

Article

Insights into Red Sea Brine Pool Specialized Metabolism Gene Clusters Encoding Potential Metabolites for Biotechnological Applications and Extremophile Survival

Laila Ziko ¹, Mustafa Adel ^{1,2}, Mohamed N. Malash ^{2,3} and Rania Siam ^{2,*}

¹ Graduate Program of Biotechnology, School of Sciences and Engineering, American University in Cairo, New Cairo, Cairo 11835, Egypt; laila.adel@aucegypt.edu (L.Z.); mustafa.adel@aucegypt.edu (M.A.)

² Biology Department, School of Sciences and Engineering, American University in Cairo, New Cairo, Cairo 11835, Egypt; mnmalash@aucegypt.edu

³ Microbiology and Immunology Department, Faculty of Pharmacy, Ahram Canadian University, Giza 12581, Egypt

* Correspondence: rsiam@aucegypt.edu; Tel.: +20-2-2615-2907

Received: 25 March 2019; Accepted: 19 April 2019; Published: 8 May 2019



Abstract: The recent rise in antibiotic and chemotherapeutic resistance necessitates the search for novel drugs. Potential therapeutics can be produced by specialized metabolism gene clusters (SMGCs). We mined for SMGCs in metagenomic samples from Atlantis II Deep, Discovery Deep and Kebrit Deep Red Sea brine pools. Shotgun sequence assembly and secondary metabolite analysis shell (antiSMASH) screening unraveled 2751 Red Sea brine SMGCs, pertaining to 28 classes. Predicted categorization of the SMGC products included those (1) commonly abundant in microbes (saccharides, fatty acids, aryl polyenes, acyl-homoserine lactones), (2) with antibacterial and/or anticancer effects (terpenes, ribosomal peptides, non-ribosomal peptides, polyketides, phosphonates) and (3) with miscellaneous roles conferring adaptation to the environment/special structure/unknown function (polyunsaturated fatty acids, ectoine, ladderane, others). Saccharide (80.49%) and putative (7.46%) SMGCs were the most abundant. Selected Red Sea brine pool sites had distinct SMGC profiles, e.g., for bacteriocins and ectoine. Top promising candidates, SMs with pharmaceutical applications, were addressed. Prolific SM-producing phyla (Proteobacteria, Actinobacteria, Cyanobacteria), were ubiquitously detected. Sites harboring the largest numbers of bacterial and archaeal phyla, had the most SMGCs. Our results suggest that the Red Sea brine niche constitutes a rich biological mine, with the predicted SMs aiding extremophile survival and adaptation.

Keywords: specialized metabolism gene clusters; Red Sea brine pools; extremophiles

1. Introduction

In the era of antibiotic resistance and a concern for a post-antibiotic era there is a pressing need to combat resistance and discover novel antibiotics [1–3]. In the USA alone, around 2 million people a year acquire a bacterial infection that is resistant to all available antibiotics [4]. Additionally, anticancer chemotherapeutic resistance is another recent biomedical challenge, which arises either intrinsically or extrinsically, following therapy [5]. Therefore, it is a necessity to search for new chemotherapeutics [3–5].

Nature is considered a mine to explore for small molecules, which may be used as new therapeutic drug leads. Until 2014, a large portion of the rising number of drugs was attributed to natural products [6]. Many organisms and microbes produce specialized metabolites that have a plethora

of functions [7]. These microbes' unique metabolic activities result in the production of specialized metabolites that facilitate their survival in highly competitive environments [7]. Specialized metabolites are mostly coded in the host genomes as clusters of genes, the specialized metabolism gene clusters (SMGCs), which occupy large chunks within prokaryotic genomes [8]. Examples of SMGCs are those encoding for polyketide synthases and non-ribosomal peptide synthases [9]. Compounds with novel chemistry have been identified by prior detection of these biosynthetic gene clusters, such as salinilactam antibiotic (Figure S1A) [10,11]. Some SMGCs are responsible for producing bioactive compounds and have guided the discovery of novel antibiotics [12]. For example, the search for SMGCs in the human microbiome led to the discovery of a new antibiotic named lactocillin (Figure S1B), which is a small molecule produced by a prevalent strain in the human vagina [12]. Several methods were previously developed to detect specialized metabolites and SMGCs, from individual microbes and from communities of microbes [8]. To mine for SMGCs in microbes inhabiting different environments, metagenomic approaches are employed [8]. One approach utilizes environmental DNA isolation, followed by shotgun sequencing and then computational search for SMGCs by bioinformatic tools such as the antibiotics and secondary metabolite analysis shell (antiSMASH) pipeline [9,13,14].

An interesting niche to discover new bioactive compounds is the marine environment, which hosts ample and highly diverse microbes [15]. Previous studies showed that marine organisms produce a large collection of compounds with diverse activities including anticancer and antibacterial agents, such as streptochlorin and abyssomicin C (Figure S1C,D) [16]. Deep-sea microbes are predicted to have a huge biosynthetic versatility as they are able to produce more specialized metabolites of novel chemistry than microbes inhabiting surface seawater [17]. More than 300 new natural products have been isolated from organisms inhabiting the deep sea [17]. Some microorganisms living in extreme deep-sea environments are considered extremophiles. Microbes thriving in more than one extreme condition, i.e. polyextremophiles may thrive under relatively high pressure, salinity, extreme temperatures, scarce nutrients and different pH conditions [18]. The Red Sea brine pools, including Atlantis II Deep (ATII), Discovery Deep (DD) and Kebrit Deep (KD), are known for polyextremophilic conditions [19–22]. The largest of the 25 Red Sea brine pools is ATII, which is characteristic by hypersalinity (252 psu), high temperature (~67.1 °C) and high metal content [20,23–25]. The differential salinity, temperature and pH of the brine water allows the vertical stratification into different layers; an upper convective Layer (ATII UCL), a lower convective Layer (ATII LCL) and brine-seawater interface (ATII INF) [21,26]. DD brine pool is geochemically similar to ATII, as they are in close proximity and have subsurface connections [21,23,24,26]. However, the conditions in the DD brine pool are less harsh than in ATII brine pool [20]. Although KD is not a hot brine, it is characterized by high H₂S concentration (maximum of 14 mg sulfur/L) [20,23,24,27]. Beneath the brine pools are sediments that are rich in heavy metals [23,28,29].

The metabolomes of extremophiles are currently being explored for the discovery of new drugs, and lead to the discovery of new compounds such as salinosporamide K (Figure S1E) [18,30]. Furthermore, Red Sea *Bacillus paralicheniformis* strains were found to harbor more putative biosynthetic gene clusters, and code for potentially bioactive metabolites, such as non-ribosomal peptides [31]. Isolates from Red Sea coastal sediments were reported to have potential antimicrobial activity [32]. Several extremozymes have been isolated and characterized from the ATII Red Sea brine pool as well, such as a thermophilic esterase, a nitrilase and antibiotic-resistant enzymes [33–35].

In this study we investigated specialized metabolism gene clusters in Red Sea brine water and sediment samples through metagenomic mining. Metagenomes from different sites in the Atlantis II, Discovery Deep and Kebrit Deep brine pools [28,36] were comprehensively studied to determine quantitatively and qualitatively SMGCs in the selected Red Sea brine prokaryotic metagenomes. Red Sea brine SMGCs' homology to known biosynthetic gene clusters, predicted functions and structures of the encoded potential specialized metabolites are addressed, and their potential contribution in shaping the microbial communities in such harsh environments are discussed.

2. Results

2.1. Abundance and Diversity of Specialized Metabolism Gene Clusters (SMGCs) in Red Sea Brine Pools

The study workflow is depicted in Figure 1, and the detailed sample locations are available in Table S1. DNA from Red Sea brine water and sediment samples were previously shotgun sequenced [19,20,22]. In this study we assembled 12,968,227 reads, generating a total of 349,631 contigs (Table 1). The contigs (assembled metagenomes) were used to investigate the SMGCs of the Red Sea brine pools. The assembly metrics are denoted in Table 1. In order to eliminate the effects of assembly size bias in downstream analyses, the number of assembled reads for each site was used to normalize the number of detected SMGCs (Table 1).

ATII LCL had the largest, and ATII 1500 had the smallest number of assembled reads, in the water samples (3,901,967 and 316,101 assembled reads, respectively) (Table 1). Pertaining to the sediment (SDM) samples, DD SDM and ATII SDM cross assemblies comprised the most assembled reads (597,552 and 478,453 assembled reads respectively) (Table 1). NB SDM had the least number of assembled reads amongst all the sites in the dataset (92,530 assembled reads) (Table 1).

A total of 2751 SMGCs (absolute value) were detected in all the sites by antiSMASH tool (Table 2 and Table S2). The normalized SMGC values are recorded in Table S3. The deepest water samples in the Atlantis II Deep water column samples (ATII 1500 and ATII 700) and the deepest Kebrut Deep brine layer (KD brine–seawater Lower Interface, KD LINF) had the highest number of detected SMGCs; 531.48, 485.27 and 360.56 respectively. The average number of detected SMGCs was 247.03 per site (Table 2). Based on the abundance of SMGCs detected in all sites, three SMGCs cohorts were identified. Cohort 1—representing >7% of total SMGCs: cf_saccharide (80.49%), cf_putative (7.46%) and cf_fatty_acid (7.34%), 2—representing 0.29–7% of total SMGCs: terpenes (1.89%), bacteriocins (0.46%), tther (0.38%), arylpolyene (0.32%), and T3PKS (0.30%) and 3—representing 0.21–0.01% of total SMGCs: NRPS, aryl-polyene-cf_fatty_acid, microcin, lantipeptide, OtherKS, Cf_saccharide-bacteriocin, otherks-Pufa-T1pks, T2pks-Cf_fatty_acid, ectoine, ladderane-Cf_fatty_acid, Cf_fatty_acid-arylpolyene, otherks-Pufa, otherks-T1pks, hserlactone, Nrps-T1pks, T1PKS, Cf_fatty_acid-Cf_saccharide, Cf_saccharide-nrps, phosphonate and T3pks-cf_saccharide (Figure 2).

The Cf_saccharide gene cluster was the only SMGC detected in all the Red Sea sites. Selected SMGCs were unique to specific sample sites, namely, Cf_fatty_acid-Arylpolyene—was only detected in ATII 200, Otherks-Pufa and Otherks-T1pks were only detected in ATII 700, Otherks-Pufa-T1pks and T2pks-Cf_fatty_acid were only detected in ATII 1500, Cf_fatty_acid-Cf_saccharide and Cf_saccharide-nrps and Phosphonate and T3pks-cf_saccharide were only detected in ATII LCL, Cf_saccharide-Bacteriocin and Hserlactone were only detected in KD LINF and Ladderane-Cf_fatty_acid and Nrps-T1pks and T1PKS were only detected in KD brine–seawater Upper Interface (KD UINF). Additionally, among the non-Red-Sea dataset, Pufa (polyunsaturated fatty acid cluster) was unique to Guaymas Basin deep-sea hydrothermal vent plume water sample (GB VNT), thiopeptide cluster was only detected in Kueishantao shallow-sea hydrothermal vent area water sample KSW VNT and acyl_amino_acids (*N*-acyl amino acid cluster) was only detected in Loki's Castle deep-sea vent black smoker chimney microbial mat sample (LC MM) (Table 2, Tables S2 and S3).

