GMR
The Original

NGS-BASED METAGENOMIC ANALYSIS OF MICROBIAL COMMUNITIES FROM THE RED SEA: ASSEMBLY, GENE PREDICTION, AND FUNCTIONAL ANNOTATION

Salah E. M. Abo-Aba[1,2] Fakieh Moamed[1], Baoum Riham O[1]

[1*]Department of Biological Sciences; Faculty of Science; [2]Princess Dr. Najla Bint Saud Al-Saud Centre for Excellence Research in Biotechnology; Faculty of Science, King Abdulaziz University, Jeddah-21589, Saudi Arabia.

ABSTRACT:
The Red Sea is far less studied than other major oceans, despite its remarkable marine productivity. Microbiology and associated technologies have emerged as the principal areas of study in the Red Sea following extensive degradation of marine ecosystems and the overexploitation of marine resources. In this investigation, whole-genome shotgun sequencing was conducted on five samples collected from different depths in the Red Sea. Samples were taken from the Jeddah beach at 21o45'23.1 N and 39o03'38.8" E. They were then taken down to depths of 10, 15, 20, 25, and 30 meters. This study's comprehensive functional and diversity analyses provide strong evidence that group identification plays a significant role in determining the structure and function of microbial communities. CAZy analysis showed that the GT2 family was the most abundant across all samples, with 2,412 in R1, 2,511 in R2, 1,973 in R3, 2,658 in R4, and 2,390 in R5 samples. Hierarchical clustering based on Bray–Curtis distance, coupled with stacked bar charts of relative abundance for PHI-base annotations, revealed that the highest relative abundance was associated with *Pseudomonas aeruginosa*, particularly in the R3 sample. This was followed by *Salmonella enterica* at PHI level 2 and the virulence-associated gene RhIE1 (PA3950) at PHI level 3. *P. aeruginosa* is commonly detected in coastal marine environments and was shown to release OprP, as supported by immunochemical analysis of saltwater samples. The only bacterium consistently detected across all samples was *Vibrio ponticus*. Notably, the R3 sample was distinct, exhibiting the highest abundance of *Vibrio chagasii* and *Vibrio* sp. AND4. Overall, our results indicate that all samples originated from a *Vibrio*-rich environment, consistent with aquatic habitats where *Vibrio* species are known to form symbiotic or pathogenic associations with their hosts.

KEYWORDS: Metagenomics, Red sea, CAZy analysis, *Pseudomonas aeruginosa*, *Vibrio chagasii.*

1. INTRODUCTION

Metagenomics, the culture-independent analysis of microbial communities via high-throughput sequencing, has emerged as an innovative approach for uncovering the structure, function, and diversity of microbes in complex ecosystems. Metagenomic techniques allow researchers to overcome the limitations of traditional cultivation methods, thereby enabling the study of most microorganisms that cannot be cultured in laboratory settings. These communities play a critical role in global biogeochemical cycles, human health, and industrial biotechnology, highlighting the importance of examining their genetic and functional capabilities (Young Yusty & Prescilla-Ledezma, 2025).

Next-generation sequencing (NGS) technology has enabled the generation of billions of short reads at lower cost, significantly changing how we research microbial ecosystems. This high-throughput sequencing enables extensive profiling of microbial consortia from diverse settings, providing significant insights into their taxonomic composition, metabolic functions, and ecological relationships. The rapid growth of NGS has enabled faster, more accurate, and cheaper data production. DNA extraction, library preparation, sequencing, quality control, de novo assembly, gene prediction, and functional annotation are all crucial processes in a typical metagenomic workflow. Together, these activities help us learn more about complex microbial communities (Katara et al., 2024). Functional annotation using databases like KEGG, eggNOG, and CAZy reveals the metabolic potential of microbial communities. CAZymes are crucial for the metabolism of polysaccharides, linking microbes to nutrient cycles and ecological functions. This project aims to compile a comprehensive gene catalog and functional landscape of ambient microbial communities through NGS assembly, predictive genomics, and CAZy-focused profiling, facilitating insights into microbial interactions, ecological modeling, and biotechnological applications (Cuartero et al., 2025).

2. MATERIALS AND METHODS

2.1 Samples Collection

The study utilized water samples collected from various depths of the Red Sea, a natural ecosystem. Five water samples were collected for the study, conducted under sterile conditions in accordance with established protocols.

Samples were collected from the coast of Jeddah at the coordinates 21o45’23.1 N and 39o03’38.8 E, extending to the following depths: 10, 15, 20, 25, and 30 meters.

