Title

Isolation and genome sequencing of 14 *Spongia* sp. bacterial associates expands the taxonomic and functional breadth of the cultivatable marine sponge microbiome

Authors

Elham Karimi[a,1*], Rodrigo Costa[a,b,c]

Affiliations

[a)] Center of Marine Sciences-CCMAR, Faculty of Science and Technology, Campus of Gambelas, University of Algarve, 8005-139 Faro, Portugal

[b)] Institute for Bioengineering and Biosciences (IBB), Instituto Superior Tecnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal

[c)] Department of Energy, Joint Genome Institute, Walnut Creek, California, USA, and Lawrence Berkeley National Laboratory, Berkeley, California, USA

*** Corresponding author:** Elham Karimi, Sorbonne Universités/CNRS, Station Biologique de Roscoff, UMR 8227, Integrative Biology of Marine Models, CS 90074, Roscoff, France
Email: ekarimi@sb-roscoff.fr

Abstract

Marine sponges live with complex microbial consortia, which have been considered as potential sources of novel natural products. However, the usual recalcitrance of host-associated microorganisms to cultivation makes studying sponge symbionts challenging. To tackle this complexity, exploration of cultivated sponge-associated bacteria and their coding potential is unavoidable. In this study, we isolate and report the draft genome sequences of 14 bacterial strains from the marine sponge *Spongia* sp. using R2A and VXA media. The strains belong to the classes *Actinobacteria*, *Gammaproteobacteria*, *Alphaproteobacteria*, and *Cytophagia* spanning 11 formally described genera plus two potentially novel genera in the *Rhodobacteraceae* family and one potentially novel family in the *Cytophagales* order. Functional genomics revealed presumed symbiosis factors typical of specific taxonomic groups (i.e. taurine metabolism genes among the *Alphaproteobacteria*, chitinase encoding genes and eukaryotic-like proteins in the *Cytophagia* genome) while multidrug efflux pumps,

[1] **Current affiliation:** Sorbonne Universités/CNRS, Station Biologique de Roscoff, UMR 8227, Integrative Biology of Marine Models, CS 90074, Roscoff, France

also important in host-microbe interactions, were common across all genomes. Moreover, we detected 76 secondary-metabolite biosynthetic gene clusters putatively involved in the production of drug-like compounds or signalling molecules across all genomes, warranting future biotechnologically driven research into their coding potential.

Keywords

Porifera, genomics, host-microbe interactions, symbiosis, marine microbiology, secondary metabolite

1. Introduction

Marine sponges possess a complex and distinct microbial consortium considered to be involved in the provision of nutrients, chemical defence and immunity to their hosts [1]. Although marine sponges have been the source of novel bacterial species [1, 2], yet we lack clear understanding of the diversity, genome structure and metabolic potential of most sponge associates. Apart from the ecological importance and roles of sponges and its microbial community in ecosystem functioning [3], it is now known that the marine sponge microbiome ranks as one of the most prolific sources of bioactive secondary metabolites in the seas [4, 5]. Isolated bacteria permit the full study of the physiology and function of microorganisms [6]. To date, limited information is available for understanding the physiology and function of sponge symbionts and their natural product biosynthesis potential. Although recent studies approaching the coding potential of metagenome-assembled genomes (MAGs) have provided useful insights into symbiosis factors that govern sponge-prokaryote interactions [7, 8], examination of cultivated sponge associated bacteria and their genomes is still much required to address the abovementioned knowledge gaps, especially in regards with the laboratory validation of the observed genomic traits.

In the present study, we describe the general genomic features and provide brief functional information of 14 bacterial strains isolated from the marine sponge *Spongia* sp. in the northeast Atlantic. This study releases new data for the further exploration of complex sponge-microbe interactions and promotes comparative genomics studies of symbiotic microorganisms by adding up novel genome resources to public databases, including original genomes of three novel bacterial lineages unclassifiable at the genus level, the first genome sequence for the genus *Lentilitoribacter* and further genome assemblies of thus-far poorly-studied genera such as *Agarivorans*, *Sphingorhabdus* and *Tropicibacter*.

