

Diversity of tryptophan halogenases in sponges of the genus *Aplysina*

Johanna Gutleben¹, Jasper J. Koehorst², Kyle McPherson¹, Shirley Pomponi^{3,4}, René H. Wijffels^{3,5}, Hauke Smidt¹, Detmer Sipkema^{1#}

Author Affiliations:

1 Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, The Netherlands, detmer.sipkema@wur.nl

2 Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, The Netherlands

3 Bioprocess Engineering, AlgaePARC, Wageningen University & Research, 6700 AA, Wageningen, The Netherlands

4 Florida Atlantic University - Harbor Branch, 5600 U.S. 1, Fort Pierce, FL 34946, United States

5 Faculty of Biosciences and Aquaculture, Nord University, 8026 Bodø, Norway

Keywords: halogenase, host-associated microbiome, phylogenetic diversity, marine sponges, bioactive compounds, environmental enzymes

Corresponding author: Detmer Sipkema, Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, The Netherlands, detmer.sipkema@wur.nl

Abstract

Marine sponges are a prolific source of novel enzymes with promising biotechnological potential. Especially halogenases, which are key enzymes in the biosynthesis of brominated and chlorinated secondary metabolites, possess interesting properties towards the production of pharmaceuticals that are often halogenated. In this study we used a PCR-based screening to simultaneously examine and compare the richness and diversity of putative tryptophan halogenase protein sequences and bacterial community structures of six *Aplysina* species from the Mediterranean and Caribbean seas. At the phylum level bacterial community composition was similar amongst all investigated species and predominated by Actinobacteria, Chloroflexi, Cyanobacteria, Gemmatimonadetes, and Proteobacteria. We detected four phylogenetically diverse clades of putative tryptophan halogenase protein sequences, which were only distantly related to previously reported halogenases. The Mediterranean species *A. aerophoba* harbored unique halogenase sequences, of which the most predominant was related to a sponge-associated *Psychrobacter*-derived sequence. In contrast, the Caribbean species shared numerous novel halogenase sequence variants and exhibited a highly similar bacterial community composition at the OTU level. Correlations of relative abundances of halogenases with those of bacterial taxa suggest that prominent sponge symbiotic bacteria including Chloroflexi and Actinobacteria are putative producers of the detected enzymes and may thus contribute to the chemical defense of their host.

Introduction

Bioactive compounds and the role of halogenase enzymes

The need to discover novel drug candidates is high on the policy agenda due to the ongoing emergence of multi-drug resistant microbial strains (O'Neill, 2016). The call for a better supply of new drugs against a large range of infectious diseases points towards novel natural products as a yet inexhaustible source of bioactivity. Especially the marine environment proved to be a particularly rich

resource for novel bioactive compounds, and many of them are halogenated (Gribble, 2015). Carbon-halogen bonds lead to an increase in thermal and oxidative stability as well as increased permeability through biological membranes compared to their non-halogenated counterparts (Smith *et al.*, 2017). Halogenated molecules exhibit a wide range of biological activities and may have antifungal, antibacterial, antiviral, anticancer, and/or anti-inflammatory properties (Butler & Sandy, 2009; Gribble, 2015). Thus, many pharmaceutical and agrochemical compounds as well as polymers are halogenated (Bolton *et al.*, 2011; Lu *et al.*, 2012; Jeschke, 2013), such as the commercially important antibiotics chloramphenicol, vancomycin and teicoplanin (Van Pée & Zehner, 2003). In nature, halogens including chloride, bromide, fluoride or iodide get attached to organic molecular scaffolds by halogenase enzymes, which have been detected in fungi, bacteria and algae from terrestrial and marine environments (Atashgahi *et al.*, 2018; Latham *et al.*, 2018). While marine enzymes preferentially halogenate with bromide (Neubauer *et al.*, 2018), chlorinated compounds are regularly detected in terrestrial sources (Latham *et al.*, 2018). To date, more than 5000 naturally produced halogenated compounds (Gribble, 2015) and six independently evolved halogenase enzyme families have been identified (Xu & Wang, 2016). These enzyme families, namely cofactor-free haloperoxidase, vanadium-dependent haloperoxidase, heme iron-dependent haloperoxidase, non-heme iron-dependent halogenase, flavin-dependent halogenase, and S-adenosyl-L-methionine-dependent halogenase, differ in their cofactor dependence, sequence homology and mechanistic features (Xu & Wang, 2016). The most extensively characterized halogenases are flavin (FADH₂)-dependent halogenases (FDHs) (Van Pée & Patallo, 2006; Zhao *et al.*, 2015). In contrast to haloperoxidases, FDHs often halogenate regioselectively, and are often part of secondary metabolite gene clusters encoding for non-ribosomal peptide synthetases (NRPS) and polyketide synthetases (PKS) producing halogenated molecules (Walsh *et al.*, 2001; Van Pée & Unversucht, 2003; Dorrestein *et al.*, 2005; Van Pée & Patallo, 2006). Regioselective halogenation under mild reaction conditions without harmful waste generation renders FDHs promising tools for biocatalysis of halogenated compounds (Smith *et al.*, 2013; Grünschow *et al.*, 2015; Shepherd *et al.*, 2015; Menon *et al.*, 2016;

Weichold *et al.*, 2016; Latham *et al.*, 2018). The amino acid sequences of FDHs contain a highly conserved flavin binding site (GxGxxG) near the N-terminus, which binds the cofactor in order to create the reactive halogen, as well as a motif with two tryptophan residues (WxWxIP), which presumably prevents the binding of a substrate close to the flavin (Van Pée & Zehner, 2003; Van Pée & Patallo, 2006). These motifs can be detected in almost every FDH described today, and serve as identification marker for novel FDH genes (Bayer *et al.*, 2013). FDHs can be distinguished into three classes depending on their preferred substrate moieties: phenols, pyrroles or tryptophan (Murphy, 2006; Van Pée & Patallo, 2006). In most cases, the exact substrates for halogenases encoded within biosynthetic gene clusters remain to be identified, however, many of the bacterial FDHs characterized to date halogenate a range of tryptophan-derivatives and other aromatic substrates (Payne *et al.*, 2013; Frese *et al.*, 2014; Shepherd *et al.*, 2015). These flavin-dependent tryptophan halogenases (THs) exhibit a broad substrate tolerance for other electron-rich organic scaffolds, although these are halogenated with lower reaction efficiencies in *in vitro* activity tests (Weichold *et al.*, 2016; Agarwal *et al.*, 2017). Since halotryptophans occur fairly frequently in natural products (Smith *et al.*, 2017), we focussed our efforts on THs in this study. Most flavin-dependent THs described to date require the activity of an additional enzyme, a flavin reductase, to provide the reduced flavin cofactor, and are thus two-component systems (Agarwal *et al.*, 2017). Single-component halogenating enzymes possess both domains for flavin-reduction and halogenation. To date, only two single-component halogenating enzymes have been reported from marine bacteria but they remain poorly characterized (Agarwal & Moore, 2014; Agarwal *et al.*, 2014).

Sponges as source of novel halogenase sequence variants

Sponges are the most prolific marine invertebrates for the discovery of novel bioactive compounds (Blunt *et al.*, 2009; Hu *et al.*, 2011; Mehbub *et al.*, 2014; Rocha-Martin *et al.*, 2014; Sipkema, 2017; Carroll *et al.*, 2019) since many of them depend on a chemical arsenal to defend themselves against diseases, competitors and predators (Thoms *et al.*, 2006; Taylor *et al.*, 2007; Pawlik, 2011). These ancient, filter feeding animals harbour dense and diverse microbial communities including members

of the bacterial phyla Actinobacteria, Acidobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Planctomycetes, Proteobacteria, Nitrospira, Poribacteria, Tectomicrobia, Verrucomicrobia, as well as archaea and numerous microeukaryotes (Taylor *et al.*, 2007; Webster & Taylor, 2012; Thomas *et al.*, 2016; Chaib De Mares *et al.*, 2017). Some of these largely uncultured bacteria can produce highly potent bioactive natural products, many of which are halogenated (Gribble, 2010; Smith *et al.*, 2017). Amongst marine sponges, the Demosponge genus *Aplysina* represents a morphologically diverse group of species challenging to identify due to their lack of a mineral skeleton (Zea *et al.*, 2014), as well as a high degree of sequence conservation in molecular marker genes (Erpenbeck *et al.*, 2007; Cruz-Barraza *et al.*, 2012). Chemotaxonomy was suggested as additional phylogenetic marker (Erpenbeck & van Soest, 2007), since *Aplysina* species are strongly chemically defended and especially renowned for the production of more than 100 halogenated natural products (Lira *et al.*, 2011; Loh & Pawlik, 2014). Such metabolites can make up to 12% of the sponge dry weight (Turon *et al.*, 2000; Thoms *et al.*, 2006).

A survey of natural products from Caribbean *Aplysina* species suggested that the brominated alkaloids were sponge-derived rather than microbiome-derived, since metabolite profiles were highly correlated to sponge morphotype rather than location and depth (Puyana *et al.*, 2015). It should be noted, however, that this study did not assess microbial composition, and hence, an essential or auxiliary role of the microbiome in metabolite production cannot be excluded. Halogenated compounds have been reported to be located in spherulous cells of *Aplysina aerophoba*, suggesting production by the sponge, or a complex symbiotic pattern with microorganisms involved at different levels of the biotransformation pathway (Turon *et al.*, 2000). Significant correlations of the relative abundances of a member of the Chloroflexi, a deltaproteobacterium and an unidentified bacterial OTU with the concentrations of three alkaloids (aerophobin-1, aplysinamisin-1 and isofistularin-3) in *A. aerophoba* indicated that bacteria were correlated to the production of brominated alkaloids (Sacristán-Soriano *et al.* 2011, 2016). This finding corroborates observations that the majority of halogenating enzymes has been described from algae, fungi and bacteria (Xu & Wang 2016).

Halogenated natural products or the corresponding biosynthetic gene clusters could previously be associated to specific bacterial symbionts in other marine sponges (Unson *et al.*, 1994; Flatt *et al.*, 2005; Ridley *et al.*, 2005; Hochmuth & Piel, 2009; Della Sala *et al.*, 2013; Öztürk *et al.*, 2013; Smith *et al.*, 2017), but to the best of our knowledge this remains unresolved for the metabolites of *Aplysina* species, both from the Mediterranean and Caribbean Sea (Puyana *et al.*, 2015; Sacristán-Soriano *et al.*, 2016). Currently, only dehalogenation mechanisms could be directly linked to the microbial community of *A. aerophoba* (Ahn *et al.*, 2003). However, the conserved nature of key enzymes encoded in biosynthetic gene clusters allows the design of degenerate PCR primers, thus facilitating the screening and discovery of novel sequence variants of these enzymes from environmental DNA (Hornung *et al.*, 2007; Kennedy *et al.*, 2008; Borchert *et al.*, 2016). Accordingly, PCR-based surveys have led to the identification of numerous putative flavin-dependent halogenase encoding genes from different environmental samples (Erol *et al.*, 2017), multiple marine sponges (Bayer *et al.*, 2013; Öztürk *et al.*, 2013) and cultivated Actinomycetes strains (Hornung *et al.*, 2007; Gao & Huang, 2009; Liao *et al.*, 2016).