2.2 DNA Extraction

DNA extraction was performed using two methods: a manual approach and the PureLink Micobiome DNA Purification Kit (Thermo Fisher Scientific). (Park, Lim, Lee, & Kwon, 2025) used a manual technique to isolate complete genomic DNA. The five samples were centrifuged twice, and then the TES lysis buffer was added to the tubes. Vortex for 5 minutes, then add the lysosomes and incubate at 56°C for 30 minutes. Each sample was treated with 10 µl of proteinase K (10 mg/ml) and incubated in a water bath at 37°C for 30 minutes. The samples were centrifuged at 100,000 rpm for 5 minutes to extract the supernatant. In a fresh tube, add 250 µl of 4 M sodium acetate, then 250 µl of chloroform: isoamyl (1:1), and centrifuge for 5 minutes at 100000 rpm. Three layers were obtained, and the upper zone was carefully transferred to a new, clean Eppendorf tube, where 3/4 or 1 v/v isopropanol was injected, and the tube was stored at -20 °C overnight. The solution was centrifuged at 13000 rpm for 2 min; the liquid layer was removed, and the DNA was dried at room temperature before resuspension in 50 µl of distilled water. Isolated DNA (10 µl) was put into a 0.5% agarose gel with 1x TBE buffer at 100 V for 60-90 minutes and stained with ethidium bromide. The DNA extraction kit was used to extract DNA according to the manufacturer's protocol, with some modifications. Proteinase K was used instead of the supplied S2 Lysis Enhancer, and the samples were incubated at 37 °C for 15 minutes rather than 56 °C, which is optimal for high-yield recovery of microbial genomic DNA. Nanodrop spectroscopy and Qubit fluorescence were used to analyze the quality and concentration of the extracted DNA. The samples were then sent to Novogene for bioinformatics analysis.

2.3 Bioinformatic Analysis

High-quality DNA was sheared and utilized for Illumina sequencing library preparation according to the TruSeq protocol. The libraries were sequenced on an Illumina NovaSeq 6000 platform, yielding paired-end reads of 150 bp, producing an average of 6.8–8.3 Gb of raw sequence data, in line with recommendations for metagenomic studies. For gene prediction and unigene clustering, predicted coding sequences were organized into non-redundant unigenes using CD-HIT v4.8.1, applying a 95% identity threshold and 90% coverage, with details on sequence length, representative ID, and cluster size documented for each unigene (Liu et al., 2025). Unigenes were functionally annotated against multiple databases. CAZy annotations were assigned using HMMER v3.3. KEGG Orthology (KO) assignments were performed using KofamKOALA. Taxonomic annotation was obtained by matching unigenes to the NCBI non-redundant protein database (NR) using DIAMOND v2.0.11 (Thurimella et al., 2023).

2.4 Abundance Estimation and Statistical Analysis

To estimate gene abundances, clean reads were mapped back to the unigene catalog using Bowtie2 v2.4.5. The count matrices have been adjusted to show relative abundances. The pheatmap package in R was used to make heatmaps. To account for differences in gene counts between groups, a z-score transformation was used. Pearson's correlation was used to calculate distances between pairs, and complete linkage was used for hierarchical clustering. DESeq2 was used for differential abundance analysis (Wambreuse et al., 2025).

2.5 Beta- diversity and ordination analysis

Beta-diversity analyses had been performed on annotated unigene profiles to investigate functional differences at the Level-2 resolution. The Bray-Curtis distance metric was used to calculate pairwise dissimilarities among samples, accounting for differences in the relative abundance of functional categories. The resulting distance matrix was then visualized using Principal Coordinates Analysis (PCoA). The vegan package in R was used for the ordination, and the results were verified using QIIME2 workflows. The first two principal coordinates (PC1 and PC2) accounted for most of the observed variation in the dataset (Yu et al., 2025).

3. RESULTS

3.1 Metagenomic Data Quality and Assembly Statistics

On average, each sample produced between 6.8 and 8.3 Gb of clean bases, with effective retention rates exceeding 99.6%. Quality scores were consistently excellent, with Q20 values around 98% and Q30 values near 94%, surpassing the commonly accepted thresholds for next-generation sequencing quality. The assembly statistics for samples R1–R5 are summarized in **Table 1**. The total assembled lengths ranged from 145 Mb (R3) to 221 Mb (R4), and the scaffolding numbers ranged from 96,282 to 168,521. Average scaffold lengths ranged from 1.2 to 1.5 kb, comparable to values reported in previous large-scale metagenomic projects (Consortium, 2012a; Li et al., 2015). The N50 values (1.3–2.1 kb) and maximum scaffolding lengths (315–954 kb) indicate assemblies of sufficient continuity for downstream taxonomic and functional analyses. R3 exhibited the highest contiguity (N50 = 2,111 bp), though with a smaller total assembly size, while R4 showed the most significant total assembly length (221 Mb) but only moderate N50 (1,585 bp).

Table 1. Assembly statistics of metagenomic samples R1–R5. Metrics include total assembled length (bp), number of scaffolds, average scaffold length (bp), and assembly continuity measures (N50, N90, and maximum scaffold length).

SampleID	Total len (bp)	Scaftigs num	Average len (bp)	N50 (bp)	N90 (bp)	Max len (bp)
R1	200,863,501	153,450	1,308.98	1,569	586	954,585
R2	210,278,082	162,506	1,293.97	1,526	587	954,585
R3	145,534,990	96,282	1,511.55	2,111	612	315,592
R4	221,414,443	168,521	1,313.87	1,585	590	401,891
R5	175,325,274	143,054	1,225.59	1,342	577	616,480

*Assembly statistics across samples R1–R5 indicate high-quality metagenomic assemblies. Total assembly lengths ranged from 145 Mb (R3) to 221 Mb (R4), with scaffolding counts of 96,282-168,521. N50 values (1.3–2.1 kb) and maximum scaffolding lengths (316–954 kb) demonstrate sufficient continuity for downstream analyses. The consistency across samples suggests uniform sequencing quality and assembly performance, making it suitable for comparative metagenomic studies.

