hence, the absence of symbiosis-related gene clusters in their genomes (e.g., those encoding curli fibre formation and

secretion systems) likely indicates that these bacteria are not committed to forming symbiotic relations.

Secondary Metabolite Biosynthesis Clusters

Pseudovibrio spp. have received great interest as sources of novel bioactive compounds, and especially antimicrobials (Crowley et al. 2014). Not only are novel bioactive compounds interesting from a clinical perspective but they may also provide insights into possible host–symbiont relations. Sponges in part depend on the production of bioactive compounds by associated microorganisms for defence against predation and disease (Taylor et al. 2007; Webster and Taylor 2012). Recently, sponge-derived *Pseudovibrio* spp. were found to inhibit a *Bacillus* sp. with sponge-degrading activity that was isolated from the sponge genus *Tedania* (Esteves et al. 2017).

TDA is a compound that has antibacterial activity, and was found to be produced by *Pseudovibrio* sp. D323 (Penesyan et al. 2011). TDA has a strong inhibitory activity against a range of marine bacteria, and hence has been proposed to protect the host from unwanted microbial colonization (Harrington et al. 2014). It has been previously shown that *Pseudovibrio* sp. FO-BEG1 and *Pseudovibrio* sp. JE062 can produce TDA, which was linked to the presence of the *tdaA-tdaF* biosynthetic gene cluster in the respective genomes (Bondarev et al. 2013). Here we found the *tdaA-tdaF* biosynthetic cluster in 26 of 31 *Pseudovibrio* genomes suggesting that TDA production is an important feature of the genus *Pseudovibrio* (fig. 4 and supplementary table S5, Supplementary Material online). In all these cases, the TDA resistance-conferring genes *tdaR1*, *tdaR2*, and *tdaR3* (Wilson et al. 2016) were detected immediately downstream of the *tdaA-tdaF* operon. However, the *Pseudovibrio* strains that lacked TDA biosynthesis also lack TDA resistance, and as such these strains likely do not benefit, but rather are harmed, in the company of coresident TDA-producing *Pseudovibrio* spp.. Further potential for SMB was analyzed by application of the antiSMASH pipeline (Weber et al. 2015). Eighteen of 43 types of SMB clusters defined by antiSMASH were detected in the *Pseudovibrio* genomes, with bacteriocin, NRPS, siderophore, t1pks-t3pks, t3pks-t1pks, and terpene SMB clusters being detected in >5 genomes (fig. 4). At least one bacteriocin biosynthesis gene cluster was identified in all *Pseudovibrio* genomes, except for *P. axinellae* DSM 24994^T. The main role of bacteriocins, which are ribosomally synthesized antibiotic peptides, is to provide their producers with a competitive advantage by inhibiting bacterial growth (Riley and Wertz 2002; Balko 2012). Therefore, similar to TDA, bacteriocins could serve *Pseudovibrio* to establish a niche in the sponge host, and at the same time protect the host against pathogen invasion. So far, few bacteriocins have been identified in the marine environment (Desriac et al. 2010), and only one study has identified novel bacteriocins in a sponge (Phelan et al. 2013). Recently, based on genome data, bacteriocins were also predicted to be produced by members of the obligate marine genus *Pseudoalteromonas* (Bosi et al. 2017). Mutual comparison of bacteriocin biosynthesis gene clusters in *Pseudovibrio* and *Pseudoalteromonas* genomes predicted that they code for different bacteriocins (supplementary table S7, Supplementary Material online). AntiSMASH detected a terpene biosynthesis gene cluster in all *Pseudovibrio* genomes. Via mining of bacterial genomes, it has become clear that terpenes are not only regularly produced by plants and fungi but also by a large number of bacteria (Cane and Ikeda 2012). Terpenes may give *Pseudovibrio* a competitive advantage due to antimicrobial activity (Martin et al. 1970; Song et al. 2015). Alternatively, the fragrance of the terpenes might be an addition to sponge-derived secondary metabolites that are designed to repel predators (Proksch 1994; Epifanio et al. 1999; Ding et al. 2015). However, terpenes typically have limited water solubility. For the SMB clusters detected by antiSMASH in *Pseudovibrio* spp., no more than 26% of the genes in any cluster showed similarity to genes in known clusters. The diversity among SMB clusters from the same type was analyzed by investigating the total number of cluster genes that mutually shared ≥30% nucleotide identity (supplementary table S8, Supplementary Material online). Based on the number of clusters from the same type that share <50% of the genes at ≥30% nucleotide identity, we predict that NRPS, bacteriocin, terpene, and t1pks-t3pks clusters across all *Pseudovibrio* genomes each encode for at least three different products. Taken together, these findings highlight *Pseudovibrio* spp. as a largely untapped resource for the discovery of novel compounds with potentially clinical significance.

