Babeș-Bolyai University
Faculty of Biology and Geology
Doctoral school Integrative biology

On-land mud volcanoes and the presence of bacteria capable of polyaromatic hydrocarbon oxidation: microbiologic and bioinformatic approaches (summary)

PhD student:
Alexei Remizovschi

Scientific supervisor:
Prof. Habil. dr. Anca Butiuc

Cluj-Napoca,
2024

Table of contents

Introduction

Mud volcanoes and PAH-degrading bacteria

Polyaromatic hydrocarbons (PAHs) are priority soil pollutants. They are the product of both anthropic fossil fuel combustion and natural catagenetic processes. PAHs can be detected in high quantities in soils adjacent to industrialized areas. The PAHs spiked soil represent a health risk which can lead to various types of cancer (Vandenbroucke et al., 2007; Keith, 2014).

In order to bioremediate the PAH-spiked soils, one must develop new biotechnology protocols, which would imply specialized PAH-degrading bacteria. Specifically, one should isolate bacteria from the naturally PAH-spiked ecosystem.

Mud volcano can be considered a naturally PAH-spiked ecosystem. Mud volcanoes are geological structures, whose genesis is tightly linked to catagenetic processes, which occur in the depths of sedimentary layers (Mazzini and Etiope, 2017). Given this aspect, mud volcanoes might theoretically sustain a community of highly specialized PAH-oxidizing bacteria. This comes as an alternative for the industrially PAH-spiked ecosystems which might be populated by only facultative PAH-degrading bacteria (Alain et al., 2006; Abdel-Shafy et al., 2016).

Sequential and textual biological databases

Nowadays, the majority of biological databases are based on either nucleotide or proteic sequences. As an example, one can mention databases such as Genbank or Uniprot. These databases are considered to be essential for any kind of biological research (Brown et al., 2014; Consortium, 2014; Benson et al., 2017). Given the exclusive dependence on biological sequence, these types of data bases have inherent limitations. For example, overreliance on the sequence ignores the research context of these sequences. Additionally, reliance on sequence does not count the outlook of the researchers, who uploaded these very sequences.

In order to mitigate this problem, this study proposed a new type of database. It described the first textual database, which would base itself on biological lexeme. This database will try to compensate the limitation of the sequential databases. Simultaneously, this database will serve as a minimal viable product and will be mainly focused on mud volcanoes.

Study objectives and the general purpose

The main purpose of this study is to show that terrestrial mud volcanoes can sustain the growth of a PAH-degrading bacterial community. In order to attain this purpose, we set the following objectives:

- Show by means of both microbiological and genomic techniques that mud volcanoes represent ecosystems fit for the in-depth study of PAH-degrading bacteria.

- The construction of the first textual data based which would be solely focused on mud volcanoes and would validate the microbiological results of this study.

Methods and materials

Sediment sampling and the isolation of PAH-degrading bacterial community

This study aimed to isolate PAH-degrading bacteria extracted from a mud volcano sediments (Lat: 48°2'12.0588"N, Long: 27°11'46.9428"E). These sediments were extracted with a manual sampler from the various depths. In order to create a representative sample, the sediments were mixed **(Figure 1)**.

