Mining the Black Sea Microbiome for Producing Natural Bioactive Compounds *(Elena Stoica)*	"Cercetări Marine" Issue no. 48 Pages 145-154	2018

MINING THE BLACK SEA MICROBIOME FOR PRODUCING NATURAL BIOACTIVE COMPOUNDS

Elena Stoica

National Institute for Marine Research and Development "Grigore Antipa", 300 Mamaia Blvd., Constanta, Romania
E-mail: estoica@alpha.rmri.ro

ABSTRACT

Marine microbiome is now recognized as an attractive source of bioactive compounds of pharmaceutical and biotechnological interest. The polyketides (PKs) and non-ribosomal peptides (NRPs) of marine microbial origin represent a significant proportion of bioactive natural products. The ecosystems of the brackish Black Sea host unique and highly diverse microbial communities. Despite having a great biodiversity, the biotechnological potential of microorganisms inhabiting these ecosystems remain mostly unexplored and unexploited. To explore this potential, marine microbial metagenomes collected in 2009 from the Romanian Black Sea sector were sequenced and mined for PKs and NRPs encoding genes. By using 16S rRNA high throughput sequencing (HTS)-based metagenomic approach in combination with an automated mining pipelines, we were able to show for the first time the great potential of Black Sea microbiome for secondary metabolite production, with special focus on NRPs and PKs of pharmaceutical importance.
Key-Words: marine microbiome, bio-mining, metagenomics, bioactive compounds, Black Sea

AIMS AND BACKGROUND

Microorganisms are natural inhabitants of a number of ecological niches in the oceans that play key roles in the structure and function of marine ecosystems and dominate the global marine biological diversity (Rappé & Giovannoni, 2003). Due to their high metabolic diversity, marine microbes also produce a vast variety of secondary metabolites which could be used for a wide range of applications, particularly for pharmaceutical and biotechnological purposes. Their rich biodiversity contained within the Earth's oceans and seas makes them a unique and rich natural products discovery reservoir (Venter et al, 2004; Glockner et al., 2012).

Marine microbial natural product discovery was initially focused on the easily accessible microorganisms from which a range of bioactive compounds have been described. However, the largest majority of the marine microorganisms (99%) are not currently culturable under laboratory conditions, and therefore the majority of bioactive compounds produced by marine microbiomes are unknown (Streit & Schmitz, 2004; Barone et al., 2014).

Metagenomics, or the study of DNA obtained directly from the environmental samples, has been mainly used over the last decades to decipher the enormous diversity of marine microorganisms and thus provides valuable means for unlocking the natural product potential harbored in the marine microbial genomes (Zerikly & Challis, 2009). Metagenomics showed that many more marine bacteria have far greater ability to produce natural specialized metabolites than was thought from classic bioactivity screens (Trindade et al., 2015). Rapid development of automated and high-throughput sequencing technologies and bioinformatics tools in the past decade have expanded databases of marine genomic information and largely promoted the genome mining as a methodology to locate and predict biosynthetic gene clusters of new natural product. Moreover several bioinformatics tools employed for genome mining have also been successfully applied to partially assembled DNA sequences from metagenomic sources to predict natural product compounds encoded within the genomes of uncultured microbes. Sequence-based metagenomics screening or sequence mining is a powerful tool for provides opportunities to identify strains with the greatest genetic potential to yield novel secondary metabolites without the requirement for complete genome sequencing or chemical isolation (Milshteyn et al., 2014).

