3ème Conférence Internationale sur les Sciences Appliquées et l'Innovation (CISAI-2025)
Proceedings Book Series –PBS- Vol 25,49-57

Integrated Metagenomic and Metabolomic Analysis of a Hypersaline Lake Ecosystem via MG-RAST

Souad AMIOUR[#1], Karim CHEKROUD[*2]

[# *]Laboratory of Bioengineering, National Higher School of Biotechnology, Constantine, Algeria

Email 1 - s.amiour@ensbiotech.edu.dz

Email 2 - karimchekroud@gmail.com

Abstract— Hypersaline ecosystems are distributed globally and are typically characterized by limited microbial diversity, largely due to the combined influence of extreme environmental factors.

This study explores the microbial community structure and metabolic potential of Lake Djendli (DJS), a hypersaline lake located in the Batna province of eastern Algeria. We employed a culture-independent metagenomic approach, processed through the MG-RAST server, to investigate the distribution and functional profiles of prokaryotic communities in the hypersaline soils. Our results indicate that the microbial community in Lake Djendli was predominantly composed of Bacteria (67.70%), followed by Archaea (31.68%). Within the bacterial domain, the most abundant phylum was Bacteroidetes (28.06%), followed by Proteobacteria (22.65%), Firmicutes (6.47%), and Cyanobacteria (1.28%). Among the archaeal sequences, there was a clear predominance of the phylum Euryarchaeota, with 36.38% of archaeal sequences belonging to the family Halobacteriaceae (39.33%). At the genus level, Halorubrum (4.35%), Halorhodospira (4.38%), Haloarcula (2.72%), and Halorhabdus (2.14%) were detected in the DJS sample.

Functional annotation revealed that metabolic functions were predominant (59.41%), followed by genetic information processing (21.91%) and environmental information processing (13.25%). Subsystem-based classification further highlighted carbohydrate metabolism (13.50%) and cluster-based subsystems (13.32%) as the most represented functional categories. These findings provide new insights into the microbial ecology and functional capacity of hypersaline environments in arid regions.

Keywords— hypersaline soils, metagenomic , MG-RAST server, Lake Djendli

I. INTRODUCTION

Hypersaline environments are generally inhabited by a limited variety of life forms (Lopez et al. 2010) including extreme habitats with limited microbial diversity, due to the combined effects of several environmental factors (Ventosa et al. 2015), as well as aquatic and terrestrial habitats (Vera-Gargallo and Ventosa 2018).

In this study, next-generation sequencing (NGS) technology was employed to explore microbial diversity and assess metabolic potential in samples collected from the hypersaline Lake Djendli (DJS), located in the Batna province of the High Plateaus in eastern Algeria. This region is characterized by an arid climate with low rainfall.

The objective of this research is to produce a highly accurate and contiguous genome assembly to support precise annotation of microbial communities and their metabolic functions. To achieve this, a culture-independent approach known as metagenomics was used, allowing for the analysis of microbial community structure and functional potential through gene sequence identification and functional profiling. For that, MG-RAST server was used to examine many features of DJS metagenome.

II. EXPERIMENTAL METHODOLOGY

DNA extraction from soils was rated the most complex compared to other compartments. We adapted our extraction protocol by using two methods; direct and indirect to obtains an adequate and sufficient biomass (See more details in Högfors-Rönnholm et al. 2018). Genomic DNA extracted from the hypersaline soil of Sebkha Djendli was quantified using the Qubit system. Then, DNA of DJS sample was sent to Macrogen Inc (Seoul, Korea) and sequenced through Illumina HiSeqTM technology. The list of bioinformatics analysis performed is summarized in Fig.1.

ISSN: 2961-6611

All analyses were carried out in Linux, containing the following steps with tools utilized: *IDBA tool* (Peng et al. 2012) was used for assembly in order to obtain the large contigs. Binning of metagenomes was performed with *MaxBin v2.2.4* (Wu et al. 2014), and quality parameters were checked with the *HMM.essential.rb* script (http://enve-omics.ce.gatech.edu/enveomics/). In each binning run, only contigs from the assembly of an individual sample were used. Coding regions within the sequences were predicted using *FragGeneScan*. The nucleotide sequence compilation called CoupleReads.fa contains the predicted coding regions. This step identifies the most likely reading frame and translates nucleotide sequences into amino acids sequences. The predicted genes, possibly more than one per fragment, are called features. We highlighted the functional and taxonomical classifications using the *MetaGenome Rapid Annotation* with Subsystem Technology (MG-RAST) server (https://www.mg-rast.org/).