2. Data description

2.1. Sampling, Isolation, and Genome sequencing

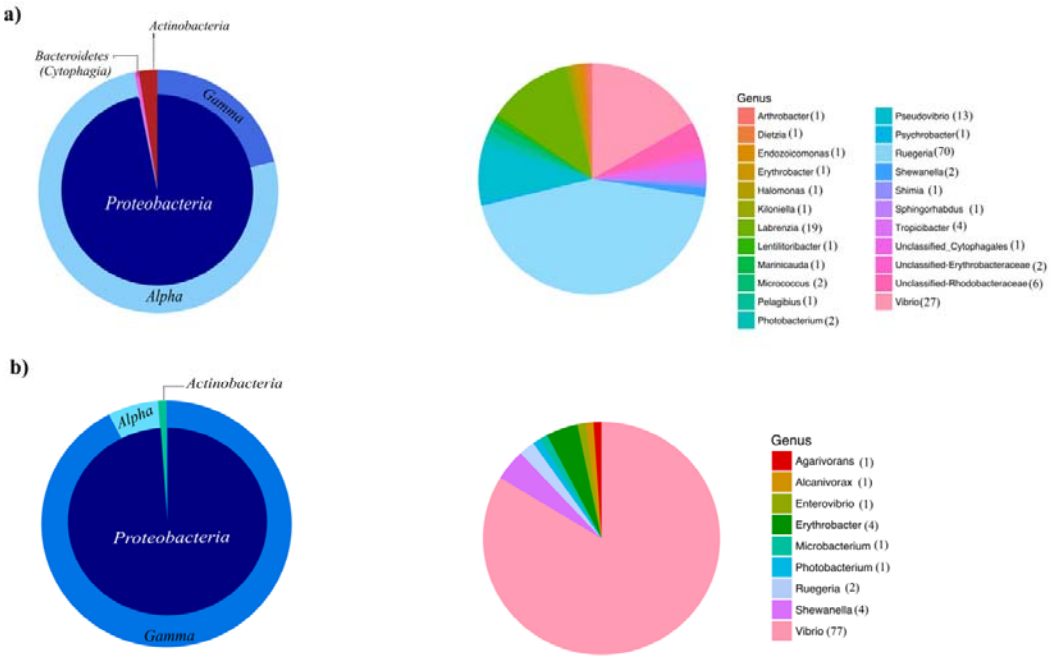
In April 2015, *Spongia* sp. specimens (4 biological replicates) were collected by SCUBA diving at 20 m open water stations in Algarve, Faro (Portugal). In the lab, microbial cell suspensions (MCSs) were prepared from sponge samples through grinding and homogenization in artificial seawater as described previously [9]. MCSs were then serially diluted (10^{-3} to 10^{-8}) and 100 µL of each dilution were spread plated in triplicates on 1/10R2A [10] and VXA [11] media. Plates were incubated at 18°C for up to 8 weeks. Bacterial growth was monitored weekly and colony forming units (CFUs) were counted. Colonies were selected randomly considering morphologically different ones as well, and pure cultures obtained by streaking colonies on the same isolation media.

In total, 252 colonies (160 from 1/10-R2A and 92 from VXA, Table S1) were selected. Genomic DNA was extracted from cell pellets derived from freshly grown cultures using the Wizard® Genomic DNA Purification Kit (Promega, Madison, USA) according to the manufacturer's instructions. Genomic DNA samples of all isolates were then subjected to 16S rRNA gene amplification by PCR followed by Sanger sequencing according to established procedures [12]. Taxonomic assignment of bacterial isolates to the genus level was performed using the classifier tool of the ribosomal database project (RDP, release 11, [13]. Identification of closest matches to all sequence queries were also obtained using the BLAST algorithm (February 2018, Table S1).

Alphaproteobacteria and *Gammaproteobacteria* strains were frequently isolated from R2A and VXA, respectively (Figure 1). Based on the taxonomic assignments obtained with RDP (Table S2), 14 strains originating from both media were chosen for genome analysis- which in most cases would likely correspond to novel genus/species from sponges (Table S1&S2). These were classified into 11 formally described and 3 potentially novel genera from *Actinobacteria* (3 genera), *Bacteroidetes* (1 putatively new genus and family), and *Proteobacteria* (10 genera from Alpha - 6 strains, with two putatively new genera- and from Gamma - 4 strains - classes) phyla (Table 1).