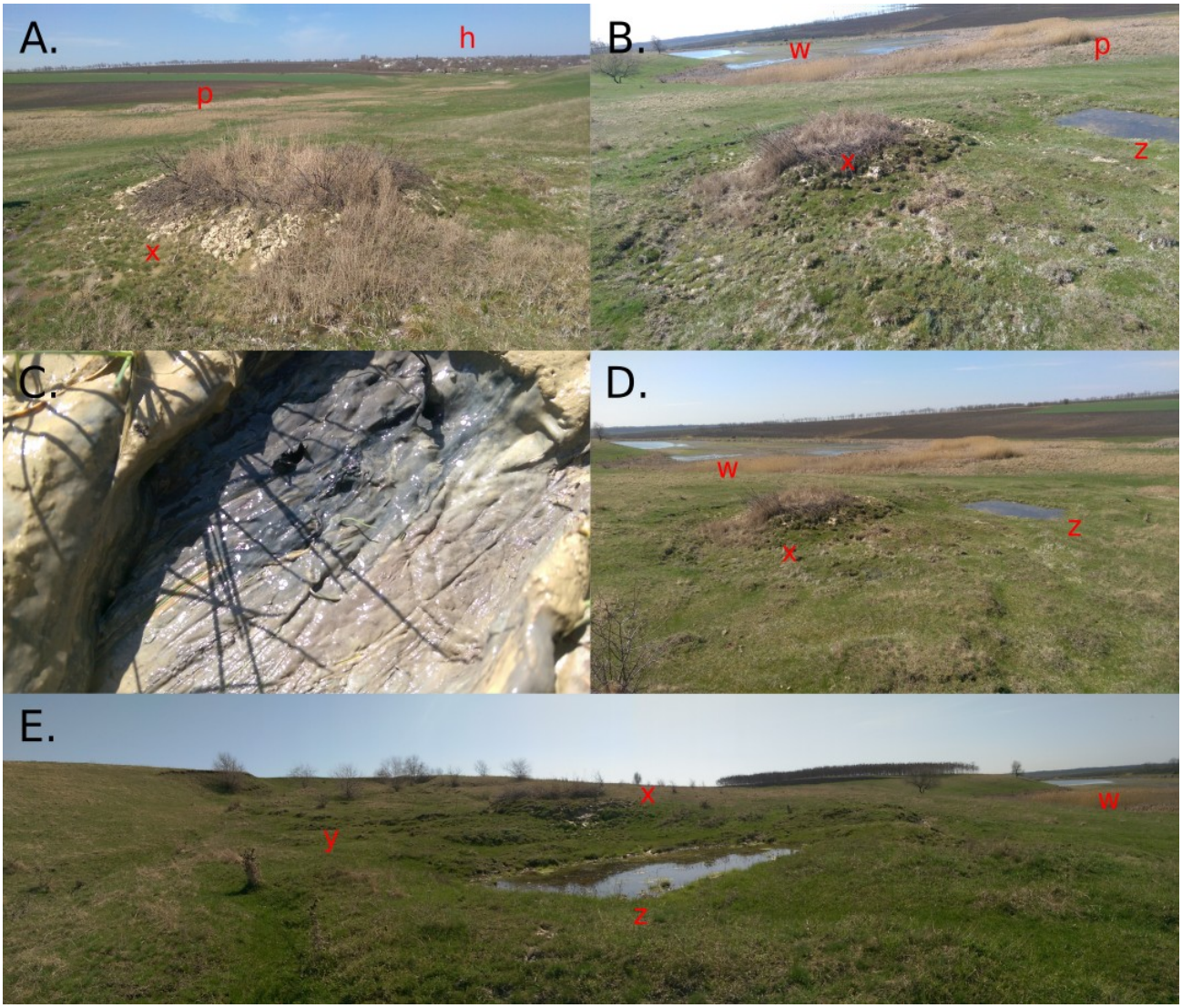

Figure 1. Investigated mud volcano. **A**, **B**, **D**, **E** – mud volcano and its localization, **C** – sediment sample. **x** – mud volcano, **p** - spring, **h** – Hancăuți village, **w** – adjacent water body, **z** - landslide, **y** – inactive mud volcanoes.

Enrichment of the PAH-degrading bacteria

Representative sediment sample was serially diluted and was used as inocula. Enrichment media (M9) with phenanthrene and pyrene as a carbon source was used as as culture media (Yin et al., 2020).

Microscopic analysis of the PAH-degrading bacteria

After the isolation of HAP-degrading consortium, an array of various microscopic analyses was performed:

- **Contrast phase microscopy – live bacteria assessment**

  Microscope: Zeiss Axio Scope.A1 (Zeiss, Germany).

- **Fluorescence microscopy (CARD-FISH/DAPI) – bacteria / archaea identification**

  Fixation: Paraformaldehyde (2%).

  Hybridization probes: *EUB3881*, *ARCH915,* and *nonEUB338*.

  Permiabilization: Lysozyme, Achromopeptidase P, Proteinase K.

  Tyramides: Alexa594.

  Microscope: Axio Imager Z2 (Zeiss, Germany).

- **Scanning electron microscopy (SEM) – identification of the physiological adaptations of the PAH-degrading bacteria**

  Dehydratation: Leica EM CPD300 (critical point drying).

  Microscope: Zeiss Merlin VP Compact (Zeiss, Germany).