Bacteria are known to produce structurally diverse secondary metabolites including nonribosomal peptides (NRPs) and polyketides (PKs). They are synthesized by multimodular enzymes nonribosomal peptide synthetases (NRPS) and type-I polyketide synthases (PKS-I). The PKS and NRPS genes are among the largest found in microbial genomes and have the molecular architectures consisting of activation (AT or A), thiolation (ACP or PCP), and condensation (KS or C) domains. In many case these secondary microbial metabolites exhibit a significant range of biological activity and act as antibacterial, antifungal, antiparasitic, antitumor and immunosuppressive agents (Wang et al., 2014). In order to discover novel PKS and NRPS secondary metabolites, several bioinformatics tools are now available to perform genome and sequence mining. The web tool Natural Product Domain Seeker (NaPDoS), which extracts and rapidly classifies KS and C domains from a wide range of sequence tags data, constitutes an approach to accurately predict pathway types, structural class of compounds that will be produced and in the case of high sequence identity the product structure itself. The NaPDoS approach employs a phylogeny based classification system that can be used to quantify and distinguish KS and C domain types from a variety of datasets including the incomplete genome assemblies typically obtained using next NGS technologies (Zeimert et al., 2012).

The ecosystems of the brackish Black Sea host unique and highly diverse microbial communities. Despite having a great biodiversity, the biotechnological potential of microorganisms inhabiting these ecosystems remain mostly unexplored

and unexploited. During the last years different state-of-the-art metagenomics and sequencing techniques have been adopted to assess marine bacterial diversity at the Romanian marine sector which generate important metagenomics datasets (Stoica et al., 2011; Stoica et al., 2012; Stoica et al., 2017). This article describes the first sequence mining of the Black Sea microbiome for nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). The study demonstrates the great potential of Black Sea bacteria for secondary metabolite production, with special focus on of medicinally relevant polyketides and nonribosomal peptides.

EXPERIMENTAL

The biosynthetic capacity of the Black Sea microbiome has been mined using a PCR-based sequence tags approach in combination with the automated mining pipeline NaPDoS (Zeimert et al., 2012; Milshteyn et al., 2014). The Natural Product Domain Seeker (NaPDoS) was used to detect and classify KS and C conservative domains of polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS) encoding genes in nucleotide and amino acid 16S rRNA metagenomic sequences generated by the ESF/N*in*E/2974/2009 project (Stoica et al., 2011). Briefly, the microbial metagenomes were collected in July 2009 from surface waters of the Black Sea at four stations (CE-1, CE-2, CE-3 and CE-5) located along the Constanta Est profile ($44^{0}10$'N) in the Romanian sector. Following the standard protocols of nucleic acids extraction (Griffiths et al., 2000) and cDNA synthesis (Gilbert et al. 2008), we performed 16S rRNA V6 amplicon pyrosequencing as described by Huber et al., 2007 and Gilbert et al., 2009. All DNA and cDNA were pyrosequenced using the GS-FLX Titanium platform at the NERC-funded Advanced Genomics Facility at the University of Liverpool (http://www.liv.ac.uk/agf/). All 16S reads were uploaded to the MG-RAST open source online server for phylogenetic and functional classification of metagenomics data [Meyer et al., 2008]. The data of 16S rRNA high throughput sequencing were deposited to Metagenome Rapid Annotation using Subsystem Technology (MG-RAST) database under the following accession numbers: 4447673.3; 4447674.3; 4447676.3; 4447677.3; 4447753.3; 4447752.3; 4447754.3; 4447756.3.

RESULTS AND DISCUSSION

*Metagenome and Metatranscriptome sequencin*g. The alpha diversity (OTUs richness) of each sample was calculated using the MG-RAST server and was generally higher for the coastal sites CE-1 (9342 OTUs) and CE-2 (8558 OTUs) than for the offshore sites CE-3 (5076 OTUs) and CE-5 (4354 OTUs) (Figure 1).The number of KS domain sequences from each dataset and sample are shown in Table 1 and Table 2. In the case of C domains, the number of sequences ranged from 1217 (CE-1) to 1627 (CE-5) for all metagenomes, and from 662 (CE-3) to 1584 (CE-5) for the metatranscriptomes.