Afterwards, Illumina MiSeq sequencing was carried out at MR DNA (CA, United States). Paired-end libraries (2×301bp) were accomplished and *de novo* assembly was performed with the NGen DNA assembly software by DNAStar, Inc. as described previously [14]. Additionally, CheckM v.1.0.18 was used to assess the completeness and contamination of the generated genomes [15]. Genome annotations were conducted using RAST [16], also using

97 COG database via WebMGA [17]. Furthermore, the genomes were analyzed for the presence
98 of gene clusters indicative of secondary metabolism using antiSMASH version 5.0.1 [18].
99 Default parameters were used for all software.



100
101 ***Figure 1.*** Taxonomic distribution of bacteria retrieved from *Spongia* sp. on **a)** 1/10 R2A **b)** VXA media.
102 The number in brackets indicate the number of isolates obtained for each genus.
103
104 ## 2.2. Basic Genome Features

105 The genomic feature information of all strains is outlined in Table 1. All genomes possessed
106 > 99% completeness (except three genomes) and <1% contamination, suggesting high data
107 reliability for further comparative genomic studies, for example.

108 Three genera including *Arthrobacter* sp. Alg241-R88, *Dietzia* sp. Alg238-R159, *Micrococcus*
109 sp. Alg238-R198 isolated on 1/10R2A were sequenced from phylum *Actinobacteria*. The size
110 of the genomes ranged from 2.59 to 4.37 Mbp with GC content ranging from 64.8 to 72.8 %.
111 Gene annotation using RAST identified 4040, 3629, and 2375 protein-coding sequences
112 (CDs) for strains Alg241-R88, Alg238-R159, and Alg238-R198, respectively. From COG
113 annotations, the genes for glycosyltransferase, acetyltransferases, ABC-type multidrug
114 transport system were assigned for these three strains (Table S3).

Besides, a novel genome from the phylum *Bacteroides* was sequenced. The genome of *Cytophagales* bacterium Alg240-R148 is 5.3 Mbp in length with an overall GC content of 38.8%. Gene annotation using RAST identified a total of 4797 coding sequences, 38 tRNAs, and 6 rRNAs. This genome is 100% complete based on CheckM estimations. The COG annotation indicated genes involved in ABC-type antimicrobial peptide transport system, ABC-type multidrug transport system, chitinases, glycosidases and tetratricopeptide repeats (Table S3).

Moreover, six strains of the class *Alphaptoteobacteria* belonging to the families *Phyllobacteriaceae, Rhodobacteraceae,* and *Sphingomonadaceae* were sequenced. The size of the genomes ranged from 2.5 to 4.78 Mbp with GC content ranging from 44.3 to 58.8%. RAST annotation indicated a range of 2446 coding sequences for *Sphingorhabdus* sp. Alg239-R122 to 5091 for *Tropicibacter* sp. Alg239-R130. In agreement with recent findings on sponge-prokaryote symbiosis factors characteristic of *Alphaproteobacteria* species [15], genes for Glutathione S-transferase, taurine catabolism dioxygenase (except for strain Alg239-R122), thiamine monophosphate synthase, Catalase (peroxidase I) were predicted for all presented *Alphaproteobacteria* strains based on COG annotation.

From the *Gammaproteobacteria* strains obtained on VXA medium, *Agarivorans* sp. Alg241-V36, *Alcanivorax* sp. Alg238-V104, *Enterovibrio* sp. Alg239-V16, and *Photobacterium* sp. Alg240-V54 were sequenced. The first three strains were 100% complete based on CheckM estimations. Genome sizes across the four genera ranged from 4.38 to 5.7 Mbp with GC content ranging from 39.8 to 61.3%. RAST annotation indicated a range of 3904 coding sequences for *Photobacterium* to 5102 for *Enterovibrio*. COG annotations revealed a range of genes involved in multidrug resistance and transport, and coding for collagenase and related proteases, putative threonine efflux proteins and Methyl-accepting chemotaxis proteins for all four strains (Table S3).

