GOLLUM: ANALYSIS OF EXTREME ECOSYSTEMS

Mireia Bernabeu-Gimeno[1,2], Jose Manuel Martí[1], Wladimiro Diaz-Villanueva[1,2], Vicente Arnau[1,2] and Carlos Peña-Garay[1]

[1] Department of Theory, Bioinformatics and Computation. Institute for Integrative Systems Biology (I2SysBio), UVEG-CSIC. Valencia, Spain.
[2] Departamento de Informática, Escuela Técnica Superior de Ingeniería (ETSE), Universidad de Valencia (UVEG). Valencia, Spain.

INTRODUCTION

An ecosystem is a biological system composed by a set of living beings and the environment where they connect. The study of genetic material obtained from environmental samples, known as Metagenomics, allows us to study the taxonomic and functional characteristics of a living community as a complex system. In ecosystems, competition between neighboring cells because of space and resources is intense and the production of antibiotics is one the most well-known competitive mechanisms. Extreme environments enhance competition between their habitants turning them into ideal places to look for possible new antibiotics.

Gollum is a project coordinated by the Institute for Integrative Systems Biology (I²SysBio) and the Canfranc Underground Laboratory (LSC) that aims at studying the ultra-oligotrophic endolithic microbial communities living in a range of different host rocks along the Somport tunnel.

METODOLOGY

1. Sampling: Somport tunnel, Canfranc

2. DNA extraction: Rock pulverization & treatment with adapted protocols

3. Sequencing: Shotgun Metagenomics

4. QC: FastQC (1) and MultiQC (2)

5. Classification: Centrifuge (3)

6. Robust contamination removal: Recentrifuge (4) and Rextract (4)

7. Sequence mapping: BWA (5)

8. Functional analysis: BLASTX against non-redundant database (7)

9. Looking for antibiotic resistance genes: BLASTn against "The Comprehensive Antibiotic Resistance Database" (6,7)

10. Looking for antibiotic biosynthesis gene clusters (BGCs): antiSMASH analysis ("knownclusterblast" option) (8)
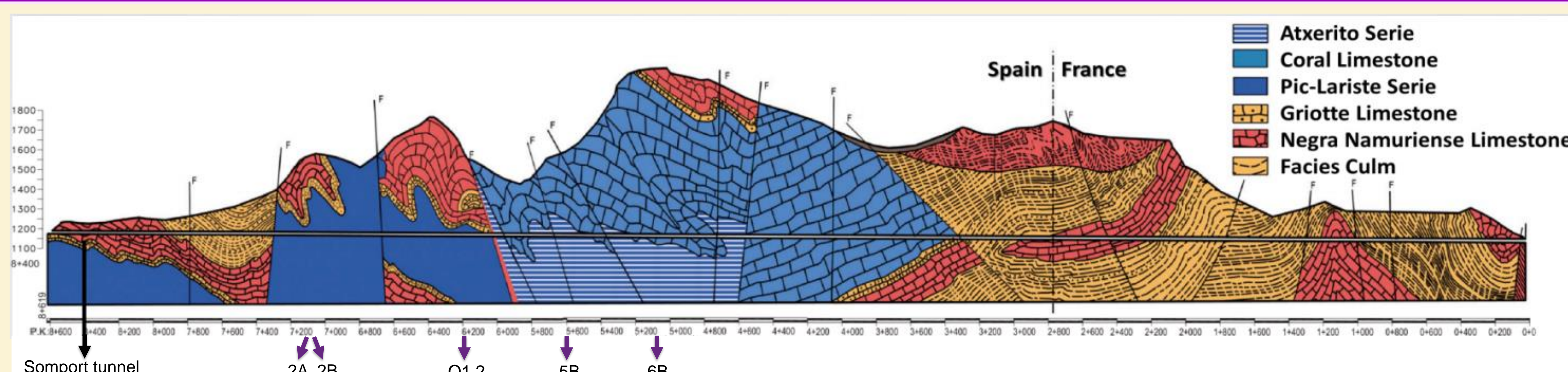