Antibiotic Resistance

Antibiotic resistance profiles were determined for 27 of 31 *Pseudovibrio* strains that were analyzed in this study (supplementary table S9, Supplementary Material online). All 27 strains were resistant to vancomycin, lincomycin, polymyxin B, and D-cycloserine; however, these resistance phenotypes could not be linked to the presence of known resistance genes. Therefore, resistance to these antibiotics is probably intrinsic. Twenty-one of 27 strains were resistant to tetracycline, which could not be linked to the presence of resistance genes in the genomes. All strains were either resistant or intermediately resistant to the β-lactam antibiotics ampicillin and penicillin, which could be explained by the class A and class B β-lactamases that were predicted to be encoded in their genomes (supplementary table S10, Supplementary Material online). The predicted class A β-lactamase encoded in 22 *P. ascidiaceicola* genomes was highly similar (≥95% amino acid identity) to $bla_{PSV\text{-}1}$ from *P. ascidiaceicola* strain 1D03 of which the resistance function has been experimentally verified by functional metagenomics (Versluis et al. 2016).

Within the *Pseudovibrio* genus, almost all class A β-lactamases clustered by species (supplementary fig. S3, Supplementary Material online). The exception was a distinct class A β-lactamase (8H04_2067) predicted in *Pseudovibrio* sp. 8H04 of which the closest experimentally verified β-lactamase was bla_{PSV-1}. However, as previously deduced, *Pseudovibrio* sp. 8H04 is expected to represent a distinct novel *Pseudovibrio* species as well. These results, together with the absence of associated mobilizing elements, suggest that the class A β-lactamases are not prone to horizontal gene transfer, and that diversity within this group of enzymes is predominately achieved by speciation. The β-lactamases in these bacteria could function as a defence against β-lactam antibiotics that are produced by other micro-organisms coinhabiting sessile marine animals. Still, these enzymes could also have other roles such as in disruption of cell signalling (Allen et al. 2009).

Environmental Distribution of *Pseudovibrio*

The 16S rRNA gene data sets in the Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra; last accessed April 14, 2016) were interrogated for the presence of the *Pseudovibrio* type strains using the IMNGS platform (Lagkouvardos et al. 2016) (fig. 5). IMNGS divides the 16S rRNA gene data sets in 105 categories according to sample origin. The highest percentage of data sets containing *Pseudovibrio* spp. was found for the category sea squirts (20.8%), followed by data sets acquired from sponges (13.7%) and symbionts (7.6%). *Pseudovibrio* spp. were only detected in 16S rRNA gene data sets from marine habitats or from generalized categories that can also contain data sets from marine habitats. This is consistent with the notion that *Pseudovibrio* spp. have been isolated exclusively from the marine environment and depend on seawater to thrive. The fact that *Pseudovibrio* spp. were detected more often in 16S rRNA gene data sets from sessile marine animals as opposed to aquatic environments suggests that indeed *Pseudovibrio* spp. form symbiotic relations. *Pseudovibrio ascidiaceicola*, *P. denitrificans*, and *P. japonicus* were each detected at >6% relative abundance in 16S rRNA gene data sets from at least two categories that represent distinct biological niches, whereas *P. axinellae* and *P. hongkongensis* were never detected at >1% relative abundance in any biological niche.