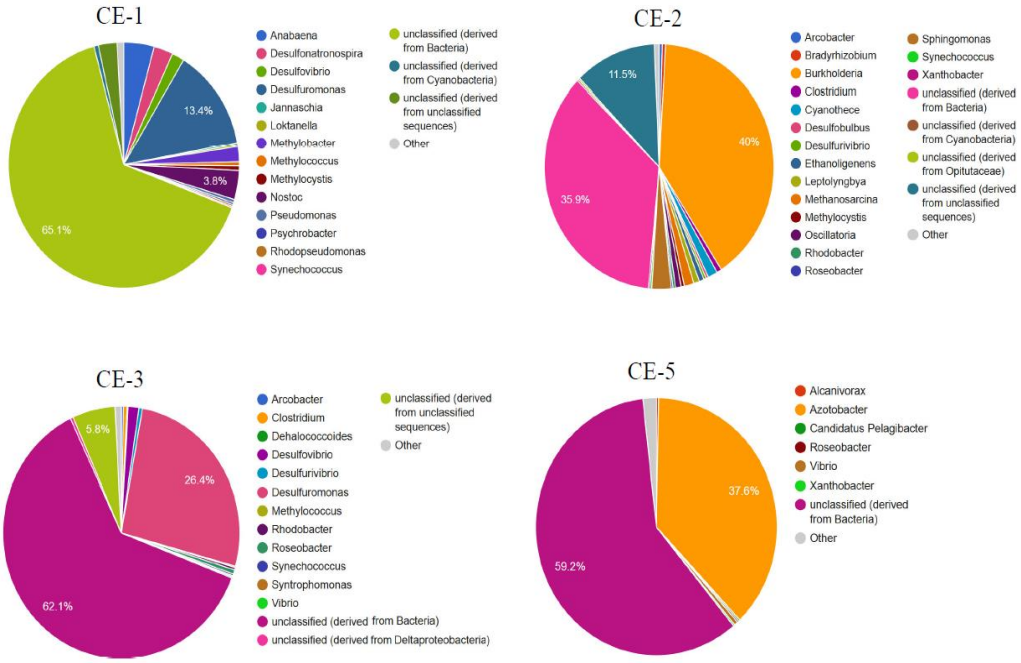


Fig. 1. The relative abundance (presented as percentage) of each bacterial genus reported for each sample analyzed by MG-RAST (Stoica et al. 2011).

Black Sea KS and C domains. Natural Product Domain Seeker (NaPDoS) was used to identify and classify ketosynthase (KS) domains from polyketide synthases and condensation (C) domains from non-ribosomal peptide synthetases from the obtained Black Sea short sequence reads. The web-based bioinformatics tool NaPDoS confirmed the sequences as KS or C domains and assigned an initial domain classification. 9 bacterial KS domain classes and 8 bacterial C domain classes were recognized by the NaPDoS classification system.

Metagenomes KS domains. A total of 1487 KS domains was identified based on blastp analysis of the metagenomic data against the NaPDoS reference KS sequences. The NaPDoS classification indicated that 816 (55%) of the Black Sea metagenome sequences were modular. 321(43%) KS domains were found within the trans-AT classes while fatty acid synthases (FASs) hit only 192 (13%) classes (Table 1). Of the remaining six C domain classes, a lower fraction was found in the Black Sea DNA metagenomes (3.7% enediyne, 2.8% typeII, 1.88% iterative, 1.27% KS1, and 0.4 PUFA and hybrid KS).

Table 1. NaPDoS KS results for metagenomic DNA data sets.

Data set	Total KS domains	# of sequence	Class*								
			FAS	TII	H	M	Tr	KS1	I	PUFA	E
CE-1	439	1217	2	9	1	362	26	0	2	2	35
CE-2	292	1506	2	13	1	81	182	0	5	2	6
CE-3	498	1469	188	13	3	163	79	16	20	1	15
CE-5	258	1627	0	8	1	210	34	3	1	1	0

**Class: Fatty acid synthase (FAS), TII- (Type II), H (Hybride); M (Modular), Tr (Trans-AT), KS1, I (Iterative), PUFA (*polyunsaturated fatty acid*), E (Enediyne)*

Metatranscriptomes KS domains. 839 KS domains were detected in all Black Sea metatranscriptomes with 313 (37%) trans-AT, 242 (29%) modular and 189 (23%) KS types (Table 2).