Our analysis resulted in the detection of 28 distinct SMGCs and the diversity of SMGCs in the Red Sea water and sediment samples was recorded (Figure 2, Table S3). The water samples harbored all of the detected 28 SMGC types and only 7 SMGC types were detected in the sediment samples. In order to compare the diversity of the SMGCs detected in our dataset, the absolute values of the different types of SMGCs were used for comparison (Table 2). KD UINF and ATII LCL sites, showed the highest diversity in SMGC types (13 different SMGC types) while DD brine–seawater interface (DD INF) had the lowest diversity (1 SMGC type) (Table 2).

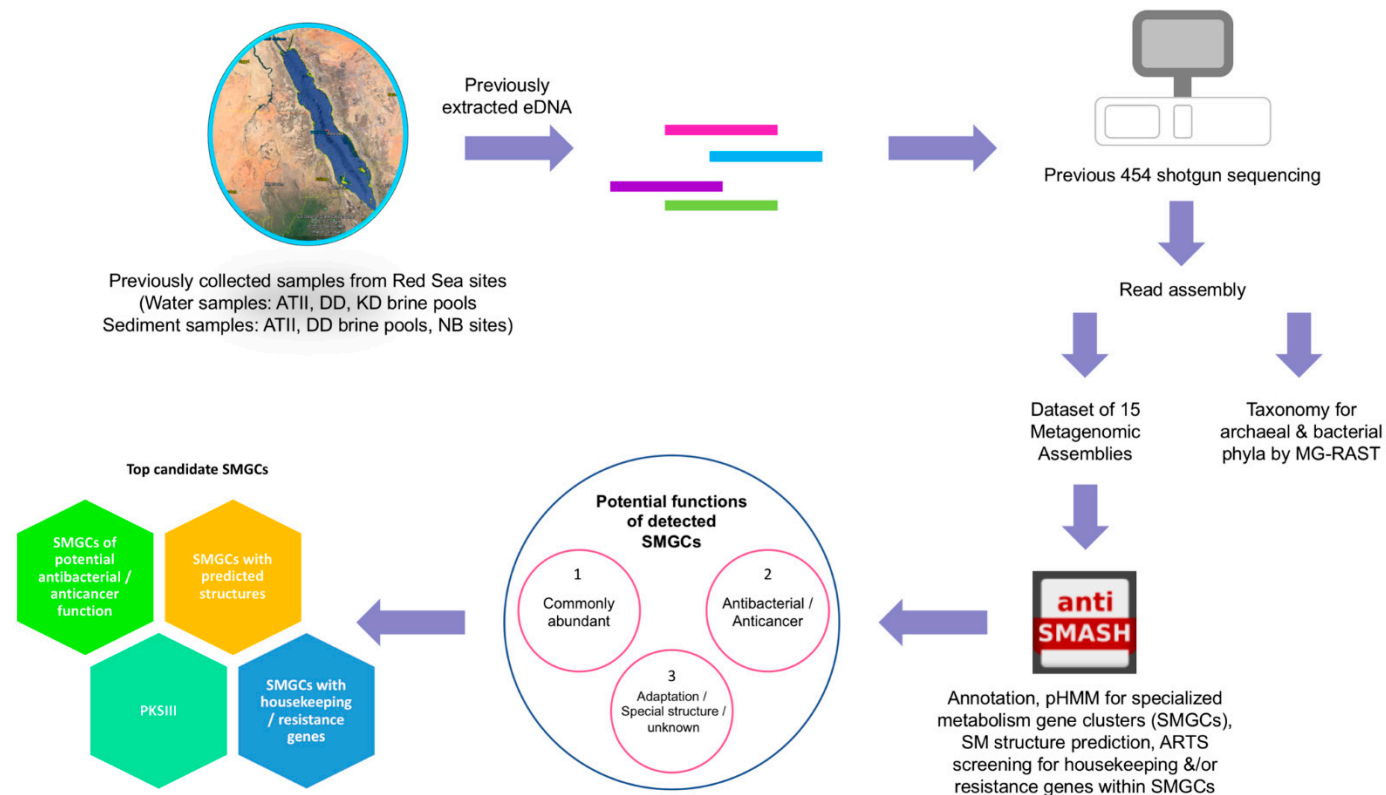


Figure 1. Study workflow for the analysis of specialized metabolism gene clusters in Red Sea brine pools. Water and sediment samples were earlier collected from Red Sea Atlantis II Deep (ATII), Discovery Deep (DD), Kebrit Deep (KD) brine pools and brine-influenced (NBI, NBII) sites. Metagenomic prokaryotic DNA was then extracted from each site and 454 shotgun sequencing was performed followed by read assembly. Taxonomic classification for archaeal and bacterial phyla was performed by protein-based phylogeny using the metagenomics rapid annotation using subsystems technology (MG-RAST) tool [37]. The metabolite analysis shell (AntiSMASH) tool was then used for annotation, for identifying specialized metabolism gene clusters (SMGCs) by translated amino acid sequence comparison with signature biosynthetic genes profile hidden Markov Model (pHMMs), and for structure prediction of the specialized metabolites [9]. The Antibiotic Resistance Target Seeker (ARTS) tool was used to detect housekeeping and/or resistance genes within the SMGCs [38]. The predicted specialized metabolites were grouped by their potential functions into three major groups. Lastly, top candidate SMGCs were identified in the Red Sea brine dataset.

Table 1. Assembly metrics denoted for each of the Red Sea brine pool and other hydrothermal vent samples.

Description	Detailed Description	Sites	Reference Sampling & Read Sequences/Assembly	Number of Reads	Number of Reads after Trimming	Number of Assembled Reads	Number of Contigs > 500 bp	Average Contig Size (bp)	Largest Contig Size (bp)
Atlantis II water column (Water above brine pool)	Atlantis II 50 m water column	ATII 50	[19]/This study	1,461,910	1,461,904	582,768	36,262	1149	21,887
	Atlantis II 200 m water column	ATII 200		1,260,578	1,260,561	530,441	34,640	1131	25,392
	Atlantis II 700 m water column	ATII 700		1,128,514	1,128,507	554,335	32,860	1285	33,783
	Atlantis II 1500 m water column	ATII 1500		833,739	833,730	316,101	20,374	1331	51,927
Atlantis II brine water	Atlantis II brine-seawater interface	ATII INF	[20]/This study	832,138	832,128	743,064	9933	1214	25,150
	Atlantis II brine Upper Convective Layer	ATII UCL		886,030	886,019	794,715	11,994	1454	103,389
	Atlantis II brine Lower Convective Layer	ATII LCL		4,104,966	4,104,994	3,901,967	19,165	2084	350,936
Discovery Deep brine water	Discovery Deep brine-seawater interface	DD INF	[20]/This study	1,095,181	1,095,157	752,025	14,144	1201	28,080
	Discovery Deep brine water	DD BR		1,111,044	1,111,032	763,387	15,306	1216	22,118
Kebrit Deep brine water	Kebrit Deep Upper brine-seawater interface	KD UINF	[20]/This study	1,562,521	1,562,512	1,020,749	24,517	1495	58,542
	Kebrit Deep Lower brine-seawater interface	KD LINF		1,510,272	1,510,262	926,337	31,983	1241	38,825
	Kebrit Deep brine water	KD BR		1,379,832	1,379,814	913,803	22,280	945	14,864
Other water metagenomes	Guaymas Basin deep-sea hydrothermal vent plume water	GB VNT	[39–41]/This study	628,619	628,569	155,841	7654	1082	17,353
	Kueishantao shallow-sea hydrothermal vent (water above vent)	KSW VNT	[42]/This study	261,446	261,399	199,237	411	4685	179,360
	Kueishantao shallow-sea hydrothermal vent (water)	K VNT	[42]/This study	444,655	444,597	338,480	2194	1843	88,498
	Juan de Fuca Ridge hydrothermal vent diffuse flow seawater	JDF VNT	[43]/This study	226,981	226,916	35,357	9366	1135	9366
Sediments	Atlantis II sediments	ATII SDM	[28,36]/This study	1,138,406	1,138,381	478,453	30,352	1194	33,674
	Discovery Deep sediments	DD SDM		1,258,290	1,258,273	597,552	38,529	1233	38,081
	Non-brine sediments	NB SDM		253,568	253,564	92,530	7292	1177	1315
Other metagenome (biofilm)	Loki's Castle deep-sea vent biofilm (microbial mat)	LC MM	[44]/This study	717,550	717,135	525,719	7897	1546	42,387
Total				22,096,240	22,095,454	14,222,861	377,153	-	-

Table 2. SMGCs abundance and the counts of archaeal and bacterial phyla in each site. SMGCs are named using the same acronyms as those used by antiSMASH tool [45]. Hyphens indicate hybrid clusters. Cf_fatty_acid: fatty acid putative cluster. Cf_saccharide: saccharide putative cluster. Hserlactone: cluster coding for homoserine lactone. NRPS: cluster coding for non-ribosomal peptide synthetase. OtherKS: cluster coding for other types of polyketide synthases. Pufa: cluster coding for poly-unsaturated fatty acids. T1pks: type I polyketide synthase. T2pks: type II polyketide synthase. T3pks: type III polyketide synthase. Acyl_amino_acids: cluster coding for N-acyl amino acid.

Detailed Description	Assembly	Number of SMGCs	Normalized Number of SMGCs *	Types of SMGCs	Number of Phyla	SMGCs Detected Uniquely Once at a Particular Site
Red Sea metagenomic samples:						
Atlantis II 1500 m water column	ATII 1500	168	531.48	9	33	Others-Pufa-T1pks, T2pks-Cf_fatty_acid
Atlantis II 700 m water column	ATII 700	269	485.27	8	32	Others-Pufa, Others-T1pks
Kebrit Deep Lower brine-seawater interface	KD LINF	334	360.56	9	33	Cf_saccharide-Bacteriocin, Hserlactone
Atlantis II 200 m water column	ATII 200	170	320.49	8	30	Cf_fatty_acid-Arylpolyene
Atlantis II brine Upper Convective Layer	ATII UCL	210	264.25	7	21	
Atlantis II 50 m water column	ATII 50	146	250.53	8	30	
Kebrit Deep Upper brine-seawater interface	KD UINF	252	246.88	13	32	Ladderane-Cf_fatty_acid, Nrps-T1pks, T1PKS
Atlantis II brine-seawater interface	ATII INF	162	218.02	6	19	
Discovery Deep sediments	DD SDM	114	190.78	4	23	
Non-brine sediments	NB SDM	16	172.92	4	9	
Kebrit Deep brine water	KD BR	149	163.05	4	23	
Atlantis II sediments	ATII SDM	70	146.30	7	26	
Atlantis II brine Lower Convective Layer	ATII LCL	524	134.29	13	25	Cf_fatty_acid-Cf_saccharide, Cf_saccharide-nrps, Phosphonate, T3pks-cf_saccharide
Discovery Deep brine-seawater interface	DD INF	94	125.00	1	20	
Discovery Deep brine water	DD BR	73	95.63	2	20	
Other metagenomic samples:						
Guaymas Basin deep-sea hydrothermal vent plume water	GB VNT	11	70.58	3	30	Pufa
Kueishantao shallow-sea hydrothermal vent (water above vent)	KSW VNT	11	55.21	4	12	Thiopeptide
Kueishantao shallow-sea hydrothermal vent (water)	K VNT	10	29.54	5	13	
Juan de Fuca Ridge hydrothermal vent diffuse flow seawater	JDF VNT	3	84.85	3	14	
Loki's Castle deep-sea vent biofilm (microbial mat)	LC MM	26	49.46	5	23	Acyl_amino_acids

* Normalized number of SMGCs are the number of SMGCs detected at each site divided by the number of assembled reads *10⁶.