3.2 Prediction of coding sequences and gene length analysis

Open reading frames (ORFs) were predicted from assembled contigs in each sample (R1–R5). Across all five metagenome assemblies (R1–R5), ORF length distributions were highly consistent, with a dominant mode centered at ~550–650 nt (approximately 180–220 amino acids) and a characteristic right-skewed tail toward longer genes **(Figure 1)**. This size profile matches expectations for prokaryotic coding repertoires, in which small- to medium-sized enzymes and membrane proteins are prevalent, while large, multi-domain proteins constitute a smaller fraction.

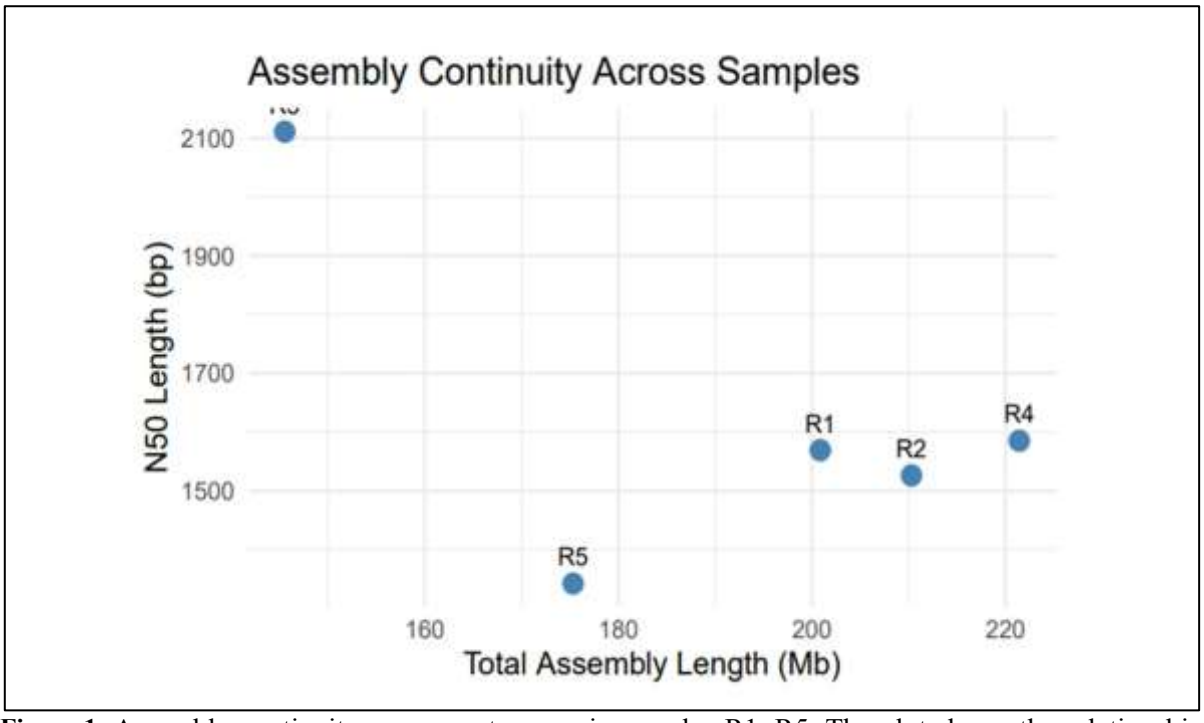


Figure 1. Assembly continuity across metagenomic samples R1–R5. The plot shows the relationship between N50 length (y-axis) and total assembly length (x-axis). Samples clustered within a similar range, with R3 displaying the highest N50 (2,111 bp) despite the lowest total assembly length, while R4 had the largest total assembly length (221 Mb) but a moderate N50.

2.3 Unigenes clustering and statistical characterization

Across all assembled samples, 490,762 unigenes were predicted. Unigene length distributions were strongly right-skewed, with a median size of 184 amino acids and a modal peak below 250 amino acids, while a small fraction of long proteins extended beyond 3,000 amino acids **(Figure 4, left)**. Cluster size analysis indicated that 63% of unigenes were singletons (Num = 1), whereas 37% belonged to multi-sequence clusters. Although most clusters were small, several large gene families grouped more than 100 redundant proteins

(Figure 4, right). This pattern reflects the combined effects of high microbial diversity and the presence of conserved or expanded gene families in complex metagenomes.