Metagenomes C domains. NaPDoS classified 22% epimerase, 18% DCL and LCL, 12% modeAA, and 11% cyc and dual C domains. The epimerase group was the largest (358) while the start group was the lowest (2% of total 1977) type of C domains (Figure 2).

Metatranscriptomes C domains. 1212 C domains were detected in the cDNA sequences. The epimerase was the highest fraction (20%) of C metatranscriptome domains. NaPDoS also classified 19% LCL, 17% dual, 16% cyc, 11% C and DCL, 34 starter and 23 modAA within the Black Sea metatranscriptomes C domains.

Table 2. NaPDoS KS results for metagenomic RNA (cDNA) data sets.

Data set	Total KS domains	# of sequence	Class*								
			FAS	TII	H	M	Tr	KS1	I	PUFA	E
CE-1	98	1196	0	0	0	25	58	12	2	0	1
CE-2	294	1414	4	4	1	80	199	0	2	0	4
CE-3	177	662	2	37	0	98	21	1	16	2	0
CE-5	270	1584	4	1	3	39	35	176	5	0	7

**Class: Fatty acid synthase (FAS), TII- (Type II), H (Hybride); M (Modular), Tr (Trans-AT), KS1, I (Iterative), PUFA (*polyunsaturated fatty acid*), E (Enediyne)*

Each sample site (CE-1, CE-2, CE-3, and CE-5) was in a different nutrient state and contained different groups of bacteria (Fig. 1) (Stoica et al. 2011). Therefore the number of KS and C domains from each site was counted to determine if differences were observable in secondary metabolite distribution. Generally, the distribution of C domains at each Black Sea offshore site was similar to the KS secondary metabolite distribution. Site CE-3 contained the highest

number of KS and C domains related to secondary metabolism. The site CE-5 had the lowest number of KS while the least amount of C domains was registered at the site CE-2. The coastal station CE-1 contained similar amount of KS and C metagenomics DNA domains except the cDNA KS domains which showed the least amount.

Phylogenetic analyses were also used to assess the similarity of KS and C sequences to those associated with experimentally characterized pathways to predict the types of products that would be produced. The meta-DNA KS sequences were closely related to 6 known polyketide synthases producing stigmatellin (71% similariy), tylosin and salinictam (73% similarity), kirromycin and leinamycin (75% similarity), and aclacinomycin and epothilone (77% similarity), respectively. The phylogenetic analyses also showed the greatest similarity (100%) to the virginiamycin trans-AT polyketide synthase (Figure 3). The amino acid sequences of KS domain were 67 to 77% similar to comparative sequences in the database responsible for production rifamycin, leinamycin, kirromycin and_virginiamycin. Many meta-DNA sequences had similarity to known C domains associated with known compounds (natural products) biosynthesis, e.g. fengycin (73%) thiocoraline (80%), calcium-dependent antibiotic (77%), yersiniabactin (78%), pristinamycin, thiocoraline and lychenicin (80%).