III. RESULTS AND DISCUSSION

All data below represents the analysis generated by the MG-RAST processing pipeline. The metagenomic sequence was compared to protein-coding gene databases; here the short or low quality sequences with ambiguous bases were not included in this analysis.

1. Statistical Analysis of the Metagenome and Predicted Features

Using shotgun metagenomic sequencing, the DJS dataset comprised 39,154,118 sequences, totaling 5,773,553,587 base pairs, with an average read length of 147 bp. Quality control (QC) filtering resulted in the removal of 9,076,694 sequences (23.18%) that did not meet the QC criteria. Among these, 6,604,204 sequences were identified as artificial duplicate reads, often consisting of short reads lacking sufficient taxonomic information.

Of the sequences that passed QC, 77,537 (0.27%) contained ribosomal RNA genes. Additionally, 13,037,285 sequences (45.70%) encoded predicted proteins with known functions, while 15,410,317 sequences (54.02%) contained predicted proteins with unknown functions.

The average guanine-cytosine (GC) content of the DJS metagenome was 54% ± 12%. Following quality control, a total of 30,077,424 sequences remained, with an average length of 150 ± 4 base pairs, resulting in a total sequence output of 4,499,689,497 base pairs (Table 1, Fig.2).

1. Taxonomic hits distribution

The DJS metagenome represented a diverse sampling of hundreds of thousands of prokaryotic organisms, primarily from the domains Bacteria and Archaea, spanning numerous genera. In this dataset, the majority of sequences (67.70%) were affiliated with the domain Bacteria, while Archaea accounted for 31.68% (Fig.3). Within the bacterial fraction, the phylum Bacteroidetes was the most dominant, contributing 28.06% of sequences. This phylum is commonly associated with hypersaline sediments and soils and is known for its ability to tolerate a wide range of salinity levels (Zhao et al., 2020). The second most abundant phylum was Proteobacteria (22.65%), a group widely recognized as containing many halophilic species commonly found in hypersaline environments (Lijuan et al., 2017), in agreement with previous findings (Gui et al., 2017).

At the genus level, Salinibacter was notably dominant, accounting for 40.97% of the identified genera. This halophilic bacterium is known to be prevalent in hypersaline ecosystems (Viver et al., 2019). In contrast, Firmicutes (6.47%) and Cyanobacteria (1.28%) were present in lower abundances, consistent with observations from solar saltern environments (Yuan et al., 2019).

Within the archaeal fraction, the phylum Euryarchaeota was predominant. Notably, 36.38% of archaeal sequences belonged to the family Halobacteriaceae, which accounted for 39.33% of the total archaeal sequences. This is consistent with previous studies on sebkha soils in the Algerian Sahara (Quadri et al., 2016). The archaeal genera Halorubrum (4.35%), Halorhodospira (4.38%), Haloarcula (2.72%), and Halorhabdus (2.14%) were also detected in the DJS sample. Similar taxonomic compositions have been reported in Sebkha Djendli (Menasria et al., 2018), other Algerian solar salterns (Sahli et al., 2020), and in the water and sediments of two saline lakes in Algeria (Boutaiba et al., 2011).

2. Functional category hits distribution

ISSN: 2961-6611

The metabolic profiles for the DJS sample based on various databases were used for annotation of the functional genes. Our findings suggested that the COG database shown had an inverse compared with NOG, where the distribution of predicted proteins involved in metabolism registered the high value (43,68 %). This was the most predominant category, followed by cellular processes and signalling (21,08%), information storage and processing (19.18%), and the last, poorly characterized proteins which were found at 16,04% (Fig.4). In the NOG database, poorly characterized proteins were the most predominant (80,57 %), and lowest presence category was detected at 5,92% engaged in metabolism. In the KO database, similar to the results obtained using the COG database, proteins involved in metabolism (59,41 %) were the most predominant, followed by genetic and environmental information processing (21,91% and 13,25% respectively). A total of 5.44% of the annotated sequences were distributed among functional categories related to cellular processes, human diseases, and organismal systems. Functional classification using the subsystem database revealed that carbohydrate metabolism was the most dominant category, accounting for 13.50% of the sequences. This was closely followed by cluster-based subsystems, which represented 13.32% (Fig. 4). Other notable categories included amino acids and derivatives (11.04%) and protein metabolism (8.46%), while all remaining functional groups contributed less than 7% each.