Conclusions

The analysis of 16S rRNA gene data sets revealed that *Pseudovibrio* spp. are predominantly found and may be specifically associated with sessile marine animals. Our analysis of a total of 31 *Pseudovibrio* genomes revealed that, except *P. hongkongensis* UST20140214-015B^T, members of the genus *Pseudovibrio* are highly similar in terms of metabolic capacity, with striking differences regarding the presence of specific SMB clusters. We identified various genomic elements that are highly conserved within the genus and are expected to give *Pseudovibrio* a competitive advantage as a symbiont of sponges or other marine animals. These conserved elements encompassed systems involved in immune evasion (T4SSs and T6SSs), SMB clusters with products that might inhibit growth of competitors (TDA, bacteriocin, and terpene biosynthesis clusters), and a range of other factors relevant for symbiotic relations such as ELPs and vitamin biosynthesis. The importance of these elements for *Pseudovibrio* survival is underscored by their conservation across *Pseudovibrio* spp. and across strains isolated from diverse geographic locations. Interestingly, based on gene composition, *Pseudovibrio* strains from the Mediterranean Sea clustered by sponge species of origin, suggesting host-specific colonization or adaptation. Taken together, the findings presented here support the hypothesis that *Pseudovibrio* spp. have evolved as symbionts of sponges and other marine invertebrates.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Bastian V.H.K. Hornung (Wageningen University & Research) for valuable input on the bioinformatics analyses. Furthermore, we thank Professor Ying Xu (Shenzhen University), Qi Li (The Hong Kong University of Science and Technology), and Yu Zhang (Shenzhen University), for proving the genome sequence data of *P. hongkongensis* UST20140214-015B^T. The authors are grateful to Dr. Vera Thiel and Dr. Herwig Heindl for sampling and isolation of the *Pseudovibrio* spp. strains AB108, AB111, AB113, and BC118.

Acknowledgements

This work was supported by the European Union through the EvoTAR project (grant agreement no. 282004) and by a grant of BE-Basic-FES funds from the Dutch Ministry of Economic Affairs (project 7.1.5).

Literature Cited

Alex A, Antunes A. 2015. Whole genome sequencing of the symbiont *Pseudovibrio* sp. from the intertidal marine sponge *Polymastia penicillus* revealed a gene repertoire for host-switching permissive lifestyle. Genome Biol Evol. 7(11):3022–3032.

Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J. 2009. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. ISME J. 3(2):243–251.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Andrews S. 2010. FASTQC: a quality control tool for high throughput sequence data. Retrieved from http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/.
Antia NJ, Harrison PJ, Oliveira L. 1991. The role of dissolved organic nitrogen in phytoplankton nutrition, cell biology and ecology. Phycologia 30(1):1–89.
Auch AF, von Jan M, Klenk HP, Goker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci. 2(1):117–134.
Ausubel F, et al. 2002. Short protocols in molecular biology. 5th ed. John Wiley & Sons, Inc., New York.
Balko AB. 2012. Characteristic, properties, prospect of application of bacteriocins. Mikrobiol Z. 74(6):99–106.
Barnhart MM, Chapman MR. 2006. Curli biogenesis and function. Annu Rev Microbiol. 60:131–147.
Biebl H, et al. 2007. Description of *Labrenzia alexandrii* gen. nov., sp. nov., a novel alphaproteobacterium containing bacteriochlorophyll a, and a proposal for reclassification of *Stappia aggregata* as *Labrenzia aggregata* comb. nov., of *Stappia marina* as *Labrenzia marina* comb. nov. and of *Stappia alba* as *Labrenzia alba* comb. nov., and emended descriptions of the genera *Pannonibacter*, *Stappia* and *Roseibium*, and of the species *Roseibium denhamense* and *Roseibium hamelinense*. Int J Syst Evol Microbiol. 57(5):1095–1107.
Bondarev V, et al. 2013. The genus *Pseudovibrio* contains metabolically versatile bacteria adapted for symbiosis. Environ Microbiol. 15(7):2095–2113.
Bosi E, et al. 2017. The pangenome of (Antarctic) *Pseudoalteromonas* bacteria: evolutionary and functional insights. BMC Genomics 18(1):93.
Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12(1):59–60.
Cane DE, Ikeda H. 2012. Exploration and mining of the bacterial terpenome. Acc Chem Res. 45(3):463–472.
Case RJ, Labbate M, Kjelleberg S. 2008. AHL-driven quorum-sensing circuits: their frequency and function among the *Proteobacteria*. ISME J. 2(4):345–349.
Caspi R, et al. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 44(D1):D471–D480.
Cerveny L, et al. 2013. Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. Infect Immun. 81(3):629–635.
Chen YH, et al. 2012. Isolation of marine bacteria with antimicrobial activities from cultured and field-collected soft corals. World J Microbiol Biotechnol. 28(12):3269–3279.
Clarke K, Gorley R. 2006. PRIMER v6: user manual/tutorial. PRIMER-E.
Claudel-Renard C, Chevalet C, Faraut T, Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res. 31(22):6633–6639.
Coil D, Jospin G, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31(4):587–589.
Coleman SA, Minnick MF. 2003. Differential expression of the invasion-associated locus B (ialB) gene of *Bartonella bacilliformis* in response to environmental cues. Microb Pathog. 34(4):179–186.
Cornelis GR. 2006. The type III secretion injectisome. Nat Rev Microbiol. 4(11):811–825.
Coulthurst SJ. 2013. The Type VI secretion system—a widespread and versatile cell targeting system. Res Microbiol. 164(6):640–654.
Crowley SP, O'Gara F, O'Sullivan O, Cotter PD, Dobson AD. 2014. Marine *Pseudovibrio* sp. as a novel source of antimicrobials. Mar Drugs 12(12):5916–5929.