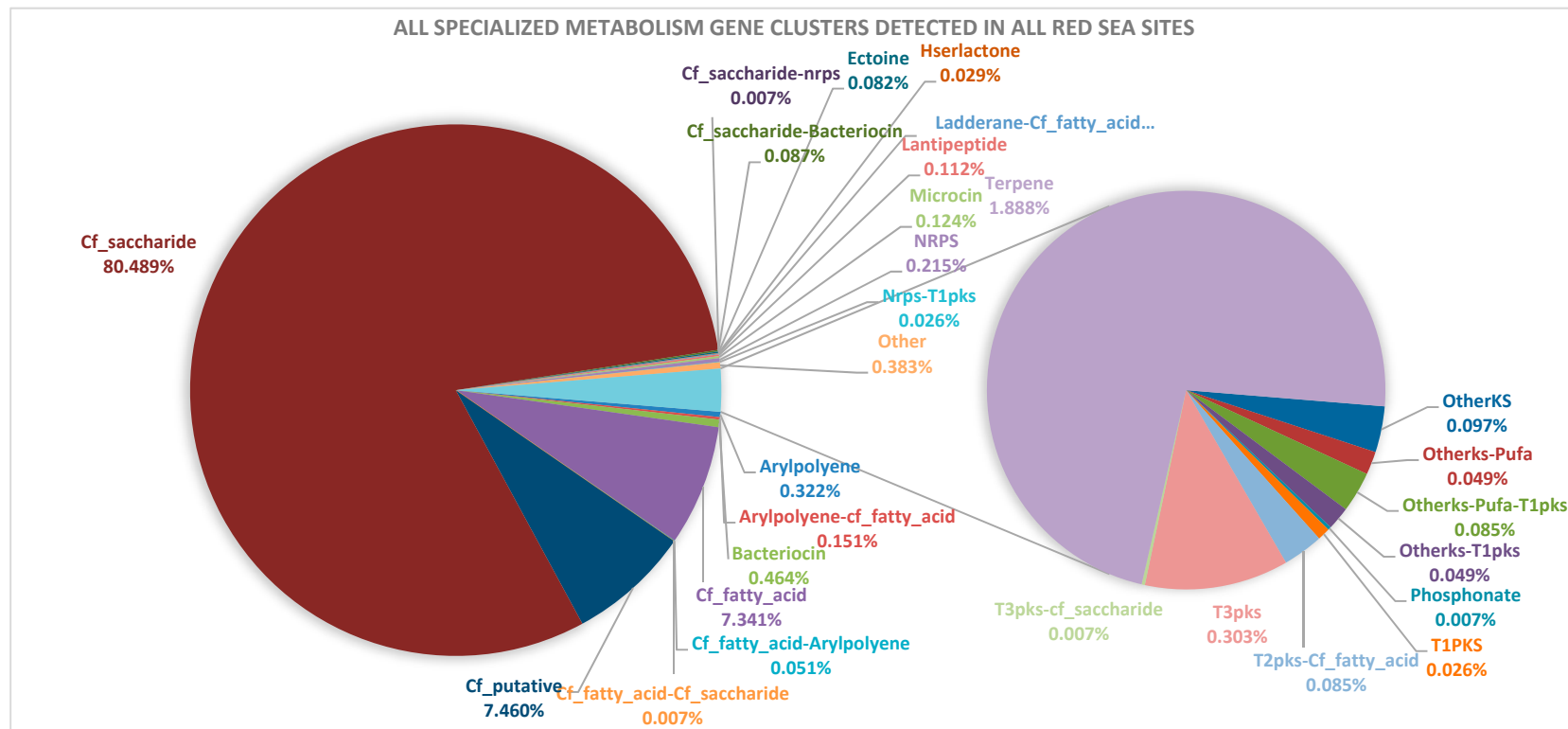


Figure 2. Overview of the specialized metabolism gene clusters encoded by the Red Sea brine pool metagenomes. The detected gene clusters are named as denoted by antiSMASH [9]. Normalized SMGC values were used.

2.2. Red Sea Brine Pool SMGCs Code for Diverse Potential Functions

We classified the predicted Red Sea SMGCs according to the potential function of their products into three main groups: (1) products of predicted functions commonly abundant in microbes, including saccharides, fatty acids, aryl polyenes and acyl-homoserine lactones, (2) subset of products with potential antibacterial and/or anticancer effects, including terpenes, ribosomal peptides, non-ribosomal peptides, polyketides and phosphonates, and (3) miscellaneous products that are predicted to confer adaptation to the environment/special structure/unknown function, including polyunsaturated fatty acids, ectoine, ladderane and others (Table 3). When applicable, the product class representative chemical structures are also depicted in Table 3, e.g., ladderane structure. Additionally, five predicted core structures of the potential products coded by the SMGCs of the Red Sea dataset were computationally predicted (Figure 3). Two core structures were predicted in the KD LINF layer for non-ribosomal peptides, both coded for putative NRPS clusters (Figure 3A). A chiral non-ribosomal peptide structure was predicted in KD UINF layer that was encoded by a putative NRPS cluster, as well as a hybrid polyketide-non-ribosomal peptide chiral structure encoded by a T1PKS-NRPS cluster (Figure 3B). The fifth structure was a polyketide predicted in ATII 1500 layer and was encoded by the putative hybrid cluster T1pks-pufa-otherks (Figure 3C). To prioritize specific SMGCs for further experimental work among all the detected SMGCs, Antibiotic Resistance Target Seeker (ARTS) analysis was used and revealed that four SMGCs harbored neighboring housekeeping and/or resistance genes (Figure 3D–F).

We identified a subset of SMGCs common in brine pool water samples and distinct from the sediment samples and the overlying water column. Six different SMGCs (Bacteriocin, cf_fatty_acid, cf_putative, cf_saccharide, other and T3PKS) were detected only in the ATII brine samples (ATII INF, ATII UCL, ATII LCL) and four SMGC types (Bacteriocin, cf_fatty_acid, cf_putative and cf_saccharide) were common among the KD samples (KD BR, KD LINF, KD UINF). Additionally, subsets of SMGCs were common in the water column overlying ATII; namely, the three SMGC types (cf_fatty_acid, cf_putative and cf_saccharide). Cf_saccharide was the only SMGC type common in both DD BR and DD INF. Note that four SMGCs (cf_fatty_acid, cf_putative, terpene and cf_saccharide) were detected in all of the sediment samples (ATII SDM, DD SDM and NB SDM) (Table S3).

When hierarchical clustering was computed for the water samples only (Figure 4A), i.e. excluding the sediment samples, most brine SMGCs reads grouped together (ATII LCL, ATII UCL, DD BR, DD INF, ATII INF, KD BR and KD LINF), except for KD UINF. Yet, ATII 50–1500 m and KD UINF water samples clustered together (Figure 4A). Hierarchical classification of the sediment samples revealed that brine pool SMGCs clustered together (ATII SDM and DD SDM) (Figure 4B). When all the samples in the dataset were used as an input for hierarchical classification (i.e. Red Sea and non-Red Sea metagenomes), the non-Red Sea SMGCs clustered together as an outgroup to Red Sea samples (Figure S2).

Table 3. The potential functions of the array of specialized metabolites encoded by Red Sea brine SMGCs. SMGCs are named as denoted by antiSMASH tool [45]. Hyphens indicate hybrid clusters. Cf_fatty_acid: fatty acid putative cluster. Cf_putative: unknown type putative cluster. Cf_saccharide: saccharide putative cluster. Hserlactone: cluster coding for homoserine lactone. Lantipeptide: cluster coding for lanthipeptide. NRPS: cluster coding for non-ribosomal peptide synthetase. OtherKS: cluster coding for other types of polyketide synthases. Pufa: cluster coding for poly-unsaturated fatty acids. T1pks: type I polyketide synthase. T2pks: type II polyketide synthase. T3pks: type III polyketide synthase.

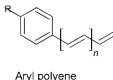
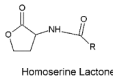

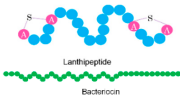

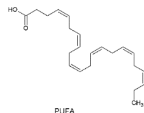


General Functional Classification:	Product (Enzyme)	Gene Cluster Names	Representative Basic Structure	Potential Function/Application of Product	Percentage of Total SMGCs	
1. Products of predicted functions commonly abundant in microbes	Saccharide	Cf_saccharide Cf_saccharide-Bacteriocin Cf_saccharide-nrps Cf_fatty_acid-Cf_saccharide T3pks-cf_saccharide	-	Microbe-host interactions e.g. lipopolysaccharides. Some saccharides that are diffusible were reported to have antibacterial activities [46,47].	80.61%	
	Fatty Acid	Cf_fatty_acid Arylpolyene-cf_fatty_acid Cf_fatty_acid-Arylpolyene Cf_fatty_acid-Cf_saccharide Ladderane-Cf_fatty_acid T2pks-Cf_fatty_acid	-	Structural functions and reported that composition can change as an adaptation to temperature and pressure in the environment [48].	7.69%	
	Aryl polyenes	Arylpolyene Arylpolyene-cf_fatty_acid Cf_fatty_acid-Arylpolyene	 Aryl polyene	Aryl polyene SMGCs found in abundance in Gram negative bacteria. Previously reported to have a protective role from damage caused by reactive oxygen species [46,49].	0.52%	
	Acyl-homoserine lactones	Hserlactone	 Homoserine Lactone	Quorum sensing [50].	0.03%	
2. Subset of products with potential antibacterial and/or anticancer effects:	Terpenes	Terpene	 Isoprene unit	A subset of the terpenes possesses antibacterial effect [51].	1.89%	
	Peptides	Ribosomal peptides	Bacteriocin Cf_saccharide-Bacteriocin Microcin Lantipeptide	 Lantipeptide	Some have antibacterial activity, and some have selective cancer cytotoxic activity [52].	0.78%
		Non-ribosomal peptides	Cf_saccharide-nrps NRPS Nrps-T1pks	-	Many non-ribosomal peptides have antibacterial (e.g., β -lactams) and anticancer (e.g. bleomycin) effects [53].	0.25%

Table 3. Cont.