Taxonomic identification of the PAH-degrading bacteria (*RNAr 16S* gene)

ZR Fungal/Bacterial DNA Miniprep kit (Zymo Research, US) was used for the DNA extraction. *RNAr 16S* gene was amplified with 27F (forward) and 1492R (reverse) primers. Amplicons were cloned into pJET1.2 vector. CloneJET PCR cloning kit was used for the cloning procedure (ThermoFisher Scientific, US). Plasmidic DNA was sliced and purified with ZR Plasmid Miniprep (Zymo Research, US) and *TaqI* enzyme (ThermoFisher Scientific, US). The obtained sequences were subjected to Sanger sequencing.

Transcriptome analysis

Enrichment culture was used for the transcriptomic analysis. Microscopical explorative assessment showed the numerical dominance of one type of bacteria (cocco-bacillus morphology). *RNAr 16S* gene profiling showed the dominant bacteria might be linked to *Pseudomonas* genus. Given these data, obtained transcriptome was aligned to a reference genome of the *Pseudomonas stutzeri* 19MN4 (NZCP0007509.1). QuasR library was used for the alignment procedure (Gaidatzis et al., 2014). Unwanted variance was removed by means of the RUVSeq (k=2) library (Risso, Ngai et al., 2014). EDAseq library was used for counts normalization (Risso, Schwartz et al., 2011). DESeq2 library was used for differential analysis (Love et al., 2014).

NLP data extraction algorithm and textual database

Raw data was collected from data corpus S2ORC (Lo et al., 2020). Textual data (N=118) was analyzed with in-house build algorithm (muddy_mine). This algorithm is based on open-source libraries such as spaCy, scispaCy and NCBI Taxonomy database (Honnibal and Johnson, 2015; Honnibal and Montani, 2017; Neumann at al., 2019; Schoch et al., 2020). The output generated by muddy_mine represents a list of csv tables which contain data regarding geology, chemistry and taxonomy affiliated to mud volcanoes. This output was used for the construction of muddy_db data base.

Results and discussion

PAH-degrading enrichment culture

After the microbiological analyses, PAH-degrading bacterial consortium was isolated. The consortium was composed of two distinct bacterial genera: *Pseudomonas sp.* and *Pseudoxanthomonas sp*. **(Figure 2).** In presence of PAHs, these bacteria synthesize a thick polysaccharidic matrix. This aspect was clearly seen during the microscopic assessment **(Figure 3)**.

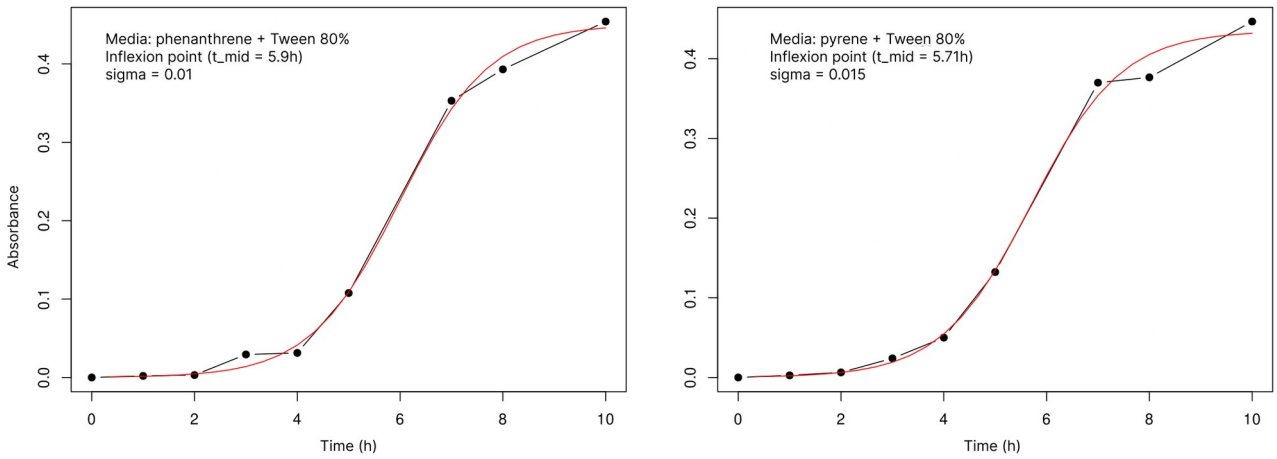

Figure 2. PAH-degrading consortia growth curves of both phenanthrene (PHE) and pyrene (PY) treatments.