2.3. Genome-wide secondary metabolite profiling

Totals of 16 (*Actinobacteria*), 3 (*Cytophagia*/*Bacteroidetes*), 38 (*Alphaptoteobacteria*), and 19 (*Gammaproteobacteria*) secondary metabolite biosynthetic gene clusters (BGCs) were detected using antiSMASH (Figure 2). For *Actinobacteria* strains, PKS (Polyketide synthases), siderophore, NRPS (Nonribosomal peptide synthetases), betalactone, terpene and ectoine BGCs were found. For the unclassified *Cytophagales* strain, one bacteriocin, terpene, and arylpolyene BGC each were detected. For all *Alphaproteobacteria* strains, PKS, bacteriocin, NRPS, terpene, homoserine lactone (hserlaton), arylpolyene, betalacton, ectoine,

and lassopeptide BGCs were predicted. These findings match those observed in earlier studies [19], particularly the presence of bacteriocin in *Alphaproteobacteria* species. For all *Gammaproteobacteria* strains, shared bacteriocin, betalactone and arylpolyene BGCs were found, though for *Enterovibrio* sp. Alg239-V16 eight different BGCs were predicted (Figure 2). These results highlight the genomic potential of the isolated bacteria for natural product discovery.

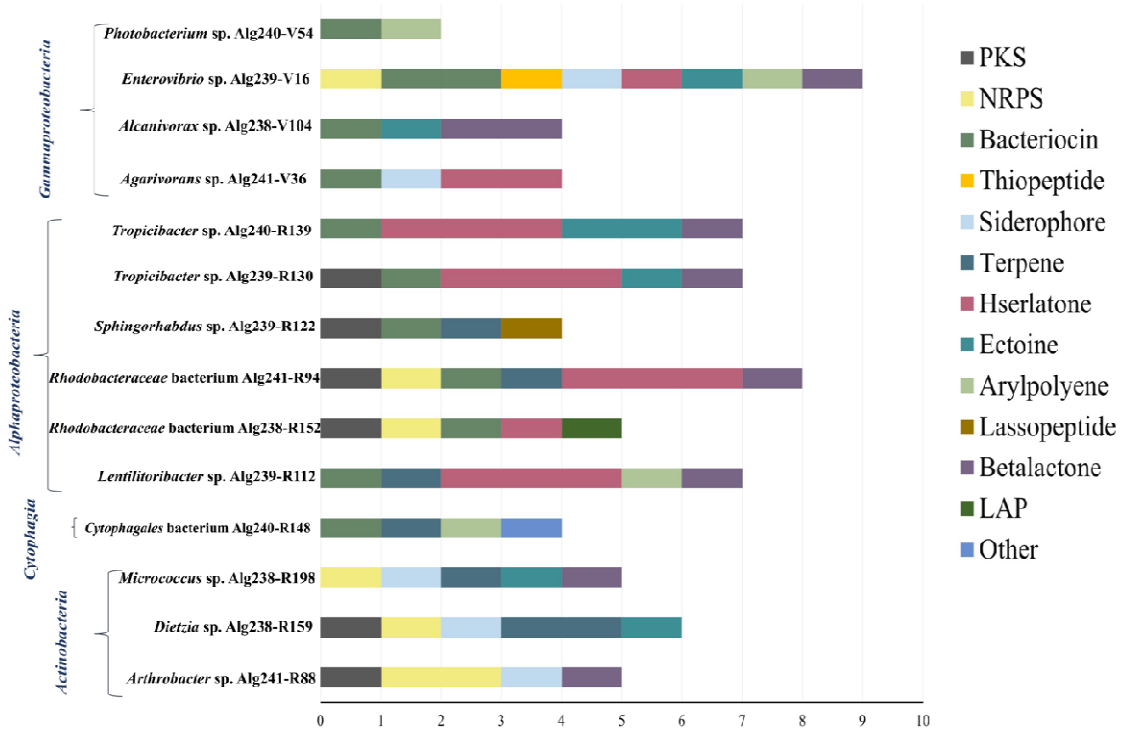


Figure 2. Secondary metabolite biosynthetic gene clusters predicted with antiSMASH across all the studied genomes.