Overall, the metabolic profiling indicated that sequences associated with carbohydrate metabolism were the most abundant subgroup in the DJS metagenome. This high representation may reflect a significant production of complex polysaccharides, commonly observed in hypersaline environments (Bhatt et al., 2013). Furthermore, the presence of amino acid derivatives such as glutamic acid, glutamine, and glycine likely reflects microbial strategies for osmotic adaptation through the biosynthesis of compatible solutes, a well-documented survival mechanism in halophilic microorganisms (Mesbah et al., 2014).

Within the subsystem database, protein metabolism was found to intersect with several other functional domains, including cofactor metabolism, nucleic acid metabolism (DNA and RNA), and membrane transport (Fig. 4). These functional categories were specifically annotated through the subsystem-based approach.

This analysis underscores the fact that the composition and functional potential of the microbial community are shaped by various environmental pressures such as temperature fluctuations, drought conditions, and salinity variations. All those extreme conditions are able to make bacteria become more susceptible to metabolic disorders. This leads to the creation of surprising mechanisms for adaptation and metabolisms (McGenity and Sorokin 2019).

IV. Conclusion

In conclusion, the metagenomic analysis of a single sample from the DJS hypersaline habitat proved to be both informative and practical, revealing complex microbial dynamics at multiple taxonomic and functional levels. Our findings demonstrate that even a single environmental sample can yield a substantial volume of genetic data, particularly regarding bacterial gene content.

The relative abundances of functional categories and predicted metabolic pathways, as inferred from the metagenomic gene pool, provided valuable insights into the structure and potential activities of the microbial community. Additionally, the integration of taxonomic profiling and functional annotation—achieved through the MG-RAST analysis pipeline across multiple databases enabled a comprehensive characterization of the microbial ecosystem present in the DJS metagenome.

References

[1] López-López, A., Yarza, P., Richter, M., Suárez-Suárez, A., Antón, J., Niemann, H., Rosselló-Móra, R. (2010) Extremely halophilic microbial communities in anaerobic sediments from a solar saltern. Environmental Microbiology 258–271, doi:10.1111/j.1758- 2229.2009. 00108.x

[2] Vera-Gargallo, B., Ventosa, A. (2018) Metagenomic insights into the phylogenetic and metabolic diversity of the prokaryotic community dwelling in hypersaline soils from the Odiel Saltmarshes (SW Spain). Genes 2018, 9, 152; doi:10.3390/genes9030152

[3] Ventosa, A., De la Haba, R.R., Sa'nchez-Porro, C., Thane Papke, R. (2015) Microbial diversity of hypersaline environments: a metagenomic approach. Current Opinion in Microbiology 2015, 25:80–87, http://dx.doi.org/10.1016/j.mib.2015.05.002

[4] Högfors-Rönnholma, E., Christelb, S., Engbloma, S., Dopson, M. (2018) Indirect DNA extraction method suitable for acidic soil with high clay content. MethodsX 5, 136–140. https://doi.org/10.1016/j.mex.2018.02.005.

[5] Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28 (11), 1420–1428. https://doi.org/10.1093/ bioinformatics/bts174

ISSN: 2961-6611

[6] Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., Singer, S.W. (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2 (26), 1–18. https://doi.org/10.1073/pnas.1402564111.

[7] Zhao, D., Zhang, S., Xue, Q., Chen, J., Zhou, J., Cheng, F., Li, M., Zhu, Y., Yu, H., Hu, S., Zheng, Y., Liu, S., Xiang, H. (2020) Abundant taxa and favourable pathways in the microbiome of Soda-saline lakes in Inner Mongolia. Frontiers in Microbiology .11:1740. https://doi.org/10.3389/fmicb.2020.01740.

[8] Lijuan, C., Changsheng, L., Qi, F., Yongping, W., Hang, Z., Yan, Z., Yongjiu, F., Huiya., L. (2017) Shifts in soil microbial metabolic activities and community structures along a salinity gradient of irrigation water in a typical arid region of China. Science of the Total Environment. http://dx.doi.org/10.1016/j.scitotenv.2017.04.105.