Differential analysis of the expressed genes in the dominant PAH-degrading bacteria - *Pseudomonas sp.*

The analysis of the transcriptome of *Pseudomonas sp.* showed that the majority of overexpressed genes are related directly or tangentially to the synthesis of the extracellular polysaccharidic matrix. For example, the genes responsible for pyocyanin synthesis (*aroK* and *pstP*) were overexpressed up to 6.1 and 3.03 times than control (**Figure 4**) (Xu et al., 2005; Das et la., 2012; Chen et al., 2012).

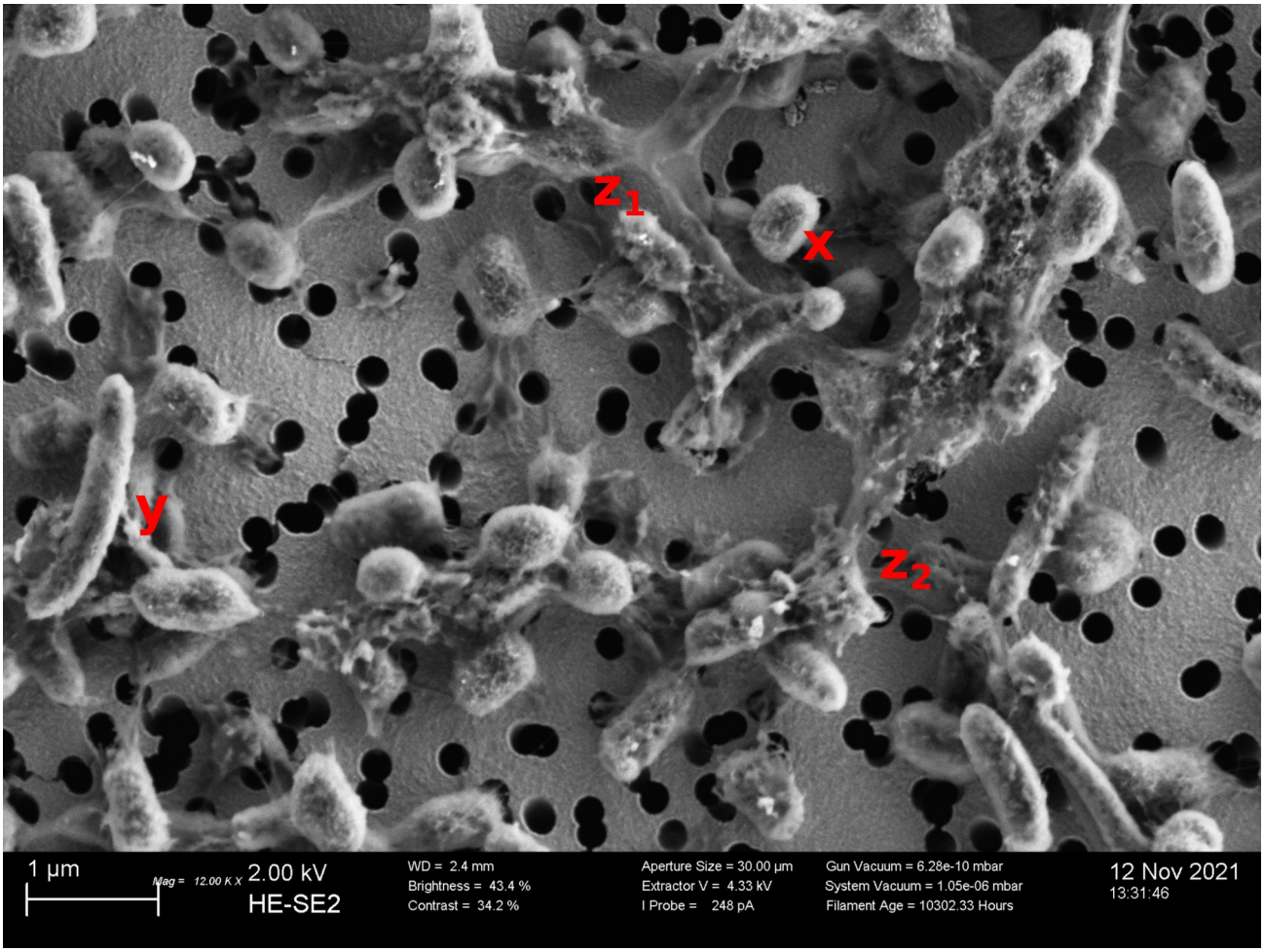

Figure 3. PAH-degrading enrichment culture visualized by scanning eletron microscope. The image shows two bacterial morphologies: cocco-bacillus (**x**, *Pseudomonas*) and bacillus (**y**, *Pseudoxanthomonas*). Bacteria are seen to be embedded into a polysaccharidic matrix (**z1**, **z2**).