Genome sequence accession number

The assembled genome sequences are available in the European Nucleotide Archive - European Molecular Biology Laboratory- EBI under the project number PRJEB28331 and can be accessed using the accession numbers given in Table 1. (http://www.ebi.ac.uk/ena/data/view/ <ACCESSION NUMBER>). Also, 16S rRNA gene sequences of the bacterial isolates (OTUs) were deposited at NCBI GenBank under the accession numbers MH818465 to MH818518 (Table S2).

Table 1.Genome features of all bacterial strains described in this study.

Genomes	Class	Order	Family	Genome size (Mbp)	Complete-ness (%)	Contami-nation (%)	coverage (X)	GC content (mol%)	Coding sequences (CDs)	rRNAs (n.b)	tRNAs (n.b)	Culture Medium	Accession Numbers
Actinobacteria													
Arthrobacter sp. Alg241-R88	*Actinobacteria*	*Micrococcales*	*Micrococcaceae*	4.37	99.71	0.73	162	64.8	4040	8	55	R2A	UNRG01000001-UNRG01000019
Dietzia sp. Alg238-R159	*Actinobacteria*	*Corynebacteriales*	*Dietziaceae*	3.9	99.41	0	207	70.1	3629	10	50	R2A	UNRI01000001-UNRI01000030
Micrococcus sp. Alg238-R198	*Actinobacteria*	*Micrococcales*	*Micrococcaceae*	2.59	98.7	0.23	306	72.8	2375	7	48	R2A	UNRH01000001-UNRH01000027
Bacteroidetes													
Cytophagales bacterium Alg240-R148	*Cytophagia*	*Cytophagales*	Unknown	5.3	100	1.79	162	38.8	4797	6	38	R2A	LS999826-LS999833
Proteobacteria													
Lentilitoribacter sp. Alg239-R112	*Alpha-proteobacteria*	*Rhizobiales*	*Phyllobacteriaceae*	3.91	99.92	0	198	44.3	3937	4	36	R2A	LS999833-LS999838
Rhodobacteraceae bacterium Alg238-R152	*Alpha-proteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	4.26	99.15	0.38	197	56.4	4264	5	43	R2A	UNRJ01000001-UNRJ01000014
Rhodobacteraceae bacterium Alg241-R94	*Alpha-proteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	4.87	99.25	0.43	156	54.8	4799	9	45	R2A	UNRF01000001-UNRF01000027
Sphingorhabdus sp. Alg239-R122	*Alpha-proteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	2.5	98.77	0.73	362	54	2446	3	39	R2A	UNRC01000001-UNRC0100002
Tropicibacter sp. Alg239-R130	*Alpha-proteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	5.2	99.15	0	108	58.8	5091	5	45	R2A	UNRB01000001-UNRB01000021
Tropicibacter sp. Alg240-R139	*Alpha-proteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	4.65	99.47	0.48	171	57.8	4642	6	48	R2A	UNRA01000001-UNRA01000014
Agarivorans sp. Alg241-V36	*Gamma-proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	4.7	100	0.89	130	44.7	4346	21	86	VXA	UNRE01000001-UNRE01000014
Alcanivorax sp. Alg238-V104	*Gamma-proteobacteria*	*Oceanospirillales*	*Alcanivoracaceae*	5.14	100	0	287	61.3	4776	4	43	VXA	UNRD01000001-UNRD01000031
Enterovibrio sp. Alg239-V16	*Gamma-proteobacteria*	*Vibrionales*	*Vibrionaceae*	5.7	100	0	120	47.7	5102	19	101	VXA	UNRK01000001-UNRK01000034
Photobacterium sp. Alg240-V54	*Gamma-proteobacteria*	*Vibrionales*	*Vibrionaceae*	4.38	97.87	0.39	110	39.8	3904	27	144	VXA	UNRL01000001-UNRL01000022

NOTES: 1- According to the MIxS standards, metadata associated with all presented genomes have been defined as follows. Habitat: Sea water, Host: Marine Sponge, Investigation type: Bacteria, Project name: isolated sponge-associated bacteria, Environment (biome): Ocean, Environment (feature): Costal water body, Environments (material): Sea water, Country: Portugal, Latitude and longitude: 36.9798□N -7.9889□W, Depth: 20 m, Collection date: 30/04/2015, Assembly method: *de novo.*with NGen software.
2- Estimated genomes completeness and contamination have been analysed using CheckM v.1.0.18.