DeLong EF. 2009. The microbial ocean from genomes to biomes. Nature 459(7244):200–206.
Desriac F, et al. 2010. Bacteriocin as weapons in the marine animal-associated bacteria warfare: inventory and potential applications as an aquaculture probiotic. Mar Drugs 8(4):1153–1177.
Ding L, et al. 2015. Bacaryolanes A-C, rare bacterial caryolanes from a mangrove endophyte. J Nat Prod. 78(12):2963–2967.
Dobretsov S, Teplitski M, Paul V. 2009. Mini-review: quorum sensing in the marine environment and its relationship to biofouling. Biofouling 25(5):413–427.
Donachie SP, Bowman JP, Alam M. 2006. *Nesiotobacter exalbescens* gen. nov., sp. nov., a moderately thermophilic alphaproteobacterium from an Hawaiian hypersaline lake. Int J Syst Evol Microbiol. 56(Pt 3):563–567.
Dutta C, Paul S. 2012. Microbial lifestyle and genome signatures. Curr Genomics 13(2):153–162.
Ebert S, Rieger PG, Knackmuss HJ. 1999. Function of coenzyme F420 in aerobic catabolism of 2, 4, 6-trinitrophenol and 2, 4-dinitrophenol by *Nocardioides simplex* FJ2-1A. J Bacteriol. 181(9):2669–2674.
Eicher SC, Dehio C. 2012. *Bartonella* entry mechanisms into mammalian host cells. Cell Microbiol. 14(8):1166–1173.
Ekseth OK, Kuiper M, Mironov V. 2014. orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics 30(5): 734–736.
Enticknap JJ, Kelly M, Peraud O, Hill RT. 2006. Characterization of a culturable alphaproteobacterial symbiont common to many marine sponges and evidence for vertical transmission via sponge larvae. Appl Environ Microbiol. 72(5):3724–3732.
Epifanio RDA, Gabriel R, Martins DL, Muricy G. 1999. The sesterterpene variabilin as a fish-predation deterrent in the Western Atlantic sponge *Ircinia strobilina*. J Chem Ecol. 25:2247–2254.
Esteves AIS, Cullen A, Thomas T. 2017. Competitive interactions between sponge-associated bacteria. FEMS Microbiol Ecol. 93:fix008.
Fan L, Liu M, Simister R, Webster NS, Thomas T. 2013. Marine microbial symbiosis heats up: the phylogenetic and functional response of a sponge holobiont to thermal stress. ISME J. 7(5):991–1002.
Fan L, et al. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proc Natl Acad Sci U S A. 109(27):E1878–E1887.
Fiebig A, et al. 2013. Genome of the R-body producing marine alphaproteobacterium *Labrenzia alexandrii* type strain (DFL-11(T)). Stand Genomic Sci. 7(3):413–426.
Fiore CL, Jarett JK, Olson ND, Lesser MP. 2010. Nitrogen fixation and nitrogen transformations in marine symbioses. Trends Microbiol. 18(10):455–463.
Flemer B, et al. 2012. Diversity and antimicrobial activities of microbes from two Irish marine sponges, *Suberites carnosus* and *Leucosolenia* sp. J Appl Microbiol. 112(2):289–301.
Fukunaga Y, et al. 2006. *Pseudovibrio ascidiaceicola* sp. nov., isolated from ascidians (sea squirts). Int J Syst Evol Microbiol. 56(Pt 2): 343–347.
Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43(Database issue):D261–D269.
Geng HF, Belas R. 2010. Expression of tropodithietic acid biosynthesis is controlled by a novel autoinducer. J Bacteriol. 192(17):4377–4387.
Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 9(1):207–216.
Goris J, et al. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 57(Pt 1):81–91.
Graca AP, et al. 2013. Antimicrobial activity of heterotrophic bacterial communities from the marine sponge *Erylus discophorus* (*Astrophorida*, *Geodiidae*). PLoS One 8:e78992.