[9] Gui, H., Purahong, W., Hyde, K.D., Xu, J., Mortimer, P.E. (2017) The Arbuscular Mycorrhizal Fungus Funneliformis mosseae Alters Bacterial Communities in Subtropical Forest Soils during Litter Decomposition. Front. Microbiol. 8:1120. https://doi.org/10.3389/fmicb.2017.01120.

[10] Viver, T., Orellana, L.H., Díaz, S., Urdiain, M., Ramos-Barbero, M.D., GonzálezPastor, J.E., Oren, A., Hatt, J.K., Amann, R., Antón, J., Konstantinidis, K.T., Rosselló-Móra, R. (2019) Predominance of deterministic microbial community dynamics in salterns exposed to different light intensities. Environ. Microbiol. 21 (11), 4300–4315. https://doi.org/10.1111/1462- 2920.14790

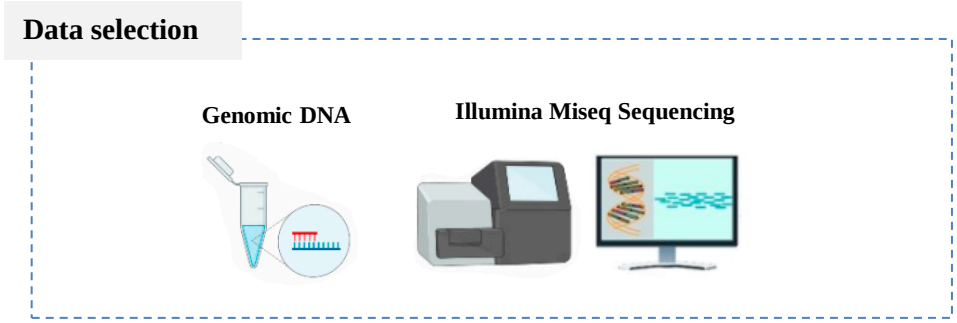
[11] Yuana, K., Chena, X., Chenb, P., Huangc, Y., Jiangd, J., Luana, T., Chena, B., Wang, X. (2019) Mercury methylation-related microbes and genes in the sediments of the Pearl River Estuary and the South China Sea. Ecotoxicology and Environmental Safety. https://doi.org/10.1016/j.ecoenv.2019.109722.

[12] Boutaiba, S., Hacene, H., Bidle, K.A., Maupin-Furlow, J.A (2011) Microbial diversity of the hypersaline sidi ameur and himalatt salt lakes of the algerian sahara. J. Arid Environ. 75 (10), 909–916. https://doi.org/10.1016/j.jaridenv.2011.04.010.

[13] Bhatt, V.D., Dande, S.S., Patil, N.V., Joshi, C.G. (2013) Molecular analysis of the bacterial microbiome in the forestomach fluid from the dromedary camel (Camelus dromedarius). Mol Biol Rep, 40:3363–3371. DOI 10.1007/s11033-012-2411-4.

[14] Mesbah, N.M., Hänelt, I., Zhao, B., Müller, V. (2014) Microbial Adaptation to Saline Environments: Lessons from the Genomes of Natranaerobius thermophilus and Halobacillus halophilus. In: R. Thane Papke and A. Oren, Halophiles Genetics and Genomes, ISBN: 978-1-908230-42-3, pp 108.

[15] McGenity, T.J., Sorokin, D.Y. (2019) Methanogens and methanogenesis in hypersaline environments. Springer Nature. Biogenesis of Hydrocarbons, Handbook of Hydrocarbon and Lipid Microbiology, https://doi.org/10.1007/978-3-319-78108-2_12