Acknowledgments

The authors would like to thank Jorge M.S. Gonçalves' research team at CCMAR-UAlg for sampling the sponges from the sea. We thank Telma Franco for her help during sample processing and Marta Valente for Sanger-sequencing of 16S rRNA genes. EK would like sincerely to thank Alireza Asvadi for his encouragement to write up the study. This work was performed when EK was PhD student supported by a full scholarship from the Erasmus Mundus Programme/SALAM EMA2 lot7/SALA1206422. Material and resources for the lab experiments were funded by FCT- Fundação para a Ciência e a Tecnologia (Portugal) through the project grant PTDC/BIA-MIC/3865/2012, conceded to RC. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Elham Karimi: conceived the study, performed experiments, curated and analysed the data, wrote the original manuscript draft. **Rodrigo Costa**: conceived the study, provided materials, reviewed and edited the original manuscript draft. Both authors reviewed and edited the final manuscript draft.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Ethics statement

This study relied on *in situ* sampling of microorganisms from marine invertebrates without a nervous system, and as such was exempt from ethical approval procedures according to the current Portuguese legislation (Decreto-Lei nº 113/2013). This study did not occur within privately owned or protected areas. This study did not involve endangered or protected species. The sampling methodology privileged minimally invasive handling procedures, following the guidelines of the European Directive 2010/63/EU.

Supplementary Material

Table S1. 16S rRNA gene-based classification (genus level) and closest 16S rRNA gene relatives of bacterial isolates obtained from *Spongia* sp. on 1/10R2A and VXA media.

Table S2. Hierarchical, 16S rRNA gene-based taxonomic affiliation of bacterial isolates retrieved from *Spongia* sp. on **a)** 1/10R2A and **b)** VXA media, with their respective grouping

into OTUs at 98% 16S rRNA gene identity.

Table S3. COG annotation of sponge-associated bacterial genomes presented in this study.

References:

[1] T. Thomas, L. Moitinho-Silva, M. Lurgi, J.R. Björk, C. Easson, C. Astudillo-García, J.B. Olson, P.M. Erwin, S. López-Legentil, H. Luter, A. Chaves-Fonnegra, R. Costa, P.J. Schupp, L. Steindler, D. Erpenbeck, J. Gilbert, R. Knight, G. Ackermann, J. Victor Lopez, M.W. Taylor, J.M. Montoya, U. Hentschel, N.S. Webster, Diversity, structure and convergent evolution of the global sponge microbiome, Nat Commun, 7 (2016) 11870, doi: 10.1038/ncomms11870.
[2] C.P. Rua, A.E. Trindade-Silva, L.R. Appolinario, T.M. Venas, G.D. Garcia, L.S. Carvalho, A. Lima, R. Kruger, R.C. Pereira, R.G. Berlinck, R.A. Valle, C.C. Thompson, F. Thompson, Diversity and antimicrobial potential of culturable heterotrophic bacteria associated with the endemic marine sponge Arenosclera brasiliensis, PeerJ, 2 (2014) e419, doi: 10.7717/peerj.419.
[3] L. Pita, L. Rix, B.M. Slaby, A. Franke, U. Hentschel, The sponge holobiont in a changing ocean: from microbes to ecosystems, Microbiome, 6 (2018) 46, doi: 10.1186/s40168-018-0428-1.
[4] U. Hentschel, J. Piel, S.M. Degnan, M.W. Taylor, Genomic insights into the marine sponge microbiome, Nat. Rev. Microbiol., 10 (2012) 641-654, doi: 10.1038/nrmicro2839.
[5] M.C. Wilson, T. Mori, C. Ruckert, A.R. Uria, M.J. Helf, K. Takada, C. Gernert, U.A. Steffens, N. Heycke, S. Schmitt, C. Rinke, E.J. Helfrich, A.O. Brachmann, C. Gurgui, T. Wakimoto, M. Kracht, M. Crusemann, U. Hentschel, I. Abe, S. Matsunaga, J. Kalinowski, H. Takeyama, J. Piel, An environmental bacterial taxon with a large and distinct metabolic repertoire, Nature, 506 (2014) 58-62, doi: 10.1038/nature12959.
[6] J. Gutleben, M. Chaib De Mares, J.D. van Elsas, H. Smidt, J. Overmann, D. Sipkema, The multi-omics promise in context: from sequence to microbial isolate, Crit Rev Microbiol, 44 (2018) 212-229, doi: 10.1080/1040841X.2017.1332003.
[7] I. Burgsdorf, B.M. Slaby, K.M. Handley, M. Haber, J. Blom, C.W. Marshall, J.A. Gilbert, U. Hentschel, L. Steindler, Lifestyle evolution in cyanobacterial symbionts of sponges, Mbio, 6 (2015) e00391-00315, doi: 10.1128/mBio.00391-15.
[8] E. Karimi, B.M. Slaby, A.R. Soares, J. Blom, U. Hentschel, R. Costa, Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges, FEMS Microbiol Ecol, 94 (2018) fiy074-fiy074, doi: 10.1093/femsec/fiy074.
[9] E. Karimi, Metagenomics and functional genomics of bacterial symbionts of Spongia (Porifera, Dictyoceratida) specimens from the Algarvian shore (South Portugal), Algarve University, http://hdl.handle.net/10400.1/10819, 2018.
[10] D.J. Reasoner, E.E. Geldreich, A new medium for the enumeration and subculture of bacteria from potable water, Appl Environ Microbiol, 49 (1985) 1-7.
[11] M. Sait, P. Hugenholtz, P.H. Janssen, Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation□independent surveys, Environ Microbiol, 4 (2002) 654-666, doi: 10.1046/j.1462-2920.2002.00352.x.
[12] A.I.S. Esteves, C.C.P. Hardoim, J.R. Xavier, J.M. Goncalves, R. Costa, Molecular richness and biotechnological potential of bacteria cultured from Irciniidae sponges in the north-east Atlantic, FEMS Microbiol Ecol, 85 (2013) 519-536, doi: 10.1111/1574-6941.12140.

[13] J.R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, R.J. Farris, A. Kulam-Syed-Mohideen, D.M. McGarrell, T. Marsh, G.M. Garrity, The Ribosomal Database Project: improved alignments and new tools for rRNA analysis, Nucleic Acids Res., 37 (2009) D141-D145, doi: 0.1093/nar/gkn879.
[14] E. Karimi, J.M.S. Gonçalves, M. Reis, R. Costa, Draft Genome Sequence of Microbacterium sp. Strain Alg239_V18, an Actinobacterium Retrieved from the Marine Sponge *Spongia* sp, Genome Announc, 5 (2017), doi: 10.1128/genomeA.01457-16.
[15] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res., 25 (2015) 1043–1055, doi: 10.1101/gr.186072.114.
[16] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, The RAST server: rapid annotations using subsystems technology, BMC Genomics, 9 (2008) 1, doi: 10.1186/1471-2164-9-75.
[17] S. Wu, Z. Zhu, L. Fu, B. Niu, W. Li, WebMGA: a customizable web server for fast metagenomic sequence analysis, BMC Genomics, 12 (2011) 444, doi: 10.1186/1471-2164-12-444.
[18] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S.Y. Lee, M.H. Medema, T. Weber, antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline, Nucleic Acids Res., 47 (2019) W81-W87, doi: 10.1093/nar/gkz310.
[19] E. Karimi, T. Keller-Costa, B.M. Slaby, C.J. Cox, U.N. da Rocha, U. Hentschel, R. Costa, Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable Alphaproteobacteria, Sci. Rep., 9 (2019) 1999, doi: 10.1038/s41598-019-38737-x.