Han MQ, Li ZY, Zhang FL. 2013. The ammonia oxidizing and denitrifying prokaryotes associated with sponges from different sea areas. Microbial Ecol. 66(2):427–436.
Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol. 10:497–506.
Harrington C, et al. 2014. Characterisation of non-autoinducing tropodithietic Acid (TDA) production from marine sponge *Pseudovibrio* species. Mar Drugs 12(12):5960–5978.
Heindl H, Wiese J, Thiel V, Imhoff JF. 2010. Phylogenetic diversity and antimicrobial activities of bryozoan-associated bacteria isolated from Mediterranean and Baltic Sea habitats. Syst Appl Microbiol. 33(2):94–104.
Hendrickson EL, Leigh JA. 2008. Roles of coenzyme F420-reducing hydrogenases and hydrogen- and F420-dependent methylenetetrahydromethanopterin dehydrogenases in reduction of F420 and production of hydrogen during methanogenesis. J Bacteriol. 190(14):4818–4821.
Hosoya S, Yokota A. 2007. *Pseudovibrio japonicus* sp. nov., isolated from coastal seawater in Japan. Int J Syst Evol Microbiol. 57(Pt 9): 1952–1955.
Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
Karpenahalli MR, Lupas AN, Soding J. 2007. TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. BMC Bioinformatics 8:2.
Kennedy J, et al. 2009. Isolation and analysis of bacteria with antimicrobial activities from the marine sponge *Haliclona simulans* collected from Irish waters. Mar Biotechnol (NY) 11(3):384–396.
Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 64(Pt 2):1825–1825.
Koehorst JJ, Saccenti E, Schaap PJ, Martins dos Santos VAP, Suarez-Diez M. 2016a. Protein domain architectures provide a fast, efficient and scalable alternative to sequence-based methods for comparative functional genomics. F1000Research 5:1987.
Koehorst JJ, van Dam JCJ, van Heck RGA, et al. 2016b. Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. Sci Rep. 6:38699.
Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 102(7):2567–2572.
Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305(3):567–580.
Lafi FF, Garson MJ, Fuerst JA. 2005. Culturable bacterial symbionts isolated from two distinct sponge species (*Pseudoceratina clavata* and *Rhabdastrella globostellata*) from the Great Barrier Reef display similar phylogenetic diversity. Microb Ecol. 50(2):213–220.
Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35(9):3100–3108.
Lagkouvardos I, et al. 2016. IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci Rep. 6:33721.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359.
Lee OO, Chui PY, Wong YH, Pawlik JR, Qian PY. 2009. Evidence for vertical transmission of bacterial symbionts from adult to embryo in the Caribbean sponge *Svenzea zeai*. Appl Environ Microbiol. 75(19):6147–6156.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Liu MY, Kjelleberg S, Thomas T. 2011. Functional genomic analysis of an uncultured delta-proteobacterium in the sponge *Cymbastela concentrica*. ISME J. 5(3):427–435.
Low HH, et al. 2014. Structure of a type IV secretion system. Nature 508(7497):550–553.
Ludwig W, et al. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32:1363–1371.
Martin DG, Slomp G, Mizsak S, Duchamp DJ, Chidester CG. 1970. The structure and absolute configuration of pentalenolactone (PA 132). Tetrahedron Lett. 11(56):4901–4904.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1):10–12.
McArthur AG, et al. 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 57(7):3348–3357.
Mccormick JRD, Morton GO. 1982. Identity of co-synthetic factor 1 of *Streptomyces Aureofaciens* and fragment Fo from co-enzyme F420 of *Methanobacterium* species. J Am Chem Soc. 104(14): 4014–4015.
McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 10(1):13–26.
Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol Biol Evol. 30(5):1218–1223.
Menezes CBA, et al. 2010. Microbial diversity associated with algae, ascidians and sponges from the north coast of Sao Paulo state, Brazil. Microbiol Res. 165(6):466–482.
Miller MB, Bassler BL. 2001. Quorum sensing in bacteria. Annu Rev Microbiol. 55:165–199.
Mitchell A, et al. 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 43:D213–D221.
Mittl PR, Schneider-Brachert W. 2007. Sel1-like repeat proteins in signal transduction. Cell Signal. 19(1):20–31.
Muscholl-Silberhorn A, Thiel V, Imhoff JF. 2008. Abundance and bioactivity of cultured sponge-associated bacteria from the Mediterranean sea. Microb Ecol. 55(1):94–106.
Nasser W, Reverchon S. 2007. New insights into the regulatory mechanisms of the LuxR family of quorum sensing regulators. Anal Bioanal Chem. 387(2):381–390.
Nguyen MTHD, Liu M, Thomas T. 2014. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Mol Ecol. 23(6):1635–1645.
O'Halloran JA, et al. 2013. *Pseudovibrio axinellae* sp. nov., isolated from an Irish marine sponge. Int J Syst Evol Microbiol. 63:141–145.
O'Halloran JA, et al. 2011. Diversity and antimicrobial activity of *Pseudovibrio* spp. from Irish marine sponges. J Appl Microbiol. 110:1495–1508.
Penesyan A, et al. 2011. Identification of the antibacterial compound produced by the marine epiphytic bacterium *Pseudovibrio* sp. D323 and related sponge-associated bacteria. Mar Drugs 9(8): 1391–1402.
Peschke U, Schmidt H, Zhang HZ, Piepersberg W. 1995. Molecular characterization of the lincomycin-production gene-cluster of *Streptomyces Lincolnensis*-78-11. Mol Microbiol. 16(6):1137–1156.
Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786.
Phelan RW, et al. 2013. Subtilomycin: a new lantibiotic from *Bacillus subtilis* strain MMA7 isolated from the marine sponge *Haliclona simulans*. Mar Drugs 11(6):1878–1898.
Proksch P. 1994. Defensive roles for secondary metabolites from marine sponges and sponge-feeding nudibranchs. Toxicon 32(6): 639–655.