General Functional Classification:	Product (Enzyme)	Gene Cluster Names	Representative Basic Structure	Potential Function/Application of Product	Percentage of Total SMGCs
Polyketides	(Type I Polyketide synthase)	Nrps-T1pks Otherks-Pufa-T1pks Otherks-T1pks T1PKS	-	A subset are responsible for antibiotic synthesis e.g. the type I polyketide synthase (PKSI) producing erythromycin [54].	0.2%
	(Type II Polyketide synthase)	T2pks-Cf_fatty_acid	-	Some type II polyketide synthase (PKSII) enzymes produce aromatic polyketide antibiotics e.g. oxytetracycline [54].	0.09%
	(Type III Polyketide synthase)	T3pks T3pks-cf_saccharide	-	Type III Polyketide synthase (PKSIII) enzymes are capable of producing an array of compounds including pyrones—a subset of pyrones were previously reported to have antibacterial and anticancer effects [55].	0.31%
3. Miscellaneous: products are predicted to confer adaptation to the environment/special structure/unknown function:	Phosphonates	Phosphonate	 Phosphonate	Some natural phosphonates are antibacterials e.g. fosfomycin. Also have structural functions [56].	0.01%
	Others	Cf_putative Other OtherKS Otherks-Pufa Otherks-Pufa-T1pks Otherks-T1pks	-	Some code for biosynthetic gene clusters of types that are still unknown [9].	8.13%
	Polyunsaturated fatty acids	Otherks-Pufa Otherks-Pufa-T1pks	 PUFA	Polyunsaturated fatty acids (PUFAs) are membrane adaptations to piezophiles, thermophiles and psychrophiles to prevent membrane crystallization [57,58].	0.14%
	Ectoine	Ectoine	 Ectoine	Halophilic adaptation & effective in vitro in preventing protein misfolding characteristic in diseases e.g. Alzheimer's [59].	0.08%
	Ladderane	Ladderane-Cf_fatty_acid	 Ladderane	Unique component of anammoxosome membrane in anammox (anaerobic ammonium oxidizing) bacteria and potential biofuel [60].	0.05%

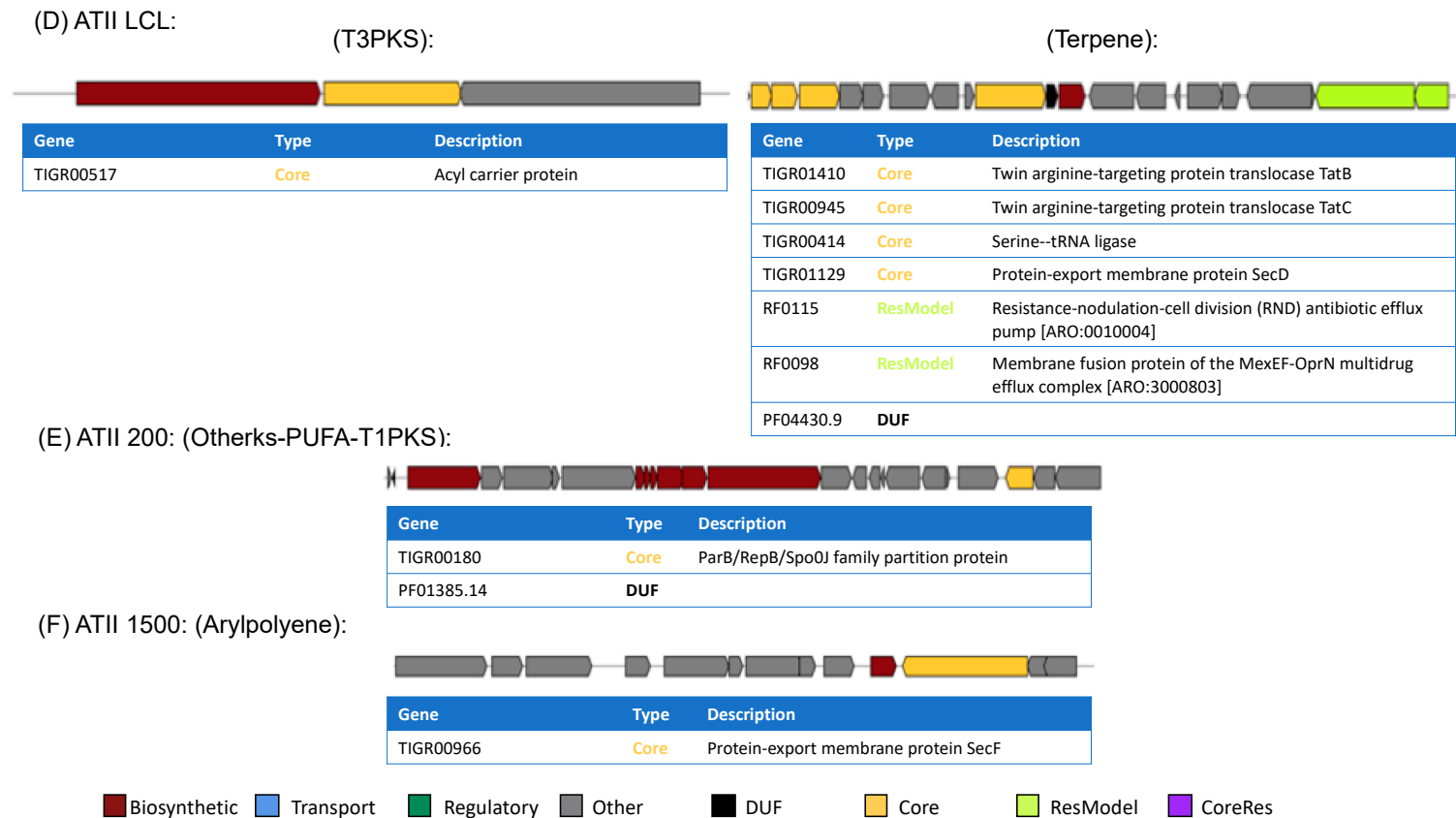


Figure 3. Promising Red Sea brine specialized metabolite candidates. Five core structures for specialized metabolites were predicted using antiSMASH 4.0 [45]: (A) KD brine–seawater Lower Interface (KD LINF) site had two non-ribosomal peptides predicted to be synthesized by two non-ribosomal peptide synthetase (NRPS) clusters. (B) KD brine–seawater Upper Interface (KD UINF) site had a non-ribosomal peptide predicted to be synthesized by a NRPS cluster and a hybrid polyketide-non-ribosomal peptide predicted to be synthesized by T1PKS-NRPS. (C) ATII 1500 site had a polyketide predicted to be produced by a T1pks-pufa-otherks hybrid cluster. Four top promising candidate SMGCs were detected by ARTS [38]: (D) ATII brine Lower Convective Layer (ATII LCL) site had two clusters encoding Terpene and T3PKS. (E) ATII 200 site had a cluster encoding aryl polyene. (F) ATII 1500 site had a cluster encoding Otherks-PUFA-T1PKS (PUFA: polyunsaturated fatty acid). ser: serine, nrp: non-ribosomal peptide, gln: glutamine, gly: glycine, asp: aspartate, thr: threonine, pk: polyketide, ohmal: hydroxy malate.

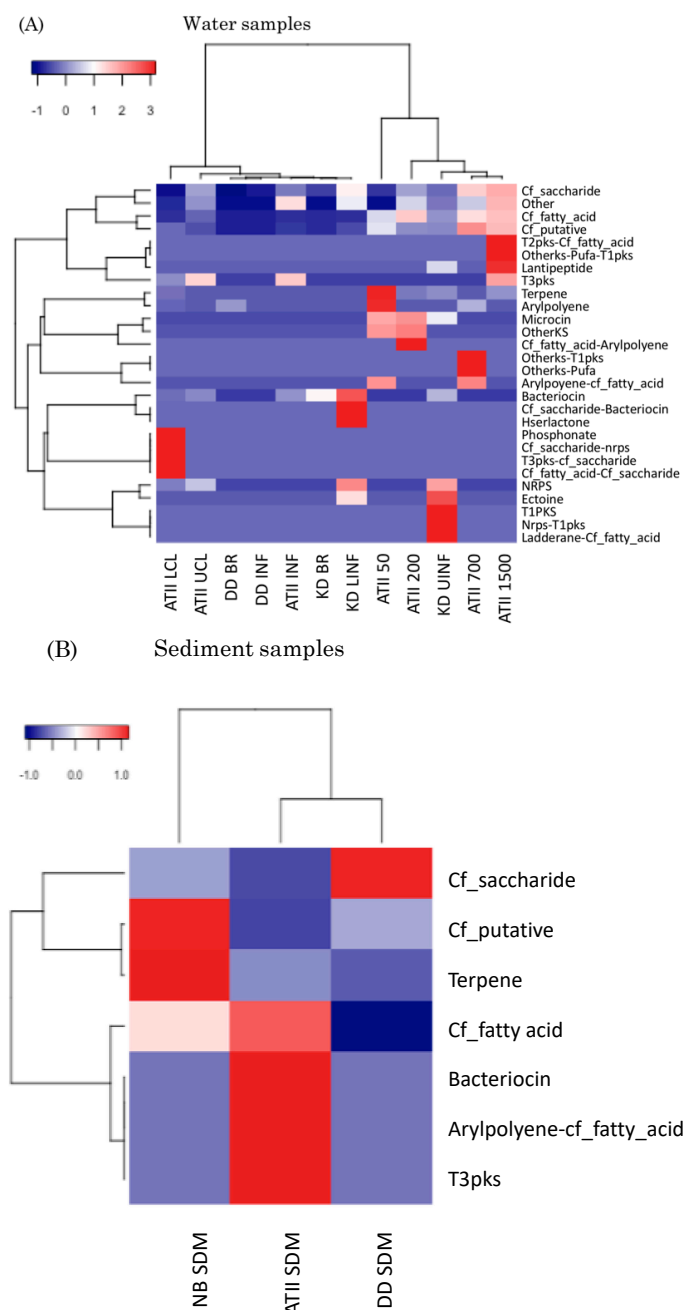


Figure 4. Heat maps representing hierarchical classification of the SMGCs detected in Red Sea brine pool metagenomes. (A) Heat map for the Red Sea water samples SMGCs. (B) Heat map for the Red Sea sediment samples SMGCs. Hierarchical clustering was based on the relative abundance of normalized numbers of SMGCs detected at each site.

2.3. Red Sea Brine Prokaryotic Diversity in Relation to Specialized Metabolism Genes

To correlate the microbial communities and how their taxonomic profiles contribute to specialized metabolites production, archaeal and bacterial phyla were investigated and were related to the SMGCs found in each site (Figure 5, Tables S4–S6). The abundance of major archaeal and bacterial phyla in the dataset was previously reported, albeit for the sequence reads not for the assembled contigs [22,36]. However, in this study we focused on bacterial phyla with the potentiality to produce bioactive products; as certain bacterial phyla were reported to produce bioactive compounds such as antibacterial and anticancer agents.

A total of 5 archaeal phyla and 28 bacterial phyla were detected in all of the Red Sea assemblies analyzed (Figure 5, Table 2, Tables S4 and S5). ATII 1500 and KD LINF showed the most diverse archaeal and bacterial phyla sites and each harbored all of the detected archaeal and bacterial phyla (33 distinct phyla), while NB SDM showed the least number of bacterial phyla (9 distinct phyla). The following phyla were ubiquitously detected in all the sites included in the dataset: Proteobacteria, Cyanobacteria, Acidobacteria, Actinobacteria, Bacteroidetes, Chlorobi and Firmicutes. Euryarchaeota and Thaumarchaeota were present in all the water samples (Figure 5, Table S4). Sediment samples harbored viruses (33.81%–52.72%) and Cyanobacteria (19.46%–38.18%) in addition to Proteobacteria (5.12%–31.51%), as major taxa. The percentage of the viruses was lower in all the water samples (0.23%–3.96%). DD samples harbored mainly Thaumarchaeota (17.44%–21.50%) and Proteobacteria (51.63%–52.33%) as major taxa. ATII 50 harbored mainly Cyanobacteria (29.97%), while ATII 200-1500 harbored Thaumarchaeota (29.12%–33.08%) and ATII 50-1500 harbored Proteobacteria (48.86%–57.97%) as major taxa. ATII brine layers mainly harbored Proteobacteria (92.44%–97.16%) as the major taxa. KD BR mainly harbored Euryarchaeota (54.59%) while KD UINF-LINF harbored Thaumarchaeota (26.68%–66.62%), and all KD layers harbored Firmicutes (6.06%–16.57%) and Proteobacteria (7.07%–45.04%) as major taxa. The most abundant archaeal and bacterial genera were recorded for each site (Table S6). Additionally, the prokaryotic phyla unique to Red Sea samples were recorded, prokaryotic phyla common to Red Sea and other hydrothermal vent metagenomes as well as the prokaryotic phyla unique to the other hydrothermal vent metagenomes (Table S10).