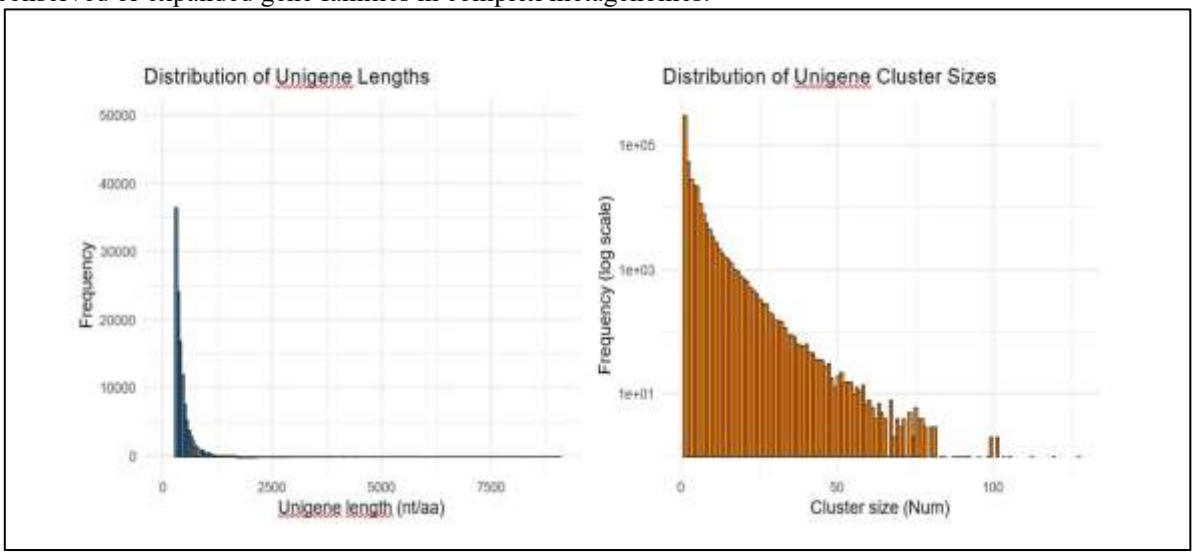
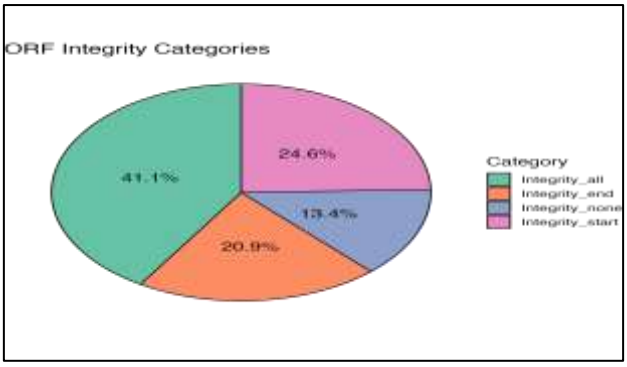


Figure 2. *Characteristics of predicted unigenes across the metagenome. The left panel shows the distribution of unigene lengths (nt/aa). Most unigenes are short, with a modal size below 250 aa and a right-skewed tail extending up to ~9,000 aa. The right panel shows the distribution of unigene cluster sizes (Num), displayed on a log-scaled y-axis. The majority of clusters are singletons (Num = 1), whereas a minority of gene families include dozens to more than one hundred redundant proteins. Together, these distributions highlight the predominance of unique protein-coding sequences alongside a subset of highly conserved or expanded gene families typical of complex microbial communities.*

3.4 Distribution of Raw Counts Across Replicates

The boxplot analysis of raw measurements across replicates (R1–R5) revealed substantial variation in the distribution of the data **(Figure 3)**. While most replicates showed values in the 103–104 range, one replicate (R1) contained an extreme outlier exceeding 108, suggesting a strong imbalance in sample counts. Such variation may arise from technical bias (e.g., sequencing depth or library preparation inconsistencies) or reflect underlying biological differences. Normalization of data before downstream statistical analysis will be essential to mitigate these effects and ensure accurate interpretation of replicate comparisons.



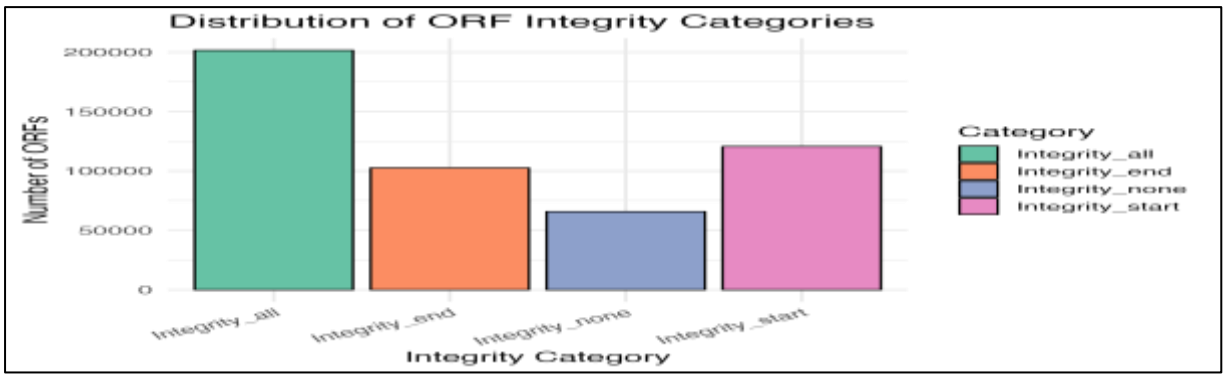


Figure 3. *Integrity of predicted ORFs across the metagenome. ORFs were classified based on the completeness of start and stop codons. (Left) Pie chart showing proportions of truncated and complete ORFs. (Right)*

Bar chart showing absolute counts for each integrity class. Approximately 41% of predicted ORFs were complete with both start and stop codons, while the remainder were partially truncated at either the start (25%), the end (21%), or at both ends (13%).