Purwantini E, Gillis TP, Daniels L. 1997. Presence of F420-dependent glucose-6-phosphate dehydrogenase in *Mycobacterium* and *Nocardia* species, but absence from *Streptomyces* and *Corynebacterium* species and methanogenic Archaea. FEMS Microbiol Lett. 146(1):129–134.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842.
R development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http//www.R-project.org.
Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 106(45):19126–19131.
Riesenfeld CS, Murray AE, Baker BJ. 2008. Characterization of the microbial community and polyketide biosynthetic potential in the palmerolide-producing tunicate *Synoicum adareanum*. J Nat Prod. 71(11):1812–1818.
Riley MA, Wertz JE. 2002. Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol. 56:117–137.
Romano S, et al. 2016. Comparative genomic analysis reveals a diverse repertoire of genes involved in Prokaryote-Eukaryote interactions within the *Pseudovibrio* genus. Front Microbiol. 7:387.
Ruger HJ, Hofle MG. 1992. Marine star-shaped-aggregate-forming bacteria: *Agrobacterium atlanticum* sp. nov., *Agrobacterium meteori* sp. nov., *Agrobacterium ferrugineum* sp. nov., nom. rev., *Agrobacterium gelatinovorum* sp. nov., nom. rev., and *Agrobacterium stellulatum* sp. nov., nom. rev. Int J Syst Bacteriol. 42(1):133–143.
Santos-Gandelman JF, et al. 2013. Characterization of cultivable bacteria from Brazilian sponges. Mar Biotechnol. 15(6):668–676.
Santos OCS, et al. 2010. Isolation, characterization and phylogeny of sponge-associated bacteria with antimicrobial activities from Brazil. Res Microbiol. 161(7):604–612.
Schmitt S, Angermeier H, Schiller R, Lindquist N, Hentschel U. 2008. Molecular microbial diversity survey of sponge reproductive stages and mechanistic insights into vertical transmission of microbial symbionts. Appl Environ Microbiol. 74(24):7694–7708.
Schneemann I, Wiese J, Kunz AL, Imhoff JF. 2011. Genetic approach for the fast discovery of phenazine producing bacteria. Mar Drugs 9(5):772–789.
Schwedt A, et al. 2015. Substrate use of *Pseudovibrio* sp. growing in ultra-oligotrophic seawater. PLoS One 10(3):e0121675.
Sertan-de Guzman AA, et al. 2007. *Pseudovibrio denitrificans* strain Z143-1, a heptylprodigiosin-producing bacterium isolated from a Philippine tunicate. FEMS Microbiol Lett. 277(2):188–196.
Sharp KH, Eam B, Faulkner DJ, Haygood MG. 2007. Vertical transmission of diverse microbes in the tropical sponge *Corticium* sp. Appl Environ Microbiol. 73(2):622–629.
Shieh WY, Lin YT, Jean WD. 2004. *Pseudovibrio denitrificans* gen. nov., sp. nov., a marine, facultatively anaerobic, fermentative bacterium capable of denitrification. Int J Syst Evol Microbiol. 54(Pt 6): 2307–2312.
Siegl A, et al. 2011. Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. ISME J. 5(1):61–70.
Sipkema D, et al. 2015. Similar sponge-associated bacteria can be acquired via both vertical and horizontal transmission. Environ Microbiol. 17(10):3807–3821.
Šmilauer P, Lepš J. 2014. Multivariate analysis of ecological data using Canoco 5. Cambridge, UK: Cambridge University Press.
Song C, et al. 2015. Exploring the genomic traits of fungus-feeding bacterial genus *Collimonas*. BMC Genomics 16:1103.
Stover CK, et al. 2000. A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. Nature 405(6789): 962–966.
Su J, et al. 2013. Phylogenetically diverse *ureC* genes and their expression suggest the urea utilization by bacterial symbionts in marine sponge *Xestospongia testudinaria*. PLoS One 8(5):e64848.
Subramoni S, Venturi V. 2009. LuxR-family 'solos': bachelor sensors/regulators of signalling molecules. Microbiology 155(Pt 5):1377–1385.
Sun SL, et al. 2011. Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. Nucleic Acids Res. 39(Database):D546–D551.
Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22(12): 1540–1542.
Sweet M, Bulling M, Cerrano C. 2015. A novel sponge disease caused by a consortium of micro-organisms. Coral Reefs 34(3):871–883.
Taylor MW, Radax R, Steger D, Wagner M. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. Microbiol Mol Biol Res. 71(2):295–347.
Thiel V, Imhoff JF. 2003. Phylogenetic identification of bacteria with antimicrobial activities isolated from Mediterranean sponges. Biomol Eng. 20(4-6):421–423.
Thomas T, et al. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. ISME J. 4(12):1557–1567.
Tindall BJ, Rosselló-Móra R, Busse H-J, Ludwig W, Kämpfer P. 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol. 60(Pt 1):249–266.
Tomich M, Planet PJ, Figurski DH. 2007. The tad locus: postcards from the widespread colonization island. Nat Rev Microbiol. 5(5): 363–375.
van Dam JC, Koehorst JJ, Schaap PJ, Martins Dos Santos VA, Suarez-Diez M. 2015. RDF2Graph a tool to recover, understand and validate the ontology of an RDF resource. J Biomed Semantics 6:39.
van Dongen S. 2000. Graph clustering by flow simulation [PhD thesis]. The Netherlands: Utrecht University.
Versluis D, McPherson K, van Passel MWJ, Smidt H, Sipkema D. 2017. Recovery of previously uncultured bacterial genera from three Mediterranean sponges. Mar Biotechnol (NY) 19(5): 454–468.
Versluis D, et al. 2016. Sponge microbiota are a reservoir of functional antibiotic resistance genes. Front Microbiol. 7:1848.
Vizcaino MI. 2011. The chemical defense of *Pseudopterogorgia americana*: a focus on the antimicrobial potential of a *Pseudovibrio* sp. [PhD thesis]. Charleston, United States: Medical University of South Carolina.
Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.
Weber T, et al. 2015. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43(W1):W237–W243.
Webster NS, Hill RT. 2001. The culturable microbial community of the Great Barrier Reef sponge *Rhopaloeides odorabile* is dominated by an alpha-Proteobacterium. Mar Biol. 138(4):843–851.
Webster NS, Negri AP, Webb RI, Hill RT. 2002. A spongin-boring alpha-proteobacterium is the etiological agent of disease in the Great Barrier Reef sponge *Rhopaloeides odorabile*. Mar Ecol Prog Ser. 232:305–309.
Webster NS, Taylor MW. 2012. Marine sponges and their microbial symbionts: love and other relationships. Environ Microbiol. 14(2): 335–346.
Webster NS, Thomas T. 2016. The sponge hologenome. Mbio 7(2):e00135–e00116.
Wilkinson CR, Fay P. 1979. Nitrogen-fixation in coral-reef sponges with symbiotic *Cyanobacteria*. Nature 279(5713):527–529.