In addition to pyocyanin related genes, there was a clear overexpression of *cobA* gene. This gene is related to cobalamin synthesis and by extension influences extracellular polysaccharidic accumulation (Warren et al., 2002; Crespo et al., 2018). Gene *cobA* was overexpressed up to 2-4 times more than control. In addition to cobalamin, differential analysis suggested that riboflavin and thiamine cofactors might significantly influence biofilm synthesis (Webb and Downs, 1997; Fassbinder et al., 2000). Data showed that *ribA* and *thiL* genes were extremely underexpressed. Gene *ribA* was underexpressed up to 16 times less than control. Gene *thiL* was

underexpressed up to 32 times less than control. In addition to vitamin related genes, data showed that genes (*cysE* and *carA*) related to cell motility were downregulated (up to 2-16 less times than control) (Sturgill et al., 2004; Butcher et al., 2016) (**Figure 4**).

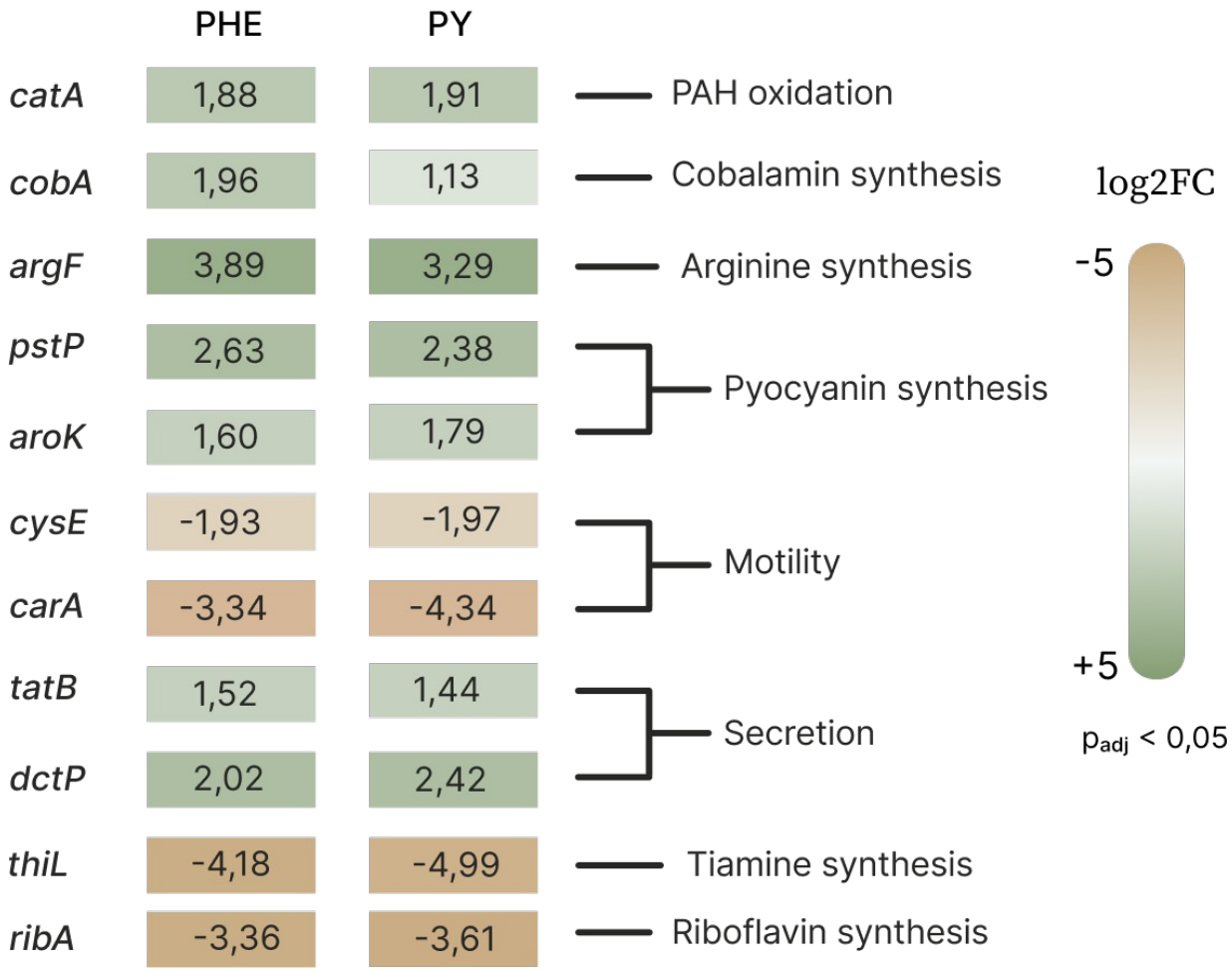

Figure 4. Differential analysis of the genes expressed in *Pseudomonas sp.* log2FC - log2 fold change; PHE – phenanthrene; PY- pyrene.

Textual database / web application – muddy_db