ISSN: 2961-6611

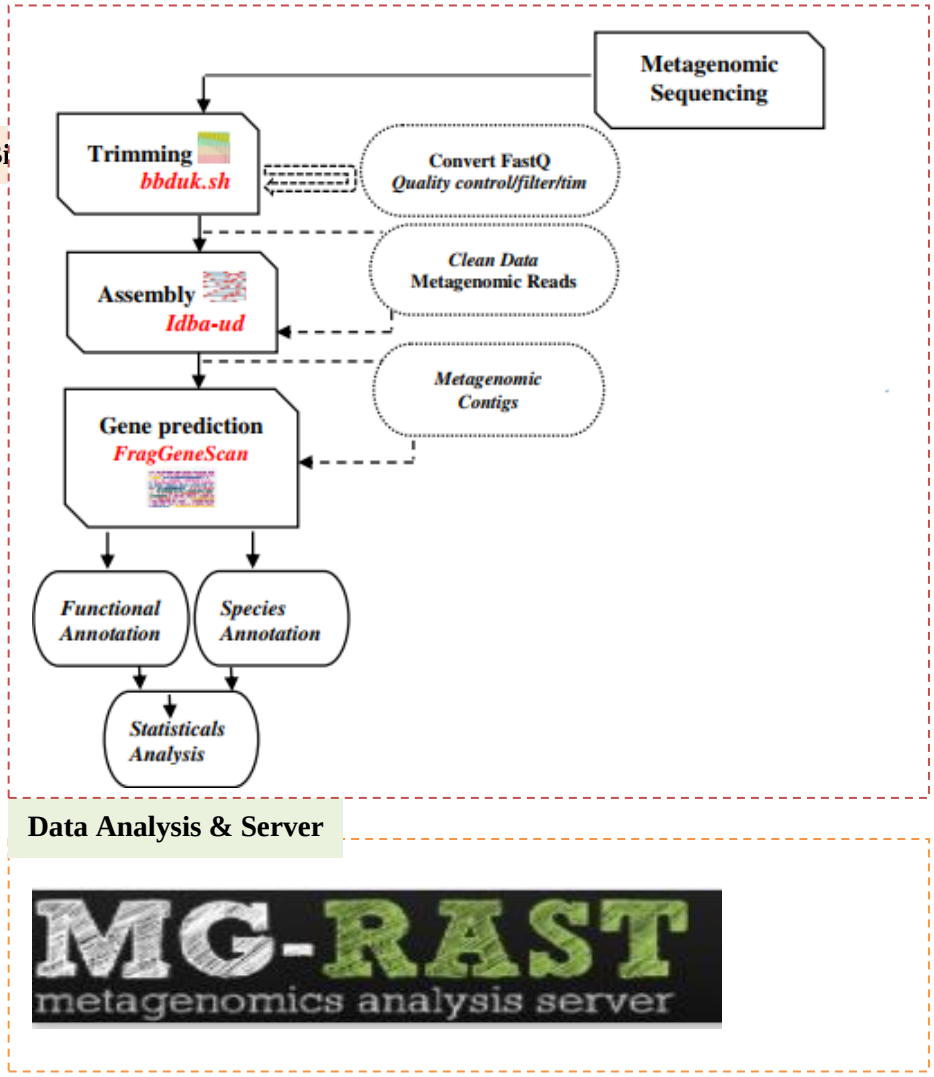


Fig.1: Workflow illustrating the analytical steps of the metagenomics study. Following sequencing, a series of bioinformatics tools were utilized, and various analyses were performed using the MG-RAST server to obtain taxonomic and functional profiles, including functional redundancy within each dataset.

Table1: Statistical Analysis of the Metagenome Using MG-RAST

ISSN: 2961-6611

Analysis Statistics

Upload: bp Count	5,773,553,587 bp
Upload: Sequences Count	39,154,118
Upload: Mean Sequence Length	147 ± 9 bp
Upload: Mean GC percent	54 ± 12 %
Artificial Duplicate Reads: Sequence Count	6,604,204
Post QC: bp Count	4,499,689,497 bp
Post QC: Sequences Count	30,077,424
Post QC: Mean Sequence Length	150 ± 4 bp
Post QC: Mean GC percent	54 ± 12 %
Processed: Predicted Protein Features	22,788,145
Processed: Predicted rRNA Features	40,571
Alignment: Identified Protein Features	7,411,103
Alignment: Identified rRNA Features	13,379