Wilson MZ, Wang R, Gitai Z, Seyedsayamdost MR. 2016. Mode of action and resistance studies unveil new roles for tropodithietic acid as an anticancer agent and the gamma-glutamyl cycle as a proton sink. Proc Natl Acad Sci U S A. 113(6):1630–1635.

Xu Y, Li Q, Tian RM, Lai QL, Zhang Y. 2015. *Pseudovibrio hongkongensis* sp. nov., isolated from a marine flatworm. Anton Leeuw Int J G. 108(1):127–132.

Yilmaz P, et al. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 42(D1): D643–D648.

Zan J, et al. 2012. A complex LuxR-LuxI type quorum sensing network in a roseobacterial marine sponge symbiont activates flagellar motility and inhibits biofilm formation. Mol Microbiol. 85(5): 916–933.

Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18(5):821–829.

Zhang X, et al. 2013. Physiological characterization of aerobic culturable bacteria in the intestine of the sea cucumber *Apostichopus japonicus*. J Gen Appl Microbiol. 59(1):1–10.

Zhang Y, Li Q, Tian RM, Lai QL, Xu Y. 2016. *Pseudovibrio stylochi* sp. nov., isolated from a marine flatworm. Int J Syst Evol Microbiol. 66(5):2025–2029.

Associate editor: Rebecca Zufall