Data for muddy_db was generated with the NLP algorithm muddy_mine. Database muddy_db represents the first iteration of a textual database, based exclusively on the mud volcano biology. The data for muddy_db was extracted from all the available mud volcano related open-access articles published in the last 20 years. In comparison to classic sequential databases, muddy_db provides the frequency of biologically-related

lexemes. This frequency can suggest the context and the importance of this lexemes in a given article.

The main page of muddy_db application (MAP) represents an interactive map where user can see the geographical localization of the academically described mud volcanoes. Each MAP entry contains data regarding: mud volcano geographical coordinates, title and the authors of the article and the PMID of the affiliated article. ARTICLES page contains all the information regarding the articles used in muddy_db creation. BACTERIA and ARCHAEA pages contain taxonomical data found in the articles. MUD VOLCANO, GEOLOGY and METHODS pages present tables with lexemes related to geology, typical mud volcano terminology and experimental methods mentioned in the mined articles **(Figure 5)**.

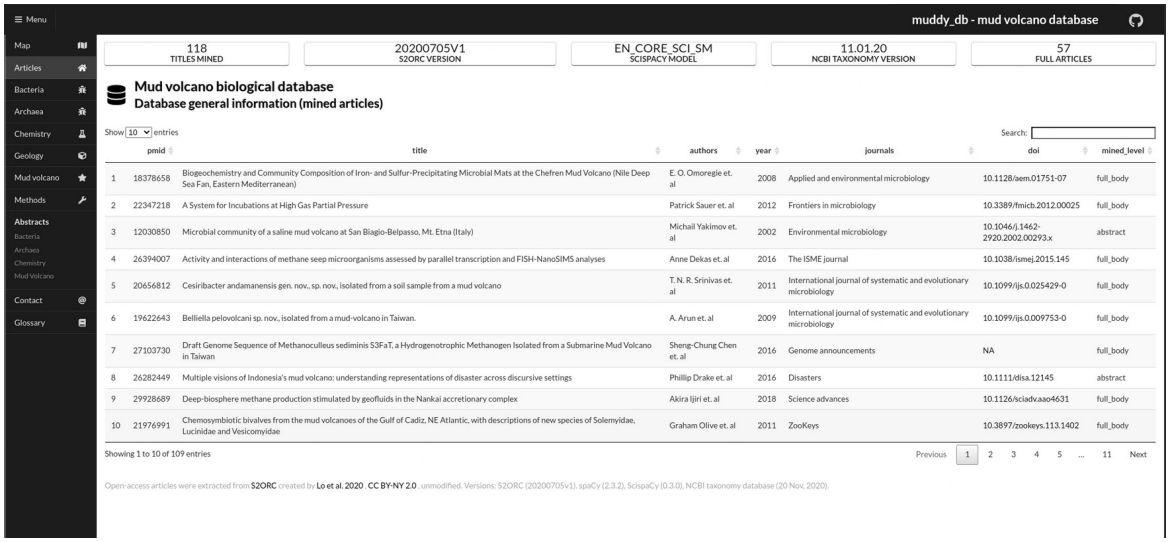

Figura 5. Database / web application muddy_db. The first textual mud volcano database.