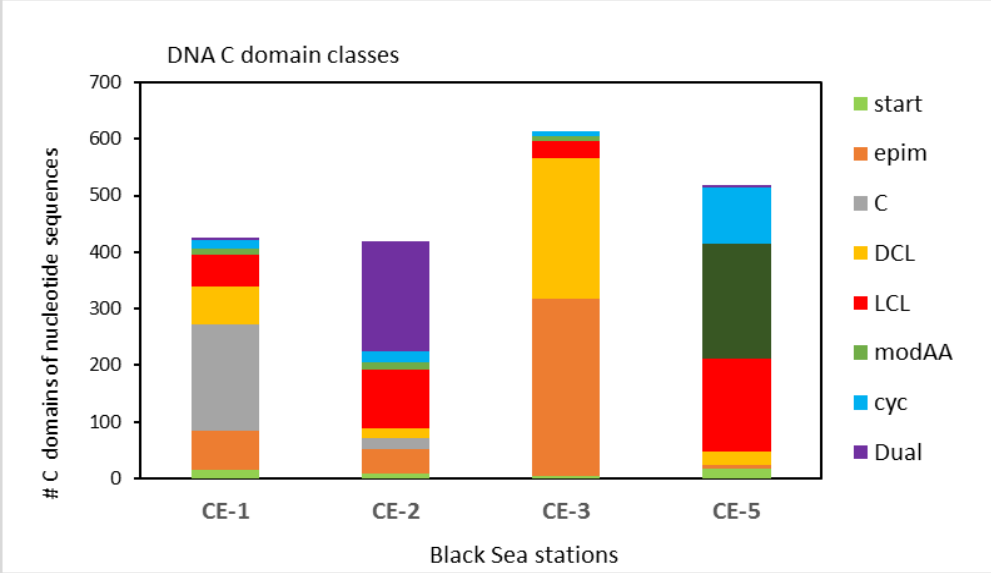


Fig. 2. Fraction of metagenome C domains at the Black Sea stations.

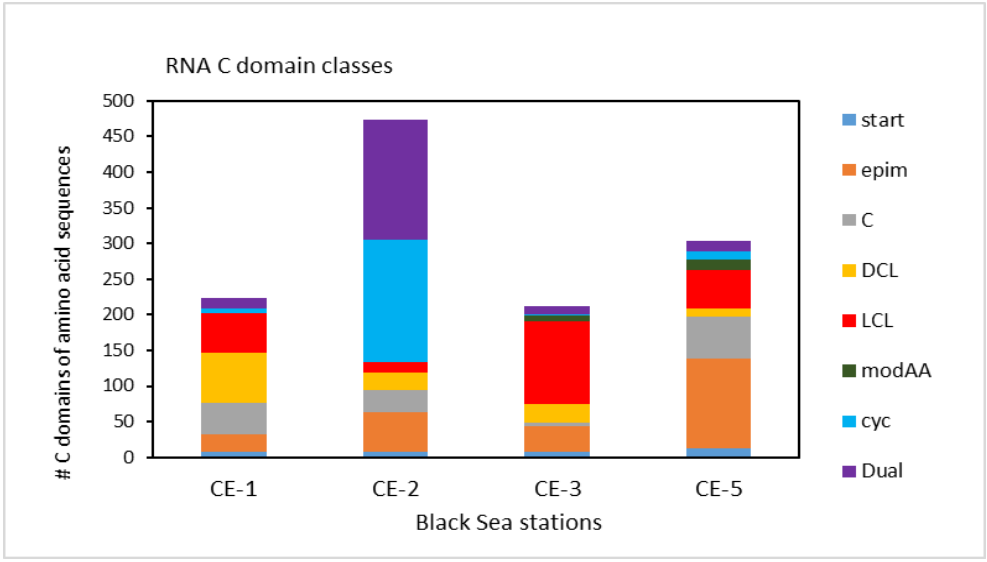
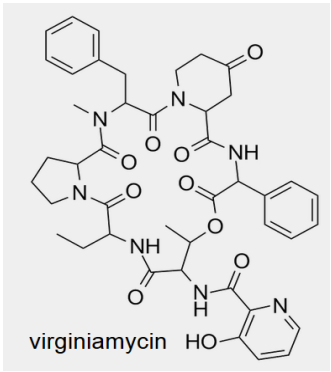
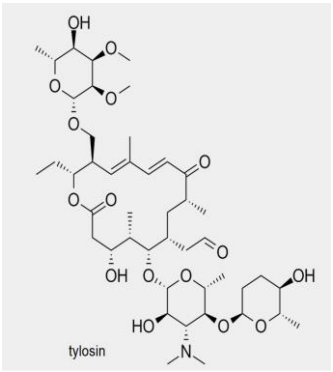
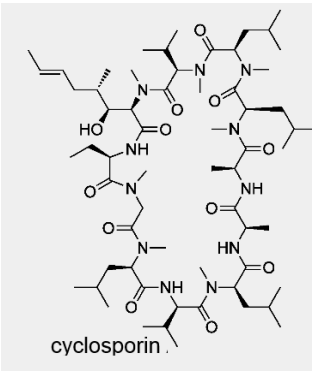
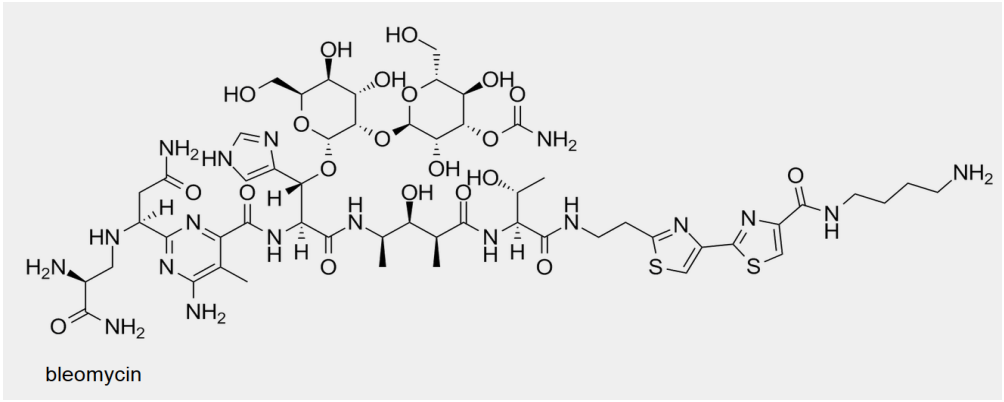