Aim

With this study we aimed to investigate the phylogenetic diversity and distribution of flavin-dependent tryptophan halogenase protein sequence variants within six species of the marine sponge genus *Aplysina*. By using 16S rRNA and TH gene amplicon sequencing we aimed to determine the resident sponge bacterial community of different *Aplysina* species and explore potential links between the microbial populations and the associated halogenase sequences. This will increase our understanding of the putative producers of halogenated secondary metabolites in sponges.

Materials and Methods

Sponge Collection

Three individuals each of the Caribbean sponge species *Aplysina archeri* (AAr), *Aplysina cauliformis* (ACa), *Aplysina fistularis* (AFi), *Aplysina fulva* (AFu) and *Aplysina lacunosa* (ALa) were collected around Bonaire by SCUBA diving at depths between one and ten meters on 8 and 9 October 2012 (Table 1). A sampling permit was given to Detmer Sipkema by the government of Bonaire. Sampled sponge individuals grew at least 1 m apart from each other. The sponge species were identified in the field by Dr. Shirley Pomponi. After sampling, the sponge fragments were rinsed three times with sterile artificial seawater (ASW, 33 g/L Reef Crystals, Blacksburg, VA, USA), cut into 1 cm³ pieces and stored in RNAlater (Sigma Aldrich) at -20°C. Additionally, three individuals of the Mediterranean sponge species *Aplysina aerophoba* (AAe) were collected by SCUBA diving at Cala Montgo, Spain (N 42.114, E 3.168), on 19 October 2010 and 15 January 2012. The collection of *A. aerophoba* samples was conducted in strict accordance with Spanish and European regulations within the rules of the Spanish National Research Council with the approval of the Directorate of Research of the Spanish Government. After sampling, the sponges were transported to the laboratory and rinsed three times with sterile artificial seawater (ASW) before grinding with a sterilized mortar and pestle. To obtain a homogenous cell suspension, two volumes of ASW were added. Cell suspensions were aliquoted and mixed with sterile glycerol in ASW for a final concentration of 17% glycerol. Samples were frozen at -20°C and stored at -80°C until DNA extraction (Sipkema *et al.*, 2011).

DNA isolation

DNA was extracted using the FastDNA Spinkit for Soil (MP Biochemicals, Santa Ana, CA, USA) according to manufacturer's instructions with the following modification for the first step: instead of using 500 mg of soil, 750 µl of cryopreserved *Aplysina aerophoba* cell suspension was centrifuged at 14,000 g for 10 min and the pellet was used for the extraction. For the Caribbean sponge species, 500 mg of wet-weight sponge tissue was rinsed in sterile ASW, cut into small pieces, and used for DNA extraction.

PCR amplification of tryptophan halogenase genes

A PCR-based method was used to screen the sponge samples for the presence and identity of potential flavin-dependent tryptophan halogenase genes. In a preliminary screening among previously reported primers, the degenerate halogenase gene targeted primers SZ002 and SZ005 (Zehner *et al.*, 2005) were found the most suitable to amplify an approximately 500 bp DNA fragment. Barcoded PCR amplicons (1 specific barcode for each sample) were obtained through a two-step PCR reaction. For the first PCR reaction, the halogenase primers (SZ002 and SZ005) included linker sequences (341F and 806R): 341F-**SZ002** 5'-CMTAYGGGRBGCASCAG-**TCGGYGTSGGCGARGCGACCRTCCC**-3' and 806R-**SZ005** 5'-GGACTACNNGGGTATCTAAT-**GCCGGAGCAGTCGAYGAASAGGTC**-3'. The linker sequences represent the binding regions for the barcoding primers in the second PCR. As we routinely use barcoded 16S rRNA gene targeted primers for prokaryotic community composition analysis, we employed those for barcoding of the less routinely used TH gene-targeted primers. Thus, the more widely used barcoding primers could be applied to barcode halogenase gene amplicons.

The first PCR amplification was performed in a volume of 50 μ l using 10 μ l 5x GoTaq buffer, 2 μ l 10 mM dNTP mixture, 0.5 μ l 5U/ μ l GoTaq DNA polymerase (Promega, Madison, WI, USA), 3 μ l 10 μ M solution of both primer 341F-SZ002 and primer 806R-SZ005, 22.5 μ l nuclease-free water and 1 μ l template DNA (10-20ng/ μ l) for each of the samples listed in Table 1. PCR conditions were: initial denaturation (94°C for 5 min), followed by 35 cycles of denaturation (94°C for 30 s), annealing (60°C for 40 s), elongation (72°C for 50 s), and a final extension (72°C for 5 min). Amplification products were visualized on a 1.25% (w/v) agarose gel and purified using the Millipore DNA Gel Extraction Kit (Millipore, Billerica, MA, USA). A second barcoding PCR was performed as described above, except that a pyrosequencing adapter A (CCATCTCATCCCTGCGTGTCTCCGACTCAG) and 18 different barcodes of 10 nucleotides length connected to the 341F linker sequence were used as forward primer and pyrosequencing adapter B (CCTATCCCCTGTGTGCCTTGGCAGTCTCAG) connected to the 806R linker sequence as reverse primer. Furthermore, the number of amplification cycles was reduced to 15. PCR

products were visualized on a 1% (w/v) agarose gel, and the bands of PCR products were excised from the gel and purified as described above. The amplified fragments with adapter and barcodes were quantified using a Qubit™ fluorometer (Invitrogen) and mixed in approximately equal concentrations (4×10^5 copies μL^{-1}) to ensure equal representation of each sample in the pool. A 454-sequencing run was performed on a GS FLX Standard PicoTiterPlate (70 × 75) using a GS FLX pyrosequencing system according to the manufacturer's instructions (Roche, Mannheim, Germany) at the Technical University of Copenhagen. Pyrosequencing data of halogenase genes were deposited at the NCBI Sequence Read Archive under sample accession numbers SRR7853828 to SRR7853845.

PCR amplification of 16S rRNA genes

Barcoded amplicons of bacterial 16S rRNA genes for all sponge samples were amplified from the extracted DNA. PCR reactions were performed in a volume of 100 μL containing 20 μL High Fidelity Buffer (ThermoFisher Scientific, Waltham, MA, USA), 2.5 μL 10 μM 338R-I, 2.5 μL 10 μM 338R-II reverse primer (Daims *et al.*, 1999), 2 μL 10 mM dNTP mixture, 1 μL 2 U/ μL Phusion Hot start II DNA polymerase, 65 μL nuclease free water. 5 μL 27F-DegS forward primer (van den Bogert *et al.*, 2011) with Titanium adapter A and a sample-specific barcode (8nt) (Hamady *et al.*, 2008) attached to the 5'- end as well as 2 μL template DNA (10-20 ng/ μL) were added to each reaction. Amplification conditions were: initial denaturation (98°C for 30 s), followed by 30 cycles of denaturation (98°C for 10 s), annealing (56°C for 20 s), elongation (72°C for 20 s), and a final extension (72°C for 10 min). Amplification products were visualized, purified, pooled and sequenced as described above. Pyrosequencing data of 16S rRNA genes were deposited at the NCBI Sequence Read Archive under sample accession numbers SRR7853935 to SRR7853950.

Halogenase gene amplicon data analysis

Halogenase gene pyrosequencing data were demultiplexed using QIIME version 1.9.0 (Caporaso *et al.*, 2010). Sequences that i) were shorter than 200 bp or longer than 1000 bp, ii) contained more than one mismatch in the forward or reverse primer sequences, iii) contained ambiguous bases or iii) were represented with less than three reads were removed. Chimeric sequences were detected

using the usearch61 algorithm (Edgar, 2010) and removed. Two samples (AAe2 and AFi3) did not pass quality criteria of the 16S rRNA gene data analysis and were thus also removed from the halogenase gene analyses. The remaining sequences were translated into the three forward open reading frames using the transeq algorithm (Blankenberg *et al.*, 2007) as implemented in Galaxy (Afgan *et al.*, 2016), and ORFs containing stop codons were removed using customized Bash and R scripts (https://github.com/mibwurrepo/Gutleben_et.al_Halogenases_Aplysinas). Sequences were clustered at 95% amino acid sequence identity based on the average protein sequence identity in genomes of the same bacterial species (Rodriguez-R & Konstantinidis, 2014; Chaib De Mares *et al.*, 2018) using the uclust algorithm (Edgar, 2010). The most abundant sequence per cluster was retained as representative sequence.

For identification of putative flavin-dependent TH sequences a reference database was created, by subsetting the UniProt/SwissProt database (Bairoch, 2002; Bateman *et al.*, 2017) to 5427 “halogenase” protein entries, since similarity searches to smaller databases return more sensitive results (Jagtap *et al.*, 2013; Pearson, 2014). Of these, 75 were manually annotated and reviewed entries (SwissProt). Representative amino acid sequences were aligned (blastp) against i) the entire UniProt/SwissProt database (release 2018_02) and ii) the halogenase database using the Diamond alignment tool (Buchfink *et al.*, 2014). Protein families (Pfam) were assigned using the InterProScan pipeline 5.17 (Quevillon *et al.*, 2005) based on an e-value cutoff of 10^{-6} .

Amino acid sequences that aligned significantly (e-value <0.001 (Pearson 2014)) against an entry in the halogenase database were retained for phylogenetic analyses. Sequences were aligned using the ClustalW algorithm (Larkin *et al.*, 2007), together with the most closely related database entries, four reference sequences (tryptophan 5-halogenase PyrH (*Streptomyces rugosporus*, A4D0H5), flavin-dependent TH RebH (*Lechevalieria aerocolonigenes*, Q8KHZ8), flavin-dependent TH PrnA (*Pseudomonas fluorescens*, P95480), halogenase ClaH (*Streptomyces uncialis*, G3K6J6)), putative TH protein sequences previously found in the sponge *C. crambe* (Öztürk *et al.*, 2013) and the outgroup

sequences NADH-dependent flavin oxidoreductase BaiH (*Clostridium scindens*, P32370) and NADPH-flavin oxidoreductase Frp (*Vibrio harveyi*, Q56691). The halogenase sequences from *A. aerophoba* obtained by Bayer et al. (2013) could not be included in this analysis since they covered a different region of the gene.