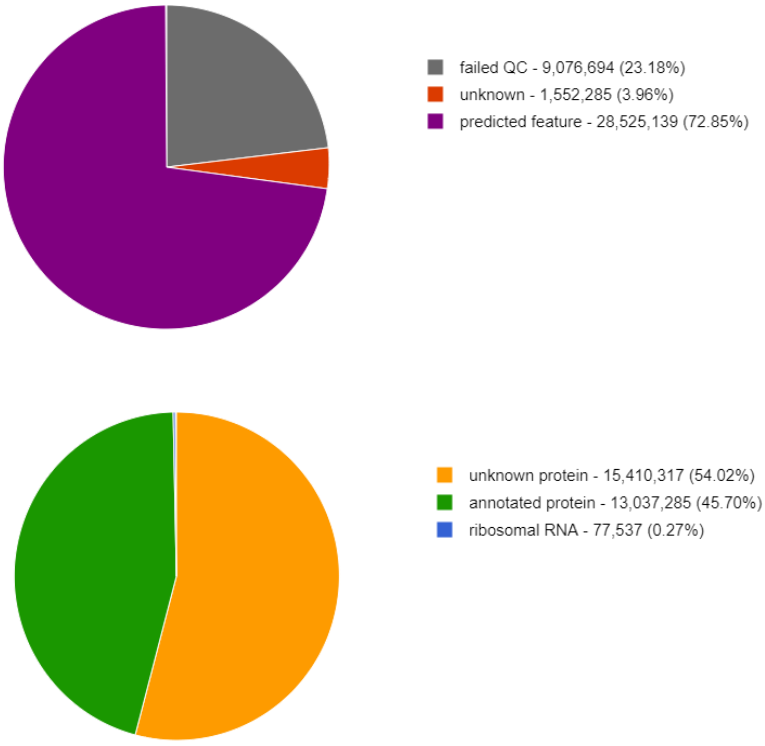


Fig.2: The charts display a sequence breakdown at the top and predicted features at the bottom

ISSN: 2961-6611

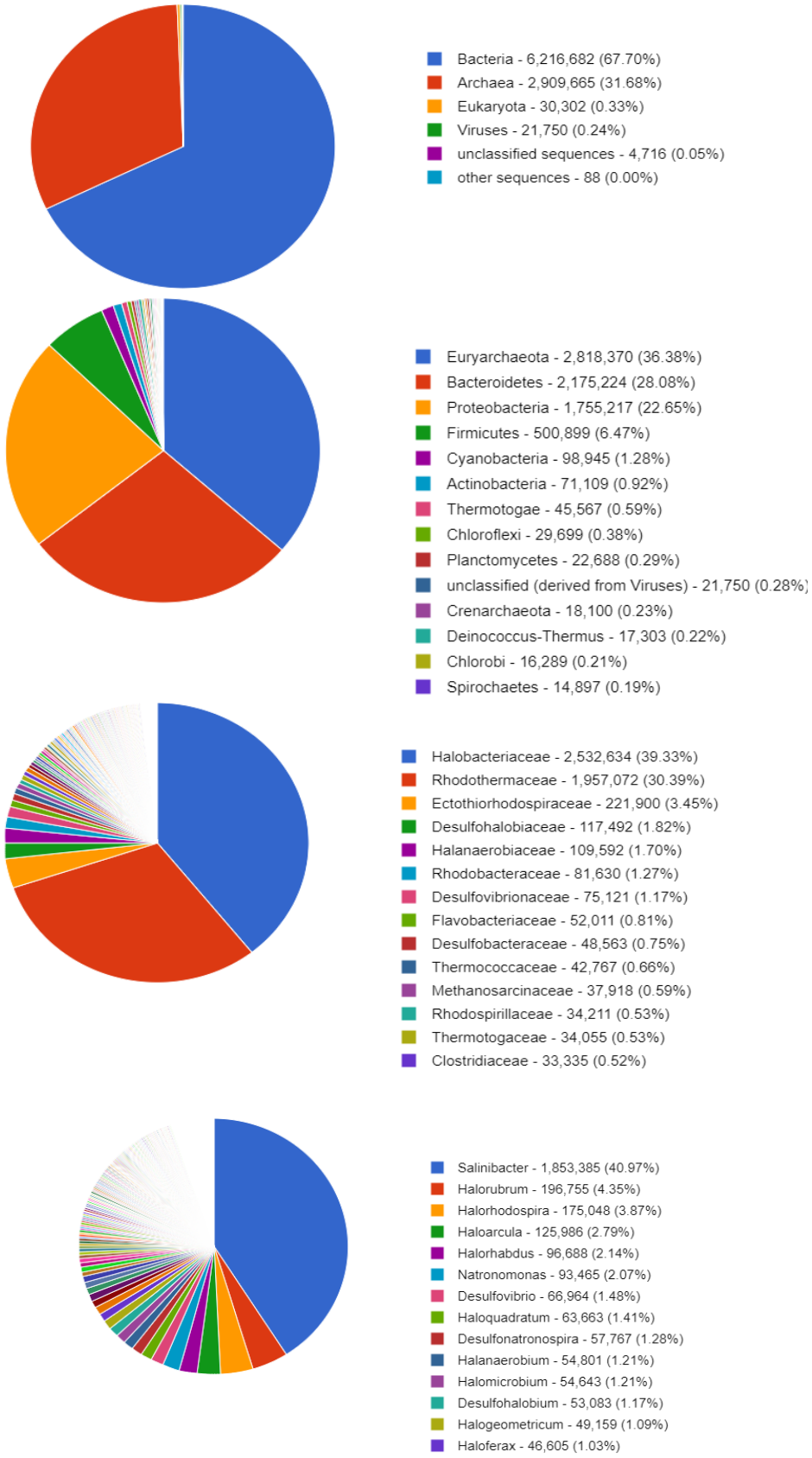
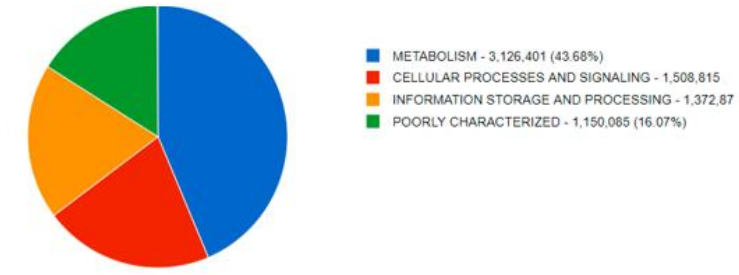