2.4. Rare Leucine Codons within Red Sea Brine SMGCs and Low Similarity to Known Clusters with Characterized Products

The counts of TTA codons were denoted for all the Red Sea SMGCs, excluding saccharides and fatty acids (Table S7). All the Red Sea sites harbored TTA rare leucine codons except DD INF layer, with a total of 7684 (absolute value), and an average of 512. The TTA codon counts were particularly high (>1000) in ATII 50, KD UINF and ATII 700 water layers.

To detect homologous gene clusters to the SMGC hits found in the Red Sea brine pool dataset we used 1-ClusterBlast, 2- KnownClusterBlast and 3- SubClusterBlast algorithms which are embedded in antiSMASH pipeline, for all the SMGCs except for *cf_saccharides* and *cf_fatty_acids*, (Table S8) [45]. Those algorithms detect homologous gene clusters and thus identify taxonomic and functional characteristics for the SMGCs [45]. Out of the 301 SMGCs, 204 were found to be of significant homology to gene clusters in the database, 30 SMGCs had identified homologous known gene clusters while only 4 SMGCs showed significant hits with homologous subclusters (Table S8). 77 of the 301 SMGCs had no significant homologous gene clusters, according to the ClusterBlast module embedded in antiSMASH (Table S8). The percentage of genes with similarity to the Red Sea SMGCs ranged between 2%–43% for the homologous gene clusters, 2–77% for the homologous known gene clusters and 20–66% for the homologous subclusters.

3. Discussion

3.1. Saccharide and Putative SMGCs Are the Most Abundant Groups in the Red Sea Brine Dataset

Metagenomic read sequences were previously reported for ATII water column samples [19], brine ATII, DD and KD water samples [20], as well as for the sediment samples [36]; 16S rRNA and taxonomic analysis of the different sediment sections were previously discussed [28]. In this study we utilized the assembled metagenomes for all the previously mentioned samples, in order to detect and thoroughly analyze the specialized metabolism gene clusters encoded in the Red Sea brine pool metagenomic dataset.

A total of 2751 SMGCs belonging to 28 different SMGC types, were detected in all 15 assembled metagenomes (Figures 1 and 2). The average number of detected clusters/Mb in the Red Sea dataset was 0.38 SMGC/Mb, ranging from 0.13 SMGC/Mb (NB SDM) to 0.67 SMGC/Mb (KD LINF) (Tables 1 and 2).

The average number of detected SMGC/Mb in more than 1000 studied bacterial genomes was reported to be 2.4 [46], ~6 times higher than that of the Red Sea brine metagenomic dataset. The number of SMGCs detected by antiSMASH is linear to the bacterial genome size [46], however, metagenomic data analysis is inherently more challenging than genomic data, owing to the short reads followed by assembly [61]. This could explain the relatively limited number of detected SMGCs. Recent studies detected SMGCs in genomes of new soil-residing *Pseudovibrio* bacterial strains using antiSMASH [62,63].

The most abundant SMGC class detected in the Red Sea samples coded for saccharides (80.49% of total SMGCs) (Figure 2). This is in concordance with the study of Cimermanic et al., wherein saccharides comprised the largest detected SMGC type (40% of total detected SMGCs) [46]. The second largest group of SMGCs was the cf_putative SMGCs (7.46% of the total SMGCs) (Table S2). Such putative gene clusters are unknown biosynthetic gene clusters, with no specific category assigned to them. It is likely that these novel gene clusters constitute a specific group of ‘SMGC dark matter’ [9], that we aimed to preliminarily categorize.

Selected SMGCs were unique to selected Red Sea sites, e.g. terpene SMGC was common to all SDM samples (Table S3), suggesting that each group of sites possesses a distinct SMGC ‘signature’ profile. Interestingly, ATII LCL has the harshest physicochemical conditions [33] and it harbors the most diverse SMGCs (Table 2), which shows the importance of studying SMGCs in extreme environments.

3.2. Preliminary Evidence of Potential Products with Pharmaceutical Applications

In order to prioritize the search for novel antibacterial and anticancer compounds among the detected orphan SMGCs, we recommend characterizing the following SMGCs in the future: 1, SMGCs coding for terpenes, peptides, polyketides and phosphonates (Table 3, group 2 SMGCs). 2, SMGCs having products with predicted structures (Figure 3A–C), wherein the SMGCs can be expressed, the structures elucidated and their bioactivities characterized [46]. 3, Type III PKSs that are easier to clone than other PKSs, and are capable of producing products with diverse structures [54]. 4, SMGCs likely to encode for novel metabolites, perhaps with antibacterial activity (Figure 3D–F): terpene, type III PKS, aryl polyene and lastly otherks-PUFA-T1PKS. The latter SMGC category should be studied, as core housekeeping genes and/or resistance genes were detected in close proximity to them. Recently, targeted mining genomes was successfully conducted through the detection of resistance and housekeeping genes within SMGCs [38]. Therefore, it is important to further explore those former four SMGCs categories as promising candidate Red Sea brine SMGCs.

Structural prediction of Red Sea SMGCs products revealed chiral and non-chiral non-ribosomal peptides as well as a polyketide (Figure 3D–F). Only five specialized metabolite core structures were predicted with antiSMASH. Perhaps the Red Sea brine microbiota are evolutionarily distant from the well-characterized genes in the published databases, hindering further product structural and functional prediction. The few predicted Red Sea SMGCs chemical structures (Figure 3D–F) remain to be verified by experimentation e.g. mass spectrometry and comparative metabolomic studies [8].

Rare TTA leucine codons are especially enriched in specialized metabolism genes and cell differentiation genes; however, they should be optimized for successful expression [45]. The TTA codon counts were particularly high within ATII 50, KD UINF, and ATII 700 SMGCs (Table S7). This is a possible translational impediment in the case of SMGCs in heterologous hosts, especially for those three sites, and should be accounted for prior to expression.

This metagenome mining study has several limitations because: 1, biosynthetic genes might have been missed as antiSMASH searches for partial or complete gene clusters rather than individual genes, 2, small contigs were excluded as the contig length cut-off was 1000 bp, and 3, antiSMASH detects SMGCs using a rule-based approach based only on the known pathways, so it is likely that we obtained less hits, e.g., SMGCs utilizing un-characterized pathways [64].

Functional screening of the ATII LCL fosmid library lead to the detection of orphan biosynthetic gene clusters that conferred antibacterial and anticancer effects [65]. Hence both functional [65] and

computational detection of SMGCs in Red Sea brine pools corroborate that investigating the Red Sea brine pool niche for SMGCs of pharmaceutical application, is a promising approach.

3.3. Red Sea Brine SMGCs form a Unique Cluster

Hierarchical clustering revealed that almost all brine water samples (ATII LCL, ATII UCL, DD BR, DD INF, ATII INF, KD BR and KD LINF) formed a unique SMGC brine cluster (Figure 4A), while KD UINF and ATII 200–1500 water samples clustered together. The physical conditions in KD UINF resembles deep Red Sea water, as the conditions are not as harsh as other brine sites [20,27]. Also, the SMGC percentage composition of KD UINF was closer to ATII 50–1500 m water sites, e.g. cf_saccharide abundance is similar in KD UINF and ATII water samples (77.78% in KD UINF and 58.22%–76.79% in ATII water samples), as opposed to 92.51% and 95.30% in the other KD samples. The detailed percent composition of each SMGC per Red Sea brine site is included (Table S9). Brine sediments had similar SMGC profiles, perhaps due to similar environmental conditions as opposed to the NB site, the latter having the least harsh conditions (Figure 4B) [28]. Upon hierarchical classification of all of the included sites, the Red-Sea metagenomic samples clustered together, contrast with the non-Red Sea SMGC profile, re-emphasizing the possibility of a Red Sea SMGC profile signature (Figure S2).

3.4. Environment–Microbe Interaction, Antagonistic Stressors and Extremophile Survival Implicated by Red Sea Brine SMGCs

Certain sites harbored unique SMGCs (Table 2) such as ladderane, that was detected only in KD UINF. Ladderane is exclusively present in the membranes of anaerobic ammonia-oxidizing (anammox) bacteria [66]. Noting that anammox bacteria thrive in the presence of sulfide influx and can perform denitrification [67], and KD is characterized by high H₂S content [23], thus denitrification may be coupled to sulfide in KD UINF anammox bacteria.

Bacteriocin SMGCs were detected in all ATII and KD brine water samples (Table S3), which are putative antagonistic stressors. An earlier study detected numerous halocins–bacteriocins and archaeocins produced by halophiles and were hypothesized to play a role in ensuring microbial diversity in hypersaline environments [68]. Microbes inhabiting ATII and KD niches seem to be utilizing similar mechanisms for survival. KD samples also hosted ectoine synthase-coding genes in the brine–seawater interface layers (Table S3). Ectoine is a stress solute that enables halophiles to withstand high salt content [59], and thus “salt-out” mechanism is likely to account for KD microbes survival in such high salinity [69].

Terpene SMGCs were detected in all SDM samples (Table S3). A previous study suggested that diterpenoids produced by the hyperthermophilic *Chloroflexus aurantiacus* may function in modulating membrane fluidity [70]. It is likely that terpenes provide the correct cell membrane fluidity for extremophiles [70]. Further experimentation is needed to shed light on the particular structure and function of the Red Sea terpenes. Terpenes with novel chemistries were isolated from a fungus inhabiting deep-ocean sediments, and conferred anticancer activity [71]. Terpene SMGCs were most enriched in the ATII 50 site, pointing to the possibility that shallow seawater favors terpene production. ATII brine water layers all harbored PKSIII SMGCs (Table S3). PKSIII enzymes are capable of producing different chemical structures, some of which have important bioactivities [54]. Maybe different stress factors within the same brine, ATII, have interfered in the overlaying water and underlying sediment microbiomes to produce the terpene SMGC signature, while causing the brine water layer microbiomes to have the PKSIII SMGC signature.

3.5. Prolific Specialized-Metabolite-Producing Phyla Detection and Red Sea Brine Pool SMGC Dark Matter Analysis

As expected, all Red Sea sites harbored Proteobacteria, Actinobacteria and Cyanobacteria (Table S5). Yet these phyla are known to produce a huge repertoire of specialized metabolites [46,72–74]. Certain bacterial taxa produce above average SMGCs/Mb, e.g., *Myxococcus*, *Streptomyces* and *Gloeobacter*

—which belong to the aforementioned phyla, respectively [46]. Our analysis of the Red Sea brine dataset revealed that not all the phyla were shared in neighboring sites, even when their SMGCs clustered together, e.g., KD BR and KD LINF. Additionally, not all neighboring sites clustered together based on the SMGCs even if they harbored similar phyla, e.g., KD UINF and KD LINF (Figures 4 and 5). These observations hint at a debatable relationship between SMGCs and phylogeny, similar to that reported for genomic datasets [46].