The resulting alignment was manually refined and trimmed to the amplified regions excluding the primers using Jalview (Waterhouse *et al.*, 2009). A maximum likelihood phylogenetic tree was calculated using RaXML HPC (Stamatakis, 2006) under the PROTGAMMAWAG substitution model, and 100 bootstrap replicates were used to evaluate clusters. The best-scoring tree was visualized using iTol (Letunic & Bork, 2016).

R (version 3.4.3) (Sasaki *et al.*, 2005) was used for diversity analyses of amino acid sequences identified as putative tryptophan halogenases. For visualization and interpretation, relative abundance information was used for interpretation and was visualized using the ggplot2 v.2.2.1 package (Wickham, 2016). Weighted UniFrac dissimilarities (Lozupone *et al.*, 2011) were calculated and ordinated using Principal Coordinates Analysis as implemented in the phyloseq package (McMurdie & Holmes, 2013). Phyloseq and the microbiome package (Lahti *et al.*, 2017) were used for calculating observed richness and Shannon index diversity. Faith's phylogenetic diversity was calculated using the package picante (Kembel *et al.*, 2010). Venn diagrams were calculated and visualized using online tool jvenn (Bourtzis *et al.*, 1996). Caribbean core halogenases were defined as being present in at least one sample from all the Caribbean species and identified using jvenn.

16S rRNA gene amplicon data analysis

Bacterial 16S rRNA gene pyrosequencing data was analyzed using mothur v.1.39.5 (P. D. Schloss *et al.*, 2009) by following the 454 Standard operating procedure

(https://www.mothur.org/wiki/454_SOP). In brief, sequences were demultiplexed, denoised, and

sequences with i) more than two mismatches in the primers, ii) more than one mismatch in the

barcode, iii) more than 8 homopolymer were discarded (*trim.flows* and *trim.seqs* commands). Reads

were reduced to unique sequences (*unique.seqs*) and aligned to the SILVA SSU 128 database (Quast *et al.*, 2013) (*align.seqs*: flip=t). Aligned reads were kept (*screen.seqs*: optimize=start-end, criteria=98, minlength=250), and empty alignment columns were removed (*filter.seqs*: vertical=T, trump=.). Read counts for sequences that were within $\geq 99\%$ sequence similarity to a more abundant sequence were merged (*pre.cluster*: diffs=2). Chimeric sequences were detected with Vsearch (*chimera.vsearch*) (Rognes *et al.*, 2016) and removed (*remove.seqs*). Taxonomy was assigned using the SILVA SSU 128 database (*classify.seqs*: cutoff=80) (Wang *et al.*, 2007). Sequences that were not classified at Domain level as well as chloroplast sequences were removed (*remove.lineage*). Uncorrected pairwise distances between aligned sequences were calculated (*dist.seqs*: cutoff=0.15), OTUs were generated based on 97% sequence identity (*cluster*: method=opti, cutoff=0.03), and files were converted to .shared format (*make.shared*: list=, group=.). Taxonomy was assigned to OTUs (*classify.otu*: list=, name=, taxonomy=, label=0.03), and representative sequences for each OTU were picked (*get.oturep*: phylip=, list=, fasta=, label=0.03, sorted=size). Further OTU table processing was done with Bash and R scripts (https://github.com/mibwurrepo/Gutleben_et.al_Halogenases_Aplysinas). Relative abundance information was used for visualization and interpretation. Calculation of community metrics and UniFrac dissimilarities were performed as described above.

Core taxa were defined as being present in at least one sample from all analyzed sponge species, or from all Caribbean sponge species (Caribbean core), and were identified using jvenn. Correlations between relative halogenase and bacterial abundances were expressed as Spearman coefficients for all taxa and all halogenase genes, as well as for the Caribbean core taxa and core halogenase genes. Only coefficients $>\pm 0.5$ and with $p < 0.05$ were considered significant, and only taxa and halogenases shared by all Caribbean species were included in this analysis to maximize statistical power.

Heatmaps were generated using the pheatmap v1.0.8 package (Kolde, 2012). Analyses are available as R Markdown (https://github.com/mibwurrepo/Gutleben_et.al_Halogenases_Aplysinas).

Results

Identification of putative tryptophan halogenases

To capture the diversity of PCR-amplified halogenase genes in *Aplysina* species, 454-pyrosequencing was performed. A total of 37 374 DNA reads, representing 3653 unique protein sequences was obtained from TH gene targeted amplicon sequencing. After clustering the unique sequences at 95% amino acid sequence identity, 1918 protein sequence clusters were retained with a maximum of 109 sequences per cluster. Detailed results per sample are given in Table 1.

Out of the representative (most abundant) sequences of these 1918 clusters, 1654 (86.24%) resulted in a hit against the entire UniProt protein database. However, only 40 sequences aligned significantly (e-value <0.001 (Pearson, 2014)) against two flavin-dependent TH entries present in the UniProt database with low sequence identities (<45% amino acid identity) and low bitscores (mean=56). Thus, the 1918 sequences were blasted against a manually curated “halogenase” protein database containing the halogenase protein sequence subset (5427 entries) from UniProt. In total, 86 sequences (4.5%) resulted in a significant (e-value <0.001) hit against one of 16 entries from the “halogenase” database with high amino acid sequence similarities (32.9%-100%) and high bitscores (mean=173 (Table 2)) and were thus identified as putative TH protein sequence fragments. Of these 86 sequences, 25 had 67-80% amino acid sequence identity to a TH from marine gammaproteobacterium HTCC2080, an abundant oligotrophic marine microorganism belonging to the NOR5/OM60 clade (Cho & Giovannoni, 2004; Thrash *et al.*, 2010). Another 19 sequences matched most closely (82-100%) to a TH fragment from *Psychrobacter* sp. D8, a gammaproteobacterium isolated from the sponge *Crambe crambe* (Öztürk *et al.*, 2013). The closest database match of another 15 sequences, although with lower (36-51%) amino acid sequence identity, was a TH from the cyanobacterium *Calothrix* sp. NIES-2100 (Hirose *et al.*, 2017).

Additionally, the 86 protein sequences with significant “halogenase” database hits were also investigated for protein domains using InterProScan (Finn *et al.*, 2017). This confirmed for 84

sequences that they belong to the “Tryptophan halogenase” protein family (PF04820) and all contained a flavin-TH domain (IPR006905). Only two short protein sequences (S1691 and S1792), which were identified as TH by blastp, did not contain a detectable domain.

Phylogenetic analyses of putative tryptophan halogenases

Phylogenetic analyses of the 86 identified putative TH protein sequences resulted in four distinct clades, and only two sequences (S1874 and S2870) did not cluster within these four clades (Figure 1). Except for *A. aerophoba*, all analyzed sponge species contained halogenases from all four phylogenetic clades. The largest clade (Clade 1) contained 30 sequences, followed by Clade 4 (25 sequences), Clade 3 (19 sequences) and Clade 2 (10 sequences).

Only one species-specific halogenase clade could be identified, namely an *A. lacunosa*-specific subclade within Clade 2 (7 out of 10 sequences, Figure 1). The sequences within this subclade were all most closely related to a TH from a Dehalococcales bacterium (A0A2E9M8J4). Within Clade 3 (19 sequences), the halogenase most abundant (88%) in *A. aerophoba* (S1779) was closely related to a putative halogenase sequence (K7XJ2) previously detected in Mediterranean *C. crambe* as well as to a halogenase from *Psychrobacter* sp. D8 isolated from the same sponge (Öztürk *et al.*, 2013). Especially Clade 4 (25 sequences) was only distantly related to previously reported halogenases from Verrucomicrobia and Cyanobacteria and contained mainly sequences from Caribbean *Aplysina* spp., indicating a large number of novel halogenases within these sponges and their microbiota. The well-characterized flavin-dependent TH gene products RebH (*L. aerocolonigenes*, Q8KHZ8 (Yeh *et al.*, 2007)) and PrnA (*P. fluorescens*, P95480 (Dong *et al.*, 2005)) clustered outside the sponge-derived clades and were closely related to only one not abundant (<1%) sequence from *A. aerophoba* (S2870). While the Caribbean species shared many halogenases, it is noteworthy that sequences from the Mediterranean *A. aerophoba* were closely related to Caribbean halogenase sequences, albeit in no case identical. Furthermore it could be observed that some halogenases were highly

abundant in only one sponge species such as S2941 in *A. cauliformis* or S3275 in *A. fulva*, while others such as S0310 or S3006 were detected in similar abundances in all Caribbean species.

Bacterial and halogenase diversity

A total of 122 142 high-quality bacterial 16S rRNA gene sequences were obtained from all sponge samples. These sequences clustered into 1993 OTUs at 97% sequence similarity. Bacterial composition at phylum level was similar for all *Aplysina* species, with Acidobacteria, Actinobacteria, Chloroflexi, Cyanobacteria, Gemmatimonadetes, SBR1093 and Proteobacteria representing the most predominant out of 20 detected bacterial phyla (Supplementary Figure S1). The 50 overall most abundant OTUs cumulatively accounted for a relative abundance between 59% (ALa2) and 90% (AFi2) in individual samples (Supplementary Figure S2). Each sponge species exhibited a large number of unique OTUs, while 39 OTUs were shared between all species. These core OTUs comprised between 6% and 45% relative abundance in the sponge samples and contained seawater-derived clades, such as Chloroflexi SAR202 (Morris *et al.*, 2004), as well as putative sponge-symbiotic bacteria such as Rhodospirillales (Karimi *et al.*, 2018), Acidobacteria and *Nitrospira* (Schmitt *et al.*, 2012).

All Caribbean *Aplysina* species shared 85 OTUs, amongst which were a predominant sponge-associated member of the Cyanobacteria (Otu0001, 1-67% relative abundance in individual samples), an unclassified bacterial OTU (Otu0010, 0.6%-8.2% relative abundance), a member of the actinobacterial OM1 clade (Otu0007, 0.6%-8.2% relative abundance) and an SBR1093 OTU (Otu0019, 0.5%-4.3% relative abundance). In total, 37 of the 50 overall most abundant OTUs were shared by all of the Caribbean species. Due to the large overlap in bacterial community structure of the Caribbean species, only *A. fistularis* clearly separated from the other Caribbean *Aplysina* spp. due to the high relative abundance of the cyanobacterial Otu0001 (>63% relative abundance). In addition, the Mediterranean *A. aerophoba* could be clearly separated from the other species in ordination plots (Figure 2 E).