3.5 Functional/Taxonomic Annotation by CAZy (Carbohydrate-Active Enzymes)

To assess the genetic potential for carbohydrate metabolism, we quantified the number of annotated CAZy genes at two hierarchical levels. At the class level **(Figure 4A)**, all samples (R1–R5) exhibited substantial gene repertoires dominated by GH and GT categories, followed by CBM, while CE, AA, and PL were comparatively less abundant. This overview illustrates the broad functional architecture of carbohydrate-active enzymes in the dataset. At the family level **(Figure 4B)**, the analysis provided finer resolution, revealing that families such as GT2, GT4, GH13, and CBM50 were particularly enriched, while other families contributed smaller but consistent numbers across replicates.

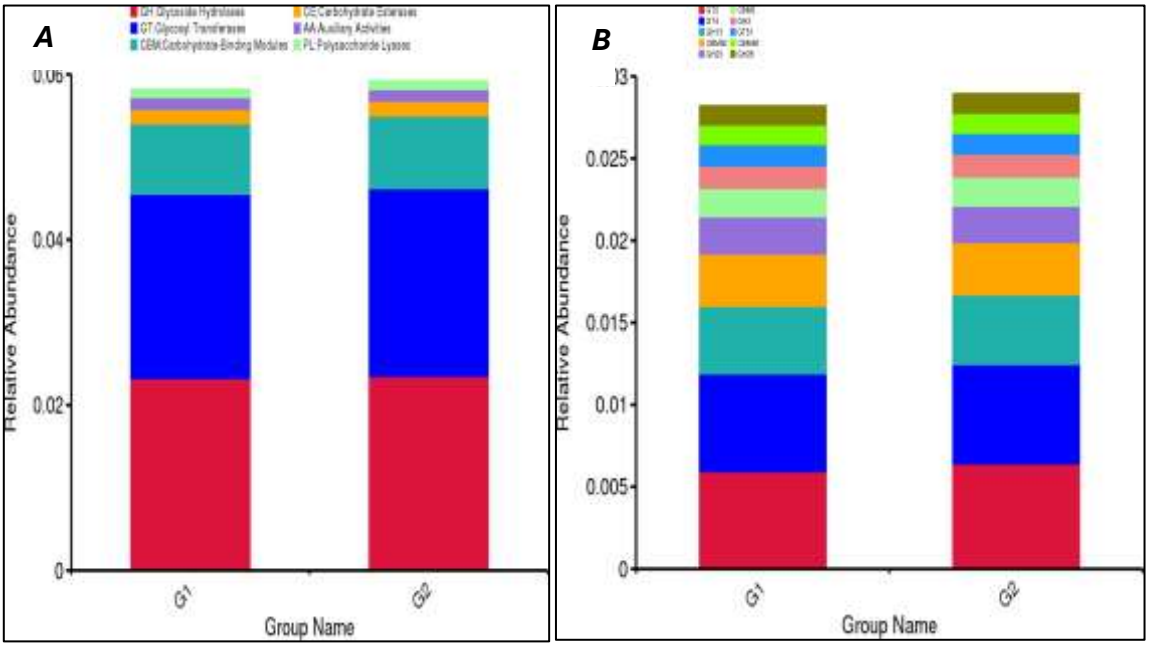


Figure 4. *Relative abundance of the top ten carbohydrate-active enzyme (CAZy) class (A) and families(B) across replicates (R1–R5). The stacked bar plot shows the proportional contributions of the following families: GT2, GT4, GH13, CBM50, GH23, GH3, GT51, CBM48, GH28, and CBM5. Among these, GT2, GT4, GH13, and CBM50 were the most dominant families, while others contributed at lower but consistent levels across replicates.*

3.6 Assessment of CAZy Unigene Distribution Across Functional Classes

The unigene-level annotation against the CAZy database identified a substantial number of sequences associated with carbohydrate-active enzymes **(Figure 5)**. Glycoside Hydrolases (GH) and Glycosyl Transferases (GT) represented the dominant categories, reflecting the substantial metabolic investment of the community in the breakdown and synthesis of complex carbohydrates. Carbohydrate-Binding Modules (CBM) were also well represented, indicating a complementary role in substrate recognition and enzyme–substrate interactions. In contrast, Carbohydrate Esterases (CE), Auxiliary Activities (AA), and Polysaccharide Lyases (PL) were less abundant, suggesting more specialized or secondary functions within the overall enzymatic repertoire. This class-level overview provides a global assessment of the functional potential of the microbial community and establishes a foundation for subsequent, higher-resolution analyses at the family level. The enrichment of GH and GT categories is consistent with a microbiome specialized for carbohydrate metabolism, while the presence of CBMs highlights structural adaptations to facilitate polysaccharide degradation.