Members of the most abundant bacterial and archaeal genera (Table S6) were reported to be living in similar niches or having biotechnological applications, e.g., *Nitrosopumilus maritimus* survives in minimal ammonia available in its marine environment [75], and *Prochlorococcus MIT9313* produces lanthipeptides [76]. Only 10% of the SMGCs were homologous to experimentally characterized gene clusters (Table S8), indicating a huge opportunity to further investigate the Red Sea brine pool SMGC dark matter. Several homologous gene clusters pertained to aquatic microbes (e.g., *Mizugakiibacter sediminis*) [77], while others pertained to halophilic microbes (e.g., *Verrucomicrobia bacterium*) [78]. SMGC similarity may contribute to microbial evolution in similar ecological niches [79], perhaps leading to a future halophilic SMGC signature profile.

When we compared the phyla in the Red Sea samples to the other included marine hydrothermal vents (GB VNT, KSW VNT, K VNT, JDF VNT, LC MM), we identified 145 unique prokaryotic phyla in the Red Sea samples (Table S10). The unique phyla to the Red-Sea, the unique phyla to the other marine hydrothermal vent metagenomes and the common phyla in all sites used are presented (Table S10).

4. Materials and Methods

4.1. Sampling, DNA Extraction and Sequencing

The overall study workflow that we employed is depicted in Figure 1. The Red Sea samples analyzed included ATII brine water samples (ATII INF: ATII brine–seawater Interface, ATII UCL: ATII brine Upper Convective Layer, ATII LCL: ATII brine Lower Convective Layer), ATII water column samples (ATII 50: ATII 50 m depth, ATII 200: ATII 200 m depth, ATII 700: ATII 700 m depth, ATII 1500: ATII 1500 m depth), DD brine water samples (DD BR: DD brine water, DD INF: DD brine–seawater interface) and KD brine water samples: (KD BR: KD brine water, KD LINF: KD brine–seawater Lower Interface, KD UINF: KD brine–seawater Upper Interface). All the sites are described in [19,20]. Additionally, sediment samples included sediments underlying ATII brine pool (ATII SDM) were included, comprising seven distinct layers, sediments underlying DD brine pool (DD SDM), comprising seven distinct layers and sediments underlying two brine-influenced sites (NB SDM) [28]. In total, 28 different samples were analyzed, including 12 water samples and 16 sediment samples, that were previously described [19,20,28].

The samples were previously collected in April 2010 on *Aegaeo* research vessel that was the second leg of the Red Sea expedition of KAUST/WHOI/HCMR (KAUST: King Abdullah University of Science and Technology, WHOI: Woods Hole Oceanographic Institution, HCMR: Hellenic Center for Marine Research) [19,20,28]. Further information on sample locations are tabulated in Table S1. The water samples were sequentially filtered through 3.0, 0.8 and 0.1 μm filters. DNA was then extracted from the 0.1 μm filter [28]. A Genome Sequencer (GS FLX) pyrosequencer was used for sequencing the DNA samples by Titanium pyrosequencing kit (454 Life Sciences). Quality control for the obtained reads was done by PRINSEQ-lite v0.20.4 [80] and CD-HIT-454 [81].

Five other non-Red-Sea marine metagenomes were additionally selected, pertaining to marine hydrothermal vent samples of varying temperatures. GB VNT: Guaymas Basin deep-sea hydrothermal vent plume water sample (3 °C) [39–41]. Water samples from Kueishantao shallow-sea hydrothermal vent area: KSW VNT: vent sample (30 °C), K VNT: water directly above the vent (49 °C) [42]. JDF VNT: Juan de Fuca Ridge hydrothermal vent diffuse flow seawater sample, particularly from the Hulk vent (125 °C) [43]. LC MM: Loki’s Castle deep-sea vent black smoker chimney biofilm sample (306 °C) [44].

4.2. Bioinformatics Assembly

The assembly files were generated using Newbler assembler v2.6 [82]. For the Red Sea samples, default parameters were used for overlap layout consensus, except extension over read tips, that was opted to cope with the low coverage in direct DNA 454 shotgun sequencing runs on the metagenomic samples. The reads were limited to one contig in output. For the non-Red Sea samples, default parameters were used.

Distinct assembly files were generated for all the water samples, from particular sites and depths [19,20]. However, sediment (SDM) reads from the same brine pool were cross-assembled, i.e., ATII SDM cross-assembly comprised samples obtained from seven different ATII sediment depths, and DD SDM cross-assembly from seven different DD sediment depths [36]. NB sediments cross-assembly was derived from samples obtained from two brine-influenced sites [28]. Although NB SDM contains sediments from two distinct neighboring sites that are brine-influenced, which are spatially different, they were pooled to provide an outgroup for the brine sediments and serve as non-brine sediment samples. This resulted in the generation of 15 assembly files that were further used for downstream analyses. In addition, five additional assembly files were constructed for the non-Red-Sea samples from other marine hydrothermal vent sites (JDF VNT, KSW VNT, K VNT, LC MM, GB VNT).

4.3. Annotation, SMGCs Analyses and Hierarchical Classification

The metagenomic assembly files were run on antiSMASH using contigs of size equal to or larger than 1000 bp. Annotation was performed using the Prodigal gene finding option for metagenomes. The SMGCs were detected by comparison of translated amino acid sequences with pHMM signature for biosynthetic genes using the ClusterFinder algorithm [9]. The full SMGC detection analysis was performed during 08/2015 (i.e. by antiSMASH version 3.0 [9]). All the counts of detected SMGCs were normalized by dividing each value with the number of assembled reads at each site $\times 10^6$ and used for subsequent analyses. To account for the variability in the sequencing depth, normalization was performed.

In order to detect homologous gene clusters to all Red Sea brine SMGCs—excluding putative clusters for saccharides and fatty acids—the algorithms ClusterBlast, KnownClusterBlast and SubClusterBlast were used. The first algorithm compares the SMGCs with all gene clusters in Genbank database of NCBI (National Center for Biotechnology Information) and its output is useful to identify the closest organism based on SMGCs [45]. The second algorithm compares the SMGCs with the gene clusters that are experimentally characterized in MIBiG database, thus giving an indication on the products likely to be synthesized by the SMGCs [45]. The third algorithm detects operons that are conserved and with a known function and gives more specific information on the product synthesized by the SMGCs [45]. This analysis was performed during 11/2017. Putative clusters coding for the production of saccharides and fatty acids were excluded because the scope of the study was to investigate products of potential ecological functions and/or pharmaceutical applications. The TTA codons were also recorded for each site and the absolute counts and normalized values were denoted for each assembly file. Additionally, predicted core structures of the specialized metabolites were recorded. All the aforementioned analyses were done using antiSMASH version 4.0 [45]. The contigs coding for products other than saccharides and fatty acids were screened for housekeeping and resistance genes within the SMGCs using the ARTS program [38].

Hierarchical classification for all the SMGCs detected in all the assembly files, was performed using R version 3.3.1 (R Development Core Team 2016). Distinct heat maps were generated for water and sediment samples using as an input the normalized SMGC count for each of the sites.

4.4. Taxonomic Classification

Taxonomical trees for the archaeal, bacterial and viruses phyla were generated by metagenomics rapid annotation using subsystems technology (MG-RAST) tool using the assembled contigs as the

input sequences, by comparing them to the non-redundant protein database M5NR (maximum e-value $1e^{-5}$, minimum identity of 70%, minimum alignment length of 50 measured in amino acids for protein and bp for RNA databases) [37,83]. Based on the protein-based phylogeny, the detected phyla were denoted for each site, with focus on the archaeal and bacterial phyla of relative abundance $\geq 0.5\%$ in at least one of the assemblies.

5. Conclusions

Our study highlights the importance of Red Sea brine pool water and sediment microbes and their potential capability in producing specialized metabolites. ATII, DD and KD brine pool sites included in the study are thus worthy of bioprospecting (Figure 2, Table 3). The diverse potential functions of the detected SMGCs' products in the Red Sea dataset, varying from halophilic adaptations to polyketides and peptides of potential antibacterial activity, renders it an attractive mine for such exploration. Our data provides insights on Red Sea brine SMGCs particularly focusing on antibacterial and anticancer exploration. Promising SMGCs code for products with reported antibacterial and anticancer effects, namely terpenes, peptides, polyketides and phosphonates. Also interesting are SMGCs with predicted structures, and SMGCs harboring housekeeping and/or resistance genes. Cloning such genes clusters would provide information on new 'cryptic' gene clusters that might be responsible for synthesis of novel natural products, and improve our understanding of the evolution of extremophiles and their adaptation mechanisms to such extreme environments [84]. Future studies are required to clone and express those SMGCs, in order to elucidate the novel chemical entities that possibly serve as antibacterial and/or anticancer drugs, among other different potential functions.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1660-3397/17/5/273/s1>, Figure S1: Chemical structures of some examples of earlier characterized antibiotic and anticancer specialized metabolites: (A) salinilactam, (B) lactocillin, (C) streptochlorin, (D) abyssomicin C and (E) salinosporamide K. Figure S2: Heat map representing hierarchical classification of the SMGCs detected in all the metagenomes in the dataset. Table S1: The sampling locations of each of the sites in the dataset. Table S2: The SMGCs detected at each site (absolute values) with the name of the SMGCs detected, the count and total numbers for each site and for each SMGC are denoted. Table S3: The normalized number of SMGCs detected at each site (normalized: number of SMGCs detected/ number of assembled reads at each site $\times 10^6$), with the name of the SMGCs detected, the count and total numbers for each site and for each SMGC are denoted. In dark blue are the cells with unique SMGCs –only detected once in all the data. Similarly colored cells indicate the SMGCs common to groups of sites (light pink: ATII water column depths, light blue: sediments, yellow: ATII brine layers, bright pink: KD brine layers). Table S4: Archaeal and Bacterial and viral phyla detected for each of the Red Sea brine assembly files. The relative abundance is shown whenever detected to be $\geq 0.5\%$ in at least one of the assemblies. Table S5: All the detected archaeal and bacterial phyla in all the Red Sea brine sites by MG-RAST. The relative abundance is shown for all the phyla and values > 0 are highlighted in red. Table S6: The most abundant archaeal and bacterial genera in all the sites in the dataset. Table S7: Table showing the rare leucine TTA codon absolute and normalized count in each site, for all the Red Sea brine SMGCs excluding the saccharides and fatty acids. Table S8: The best hit homologous gene clusters pertaining to all the Red Sea brine SMGCs detected in the study by antiSMASH excluding *cf_saccharides* and *cf_fatty_acids* in all the sites included in the Red Sea dataset as computed by ClusterBlast algorithm of antiSMASH [1]. The homologous known gene clusters were detected are also denoted, as well as homologous subclusters. Blue: SMGCs identified in 11/2017 upon re-running the contigs with hits that did not appear in 8/2015. Green: SMGCs identified before in 8/2015 but not when re-run in 11/2017. Table S9: Table showing the percentage of each SMGC as compared to the total SMGCs detected per site among the Red Sea brine samples. Table S10: All the detected archaeal and bacterial genera in all sites included in the dataset.

Author Contributions: L.Z. has conducted the computational analyses and data visualization. M.A. and M.N.M. have performed the assembly of the metagenomic sequencing data. R.S. has supervised this work. L.Z. and R.S. wrote the manuscript. R.S., L.Z. and M.A. have edited the manuscript, discussed the results and commented on the manuscript. All authors read and approved the manuscript.

Funding: This work was funded by an AUC Faculty Support Grant to R.S.

Acknowledgments: The authors appreciate the DNA preparation work conducted by Ahmed Abdelaziz and Amged Ouf of the American University in Cairo. Many thanks to Ali H. A. Elbeherly for helping us with Table S1. We appreciate the help of Amr Tayel (Chemistry department, American University in Cairo) with the chemical structure figures in Figure S1.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: All the metagenomes included in this study are publicly available on Sequence Read Archive website (SRA: <https://www.ncbi.nlm.nih.gov/sra>). For each file, the bio-project and sample accession numbers are presented in Table S1.