In order to investigate connections between the halogenase diversity and the sponge-associated bacterial diversity, alpha and beta diversity indices were calculated for the 86 putative halogenase sequences and the trends were compared to the diversity indices calculated for the bacterial 16S rRNA gene sequences retrieved from the *Aplysina* samples (Figure 2 and Supplementary Figure S3). Phylogenetic Diversity, Observed Richness and Shannon Diversity indices for halogenases and bacterial 16S rRNA gene sequences followed a similar trend for all species (Figure 2 A, D), however, correlations of diversity indices were only significant for Observed Richness (Supplementary Figure S3). Overall, a high phylogenetic diversity of both bacterial OTUs and halogenase sequences was detected for *A. aerophoba*, *A. cauliformis* and *A. lacunosa*, while *A. fistularis* exhibited the least phylogenetically diverse gene repertoire (Figure 2 A). The observed richness of *A. aerophoba* for both halogenases and bacterial 16S rRNA gene sequences was lower than for the Caribbean species, however the phylogenetic diversity of both measured variables was high, indicating a large phylogenetic breadth. These differences were not statistically significant (anova $p_{adj} > 0.05$). Beta diversity analyses showed no discernible species separation as well as a large spread in ordination space for the Caribbean species, while the *A. aerophoba* samples clustered apart from the other species and extremely close to each other, indicating an almost identical halogenase profile as well as bacterial 16S rRNA gene profile (Figure 2 B, E). Overall, the first two axes explained 84% of variation in the halogenase dataset, and 63% in the 16S rRNA gene dataset, where all Caribbean species spread along the first axis, separating only *A. aerophoba* along the second axis (Figure 2 E). This analysis revealed an overlap in the halogenase composition of the Caribbean sponges, who shared four halogenase sequences (S3006, S0310, S3103, S3275). Additionally, each sponge species harbored between four (*A. fulva*) and 13 (*A. archeri*) species-specific halogenase sequences (Figure 2 C).

Co-correlation of bacterial 16S rRNA gene and halogenase relative abundances

Since diversity calculations indicated an interrelation between halogenase and bacterial 16S rRNA gene diversity, Spearman correlations on the relative abundances of halogenase amino acid sequences with bacterial 16S rRNA genes were calculated for the 50 most predominant bacterial OTUs. All 50 most predominant OTUs exhibited significant ($p < 0.05$, Spearman $r > \pm 0.5$) correlations with two or more of the 86 putative halogenases (Supplementary Figure S4). It should be noted, however, that the large number of unique and distinct halogenases as well as bacterial OTUs derived from the Mediterranean species *A. aerophoba* caused a prominent block of strong correlations (Supplementary Figure S4), and these results should thus be interpreted carefully. Therefore, we

further evaluated only the relative abundance co-correlations of bacterial 16S rRNA genes and halogenases that were shared between all Caribbean species (Figure 3).

Of the 85 shared bacterial OTUs, 37 were significantly ($p < 0.05$, Spearman $r > 0.5$) positively correlated with the four shared putative halogenases. The shared bacterial OTUs constituted between 28% and 80% (mean=46%) of the total relative abundance per sample, while the shared halogenases represent between 10% and 71% (mean=30%) relative abundance. The majority of the correlated OTUs were affiliated to the phyla Chloroflexi (16 OTUs), Proteobacteria (8 OTUs, Alpha-, Delta- and Gammaproteobacteria), Actinobacteria (3 OTUs) and Tectomicrobia (2 OTUs) (Figure 3), all of which are renown for the production of halogenating enzymes (Bayer *et al.*, 2013, 2018; Öztürk *et al.*, 2013; Liao *et al.*, 2016; Smith *et al.*, 2017). For three halogenases, one specific bacterial OTU was found especially highly correlated ($r > 0.72$), suggesting this distinct bacterial taxon as potential halogenase producer: The highly predominant halogenases S0310 and S3275, both phylogenetically associated to clade 4, were suggested to be produced by the Chloroflexi Otu0069, member of the sponge-associated lineage TK10 (Schmitt *et al.*, 2011; Burgsdorf *et al.*, 2014), and an unclassified Chloroflexi Otu0228 with 98% 16S rRNA gene sequence identity to a sponge-associated member of the Chloroflexi (GenBank FJ481334, (Montalvo & Hill, 2011)), respectively. The less abundant halogenase S3103, phylogenetically placed in clade 1, was most highly correlated to an unclassified Nitrospinae Otu0055, displaying 97% 16S rRNA gene sequence identity to an uncultured sponge-associated bacterium (GenBank FJ900348, (Kamke *et al.*, 2010)). Despite this analysis being of only predictive nature, the co-occurrence of certain halogenases with specific bacterial taxa can narrow down the range for searching potential microbial producers of these proteins.

Discussion

The marine environment is a prolific source of novel enzymes with potential industrial applications. To screen marine samples for such enzymes, metagenomics approaches are promising, whereby all extracted DNA is investigated with DNA sequence-based methods (Hugenholtz & Tyson, 2008; Kennedy *et al.*, 2008; Vakhlu *et al.*, 2008; Wilson & Piel, 2013; Barone *et al.*, 2014; Loureiro *et al.*, 2018). For genes where suitable PCR primers can be designed, a PCR-based screening approach is especially useful to rapidly explore diversity of the gene of interest in a larger number of environmental samples. Such approaches have previously been applied for the discovery of novel natural product biosynthesis genes (Zhao *et al.*, 2008; Milshteyn *et al.*, 2014; Amos *et al.*, 2015; Müller *et al.*, 2015; Borchert *et al.*, 2016), lipases (Wang *et al.*, 2010) or alcohol dehydrogenases (Itoh *et al.*, 2014), to just name a few examples from a broad range of functional genes discovered in environmental samples (Kotik, 2009). In this study, the PCR-based screening approach resulted in the discovery of a large phylogenetic breadth of previously undescribed, putative halogenase protein fragments, which may be involved in the biosynthesis of numerous brominated natural products found in Mediterranean and Caribbean *Aplysina* species.

Sponges harbor numerous novel putative tryptophan halogenases

Out of all detected protein sequences clustered at 95% amino acid sequence identity, 86.24% had a match in the entire UniProt database, hinting at still a large number of unknown protein sequences in marine sponges. Additionally, only 86 out of 1918 sequences could successfully be identified as putative halogenases. This could indicate a low specificity of the primers applied, which is a regularly observed problem for primers targeting functional genes in microbial communities (Mohamed *et al.*, 2010; Pereyra *et al.*, 2010; Gaby & Buckley, 2012; Bonilla-Rosso *et al.*, 2016) due to relatively low degrees of sequence conservation in comparison to the 16S rRNA gene.

On the other hand, the number (86) of putative THs is similar for sponges and other environmental samples such as freshwater, marine and soil environments or cultivated Actinomycetes strains (18

(Öztürk *et al.*, 2013), 36 (Bayer *et al.*, 2013), 38 (Erol *et al.*, 2017), 103 (Hornung *et al.*, 2007), 254 (Neubauer *et al.*, 2018)). The low sequence identities (mean=44%), and the match to only few (16) “halogenase” entries from the UniProt database, further indicate a large, yet untapped resource of halogenases in marine sponges. Only one sequence (S2870 in *A. aerophoba*) with low abundance (<1%) was related with <44% amino acid sequence similarity to two well-characterized flavin-dependent THs, PrnA and RebH (Figure 1). PrnA is encoded in the biosynthetic gene cluster for the production of the antifungal compound pyrrolnitrin in *Pseudomonas fluorescens* (Harris *et al.*, 1985), whereas RebH is part of the biosynthetic gene cluster for the anticancer compound rebeccamycin in *Lechevalieria aerocolonigenes* (Onaka *et al.*, 2003).

Due to low sequence identities to well-studied halogenases, the putative THs could not be confidentially assigned to any of the known structural classes of halogenated compounds. These enzymes probably represent novel types, which halogenate different structural moieties compared to previously reported halogenases. Such a TH with an unusual substrate preference has been discovered in the metagenome of the sponge symbiont *Candidatus Entotheonella sarta* (Smith *et al.*, 2017). This halogenase displays only between five and 27% amino acid sequence identity to the halogenases from this study (data not shown) and represents another recent example of the large, yet untapped genetic resources for novel halogenases in marine sponges. Our phylogenetic analyses resulted in four distinct clades, potentially separating functionally divergent groups of halogenases. These findings expand on previously reported results for *A. aerophoba* (Bayer *et al.*, 2013) and *C. crambe* (Öztürk *et al.*, 2013) that showed three distinct sponge-specific clades of halogenases. Each analyzed sponge species harbored sequences from all four clades, hinting at a potentially highly diverse spectrum of halogenated molecules that can be synthesized within each sponge. Rua *et al.* (2018) hypothesized that larger microbiome diversity influences the potential of bioactive compound production in sponges. In our study we found that a high phylogenetic diversity and richness of microbial taxa corresponded to a high diversity of halogenases. The Mediterranean *A. aerophoba* harbored a bacterial community distinct from that of the Caribbean species, which is consistent with

previous investigations (Thomas *et al.*, 2016; Chaib De Mares *et al.*, 2017), and while its halogenases were closely related to Caribbean ones, they were in no case identical.

All except for two of the detected novel putative halogenases as well as their closest relatives exhibited active binding sites for flavin and L-tryptophan. Since these enzymes are known to tolerate a wide range of organic scaffolds (Agarwal *et al.*, 2017) and since all FDHs known to date can also function as brominases (Xu & Wang, 2016) they are potentially halogenating completely unknown organic substrates. Thus, these enzymes may contribute to the production of the more than 100 halogenated natural products reported from *Aplysina* species (Turon *et al.*, 2000; Thoms *et al.*, 2006; Lira *et al.*, 2011; Loh & Pawlik, 2014; Puyana *et al.*, 2015). However, further studies are necessary to unravel the exact functions of these enzymes as no closely related enzymes have been functionally characterized (Figure 1).

Prominent sponge symbionts are potential halogenase producers

In an attempt to predict the potential bacterial producers of the detected halogenase genes, we correlated the relative abundances of the four 16S rRNA gene OTUs and the 85 putative halogenases shared by the Caribbean *Aplysina* species. We hypothesized that the shared halogenases are produced by bacteria that can be found amongst the shared bacterial taxa in the different sponge species. Furthermore, an increased relative abundance of the producers is expected to be reflected in an increased number of detected halogenase genes. Two halogenases (S0310 and S3275) were the most highly correlated with sponge-associated Chloroflexi (Otu0069 and Otu0228 respectively, Figure 3). These clades include predominant sponge symbionts with the genomic repertoire for chemical defense (Slaby *et al.*, 2017; Bayer *et al.*, 2018). Chloroflexi were previously found to contain halogenases (Bayer *et al.*, 2013) and were linked to the production of brominated compounds in *A. aerophoba* (Sacristán-Soriano *et al.*, 2011, 2016). We furthermore detected strong correlation of a halogenase sequence to a Nitrospinae OTU, however, little is known about this taxon in sponges to date. Nonetheless, Nitrospinae are closely related to the candidate phylum Tectomicrobia, which

occurs in *Aplysina* species (Chaib De Mares *et al.*, 2018) and of which some members are renowned for their large secondary metabolism gene repertoire (Wilson *et al.*, 2014; Smith *et al.*, 2017).