Figure 1. Sampling sites along Somport tunnel, Canfranc.
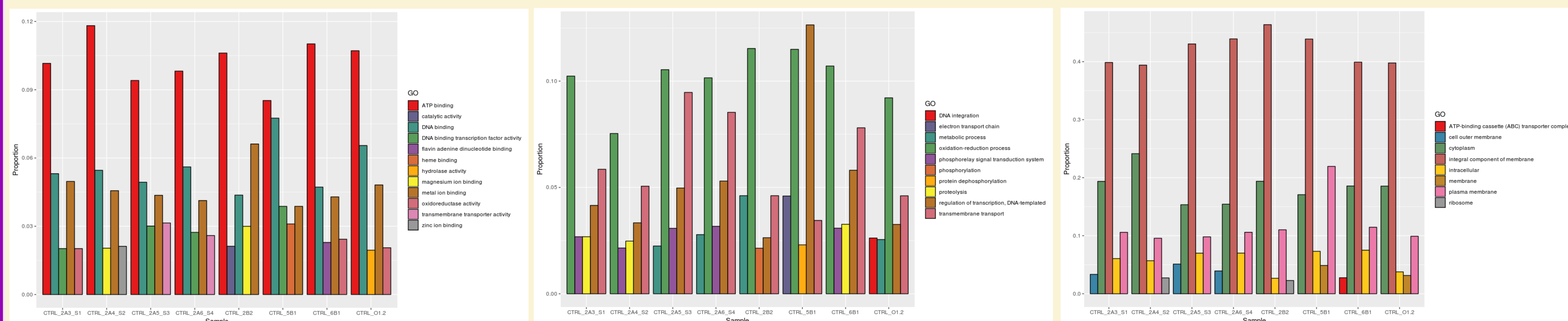
RESULTS

Figure 2. Over-represented GO (gene ontology) terms in each sample derived from the "Molecular function", "Biological process" and "Cellular component" categories.
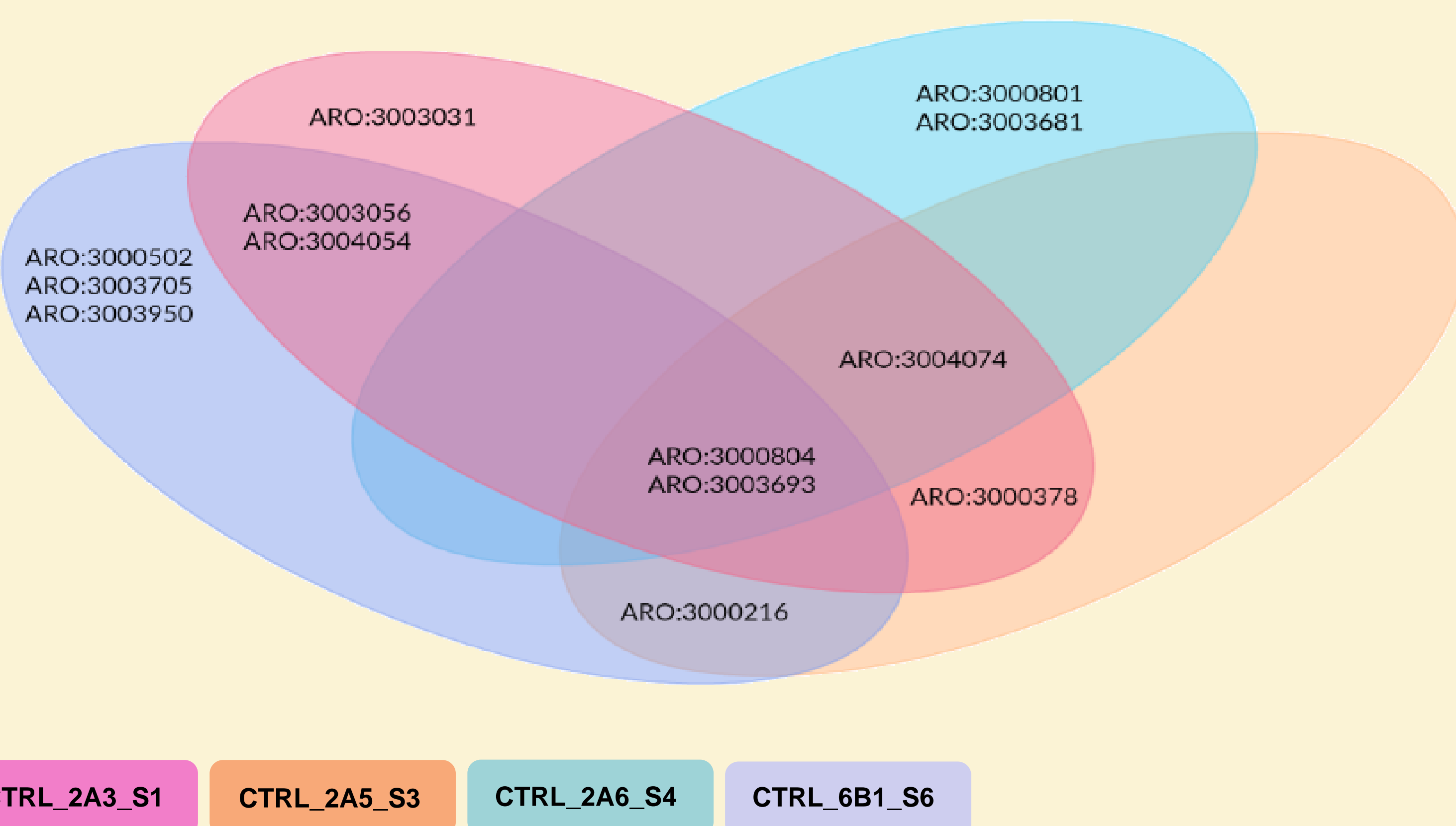