Fig. 3. Fraction of metatranscriptome C domains at the Black Sea stations.









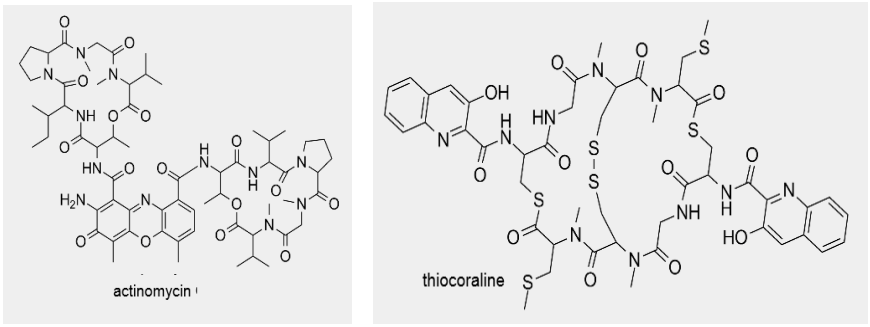


Fig. 4. Example of known PKs and NRPs secondary metabolites (natural products) that can be produced by the Black Sea microbiome.

Black Sea condensation domains also reveals a high degree of sequence similarity (89%) with known LCL sequences producing cyclomarin and bleomycin. The amino acid sequences of C domain were 73 to 89% similar to comparative Phylogenetic analysis of the sequences in the database related to secondary metabolism (Fig. 3). However majority of the Black Sea sequences (92% KS and 94% C) showed a low similarity (< 60%) to known bacterial secondary metabolite KS and C domains, suggesting that they are encoding for new compounds, which has not previously been described.

CONCLUSIONS

The study presented here provides the first search for KS and C domains from the Black Sea surface waters off the coast of Romania. Screening of the metagenomes revealed a wide diversity of KS and C domains 2326 KS and 3189 C domains. These sequences were classified by NapDoS system and BLAST, showing a close enough similarity (71 to 100%) to comparative sequences in the database responsible for production a large variety of secondary metabolites with pharmacological properties including, pristinamycin, thiocoraline, bleomycin, cyclomarin, and virginiamycin (Fig. 4). However, the high degree of divergence suggests that they probably represent biosynthetic pathways that make novel natural products. Overall, these results demonstrate that there is an enormous potential to discover natural products in the Black Sea. Further systematic studies are needed to fully know the secondary metabolites produced by the Black Sea microbiome.