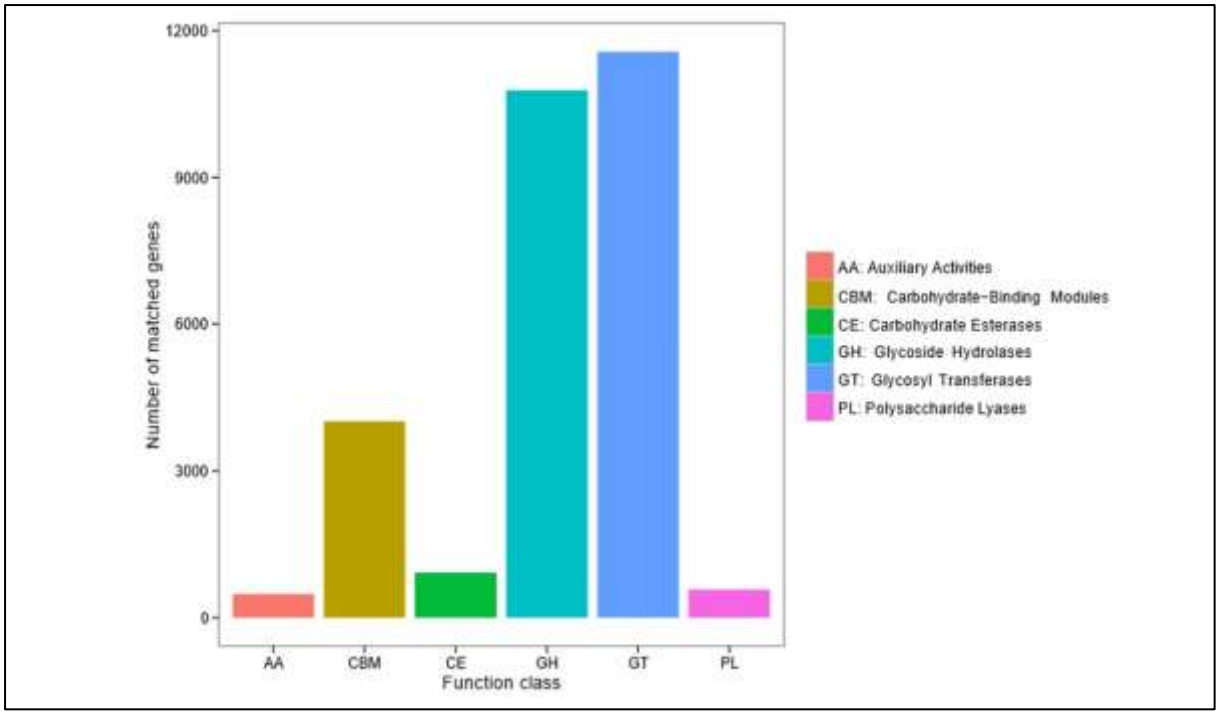


Figure 5*. Number of unigenes annotated to major CAZy functional classes. Gene counts are shown for six categories: GH (Glycoside Hydrolases), GT (Glycosyl Transferases), CBM (Carbohydrate-Binding Modules), CE (Carbohydrate Esterases), AA (Auxiliary Activities), and PL (Polysaccharide Lyases). GH and GT dominate the unigene repertoire, followed by CBM, whereas CE, AA, and PL are comparatively less abundant.*

3.7 Functional Annotation of Unigenes Based on the eggNOG Database.

Functional annotation of unigenes using the eggNOG database revealed significant differences in gene distribution between groups **(Figure 6)**. At the orthologous group (OG) level (panel a), several conserved clusters, such as COG0583, COG0840, and COG2199, were consistently abundant, indicating their core roles in microbial community function. Level 2 categorization (panel b) provided more profound insights into functional specialization, highlighting transcription regulators, DNA-binding transcription factors, signal transduction proteins, and transport systems (e.g., ABC transporters), suggesting differential regulation and adaptive capacity between groups. At the broader Level 1 classification (panel c), the majority of unigenes were assigned to key metabolic and cellular processes, including amino acid and carbohydrate metabolism, energy production, cell wall biogenesis, and translation. The consistent clustering of replicates within groups supports the robustness of these observations. Collectively, these results suggest that while both groups share a core set of conserved functions, G1 and G2 differ in functional emphasis, particularly in transcriptional regulation and metabolic processes, which may
underlie the distinct community structures observed in diversity analyses.