Figure 3. Venn diagram of ARO (antibiotic resistance organisms) identifiers with the highest relative frequency in the four *CTRL_** (control-subtracted) subsamples with a higher number of results.

Sample(s)	Taxid	Most similar known cluster
CTRL_2A3_S1	1500687	2,4-Diacetylphloroglucinol_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_2A6_S4	47883	Pyocyanine_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_2A6_S4	200451	Poaeamide_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_2A6_S4-CTRL_6B1_S6	380021	Pyoluteorin_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_2A6_S4-CTRL_6B1_S6	380021	Pyrrolnitrin_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_2A6_S4-CTRL_6B1_S6	380021	Rhizoxins_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_2A6_S4-CTRL_6B1_S6	380021	2,4-Diacetylphloroglucinol_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_6B1_S6	587753	Pyrrolnitrin_biosynthetic_gene_cluster
CTRL_2A3_S1-CTRL_2A5_S3-CTRL_6B1_S6	1611770	Pyrrolnitrin_biosynthetic_gene_cluster
CTRL_2A4_S2-CTRL_2A5_S3-CTRL_2A6_S4	47878	Poaeamide_biosynthetic_gene_cluster
CTRL_2A5_S3	536	Violacein_biosynthetic_gene_cluster
CTRL_2A5_S3	50340	Nunapeptin_/_nunamycin_biosynthetic_gene_cluster
CTRL_2A5_S3	75588	Pyocyanine_biosynthetic_gene_cluster
CTRL_2A5_S3	1534110	Bananamides_biosynthetic_gene_cluster
CTRL_2A5_S3-CTRL_2A6_S4	169669	Poaeamide_biosynthetic_gene_cluster
CTRL_2A5_S3-CTRL_2A6_S4	321662	Bananamides_biosynthetic_gene_cluster
CTRL_2A6_S4	76758	Poaeamide_biosynthetic_gene_cluster
CTRL_5B1_S7-CTRL_6B1_S6	287	2-amino-4-methoxy-trans-3-butenoic_acid_biosynthetic_gene_cluster
CTRL_6B1_S6	86265	2,4-Diacetylphloroglucinol_biosynthetic_gene_cluster
CTRL_6B1_S6	183795	Brabantamide_biosynthetic_gene_cluster
CTRL_6B1_S6	219572	Poaeamide_biosynthetic_gene_cluster
CTRL_6B1_S6	1500686	Rhizoxins_biosynthetic_gene_cluster
CTRL_6B1_S6	1678028	Violacein_biosynthetic_gene_cluster
CTRL_6B1_S6	1881017	Syringomycin_biosynthetic_gene_cluster

Table 1. Identified BGCs with a 100% of genes that show similarity with the genes of known BGCs associated with the biosynthesis of secondary metabolites with described antibiotic activity.

CONCLUSIONS

Microorganisms in ultra-oligotrophic endolithic environments are in serious competition for nutrients and have developed different offensive and defensive strategies, like the synthesis of antibiotics.

To whom correspondence should be addressed:

mibergi@alumni.uv.es

REFERENCES

1- Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
2- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047-3048.
3- Kim, D., Song, L., Breitwieser, F., & Salzberg, S. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Research, 26(12), 1721-1729. doi: 10.1101/gr.210641.116
4- Martí, J.M. (2018) Recentrifuge: robust comparative analysis and contamination removal for metagenomics. bioRxiv, 190934. doi: 10.1101/190934
5- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754-1760. doi: 10.1093/bioinformatics/btp324
6- Altschul, S., Gish, W., Miller, W., Myers, E., & Lipman, D. (1990). Basic local alignment search tool. Journal Of Molecular Biology, 215(3), 403-410. doi: 10.1016/s0022-2836(05)80360-2
7- McArthur, A., Waglechner, N., Nizam, F., et al. (2013). The Comprehensive Antibiotic Resistance Database. Antimicrobial Agents And Chemotherapy, 57(7), 3348-3357. doi: 10.1128/aac.00419-13
8- Medema, M., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., & Fischbach, M. et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. , 39(suppl_2), W339-W346. doi: 10.1093/nar/gkr466