References

1. Williams, D.N. Antimicrobial resistance: Are we at the dawn of the post-antibiotic era? *J. R. Coll. Physicians Edinb.* **2016**, *46*, 150–156. [[CrossRef](#)] [[PubMed](#)]
2. Krishnamurthy, M.; Moore, R.T.; Rajamani, S.; Panchal, R.G. Bacterial genome engineering and synthetic biology: combating pathogens. *BMC Microbiol.* **2016**, *16*, 258. [[CrossRef](#)] [[PubMed](#)]
3. Brown, E.D.; Wright, G.D. Antibacterial drug discovery in the resistance era. *Nature* **2016**, *529*, 336–343. [[CrossRef](#)] [[PubMed](#)]
4. Martens, E.; Demain, A.L. The antibiotic resistance crisis, with a focus on the United States. *J. Antibiot. (Tokyo)* **2017**, *70*, 520–526. [[CrossRef](#)] [[PubMed](#)]
5. Holohan, C.; Van Schaeybroeck, S.; Longley, D.B.; Johnston, P.G. Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* **2013**, *13*, 714–726. [[CrossRef](#)] [[PubMed](#)]
6. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629–661. [[CrossRef](#)]
7. Demain, A.; Fang, A. The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* **2000**, *69*, 222.
8. Milshteyn, A.; Schneider, J.S.S.; Brady, S.F.F. Mining the metabiome: Identifying novel natural products from microbial communities. *Chem. Biol.* **2014**, *21*, 1211–1223. [[CrossRef](#)]
9. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H.U.; Brucoleri, R.; Lee, S.Y.; Fischbach, M.A.; Muller, R.; Wohlleben, W.; et al. antiSMASH 3.0—A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **2015**, 1–7. [[CrossRef](#)]
10. Ziemert, N.; Alanjary, M.; Weber, T. The evolution of genome mining in microbes—A review. *Nat. Prod. Rep.* **2016**, *33*, 988–1005. [[CrossRef](#)] [[PubMed](#)]
11. Udvary, D.W.; Zeigler, L.; Asolkar, R.N.; Singan, V.; Lapidus, A.; Fenical, W.; Jensen, P.R.; Moore, B.S.; Walsh, C.T. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *PNAS* **2007**, *104*, 10376–10381. [[CrossRef](#)] [[PubMed](#)]
12. Donia, M.S.S.; Cimermanic, P.; Schulze, C.J.J.; Wieland Brown, L.C.; Martin, J.; Mitreva, M.; Clardy, J.; Lington, R.G.G.; Fischbach, M.A.A.; Wieland Brown, L.C.; et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **2014**, *158*, 1402–1414. [[CrossRef](#)]
13. Schofield, M.M.; Sherman, D.H. Meta-omic characterization of prokaryotic gene clusters for natural product biosynthesis. *Curr. Opin. Biotechnol.* **2013**, *24*, 1151–1158. [[CrossRef](#)] [[PubMed](#)]
14. Boddy, C.N. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 443–450. [[CrossRef](#)]
15. Trindade, M.; van Zyl, L.J.; Navarro-Fernández, J.; Abd Elrazak, A. Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Front. Microbiol.* **2015**, *6*, 1–14. [[CrossRef](#)]
16. Ul Hassan, S.S.; Shaikh, A.L. Marine actinobacteria as a drug treasure house. *Biomed. Pharmacother.* **2017**, *87*, 46–57. [[CrossRef](#)]
17. Pettit, R.K. Culturability and Secondary Metabolite Diversity of Extreme Microbes: Expanding Contribution of Deep Sea and Deep-Sea Vent Microbes to Natural Product Discovery. *Mar. Biotechnol.* **2011**, *13*, 1–11. [[CrossRef](#)] [[PubMed](#)]
18. Giddings, L.-A.; Newman, D.J. *Bioactive Compounds from Extremophiles*; Springer: Cham, Switzerland, 2015; ISBN 3319148354.
19. Ferreira, A.J.S.; Siam, R.; Setubal, J.C.; Moustafa, A.; Sayed, A.; Chambergo, F.S.; Dawe, A.S.; Ghazy, M.A.; Sharaf, H.; Ouf, A.; et al. Core microbial functional activities in ocean environments revealed by global metagenomic profiling analyses. *PLoS ONE* **2014**, *9*. [[CrossRef](#)]

20. Abdallah, R.Z.; Adel, M.; Ouf, A.; Sayed, A.; Ghazy, M.A.; Alam, I.; Essack, M.; Lafi, F.F.; Bajic, V.B.; El-Dorry, H.; et al. Aerobic methanotrophic communities at the Red Sea brine-seawater interface. *Front. Microbiol.* **2014**, *5*. [[CrossRef](#)]
21. Antunes, A.; Ngugi, D.K.; Stingl, U. Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environ. Microbiol. Rep.* **2011**, *3*, 416–433. [[CrossRef](#)] [[PubMed](#)]
22. Elbehery, A.H.A.; Aziz, R.K.; Siam, R. Insertion sequences enrichment in extreme Red sea brine pool vent. *Extremophiles* **2017**, *21*, 271–282. [[CrossRef](#)] [[PubMed](#)]
23. Backer, H.; Schoell, M. New Deeps with Brines and Metalliferous Sediments in the Red Sea. *Nat. Phys. Sci.* **1972**, *240*, 153. [[CrossRef](#)]
24. Hartmann, M.; Scholten, J.C.C.; Stoffers, P.; Wehner, F. Hydrographic structure of brine-filled deeps in the Red Sea - New results from the Shaban, Kebrit, Atlantis II, and discovery deep. *Mar. Geol.* **1998**, *144*, 311–330. [[CrossRef](#)]
25. Anschutz, P.; Blanc, G.; Monnin, C.; Boulègue, J. Geochemical dynamics of the Atlantis II Deep (Red Sea): II. Composition of metalliferous sediment pore waters. *Geochim. Cosmochim. Acta* **2000**, *64*, 3995–4006. [[CrossRef](#)]
26. Faber, E.; Botz, R.; Poggenburg, J.; Schmidt, M.; Stoffers, P.; Hartmann, M. Methane in Red Sea brines. *Org. Geochem.* **1998**, *29*, 363–379. [[CrossRef](#)]
27. Eder, W.; Jahnke, L.L.; Schmidt, M.; Huber, R. Microbial Diversity of the Brine-Seawater Interface of the Kebrit Deep, Red Sea, Studied via 16S rRNA Gene Sequences and Cultivation Methods. *Appl. Environ. Microbiol.* **2001**, *67*, 3077–3085. [[CrossRef](#)] [[PubMed](#)]
28. Siam, R.; Mustafa, G.A.; Sharaf, H.; Moustafa, A.; Ramadan, A.R.; Antunes, A.; Bajic, V.B.; Stingl, U.; Marsis, N.G.R.; Coolen, M.J.L.; et al. Unique prokaryotic consortia in geochemically distinct sediments from red sea Atlantis II and discovery deep brine pools. *PLoS ONE* **2012**, *7*. [[CrossRef](#)] [[PubMed](#)]
29. Zierenberg, R.A. Deposition of Metalliferous Sediment Beneath a Brine Pool in the Atlantis II Deep, Red Sea. In *Gorda Ridge*; Springer: New York, NY, USA, 1990; pp. 131–142.
30. Eustáquio, A.S.; Nam, S.J.; Penn, K.; Lechner, A.; Wilson, M.C.; Fenical, W.; Jensen, P.R.; Moore, B.S. The discovery of salinosporamide K from the marine bacterium “*Salinispora pacifica*” by Genome mining gives insight into pathway evolution. *ChemBioChem* **2011**, *12*, 61–64. [[CrossRef](#)] [[PubMed](#)]
31. Othoum, G.; Bougouffa, S.; Razali, R.; Bokhari, A.; Alamoudi, S.; Antunes, A.; Gao, X.; Hoehndorf, R.; Arold, S.T.; Gojobori, T.; et al. In silico exploration of Red Sea Bacillus genomes for natural product biosynthetic gene clusters. *BMC Genomics* **2018**, *19*, 382. [[CrossRef](#)]
32. Al-Amoudi, S.; Essack, M.; Simões, M.F.; Bougouffa, S.; Soloviev, I.; Archer, J.A.C.; Lafi, F.F.; Bajic, V.B. Bioprospecting Red Sea Coastal Ecosystems for Culturable Microorganisms and Their Antimicrobial Potential. *Mar. Drugs* **2016**, *14*, 165. [[CrossRef](#)] [[PubMed](#)]
33. Mohamed, Y.M.; Ghazy, M.A.; Sayed, A.; Ouf, A.; El-Dorry, H.; Siam, R. Isolation and characterization of a heavy metal-resistant, thermophilic esterase from a Red Sea Brine Pool. *Sci. Rep.* **2013**, *3*, 1–8. [[CrossRef](#)] [[PubMed](#)]
34. Sonbol, S.A.; Ferreira, A.J.S.; Siam, R. Red Sea Atlantis II brine pool nitrilase with unique thermostability profile and heavy metal tolerance. *BMC Biotechnol.* **2016**, *16*, 14. [[CrossRef](#)]
35. Elbehery, A.H.A.; Leak, D.J.; Siam, R. Novel thermostable antibiotic resistance enzymes from the Atlantis II Deep Red Sea brine pool. *Microb. Biotechnol.* **2017**, *10*, 189–202. [[CrossRef](#)] [[PubMed](#)]
36. Adel, M.; Elbehery, A.H.A.; Aziz, S.K.; Aziz, R.K.; Grossart, H.P.; Siam, R. Viruses-to-mobile genetic elements skew in the deep Atlantis II brine pool sediments. *Sci. Rep.* **2016**, *6*, 32704. [[CrossRef](#)] [[PubMed](#)]
37. Meyer, F.; Paarmann, D.; D’Souza, M.; Olson, R.; Glass, E.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; et al. The metagenomics RAST server—A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.* **2008**, *9*, 386. [[CrossRef](#)] [[PubMed](#)]
38. Alanjary, M.; Kronmiller, B.; Adamek, M.; Blin, K.; Weber, T.; Huson, D.; Philmus, B.; Ziemert, N. The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res.* **2017**, *45*, W42–W48. [[CrossRef](#)]
39. Sheik, C.S.; Jain, S.; Dick, G.J. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ. Microbiol.* **2014**, *16*, 304–317. [[CrossRef](#)] [[PubMed](#)]