Conclusion

Mud volcanoes could be named appropriate ecosystems for PAH-degrading bacteria growth and by extension an appropriate study subject. This study proved that one can isolate HAP-degrading bacteria from mud volcano sediments. This study described for the first time the PAH-degrading *Pseudomonas* / *Pseudoxanthomonas* consortium. Additionally, the study suggested new methods for both amplification and suppression of biofilm synthesis and PAH-oxidation. Specifically, this study showed that riboflavin and thiamine have inhibitory action on biofilm synthesis. On the other hand, cobalamin can amplify the extracellular polysaccharidic synthesis and by extension bacterial PAH-oxidation.

The first textual database muddy_db described the state of mud volcano research, provided data for further meta-analysis, and validated the microbiological data generated by this study. Specifically, muddy_db showed that there was no mention of PAH-degrading *Pseudomonas* / *Pseudoxanthomonas* consortia in mud volcano literature.

Resources

All the data related to this study was archived.

- Raw transcriptome sequences (access number: PRJNA843935) are available at: https://doi.org/10.5281/zenodo.4587649.
- Transcriptome analysis pipeline is available on Github (https://github.com/TracyRage/rnaseq_pipeline)
- RNAr 16S sequences are available at: https://doi.org/10.5281/zenodo.6538838.
- All the microscopic images are available at: CARD-FISH/DAPI (https://doi.org/10.5281/zenodo.4553960) and SEM (https://doi.org/10.5281/zenodo.5812226).
- muddy_mine source code is available at: https://github.com/TracyRage/muddy_mine.
- muddy_db source code is available on Github (https://github.com/TracyRage/muddy_db).

Keywords

hydrocarbonoclastic, mud volcano, PAH, pseudoxanthomonas, textual database, thiamine, riboflavin

References

Abdel-Shafy H and Mansour M. 2016, A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian Journal of Petroleum*, **25**:(1), 107–123. https://doi.org/10.1016/j.ejpe.2015.03.011.

Alain K, Holler T, Musat F, Elvert M, Treude T and Kruger, M. 2006, Microbiological investigation of methane- and hydrocarbon-discharging mud volcanoes in the Carpathian Mountains, Romania. *Environmental Microbiology*, **8**:(4). 574–590. https://doi.org/10.1111/j.1462-2920.2005.00922.x

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD and Sayers EW. 2017, GenBank. *Nucleic Acids Research*, **46**:(D1), D41–D47. https://doi.org/ 10.1093/nar/gkx1094

Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, Tolstoy I, Tatusova T, Pruitt KD, Maglott DR and Murphy TD. 2014, Gene: a gene-centered information resource at NCBI. *Nucleic Acids Research*, **43**:(D1), D36–D42. https://doi.org/10. 1093/nar/gku1055

Butcher B, Chakravarthy S, D'Amico K, Stoos K and Filiatrault M. 2016, Disruption of the *carA* gene in *Pseudomonas syringae* results in reduced fitness and alters motility. *BMC Microbiology*, **16**:(1). https://doi.org/10.1186/s12866-016-0819-z

Chen K, Dou J, Tang S, Yang Y, Wang H, Fang H and Zhou C. 2012, Deletion of the aroKgene is essential for high shikimic acid accumulation through the shikimate pathway in *E. coli*. *Bioresource Technology*, **119**, 141–147. https://doi.org/10.1016/j.biortech.2012.05.100

Consortium U. 2014, UniProt: a hub for protein information. *Nucleic Acids Research*, **43**:(D1), D204–D212. https://doi.org/10.1093/nar/gku989

Crespo A, Blanco-Cabra N and Torrents E. 2018, Aerobic vitamin B12 biosynthesis is essentialfor *Pseudomonas Aeruginosa* class ii ribonucleotide reductase activity during planktonic and biofilm growth. *Frontiers in Microbiology*, **9.** https://doi.org/10.3389/fmicb.2018.00986

Das T, and Manefield M. 2012, Pyocyanin promotes extracellular DNA release in *Pseudomonas aeruginosa*. *PLoS ONE*, **7**:(10), e46718, https://doi.org/10.1371/journal.pone. 0046718