One highly predominant cyanobacterial OTU was present in all Caribbean species and was not significantly correlated to halogenases in this study, adding to the hypothesis that the role of Cyanobacteria in sponge-microbe symbioses of high microbial abundance sponges might be mainly related to nutrient production rather than chemical defense (Freeman *et al.*, 2013; Burgsdorf *et al.*, 2015). In low microbial abundance sponges, however, the production of brominated metabolites has been linked to cyanobacterial symbionts (Unson *et al.*, 1994; Flatt *et al.*, 2005). Thus, our results support previous findings and for the first time indicate potential microbial producers of brominated compounds found in Caribbean *Aplysina* species. However, further studies, including comparative and functional genomics of sponge symbionts are necessary to reliably link halogenated compound production to their microbial producers.

Conclusion

The plethora of previously undescribed putative flavin-dependent THs from the metagenomic DNA of Mediterranean and Caribbean *Aplysina* species unraveled here indicates a large potential for the discovery of novel halogenating enzymes from these marine invertebrates and their associated microbiomes. The separation into four phylogenetically distinct clades of halogenase protein sequences indicates that multiple classes of organic scaffolds may be halogenated by sponge-associated microbes. High bacterial diversity was in most cases indicative of a high halogenase diversity, and while the Caribbean species shared many halogenases as well as bacterial OTUs, the Mediterranean *A. aerophoba* could be clearly distinguished. Based on co-occurrence, three prominent bacterial sponge symbionts belonging to the Chloroflexi and Nitrospinae were identified as potential sources of abundant halogenases. These results may thus contribute to explaining the origin of the numerous halogenated compounds discovered in *Aplysina* species.

Author Contribution

DS initiated this project and collected the samples with SP and RW. JG and KMP conducted the molecular lab bench work. JG, JK and DS planned and executed the analysis pipeline for the sequencing data. JG wrote the paper and all authors contributed to its improvement.

Funding

This work was supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007– 2013 [REA grant agreement n8 607786 (BluePharmTrain)] and the Dutch BE-Basic Foundation [project 7.1.5].

Acknowledgements

We thank María J. Uriz from the Centre d'Estudis Avançats de Blanes (CEAB), CSIC, Blanes, Spain for providing support for sample collection and access to the laboratories for prompt processing, Søren Sørensen and Waleed Abu Al-Soud with the sequencing of the halogenase amplicons and Sudarshan Shetty and Georg Steinert for help with the data analysis.

References

- Afgan E, Baker D, van den Beek M, *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**: W3–W10.
- Agarwal V & Moore BS (2014) Enzymatic synthesis of polybrominated dioxins from the marine environment. *ACS Chem Biol* **9**: 1980–1984.
- Agarwal V, El Gamal AA, Yamanaka K, Poth D, Kersten RD, Schorn M, Allen EE & Moore BS (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat Chem Biol* **10**: 640–647.
- Agarwal V, Miles ZD, Winter JM, Eustáquio AS, Gamal AA El & Moore BS (2017) Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem Rev* **117**: 5619–5674.
- Ahn Y-B, Rhee S-K, Fennell DE, Kerkhof LJ, Hentschel U & Häggblom MM (2003) Reductive dehalogenation of brominated phenolic compounds by microorganisms associated with the marine sponge *Aplysina aerophoba*. *Appl Environ Microbiol* **69**: 4159–4166.
- Amos GCA, Borsetto C, Laskaris P, Krsek M, Berry AE, Newsham KK, Calvo-Bado L, Pearce DA, Vallin C & Wellington EMH (2015) Designing and implementing an assay for the detection of rare and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. *PLoS One* **10**: 1–15.
- Atashgahi S, Liebensteiner MG, Janssen DB, Smidt H, Stams AJM & Sipkema D (2018) Microbial Synthesis and Transformation of Inorganic and Organic Chlorine Compounds. *Front Microbiol* **9**: 1–22.
- Bairoch A (2002) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **28**: 45–48.
- Barone R, De Santi C, Palma Esposito F, Tedesco P, Galati F, Visone M, Di Scala A & De Pascale D (2014) Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. *Front Mar Sci* **1**: 1–6.
- Bateman A, Martin MJ, O'Donovan C, *et al.* (2017) UniProt: The universal protein knowledgebase. *Nucleic Acids Res* **45**: D158–D169.
- Bayer K, Scheuermayer M, Fieseler L & Hentschel U (2013) Genomic Mining for Novel FADH₂-Dependent Halogenases in Marine Sponge-Associated Microbial Consortia. *Mar Biotechnol* **15**: 63–72.
- Bayer K, Jahn MT, Slaby BM, Moitinho-Silva L & Hentschel U (2018) Marine Sponges as Chloroflexi Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade. *mSystems* **3**: 328013.
- Blankenberg D, Taylor J, Schenck I, *et al.* (2007) A framework for collaborative analysis of ENCODE data: Making large-scale analyses biologist-friendly. *Genome Res* **17**: 960–964.
- Blunt JW, Copp BR, Hu WP, Munro MHG, Northcote PT & Prinsep MR (2009) Marine natural products. *Nat Prod Rep* **26**: 170–244.
- van den Bogert B, de Vos WM, Zoetendal EG & Kleerebezem M (2011) Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. *Appl Environ Microbiol* **77**: 2071–2080.
- Bolton O, Lee K, Kim HJ, Lin KY & Kim J (2011) Activating efficient phosphorescence from purely organic materials by crystal design. *Nat Chem* **3**: 205–210.
- Bonilla-Rosso G, Wittorf L, Jones CM & Hallin S (2016) Design and evaluation of primers targeting genes encoding NO-forming nitrite reductases: Implications for ecological inference of denitrifying communities. *Sci Rep* **6**: 1–8.
- Borchert E, Jackson SA, O'Gara F & Dobson ADW (2016) Diversity of natural product biosynthetic genes in the microbiome of the deep sea sponges *Inflatella pellicula*, *Poecillastra compressa*, and *Stelletta normani*. *Front Microbiol* **7**: 1–11.
- Bourtzis K, Nirgianaki A, Markakis G & Savakis C (1996) Wolbachia infection and cytoplasmic incompatibility in *Drosophila* species. *Genetics* **144**: 1063–1073.
- Buchfink B, Xie C & Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60.
- Burgsdorf I, Erwin PM, López-Legentil S, Cerrano C, Haber M, Frenk S & Steindler L (2014) Biogeography rather than association with cyanobacteria structures symbiotic microbial communities in the marine sponge *Petrosia ficiformis*. *Front Microbiol* **5**: 1–11.
- Burgsdorf I, Slaby BM, Handley KM, Haber M, Blom J, Marshall CW, Gilbert JA, Hentschel U & Steindler L (2015) Lifestyle Evolution in

- Cyanobacterial Symbionts of Sponges. *MBio* **6**: 1–14.
- Butler A & Sandy M (2009) Mechanistic considerations of halogenating enzymes. *Nature* **460**: 848–854.
- Caporaso JG, Kuczynski J, Stombaugh J, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Publ Gr* **7**: 335–336.
- Carroll AR, Copp BR, Davis RA, Keyzers RA & Prinsep MR (2019) Marine natural products. *Nat Prod Rep* **36**: 122–173.
- Chaib De Mares M, Sipkema D, Huang S, Bunk B, Overmann J & van Elsas JD (2017) Host Specificity for Bacterial, Archaeal and Fungal Communities Determined for High- and Low-Microbial Abundance Sponge Species in Two Genera. *Front Microbiol* **8**: 1–13.
- Chaib De Mares M, Jiménez DJ, Palladino G, Gutleben J, Lebrun LA, Muller EEL, Wilmes P, Sipkema D & van Elsas JD (2018) Expressed protein profile of a Tectomicrobium and other microbial symbionts in the marine sponge *Aplysina aerophoba* as evidenced by metaproteomics. *Sci Rep* **8**: 11795.
- Cho J-C & Giovannoni SJ (2004) Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine Gammaproteobacteria. *Appl Environ Microbiol* **70**: 432–440.
- Cruz-Barraza JA, Carballo JL, Rocha-Olivares A, Ehrlich H & Hog M (2012) Integrative taxonomy and molecular phylogeny of genus *Aplysina* (Demospongiae: Verongida) from Mexican Pacific. *PLoS One* **7**: 1–15.
- Daims H, Brühl A, Amann R, Schleifer KH & Wagner M (1999) The domain-specific probe EUB338 is insufficient for the detection of all bacteria: Development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434–444.
- Dong C, Flecks S, Unversucht S, Haupt C, Van Pée KH & Naismith JH (2005) Structural biology: Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science (80-)* **309**: 2216–2219.
- Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL & Walsh CT (2005) Dichlorination of a pyrrolyl-S-carrier protein by FADH₂-dependent halogenase P1A during pyoluteorin biosynthesis. *Proc Natl Acad Sci* **102**: 13843–13848.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Erol O, Arends I & Muyzer G (2017) Discovery of Putative Halogenases in Environmental Samples Using Metagenomics. *Curr Biotechnol* **6**: 17–25.
- Erpenbeck D & van Soest RWM (2007) Status and perspective of sponge chemosystematics. *Mar Biotechnol* **9**: 2–19.
- Erpenbeck D, Duran S, Rützler K, Paul V, Hooper JNA & Wörheide G (2007) Towards a DNA taxonomy of Caribbean demosponges: A gene tree reconstructed from partial mitochondrial CO1 gene sequences supports previous rDNA phylogenies and provides a new perspective on the systematics of Demospongiae. *J Mar Biol Assoc United Kingdom* **87**: 1563–1570.
- Finn RD, Attwood TK, Babbitt PC, *et al.* (2017) InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res* **45**: D190–D199.
- Flatt PM, Gautschi JT, Thacker RW, Musafija-Girt M, Crews P & Gerwick WH (2005) Identification of the cellular site of polychlorinated peptide biosynthesis in the marine sponge *Dysidea* (Lamellodysidea) herbacea and symbiotic cyanobacterium *Oscillatoria* spongelliae by CARD-FISH analysis. *Mar Biol* **147**: 761–774.
- Freeman CJ, Thacker RW, Baker DM & Fogel ML (2013) Quality or quantity: Is nutrient transfer driven more by symbiont identity and productivity than by symbiont abundance? *ISME J* **7**: 1116–1125.
- Frese M, Guzowska PH, Voß H & Sewald N (2014) Regioselective enzymatic halogenation of substituted tryptophan derivatives using the FAD-dependent halogenase RebH. *ChemCatChem* **6**: 1270–1276.
- Gaby JC & Buckley DH (2012) A comprehensive evaluation of PCR primers to amplify the *nifH* gene of nitrogenase. *PLoS One* **7**: 1–12.
- Gao P & Huang Y (2009) Detection, distribution, and organohalogen compound discovery implications of the reduced flavin adenine dinucleotide-dependent halogenase gene in major filamentous actinomycete taxonomic groups. *Appl Environ Microbiol* **75**: 4813–4820.
- Gribble GW (2010) *ChemInform Abstract: Naturally Occurring Organohalogen Compounds - A Comprehensive Update.*
- Gribble GW (2015) A recent survey of naturally occurring organohalogen compounds. *Environ Chem* **12**: 396–405.