40. Baker, B.J.; Lesniewski, R.A.; Dick, G.J. Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J.* **2012**, *6*, 2269–2279. [[CrossRef](#)] [[PubMed](#)]
41. Lesniewski, R.A.; Jain, S.; Anantharaman, K.; Schloss, P.D.; Dick, G.J. The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J.* **2012**, *6*, 2257–2268. [[CrossRef](#)]
42. Tang, K.; Liu, K.; Jiao, N.; Zhang, Y.; Chen, C.T.A. Functional Metagenomic Investigations of Microbial Communities in a Shallow-Sea Hydrothermal System. *PLoS ONE* **2013**, *8*, e72958. [[CrossRef](#)]
43. Anderson, R.E.; Beltrán, M.T.; Hallam, S.J.; Baross, J.A. Microbial community structure across fluid gradients in the Juan de Fuca Ridge hydrothermal system. *FEMS Microbiol. Ecol.* **2013**, *83*, 324–339. [[CrossRef](#)] [[PubMed](#)]
44. Stokke, R.; Dahle, H.; Roalkvam, I.; Wissuwa, J.; Daae, F.L.; Tooming-Klunderud, A.; Thorseth, I.H.; Pedersen, R.B.; Steen, I.H. Functional interactions among filamentous Epsilonproteobacteria and Bacteroidetes in a deep-sea hydrothermal vent biofilm. *Environ. Microbiol.* **2015**, *17*, 4063–4077. [[CrossRef](#)] [[PubMed](#)]
45. Blin, K.; Wolf, T.; Chevrette, M.G.; Lu, X.; Schwalen, C.J.; Kautsar, S.A.; Suarez Duran, H.G.; de los Santos, E.L.C.; Kim, H.U.; Nave, M.; et al. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, *1854*, 1019–1037. [[CrossRef](#)] [[PubMed](#)]
46. Cimermancic, P.; Medema, M.H.H.; Claesen, J.; Kurita, K.; Wieland Brown, L.C.; Mavrommatis, K.; Pati, A.; Godfrey, P.A.A.; Koehrsen, M.; Clardy, J.; et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158*, 412–421. [[CrossRef](#)] [[PubMed](#)]
47. Weitnauer, G.; Mühlenweg, A.; Trefzer, A.; Hoffmeister, D.; Süßmuth, R.D.; Jung, G.; Welzel, K.; Vente, A.; Girreser, U.; Bechthold, A. Biosynthesis of the orthosomycin antibiotic avilamycin A: Deductions from the molecular analysis of the avi biosynthetic gene cluster of *Streptomyces viridochromogenes* Tü57 and production of new antibiotics. *Chem. Biol.* **2001**, *8*, 569–581. [[CrossRef](#)]
48. Wang, F.; Xiao, X.; Ou, H.-Y.; Gai, Y.; Wang, F. Role and regulation of fatty acid biosynthesis in the response of *Shewanella piezotolerans* WP3 to different temperatures and pressures. *J. Bacteriol.* **2009**, *191*, 2574–2584. [[CrossRef](#)] [[PubMed](#)]
49. Schöner, T.A.; Gassel, S.; Osawa, A.; Tobias, N.J.; Okuno, Y.; Sakakibara, Y.; Shindo, K.; Sandmann, G.; Bode, H.B. Aryl Polyenes, a Highly Abundant Class of Bacterial Natural Products, Are Functionally Related to Antioxidative Carotenoids. *ChemBioChem* **2016**, *17*, 247–253. [[CrossRef](#)]
50. Churchill, M.E.A.; Sibhatu, H.M.; Uhlson, C.L.; Charis, L.U.; Uhlson, C.L. Defining the structure and function of acyl-homoserine lactone autoinducers. *Methods Mol. Biol.* **2011**, *692*, 159–171.
51. Brahmshatriya, P.; Brahmshatriya, P. Terpenes: Chemistry, Biological Role, and Therapeutic Applications. In *Natural Products*; Ramawat, K., Mérillon, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 2665–2691.
52. Kaur, S.; Kaur, S. Bacteriocins as potential anticancer agents. *Front. Pharmacol.* **2015**, *6*, 1–11. [[CrossRef](#)]
53. Felnagle, E.A.; Jackson, E.E.; Chan, Y.A.; Podevels, A.M.; Berti, A.D.; McMahon, M.D.; Thomas, M.G.; Berti, D.; McMahon, M.D.; Thomas, M.G. Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Mol. Pharm.* **2008**, *5*, 191–211. [[CrossRef](#)]
54. Yu, D.; Xu, F.; Zeng, J.; Zhan, J. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **2012**, *64*, 285–295. [[CrossRef](#)] [[PubMed](#)]
55. Lim, Y.P.; Go, M.K.; Yew, W.S. Exploiting the biosynthetic potential of type III polyketide synthases. *Molecules* **2016**, *21*, 806. [[CrossRef](#)]
56. Metcalf, W.W.; van der Donk, W.A. Biosynthesis of phosphonic and phosphinic acid natural products. *Annu. Rev. Biochem.* **2009**, *78*, 65–94. [[CrossRef](#)]
57. Abd Elrazak, A.; Ward, A.C.; Glassey, J. Polyunsaturated fatty acid production by marine bacteria. *Bioprocess Biosyst. Eng.* **2013**, *36*, 1641–1652. [[CrossRef](#)] [[PubMed](#)]
58. Capece, M.C.; Clark, E.; Saleh, J.K.; Halford, D.; Heinl, N.; Hoskins, S.; Rothschild, L.J. Polyextremophiles and the constraints for terrestrial habitability. In *Polyextremophiles: Life Under Multiple Forms of Stress*; Seckbach, J., Oren, A., Stan-Lotter, H., Eds.; Springer: Dordrecht, The Netherlands, 2013; Volume 27, pp. 3–59, ISBN 978-94-007-6487-3.
59. Jorge, C.D.; Borges, N.; Bagyan, I.; Bilstein, A.; Santos, H. Potential applications of stress solutes from extremophiles in protein folding diseases and healthcare. *Extremophiles* **2016**, *20*, 251–259. [[CrossRef](#)]

60. Javidpour, P.; Deutsch, S.; Mutalik, V.K.; Hillson, N.J.; Petzold, C.J.; Keasling, J.D.; Beller, H.R. Investigation of proposed ladderane biosynthetic genes from anammox bacteria by heterologous expression in *E. coli*. *PLoS ONE* **2016**, *11*, e0151087. [[CrossRef](#)] [[PubMed](#)]
61. Medema, M.H.; Fischbach, M.A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, *11*, 639–648. [[CrossRef](#)]
62. Crits-Christoph, A.; Diamond, S.; Butterfield, C.N.; Thomas, B.C.; Banfield, J.F. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **2018**, *558*, 440–444. [[CrossRef](#)]
63. Naughton, L.M.; Romano, S.; O’Gara, F.; Dobson, A.D.W. Identification of secondary metabolite gene clusters in the *Pseudovibrio* Genus reveals encouraging biosynthetic potential toward the production of novel bioactive compounds. *Front. Microbiol.* **2017**, *8*, 1–15. [[CrossRef](#)]
64. Blin, K.; Kim, H.U.; Medema, M.H.; Weber, T. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief Bioinform.* **2017**, 1–11. [[CrossRef](#)]
65. Ziko, L.; Saqr, A.H.A.; Ouf, A.; Gimpel, M.; Aziz, R.K.; Neubauer, P. Antibacterial and anticancer activities of orphan biosynthetic gene clusters from Atlantis II Red Sea brine pool. *Microb. Cell Fact.* **2019**, *18*, 1–16. [[CrossRef](#)]
66. Van Niftrik, L.; Jetten, M.S.M. Anaerobic Ammonium-Oxidizing Bacteria: Unique Microorganisms with Exceptional Properties. *Microbiol. Mol. Biol. Rev.* **2012**, *76*, 585–596. [[CrossRef](#)]
67. Russ, L.; Speth, D.R.; Jetten, M.S.M.; Op den Camp, H.J.M.; Kartal, B. Interactions between anaerobic ammonium and sulfur-oxidizing bacteria in a laboratory scale model system. *Environ. Microbiol.* **2014**, *16*, 3487–3498. [[CrossRef](#)] [[PubMed](#)]
68. Atanasova, N.S.; Pietilä, M.K.; Oksanen, H.M. Diverse antimicrobial interactions of halophilic archaea and bacteria extend over geographical distances and cross the domain barrier. *Microbiologyopen* **2013**, *2*, 811–825. [[CrossRef](#)] [[PubMed](#)]
69. Ma, Y.; Galinski, E.A.; Grant, W.D.; Oren, A.; Ventosa, A. Halophiles 2010: Life in saline environments. *Appl. Environ. Microbiol.* **2010**, *76*, 6971–6981. [[CrossRef](#)] [[PubMed](#)]
70. De Carvalho, C.C.C.R.C.R.; Fernandes, P. Production of metabolites as bacterial responses to the marine environment. *Mar. Drugs* **2010**, *8*, 705–727. [[CrossRef](#)]
71. Du, L.; Li, D.; Zhu, T.; Cai, S.; Wang, F.; Xiao, X.; Gu, Q. New alkaloids and diterpenes from a deep ocean sediment derived fungus *Penicillium* sp. *Tetrahedron* **2009**, *65*, 1033–1039. [[CrossRef](#)]
72. Manivasagan, P.; Kang, K.-H.H.; Sivakumar, K.; Li-Chan, E.C.Y.Y.; Oh, H.-M.M.; Kim, S.-K.K. Marine actinobacteria: An important source of bioactive natural products. *Environ. Toxicol. Pharmacol.* **2014**, *38*, 172–188. [[CrossRef](#)] [[PubMed](#)]
73. Baltz, R.H. Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.* **2016**, 1–16. [[CrossRef](#)] [[PubMed](#)]
74. Thornburg, C.C.; Zabriskie, T.M.; McPhail, K.L.; Mark Zabriskie, T.; McPhail, K.L. Deep-sea hydrothermal vents: Potential hot spots for natural products discovery. *J. Nat. Prod.* **2010**, *73*, 489–499. [[CrossRef](#)] [[PubMed](#)]
75. Walker, C.B.; de la Torre, J.R.; Klotz, M.G.; Urakawa, H.; Pinel, N.; Arp, D.J.; Brochier-Armanet, C.; Chain, P.S.G.; Chan, P.P.; Gollabgir, A.; et al. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8818–8823. [[CrossRef](#)] [[PubMed](#)]
76. Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P.J.; Joewono, I.; Rusch, D.; Chisholm, S.W.; van der Donk, W.A. Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10430–10435. [[CrossRef](#)]
77. Kojima, H.; Tokizawa, R.; Fukui, M. *Mizugakiibacter sediminis* gen. nov., sp. nov., isolated from a freshwater lake. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 3983–3987. [[CrossRef](#)]
78. Romanenko, L.A.; Tanaka, N.; Kurilenko, V. V.; Svetashev, V.I.; Lyudmila, C.; Romanenko, A. *Tamlana sedimentorum* sp. nov., isolated from shallow sand sediments of the Sea of Japan. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 2891–2896. [[CrossRef](#)]
79. Nadeem, F.; Oves, M.; Qari, H.; Ismail, I. Red Sea Microbial Diversity for Antimicrobial and Anticancer Agents. *J. Mol. Biomark. Diagn.* **2015**, *7*, 1–14.
80. Schmieder, R.; Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **2011**, *27*. [[CrossRef](#)] [[PubMed](#)]

81. Niu, B.; Fu, L.; Sun, S.; Li, W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinform.* **2010**, *11*, 187. [[CrossRef](#)] [[PubMed](#)]
82. Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.-J.; Chen, Z.; et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005**, *437*, 376–380. [[CrossRef](#)] [[PubMed](#)]
83. Wilke, A.; Bischof, J.; Harrison, T.; Brettin, T.; D'Souza, M.; Gerlach, W.; Matthews, H.; Paczian, T.; Wilkening, J.; Glass, E.M.; et al. A RESTful API for Accessing Microbial Community Data for MG-RAST. *PLoS Comput. Biol.* **2015**, *11*, e1004008. [[CrossRef](#)] [[PubMed](#)]
84. Walsh, C.T.; Fischbach, M.A. Natural products version 2.0: Connecting genes to molecules. *J. Am. Chem. Soc.* **2010**, *132*, 2469–2493. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).