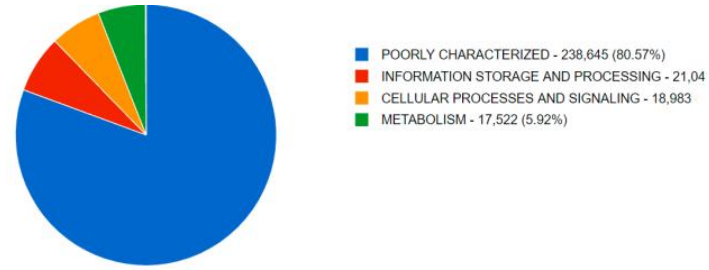
Fig. 3. Pie charts representing the taxonomic distribution of the DJS metagenome at the kingdom, phylum, family, and genus levels.

ISSN: 2961-6611

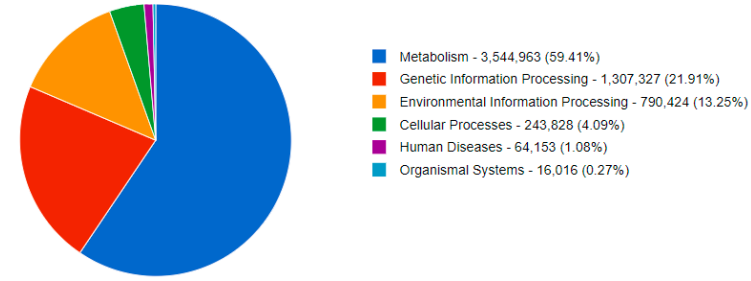
COG



NOG



KO



Subsystems

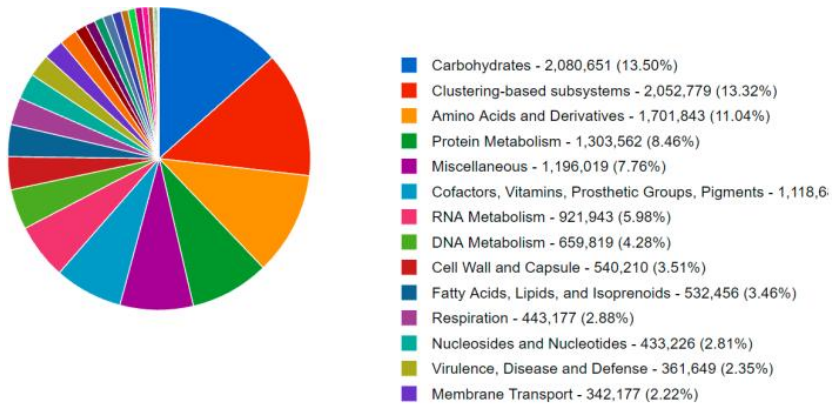


Fig. 4. Pie charts showing the distribution of various functional category hits in the DJS metagenome, based on annotations from the COG, NOG, KO, and Subsystems databases.

ISSN: 2961-6611