Acknowledgements
This study has been carried out within the NUCLEU Project (PROMARE - PN16230202/2016), funded by National Authority for Scientific Research and Innovation (ANCSI).

REFERENCES

Barone et al. 2014 Streit, W.R., Schmitz, R.A. (2004), Metagenomics - the key to the uncultured microbes. Curr. Opin. Microbiol. 7: 492-498;

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: 38;

Gilbert J.A., Field D., Huang Y., Edwards R., Li W. et al. (2008), Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. PLoS ONE 3(8): e3042;

Gilbert J.A., Field D., Swift P., Newbold L., Oliver A., et al. (2009), Seasonal succession of microbial communities in the Western English Channel using 16S rRNA-tag pyrosequencing. Env. Microb 11(12): 3132–3139;

Glockner F.O., Stal L.J., Sandaa R-A., Gasol J.M., O'gara F., Hernandez F., Labrenz M., Stoica E. et al. (2012), Marine microbial diversity and its role in ecosystem functioning and environmental change, Marine Board Position Paper 17, Calewaert, J.B and McDonough N (Eds.). Publisher: European Scientific Foundation;

Griffiths R.I., Whiteley A.S., O'Donnell A.G., Bailey M.J. (2000), Rapid method for coextraction of DNA- and rRNA-based microbial community composition. Appl Environ Microbiol 66: 5488-5491;

Huber J.A., Mark Welch D.B., Morrison H.G., Huse S.M., Neal Pr, et al. (2007), Microbial population structures in the deep marine biosphere. Science 5: 97-100;

Meyer F., Paarmann D., D'Souza M., Olson R., Glass E.M., Kubal M., Paczian T., Rodriguez A., Stevens R., Wilke A., et al. (2008), The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinform, 9: 386;

Milshteyn A., Schneider J.S., Brady S.F. (2014), Mining the Metabiome*:* Identifying Novel Natural Products from Microbial Communities. Chem. Biol, 21: 1211-1223;

Rappé M.S., Giovannoni S.J. (2003), The uncultured microbial majority. Annu. Rev. Microbiol. 57: 369-394;

Stoica E. (2012), Molecular approaches for rapid and quantitative detections of cyanobacteria and their toxins from coastal Black Sea - MARCY, in: Funded Projects by the Black Sea ERA. NET 2010 Pilot Joint Call for Research Proposals, Ed. International Centre for Black Sea Studies (ICBSS), Printnow.gr, Greece, November 2012, pp. 17-20;

Stoica E., Pavlovska M., Dykyi E., Kormas K. (2017), Next generation sequencing-based approaches to characterize microbial pathogenic community and their potential relation to the Black Sea ecosystem status, CIESM Workshop Monographs N49, Int. Commission Scientific Exploration Mediterranean Sea, ISSN: 1726-5886;

Stoica E., Wilkening Keegen J. K., Rees A., Gilbert J.A. (2011), Amplicon pyrosequencing analysis of bacterial planktonic community composition from Northwestern Black Sea, Book of Abstract, 4th FEMS Congress of European Microbiologists, Geneva, Switzerland;

Trindade M., Van Zyl L., Navarro-Fernández J., Elrazak A.A. (2015), Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. Front microbiol. 6: 890;

Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A, Wu D.Y., Paulsen I. et al. (2004), Environmental genome shotgun sequencing of the Sargasso Sea, Science 304, 66-74;

Wang D., Fewer P., Holm L., Rouhiainen L., Sivonen K. (2014), Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. Proc Natl. |Acad Sci USA. 111(25): 9259–9264;

Zerikly M., Challis G.L. (2009), Strategies for the discovery of new natural products by genome mining. Chembiochem, 10: 625-63;

Ziemert N., Podell S., Penn K., Badger J.H., Allen E., Jensen P.R. (2012), The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. PLoS One. 7(3):e34064.