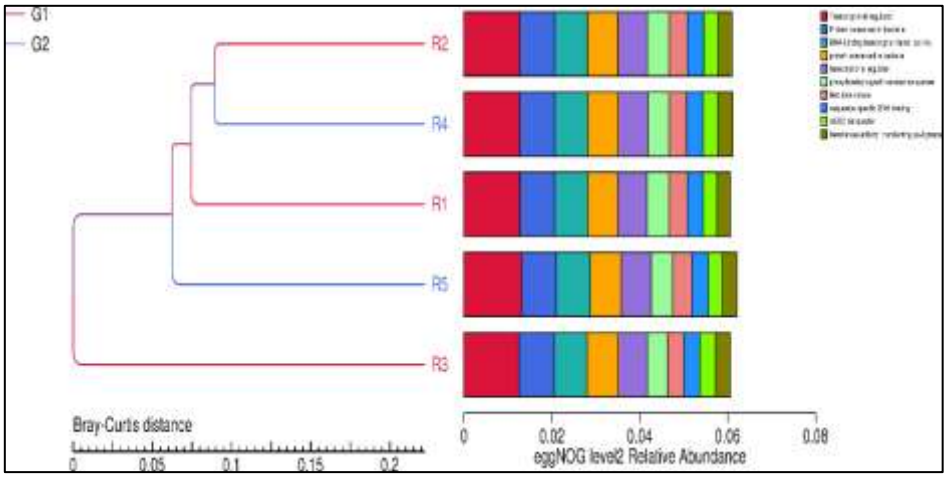


Figure 6. *The relative abundance of the eggNOG level 2 of G1 and G2 groups.*

3. DISCUSSION

The Red Sea is an ideal location for metagenomic research due to its exceptional environmental conditions and rich microbiological diversity. Microbiology and related technologies have become the primary focus of research in the Red Sea, following the widespread devastation of marine ecosystems and the overexploitation of marine resources (Altalhi et al., 2025). Meanwhile, metagenomic analysis is now possible at lower assembly depths than ever before, thanks to the development of hybrid and third-generation sequencing technologies (Garg, Patel, Rawat, & Rosado, 2024). In this investigation, whole-genome shotgun sequencing was conducted on five samples collected from different depths in the Red Sea. The study's materials consisted of water samples collected from the Red Sea, a natural environment, at various depths. For the investigation, five water samples were collected using standard procedures and sterile conditions. This study's comprehensive functional and diversity studies offer strong evidence that group identification plays a significant role in determining the structure and function of microbial communities. Specifically, functional profiles of virulence-associated genes and carbohydrate-active enzymes showed strong segregation of all samples, highlighting the biological and ecological uniqueness of these communities. Starting with CAZy analysis, the GT2 family was the most abundant in all samples, with 2,412 in R1, 2,511 in R2, 1,973 in R3, 2,658 in R4, and 2,390 in R5. Furthermore, lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56) exhibited the most relative abundance in all samples through level 2 analysis. An N-acetylglucosamine residue is transferred to a D-galactose residue in the lipopolysaccharide core of bacteria by the enzyme lipopolysaccharide N-acetylglucosaminyltransferase. It utilizes UDP-\(N\)-acetyl-α-D-glucosamine as a substrate and is a glycosyltransferase, facilitating the synthesis of complex sugars. The second enzyme in relative abundance in the CAZy analysis in the R1 sample was UDP-GlcNAc: alpha-N-acetylglucosaminyltransferase (EC 2.4.1.-), and in R2, R3, R4, and R5 it was alternating beta-1,3_4-N-acetylmannan synthase (2.4.1.-)_ UD(...). In addition, the KEGG UniGene annotation across primary functional categories analysis showed that the most relatively abundant categories in all samples were: cellular community in prokaryotes, signal transduction, replication and repair, drug resistance as antimicrobial, carbohydrate metabolism, and the endocrine system. Likewise, the eggNOG functional profiles at level 2 showed that the most abundant pathways across all annotations were transcriptional regulation, followed by protein conservation in bacteria and DNA-binding transcription factor activity. The hierarchical clustering dendrograms (Bray–Curtis distance) with stacked bar relative abundance profiles for PHI-base annotation showed that the most relative abundances were *Pseudomonas aeruginosa* (especially the R3 sample), followed by *Salmonella enterica* at level 2, and RhIE1 (PA3950) at level 3 of PHI. This finding was confirmed by Kimata, Nishino, Suzuki, & Kogure, (2004), who found that *P. aeruginosa* is frequently found in coastal marine habitats and sheds OprP, as indicated by immunochemical examinations of the saltwater data. In 1995, the outer membrane protein OprP porin of this bacterium was shown to be extensively disseminated as a dissolved component in saltwater. Three approaches were used to identify bacterial cells isolated by selective medium: the presence of two outer membrane lipoprotein genes unique to *P. aeruginosa*, oprI and oprL; the API20 NE kit; and 16S rDNA sequence analysis. These findings verified that *P. aeruginosa* accounted for the bulk of isolates from the bay. Finally, the unique bacterium in all samples was *Vibrio ponticus*. The Gram-negative, somewhat halophilic marine bacterium *Vibrio ponticus* needs Na+ ions for survival. On Spain's Mediterranean coast, it was initially separated from saltwater, mussels, and sick sea bream (Sparus aurata). Phylogenetic investigation supports the uniqueness of *Vibrio ponticus* within the genus Vibrio by placing it near the *Vibrio fluvialis–Vibrio furnissii* clade, with 16S rDNA sequence similarities of 97.1% and 97.3%, respectively (Yibar et al., 2025). The R3 sample was unique, containing the highest concentrations of *Vibrio sp. AND4* and *Vibrio chagasii*. It is well known that Vibrio strains are commonly found in aquatic environments and can establish symbiotic or pathogenic relationships with their hosts. The aerobe, mesophilic bacterium *Vibrio chagasii* R-3712, was discovered in the gut of *turbot larvae* (Urtubia et al., 2023). Furthermore, the KEGG top hit was K03406 (methyl-accepting chemotaxis protein), which is responsible for two pathways: the two-component system and bacterial chemotaxis. This was followed by K07497 (putative transposase), which had the most relative abundance in the R3 sample. It was found that large-scale genomic rearrangements and horizontal gene transfer (HGT), which are essential to the bacterium's evolution, adaptability, and pathogenicity, are mostly linked to suspected transposases (Schwarzerova et al., 2024).