Fassbinder F, Kist M and Bereswill S. 2000, Structural and functional analysis of the riboflavin synthesis genes encoding GTP cyclohydrolase II (*ribA*), DHBP synthase (*ribBA*), riboflavin synthase (*ribC*), and riboflavin deaminase/reductase (*ribD*) fromHelicobacter pyloristrain P1. *FEMS Microbiology Letters*, **191**:(2), 191–197. https://doi.org/10.1111/j.1574-6968.2000.tb09339.x

Gaidatzis D, Lerch A, Hahne F and Stadler M. 2014 QuasR: quantification and annotation of short reads in R. *Bioinformatics*, **31**:(7), 1130–1132. https://doi.org/10.1093/ bioinformatics/btu781

Honnibal M and Johnson M. 2015, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, *Association for Computational Linguistics,* https : //doi.org/10.18653/v1/d15-1162

Honnibal M and Montani I. 2017, spaCy 2: Natural language understanding with Bloom embed-dings, convolutional neural networks and incremental parsing, To appear

Keith L. 2014, The Source of U.S. EPA's sixteen PAH priority pollutants. polycyclic aromatic compounds, **35**:(2-4), 147–160. https : / / doi . Org / 10 . 1080 / 10406638 . 2014 . 892886

Lo K, Wang L, Neumann M, Kinney R and Weld D. 2020, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, *Association for Computational Linguistics*, https://doi.org/10.18653/v1/2020.acl-main.447

Love M, Huber W și Anders S. 2014, Moderated estimation of fold change and dispersion fo RNA-seq data with DESeq2. *Genome Biology,* **15**:(12). https://doi.org/10.1186/s13059-014-0550-8

Mazzini A and Etiope G. 2017, Mud volcanism: An updated review. *Earth-Science Reviews*, **168**, 81–112, **168**, 81–112. https://doi.org/10.1016/j.earscirev.2017.03.001

Neumann M, King D, Beltagy I and Ammar W. 2019, Proceedings of the 18th BioNLP Workshop and Shared Task, *Association for Computational Linguistics.* https://doi. org/10.18653/v1/w19-5034

Risso D, Ngai J, Speed T and Dudoit S. 2014, Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, **32**:(9), 896–902. https: //doi.org/10.1038/nbt.2931

Risso D, Schwartz K, Sherlock G and Dudoit S. 2011, GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, **12**:(1), 480. https://doi.org/10.1186/1471-2105-12-480

Schoch C, Ciufo S, Domrachev M, Hotton C, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan J, Sun L, Turner S and Karsch-Mizrachi I. 2020, NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database.* https://doi.org/10.1093/database/baaa062

Sturgill G, Toutain CM, Komperda J, O'Tool G and Rather PN. 2004, Role of CysE in Production of an Extracellular Signaling Molecule in *Providencia stuartii* and *Escherichia coli*: Loss of *cysE* Enhances Biofilm Formation in *Escherichia coli.*

Journal of Bacteriology, **186**:(22). https://doi.org/10.1128/jb.186.22.7610-7617.2004

Vandenbroucke M and Largeau C. 2007, Kerogen origin, evolution and structure. *Organic Geochemistry*, **38**:(5), 719–833. https://doi.org/10.1016/j.orggeochem.2007. 01.001

Webb E and Downs D. 1997, Characterization of thiL, encoding thiamin-monophosphate kinase in Salmonella typhimurium. Journal of Biological Chemistry, **272**, 25. http://dx.doi.org/10.1074/jbc.272.25.15702

Warren MJ, Raux E, Schubert HL and Escalante-Semerena JC. 2002, The biosynthesis of adenosylcobalamin (vitamin B12). *Natural Product Report*s, **19** :(4), 390–412. https : //doi.org/10.1039/b108967f

Xu H, Lin W, Huiming X, Shuwa L, Yingli Y, Hongming B, Fing Z, Xiuming B, Yanglin S and Per QM. 2005, Influence of ptsP gene on pyocyanin production in *Pseudomonas aeruginosa*. *FEMS Microbiology Letters*, **253**:(1), 103-109. http://dx.doi.org/10.1016/j.femsle.2005.09.027

Yin C, Xiong W, Qiu H, Peng W, Deng Z, Lin S and Liang R. 2020, Characterization of the phenanthrene-degrading *Sphingobium yanoikuyae* SJTF8 in heavy metal co-existing liquid medium and analysis of its metabolic pathway. *Microorganisms*, **8**:(6), 946. https: //doi.org/10.3390/microorganisms8060946