- Grüschow S, Smith DRM, Gkotsi DS & Goss RJM (2015) Halogenases. *Biocatalysis in Organic Synthesis*, (Faber, Fessner & Turner, eds). Georg Thieme Verlag, Stuttgart
- Hamady M, Walker JJ, Harris JK, Gold NJ & Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**: 235–237.
- Harris CM, Kannan R, Kopecka H & Harris TM (1985) The Role of the Chlorine Substituents in the Antibiotic Vancomycin: Preparation and Characterization of Mono- and Didechlorovancomycin. *J Am Chem Soc* **107**: 6652–6658.
- Hirose Y, Shimura Y, Fujisawa T, Nakamura Y & Kawachi M (2017) Genome sequencing of cyanobacterial culture collection at National Institute for Environmental Studies (NIES). *Unpublished*. <https://www.ncbi.nlm.nih.gov/nuccore/AP018178.1>.
- Hochmuth T & Piel J (2009) Polyketide synthases of bacterial symbionts in sponges - Evolution-based applications in natural products research. *Phytochemistry* **70**: 1841–1849.
- Hornung A, Bertazzo M, Dziarnowski A, *et al.* (2007) A genomic screening approach to the structure-guided identification of drug candidates from natural sources. *ChemBioChem* **8**: 757–766.
- Hu GP, Yuan J, Sun L, She ZG, Wu JH, Lan XJ, Zhu X, Lin YC & Chen SP (2011) Statistical research on marine natural products based on data obtained between 1985 and 2008. *Mar Drugs* **9**: 514–525.
- Hugenholtz P & Tyson GW (2008) Metagenomics. *Nature* **455**: 481–483.
- Itoh N, Kariya S & Kurokawa J (2014) Efficient PCR-Based Amplification of Diverse Alcohol Dehydrogenase Genes from Metagenomes for Improving Biocatalysis: Screening of Gene-Specific Amplicons from Metagenomes. *Appl Environ Microbiol* **80**: 6280–6289.
- Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL & Griffin TJ (2013) A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* **13**: 1352–1357.
- Jeschke P (2013) The Unique Role of Halogen Substituents in the Design of Modern Crop Protection Compounds. *Modern Methods in Crop Protection Research*, pp. 73–128.
- Kamke J, Taylor MW & Schmitt S (2010) Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *ISME J* **4**: 498–508.
- Karimi E, Slaby BM, Soares AR, Blom J, Hentschel U & Costa R (2018) Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. *FEMS Microbiol Ecol* **94**: 1–18.
- Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP & Webb CO (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.
- Kennedy J, Marchesi JR & Dobson ADW (2008) Marine metagenomics: Strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microb Cell Fact* **7**.
- Kolde R (2012) pheatmap: {{Pretty Heatmaps}}. *R Packag version 108* 1–7. <https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf>.
- Kotik M (2009) Novel genes retrieved from environmental DNA by polymerase chain reaction: Current genome-walking techniques for future metagenome applications. *J Biotechnol* **144**: 75–82.
- Lahti L, Shetty S, Blake T & Salojarvi J (2017) Tools for microbiome analysis in R. *Bioconductor*. <http://microbiome.github.io/microbiome>.
- Larkin MA, Blackshields G, Brown NP, *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Latham J, Brandenburger E, Shepherd SA, Menon BRK & Micklefield J (2018) Development of Halogenase Enzymes for Use in Synthesis. *Chem Rev* **118**: 232–269.
- Letunic I & Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242–W245.
- Liao L, Chen R, Jiang M, Tian X, Liu H, Yu Y, Fan C & Chen B (2016) Bioprospecting potential of halogenases from Arctic marine actinomycetes. *BMC Microbiol* **16**: 1–9.
- Lira NS, Montes RC, Tavares JF, Da Silva MS, Da Cunha EVL, De Athayde-Filho PF, Rodrigues LC, Da Silva Dias C & Barbosa-Filho JM (2011)

- Brominated compounds from marine sponges of the genus *Aplysina* and a compilation of their ¹³C NMR spectral data. *Mar Drugs* **9**: 2316–2368.
- Loh T-L & Pawlik JR (2014) Chemical defenses and resource trade-offs structure sponge communities on Caribbean coral reefs. *Proc Natl Acad Sci* **111**: 4151–4156.
- Loureiro C, Medema MH, van der Oost J & Sipkema D (2018) Exploration and exploitation of the environment for novel specialized metabolites. *Curr Opin Biotechnol* **50**: 206–213.
- Lozupone C, Lladser ME, Knights D, Stombaugh J & Knight R (2011) UniFrac: An effective distance metric for microbial community comparison. *ISME J* **5**: 169–172.
- Lu Y, Liu Y, Xu Z, Li H, Liu H & Zhu W (2012) Halogen bonding for rational drug design and new drug discovery. *Expert Opin Drug Discov* **7**: 375–383.
- McMurdie PJ & Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**.
- Mehub MF, Lei J, Franco C & Zhang W (2014) Marine sponge derived natural products between 2001 and 2010: Trends and opportunities for discovery of bioactives. *Mar Drugs* **12**: 4539–4577.
- Menon BRK, Latham J, Dunstan MS, Brandenburger E, Klemstein U, Leys D, Karthikeyan C, Greaney MF, Shepherd SA & Micklefield J (2016) Structure and biocatalytic scope of thermophilic flavin-dependent halogenase and flavin reductase enzymes. *Org Biomol Chem* **14**: 9354–9361.
- Milshcheyn A, Schneider JS & Brady SF (2014) Mining the metabiome: Identifying novel natural products from microbial communities. *Chem Biol* **21**: 1211–1223.
- Mohamed NM, Saito K, Tal Y & Hill RT (2010) Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. *ISME J* **4**: 38–48.
- Montalvo NF & Hill RT (2011) Sponge-Associated Bacteria Are Strictly Maintained in Two Closely Related but Geographically Distant Sponge Hosts. *Appl Environ Microbiol* **77**: 7207–7216.
- Morris RM, Rappé MS, Urbach E, Connon SA & Giovannoni SJ (2004) Prevalence of the Chloroflexi-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. *Appl Environ Microbiol* **70**: 2836–2842.
- Müller CA, Oberauer-Wappis L, Peyman A, Amos GCA, Wellington EMH & Berg G (2015) Mining for nonribosomal peptide synthetase and polyketide synthase genes revealed a high level of diversity in the Sphagnum bog metagenome. *Appl Environ Microbiol* **81**: 5064–5072.
- Murphy CD (2006) Recent developments in enzymatic chlorination. *Nat Prod Rep* **23**: 147–152.
- Neubauer PR, Widmann C, Wibberg D, Schröder L, Frese M, Kottke T, Kalinowski J, Niemann HH & Sewald N (2018) A flavin-dependent halogenase from metagenomic analysis prefers bromination over chlorination. *PLoS One* **13**: 1–21.
- O'Neill J (2016) *Tackling drug-resistant infections globally: Final report and recommendations*.
- Onaka H, Taniguchi S, Igarashi Y & Furumai T (2003) Characterization of the Biosynthetic Gene Cluster of Rebecamycin from *Lechevalieria aerocolonigenes* ATCC 39243. *Biosci Biotechnol Biochem* **67**: 127–138.
- Öztürk B, de Jaeger L, Smidt H & Sipkema D (2013) Culture-dependent and independent approaches for identifying novel halogenases encoded by *Crambe crambe* (marine sponge) microbiota. *Sci Rep* **3**: 1–9.
- P. D. Schloss, S. L. Westcott, T. Ryabin, *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Pawlik JR (2011) The Chemical Ecology of Sponges on Caribbean Reefs: Natural Products Shape Natural Systems. *Bioscience* **61**: 888–898.
- Payne JT, Andorfer MC & Lewis JC (2013) Regioselective arene halogenation using the FAD-dependent halogenase RebH. *Angew Chemie - Int Ed* **52**: 5271–5274.
- Pearson WR (2014) An Introduction to Sequence Similarity (“Homology”) Searching. *Curr Protoc Bioinforma* **1**: 1286–1292.