REFERENCES

4. Altalhi, S., Schultz, J., Jamil, T., Diercks, I., Sharma, S., Follmann, J., . . . Rosado, A. S. (2025). Decoding microbial diversity, biogeochemical functions, and interaction potentials in red sea hydrothermal vents. *Environmental Microbiome, 20*(1), 118. doi:10.1186/s40793-025-00784-5
5. Cuartero, J., Perez-Mon, C., Qi, W., Stierli, B., Frey, B., & Varliero, G. (2025). Increased carbon inputs alter soil microbial genetic potential for biogeochemical cycling in Arctic ecosystems. *Communications Earth & Environment, 6*(1), 807. doi:10.1038/s43247-025-02768-2
6. Garg, D., Patel, N., Rawat, A., & Rosado, A. S. (2024). Cutting edge tools in the field of soil microbiology. *Current Research in Microbial Sciences, 6*, 100226. doi:https://doi.org/10.1016/j.crmicr.2024.100226
7. Katara, A., Chand, S., Chaudhary, H., Chaudhry, V., Chandra, H., & Dubey, R. C. (2024). Evolution and applications of Next Generation Sequencing and its intricate relations with chromatographic and spectrometric

techniques in modern day sciences. *Journal of Chromatography Open, 5*, 100121. doi:https://doi.org/10.1016/j.jcoa.2024.100121
8. Kimata, N., Nishino, T., Suzuki, S., & Kogure, K. (2004). Pseudomonas aeruginosa isolated from marine environments in Tokyo Bay. *Microb Ecol, 47*(1), 41-47. doi:10.1007/s00248-003-1032-9
9. Liu, Y., Hrit, J. A., Chomiak, A. A., Stransky, S., Hoffman, J. R., Tiedemann, R. L., . . . Rothbart, S. B. (2025). DNA hypomethylation promotes UHRF1-and SUV39H1/H2-dependent crosstalk between H3K18ub and H3K9me3 to reinforce heterochromatin states. *Molecular Cell, 85*(2), 394-412.e312. doi:https://doi.org/10.1016/j.molcel.2024.11.009
10. Park, Y.-J., Lim, J. K., Lee, Y.-J., & Kwon, K. K. (2025). Protocol for efficient recovery of high-quality DNA from microbiome of marine invertebrates. *Journal of Microbiology, 63*(9), e2507003. doi:10.71150/jm.2507003
11. Schwarzerova, J., Zeman, M., Babak, V., Jureckova, K., Nykrynova, M., Varga, M., . . . Cejkova, D. (2024). Detecting horizontal gene transfer among microbiota: an innovative pipeline for identifying co-shared genes within the mobilome through advanced comparative analysis. *Microbiol Spectr, 12*(1), e0196423. doi:10.1128/spectrum.01964-23
12. Thurimella, K., Mohamed, A. M. T., Graham, D. B., Owens, R. M., La Rosa, S. L., Plichta, D. R., . . . Xavier, R. J. (2023). Protein Language Models Uncover Carbohydrate-Active Enzyme Function in Metagenomics. *bioRxiv*. doi:10.1101/2023.10.23.563620
13. Urtubia, R., Miranda, C. D., Rodríguez, S., Dubert, J., Barja, J. L., & Rojas, R. (2023). First Report, Characterization and Pathogenicity of Vibrio chagasii Isolated from Diseased Reared Larvae of Chilean Scallop, Argopecten purpuratus (Lamarck, 1819). *Pathogens, 12*(2), 183.
14. Wambreuse, N., Bossiroy, E., David, F., Vanwinge, C., Fievez, L., Bureau, F., . . . Delroisse, J. (2025). Carotenoid-based immune response in sea cucumbers relies on newly identified coelomocytes-the carotenocytes. *Front Immunol, 16*, 1668167. doi:10.3389/fimmu.2025.1668167
15. Yibar, A., Duman, M., Ay, H., Ajmi, N., Tasci, G., Gurler, F., . . . Saticioglu, I. B. (2025). Genomic Insight into Vibrio Isolates from Fresh Raw Mussels and Ready-to-Eat Stuffed Mussels. *Pathogens, 14*(1), 52.
16. Young Yusty, S., & Prescilla-Ledezma, A. (2025). From Culture to Metagenomics: How Methodological Advances Reshape Our Understanding of the Oral Microbiota of Venomous Snakes. *Microbiology Research, 16*(11), 233.
17. Yu, H., Li, J., Wang, Y., Zhang, T., Mehmood, T., & Habimana, O. (2025). Dysbiosis and genomic plasticity in the oily scalp microbiome: a multi-omics analysis of dandruff pathogenesis. *Front Microbiol, 16*, 1595030. doi:10.3389/fmicb.2025.1595030