- Van Pée KH & Patallo EP (2006) Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl Microbiol Biotechnol* **70**: 631–641.
- Van Pée KH & Unversucht S (2003) Biological dehalogenation and halogenation reactions. *Chemosphere* **52**: 299–312.
- Van Pée KH & Zehner S (2003) Enzymology and molecular genetics of biological halogenation. *Handbook of Environmental Chemistry*, Vol. 3, pp. 171–199.
- Pereyra LP, Hiibel SR, Prieto Riquelme M V, Reardon KF & Pruden A (2010) Detection and quantification of functional genes of cellulosedegrading, fermentative, and sulfate-reducing bacteria and methanogenic archaea. *Appl Environ Microbiol* **76**: 2192–2202.
- Puyana M, Pawlik J, Blum J & Fenical W (2015) Metabolite variability in caribbean sponges of the genus *aplysina*. *Brazilian J Pharmacogn* **25**: 592–599.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J & Glöckner FO (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* **41**: 590–596.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R & Lopez R (2005) InterProScan: Protein domains identifier. *Nucleic Acids Res* **33**.
- Ridley CP, Bergquist PR, Harper MK, Faulkner DJ, Hooper JNA & Haygood MG (2005) Speciation and biosynthetic variation in four dictyoceratid sponges and their cyanobacterial symbiont, *Oscillatoria spongelliae*. *Chem Biol* **12**: 397–406.
- Rocha-Martin J, Harrington C, Dobson ADW & O’Gara F (2014) Emerging strategies and integrated systems microbiology technologies for biodiscovery of marine bioactive compounds. *Mar Drugs* **12**: 3516–3559.
- Rodríguez-R LM & Konstantinidis KT (2014) Bypassing cultivation to identify bacterial species. *Microbe* **9**: 111–118.
- Rognes T, Flouri T, Nichols B, Quince C & Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.
- Rua CPJ, de Oliveira LS, Froes A, *et al.* (2018) Microbial and Functional Biodiversity Patterns in Sponges that Accumulate Bromopyrrole Alkaloids Suggest Horizontal Gene Transfer of Halogenase Genes. *Microb Ecol* **76**: 825–838.
- Sacristán-Soriano O, Banaigs B, Casamayor EO & Becerro MA (2011) Exploring the links between natural products and bacterial assemblages in the sponge *aplysina aerophoba*. *Appl Environ Microbiol* **77**: 862–870.
- Sacristán-Soriano O, Banaigs B & Becerro MA (2016) Can light intensity cause shifts in natural product and bacterial profiles of the sponge *Aplysina aerophoba*? *Mar Ecol* **37**: 88–105.
- Della Sala G, Hochmuth T, Costantino V, Teta R, Gerwick W, Gerwick L, Piel J & Mangoni A (2013) Polyketide genes in the marine sponge *Plakortis simplex*: A new group of mono-modular type I polyketide synthases from sponge symbionts. *Environ Microbiol Rep* **5**: 809–818.
- Sasaki T, Massaki N & Kubo T (2005) *Wolbachia* variant that induces two distinct reproductive phenotypes in different hosts. *Heredity (Edinb)* **95**: 389–393.
- Schmitt S, Deines P, Behnam F, Wagner M & Taylor MW (2011) Chloroflexi bacteria are more diverse, abundant, and similar in high than in low microbial abundance sponges. *FEMS Microbiol Ecol* **78**: 497–510.
- Schmitt S, Tsai P, Bell J, *et al.* (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564–576.
- Shepherd SA, Karthikeyan C, Latham J, Struck AW, Thompson ML, Menon BRK, Styles MQ, Levy C, Leys D & Micklefield J (2015) Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem Sci* **6**: 3454–3460.
- Sipkema D (2017) Marine biotechnology: diving deeper for drugs. *Microb Biotechnol* **10**: 7–8.
- Sipkema D, Schippers K, Maalcke WJ, Yang Y, Salim S & Blanch HW (2011) Multiple approaches to enhance the cultivability of bacteria associated with the marine sponge *Haliclona (gellius) sp.* *Appl Environ Microbiol* **77**: 2130–2140.
- Slaby BM, Hackl T, Horn H, Bayer K & Hentschel U (2017) Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J* **11**: 2465–2478.
- Smith DRM, Gruschow S & Goss RJM (2013) Scope and potential of halogenases in biosynthetic applications. *Curr Opin Chem Biol* **17**: 276–

- Smith DRM, Uria AR, Helfrich EJN, Milbredt D, Van Pée KH, Piel J & Goss RJM (2017) An unusual flavin-dependent halogenase from the metagenome of the marine sponge *Theonella swinhoei* WA. *ACS Chem Biol* **12**: 1281–1287.
- Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Taylor MW, Radax R, Steger D & Wagner M (2007) Sponge-Associated Microorganisms: Evolution, Ecology, and Biotechnological Potential. *Microbiol Mol Biol Rev* **71**: 295–347.
- Thomas T, Moitinho-Silva L, Lurgi M, *et al.* (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* **7**: 11870.
- Thoms C, Ebel R & Proksch P (2006) Activated chemical defense in Aplysina sponges revisited. *J Chem Ecol* **32**: 97–123.
- Thrash JC, Cho JC, Ferreira S, Johnson J, Vergin KL & Giovannoni SJ (2010) Genome sequences of strains HTCC2148 and HTCC2080, belonging to the OM60/NOR5 clade of the Gammaproteobacteria. *J Bacteriol* **192**: 3842–3843.
- Turon X, Becerro MA & Uriz MJ (2000) Distribution of brominated compounds within the sponge *Aplysina aerophoba*: Coupling of X-ray microanalysis with cryofixation techniques. *Cell Tissue Res* **301**: 311–322.
- Unson MD, Holland ND & Faulkner DJ (1994) A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Mar Biol* **119**: 1–11.
- Vaklu J, Sudan AK & Johri BN (2008) Metagenomics: Future of microbial gene mining. *Indian J Microbiol* **48**: 202–215.
- Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA & Patel HM (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol* **5**: 525–534.
- Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Wang Q, Wu H, Wang A, Du P, Pei X, Li H, Yin X, Huang L & Xiong X (2010) Prospecting metagenomic enzyme subfamily genes for DNA family shuffling by a novel PCR-based approach. *J Biol Chem* **285**: 41509–41516.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.
- Webster NS & Taylor MW (2012) Minireview Marine sponges and their microbial symbionts: love. *Environ Microbiol* **14**: 335–346.
- Weichold V, Milbredt D & Van Pée KH (2016) Specific enzymatic halogenation - From the discovery of halogenated enzymes to their applications in vitro and in vivo. *Angew Chemie - Int Ed* **55**: 6374–6389.
- Wickham H (2016) *Ggplot2*. Springer-Verlag New York <http://ggplot2.org>.
- Wilson MC & Piel J (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* **20**: 636–647.
- Wilson MC, Mori T, Rückert C, *et al.* (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**: 58–62.
- Xu G & Wang BG (2016) Independent evolution of six families of halogenating enzymes. *PLoS One* **11**: 1–13.
- Yeh E, Blasiak LC, Koglin A, Drennan CL & Walsh CT (2007) Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **46**: 1284–1292.
- Zea S, Henkel TP & Pawlik JR (2014) *The Sponge Guide: a picture guide to Caribbean Sponges*. 3rd ed. www.spongeguide.org.
- Zehner S, Kotsch A, Bister B, Süßmuth RD, Méndez C, Salas JA & Van Pée KH (2005) A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem Biol* **12**: 445–452.
- Zhao J, Yang N & Zeng R (2008) Phylogenetic analysis of type I polyketide synthase and nonribosomal peptide synthetase genes in Antarctic sediment. *Extremophiles* **12**: 97–105.

Zhao Y, Yan B, Yang T, Jiang J, Wei H & Zhu X (2015) Purification and crystallographic analysis of a FAD-dependent halogenase from *Streptomyces* sp. JCM9888. *Acta Crystallogr Sect Struct Biol Commun* **71**: 972–976.

Uncorrected Proof

Figure Legends

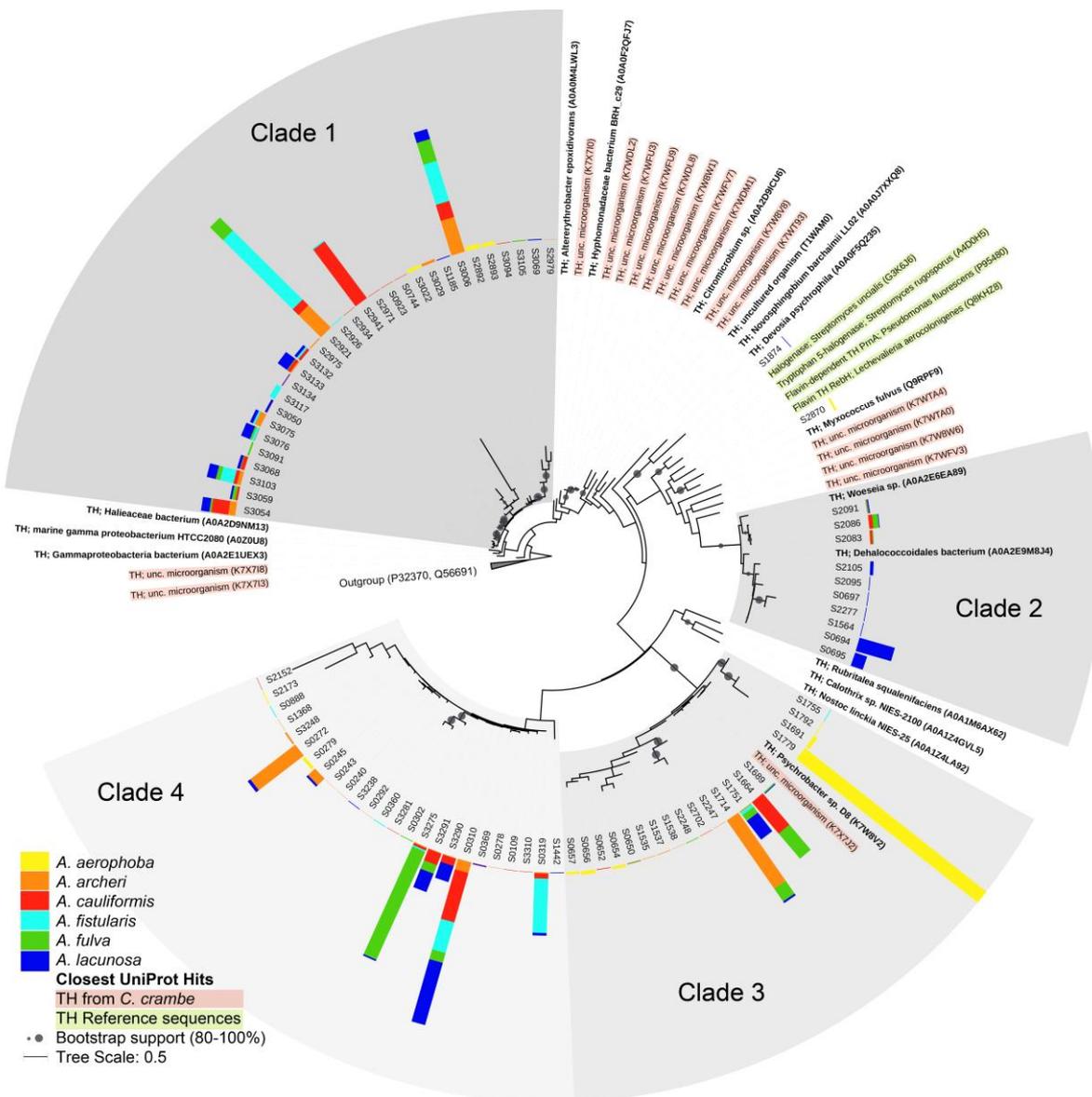


Figure 1: Maximum Likelihood Phylogeny of putative tryptophan halogenase (TH) protein sequences from six *Aplysia* species. Sequences obtained in this study are labelled with S. The tree was constructed from 247 amino acid positions. Two NADH-dependent flavin oxidoreductases (P32370 from *Clostridium scindens* and Q56691 from *Vibrio harveyi*) were used as outgroup. Four well-studied TH obtained from UniProt (green) and putative TH from the sponge *Crambe crambe* (pink) are included as reference sequences. Bars indicate relative abundance of the sequences in the sponge species, with the outer rim of the grey shading indicating 88%. Bootstrap values >80% are indicated by grey circles at the branch points. Sequences in bold refer to the closest relatives from the “halogenase” database. UniProt sequence accession numbers are given inside brackets. Tree scale corresponds to the mean expected number of amino acid substitutions per site.

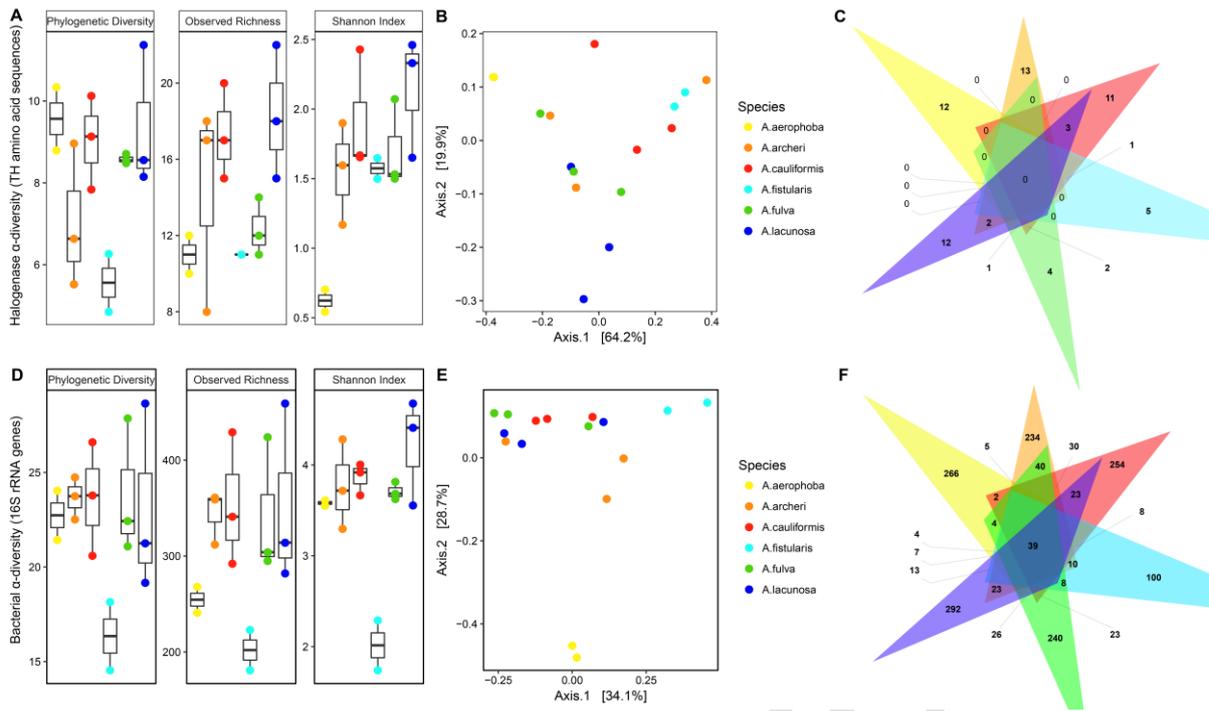


Figure 2: Halogenase and bacterial 16S rRNA gene alpha diversity indices (A,D). PCoA analysis of weighted UniFrac distances on relative abundance information of tryptophan halogenase (TH) amino acid sequences (B) and 16S rRNA genes (E). *A. aerophoba* replicates are almost identical and thus overlap in plot B. Venn diagrams of shared and unique TH sequences (C) and 16S rRNA genes (F) per sponge species.

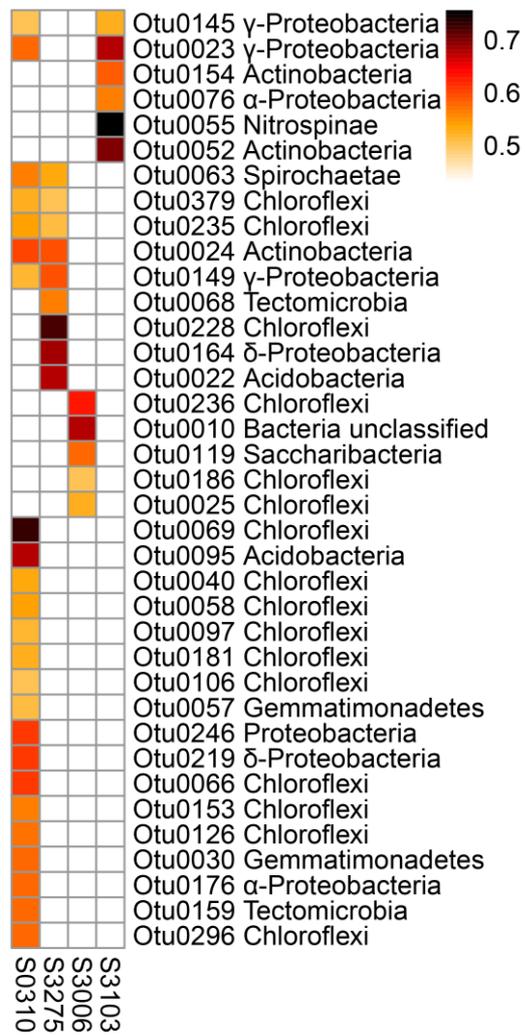


Figure 3: Heatmap displaying significant (Spearman $r > 0.5$, $p < 0.05$) co-occurrence correlations of tryptophan halogenases (columns) and bacterial 16S rRNA gene OTUs (rows) shared between all Caribbean Aplysina species. OTU IDs and phyla are given to identify bacterial taxa. Columns and rows are clustered by Euclidian distance. Colors indicate correlation coefficients ranging from 0.5 (orange) to 1 (black).

Table 1: Metadata, sample information and results of 16S rRNA gene as well as halogenase gene sequencing of the *Aplysina* species analyzed in this study. THs: tryptophan halogenases. (*) Samples did not pass quality criteria and were excluded from the analyses. Putative halogenase sequences were clustered at the amino acid sequence level.

Sample ID	Species	Sea	Latitude	Longitude	Depth [m]	Temp.	Preservation	16S rRNA genes			Halogenases	
								Reads	OTUs (97% Phyla)	Reads	Putative THs (95%)	
AAe1	<i>Aplysina aerophoba</i>	Mediterranean	42.114	3.168	8	NA	Cryopreservation	8707	268	17	2813	10
AAe2*	<i>Aplysina aerophoba</i>	Mediterranean	42.114	3.168	8	NA	Cryopreservation	NA	NA	NA	NA	NA
AAe3	<i>Aplysina aerophoba</i>	Mediterranean	42.115	3.168	12	NA	Cryopreservation	6032	241	19	1716	12
AAr1	<i>Aplysina archeri</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	9258	312	17	844	8
AAr2	<i>Aplysina archeri</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	9786	359	16	2927	18
AAr3	<i>Aplysina archeri</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	5397	361	16	1631	17
ACa1	<i>Aplysina cauliformis</i>	Caribbean	12.026	-68.251	16	29°C	RNA Later	11664	429	18	907	15
ACa2	<i>Aplysina cauliformis</i>	Caribbean	12.026	-68.251	16	29°C	RNA Later	5523	341	18	1218	20
ACa3	<i>Aplysina cauliformis</i>	Caribbean	12.026	-68.251	16	29°C	RNA Later	3890	292	14	1210	17
AFi1	<i>Aplysina fistularis</i>	Caribbean	12.094	-68.232	1	30°C	RNA Later	4789	223	16	658	11
AFi2	<i>Aplysina fistularis</i>	Caribbean	12.094	-68.232	1	30°C	RNA Later	8429	181	17	688	11
AFi3*	<i>Aplysina fistularis</i>	Caribbean	12.094	-68.232	1	30°C	RNA Later	NA	NA	NA	NA	NA
AFu1	<i>Aplysina fulva</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	9995	304	18	5324	12
AFu2	<i>Aplysina fulva</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	9988	295	17	2110	11
AFu3	<i>Aplysina fulva</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	10140	424	17	4308	14
ALa1	<i>Aplysina lacunosa</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	4896	314	16	2360	15
ALa2	<i>Aplysina lacunosa</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	7198	459	18	6976	22
ALa3	<i>Aplysina lacunosa</i>	Caribbean	12.160	-68.283	10	28°C	RNA Later	6450	282	15	4497	18

Table 2: Hit table (blastp) of sequences aligned against the manually curated halogenase protein database. *TH:

Tryptophan halogenase, THf: Tryptophan halogenase (Fragment). %ID is given as % amino acid sequence identity.

No. of sequences	Closest match UniProt accession	Protein name	Organism	%ID min.	%ID max.	Bitscore avg.	Publication
25	A0Z0U8	TH	marine Gammaproteobacterium HTCC2080	67.3	80.4	223	(Thrash <i>et al.</i> , 2010)
19	K7W8V2	THf	<i>Psychrobacter</i> sp. D8	82.1	100	194	(Öztürk <i>et al.</i> , 2013)
15	A0A1Z4GVL5	TH	<i>Calothrix</i> sp. NIES-2100	36.2	51.2	116	(Hirose <i>et al.</i> , 2017)
7	A0A2E9M8J4	TH	Dehalococcoidales bacterium	55.9	76.2	169	(Tully <i>et al.</i> , 2018)
5	A0A1Z4LA92	TH	<i>Nostoc linckia</i> NIES-25	38.2	39.9	118	(Tully <i>et al.</i> , 2018)
3	A0A2D9NM13	TH	Haliaceae bacterium	71.2	83.9	172	(Tully <i>et al.</i> , 2018)
3	A0A2E6EA89	TH	<i>Woeseia</i> sp.	75.5	76.2	245	(Bagnoud <i>et al.</i> , 2015)
1	A0A0F2QFJ7	TH	Hyphomonadaceae bacterium BRH_c29	67.7	67.7	84	(Bagnoud <i>et al.</i> , 2015)
1	A0A0F5Q235	TH	<i>Devosia psychrophila</i>	36.2	36.2	93	(Lepp <i>et al.</i> , 2015)
1	A0A0J7XXQ8	TH	<i>Novosphingobium barchaimii</i> LL02	32.9	32.9	46	(Pearce <i>et al.</i> , 2015)
1	A0A0M4LWL3	TH	<i>Altererythrobacter epoxidivorans</i>	42.6	42.6	48	(Li <i>et al.</i> , 2016)
1	A0A1M6AX62	TH	<i>Rubritalea squalenifaciens</i> DSM 18772	37.9	37.9	100	(Varghese, 2016)
1	A0A2D9ICU6	TH	<i>Citromicrobium</i> sp.	45.5	45.5	41	(Tully <i>et al.</i> , 2018)
1	A0A2E1UEX3	TH	Gammaproteobacteria bacterium	80.7	80.7	200	(Tully <i>et al.</i> , 2018)
1	Q9RPF9	TH	<i>Myxococcus fulvus</i>	53.9	53.9	182	(Hammer <i>et al.</i> , 1999)
1	T1WAM0	TH	uncultured organism	39.5	39.5	48	(Nyyssönen <i>et al.</i> , 2013)