Analyzing metagenomic datasets from extreme environments to uncover biotechnologically valuable biomolecules

Ian Alves Machado[1], Sara Silvério[2], Ricardo Franco-Duarte[3], and Cátia Santos-Pereira[2]

[1] Informatics department, University of Minho, 4710-057 Braga, Portugal
[2] Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal
[3] Centre of Molecular and Environmental Biology, University of Minho, 4710-057 Braga, Portugal

Abstract. Extreme environments, such as hypersaline environments, are home to a huge number of microorganisms adapted to challenging conditions. This study aims to explore the biotechnological potential of these microorganisms. For this purpose, metagenomic datasets obtained from samples collected in hypersaline locations were analyzed by different bioinformatics software aiming to reconstruct metagenome-assembled genomes (MAGs), that were then explored in terms of taxonomic and functional annotation aiming to identify biomolecules of biotechnological interest, such as biosurfactants and enzymes.

Keywords: Hypersaline environments, extremophiles, metagenomics, metagenome-assembled genomes.

1 Motivation and objectives

The planet has a wide biodiversity of ecosystems in which various microorganisms with different characteristics live. These microorganisms have the capacity to produce a series of compounds of biotechnological interest, particularly those inhabiting extreme environments. Due to the difficulty of culturing a large number of these microorganisms in the laboratory, sequencing methodologies coupled with bioinformatic tools have been developed to predict and reconstruct the genome of these microorganisms, and thus identify potential compounds such as enzymes, biosurfactants, antimicrobials, anticancer agents, among others, applicable in various areas, such as the pharmaceutical, environmental and food industries.

Currently, many of these habitats remain unexplored, so research into these environments promises scientific opportunities that can contribute to the creation of new technologies and therapies based on the potential biomolecules identified. It is therefore essential to explore tools and databases capable of providing a deep understanding of the sequencing data obtained from these environments.

In this work, we searched for tools and databases that allow us to study and analyze biomolecules from microorganisms inhabiting hypersaline environments.

2 State of the Art

2.1 Extreme environments and extremophiles

Extreme environments are classified as environments exposed to one or more extreme environmental parameters such as temperature, osmolarity, pH, salinity, UV radiation, pressure, that exhibit values close to the known limit of life [1]. These environments are home to a variety of microorganisms called extremophiles. They are able to thrive in conditions of high (thermophiles) or low (psychrophiles) temperatures, acidic (acidophiles) or alkaline (alkalophiles) pH, high pressures (piezophiles), anaerobic environments (anaerobes), high salinities (halophiles), among others [2]. Due to their unique adaptation characteristics, they produce a series of secondary metabolites that have a high biotechnological potential [3], including for example biofertilizers [4–7], enzymes able to degrade lignocellulosic biomass [8], antibiotics, antitumor molecules and antifungals [4, 9], and biosurfactants [10–12].

2.2 Hypersaline environments and halophiles

Saline environments are characterized by having salt concentrations at sea level (approximately 3-5% w/v of dissolved salts), while hypersaline environments have values greater than 10% [13]. They occupy approximately 44% of the planet's inland waters and provide important resources for humans and nature. According to studies, salinity is one of the main drivers of biodiversity in aquatic environments. Despite this, these ecosystems are often overlooked by most researchers. Thus, even today, little is known about these systems [14].

The microbial communities thriving in hypersaline environments are mainly dominated by microorganisms called halophiles, which have developed multiple strategies that allow their survival under high salt concentrations. These adaptations include the ability to prevent inorganic salt entry and cytoplasmic water loss, as well as to synthetize organic osmolytes [15]. Halophiles also display proteins with more salt bridges than proteins found in normal conditions, that have more acidic and less hydrophobic residues [4]. *Halobacterium salinarium* and *Natronomonas pharaonis*, for example, possess rhodopsin proteins that act as photoreceptors when the chromophore is isomerized by light, regulating ion concentration within cells [16]. *Wallemia ichthyophaga*, has shown a 3-fold thickening of the cell wall and a 4-fold increase in the size of multicellular clusters. In addition, sequencing analysis shows an increase in genes responsible for encoding proteins capable of altering the flow of solutes and causing an increase in cell wall resistance [17]. Another example is the species *Salinibacter ruber*, which belongs to the phylum Bacteroidetes. These are extreme halophilic microorganisms that use the salt-in strategy, accumulating potassium in the intracellular medium as a compatible solute [18].

2.3 The biotechnological potential of halophiles

Halophiles have been reported to produce biomolecules with applicability in different industries. Halophilic enzymes found at the bottom of the ocean have a high biotechnological potential, such as in the treatment of wastewater containing a high degree of salinity and starch. The esterases associated with deep sea environments are characterized by being lipolytic, making them suitable for the production of biodiesel, food and polyunsaturated fatty acids [19]. The ability of three microbial strains isolated from a salt lake in Iran to resist high concentrations of heavy metals was reported [20]. Halophilic microorganisms have been identified in the solar salt flats of Sfax, Tunisia, which have the ability to produce antimicrobial agents [21]. According to the literature, 65% of haloarchaea have the ability to produce the carotenoid bacterioruberin. This has a high capacity to capture free radicals and extinguish singlet oxygen, thus recovering damage caused by exposure to ultraviolet rays, being explored in the cosmetic field [22]. Halophilic microorganisms isolated from salterns in Argentina were also shown to produce biosurfactants, that have applications in numerous industries [23].

2.4 Metagenomics

Although there are countless microorganisms present in the different Earth environments, only 1% can be culturable in laboratory conditions. Being a culture-independent technique, metagenomics plays a crucial role in understanding the microbial diversity that inhabits our planet [24]. It is currently considered the most robust method for analyzing the composition of microbial communities. It consists mainly of the extraction and analysis of the total DNA, recovering genetic materials directly from the environmental sample [25]. It also entails the analysis of biological networks at multiple hierarchical levels (from metagenomes, metatranscriptomes, metaproteomes and metametabolomes) *in situ*. Several studies have used this technique, mainly focusing on terrestrial, marine and even intestinal environments [26].

Metagenomic analysis can be divided into two main approaches: sequencing and functional. In the sequencing approach, bioinformatic pipelines are developed to analyse the sequencing datasets in terms of taxonomic and functional annotation. The functional approach entails the expression of the environmental DNA into microbial hosts followed by functional screening for the desired activities [27].

2.5 Databases

Metagenomics generates a large datasets, in which bioinformatics plays a fundamental role for future analysis, mainly by pre-processing data, creating tools, pipelines and determining the reliability of the data [28]. To analyze the biosynthetic capacity of the microbial communities in terms of different classes of compounds such as antimicrobial agents, enzymes, anticancer agents and biosurfactants different databases can be explored.

Regarding enzymes, the CAZy [29], BRENDA [30], Expasy [31] and antiSMASH [32] databases have been successfully used to retrieve promising novel enzymes from metagenomic datasets. As demonstrated in a study on mud hot springs in Fiji, the AntiSMASH tool was instrumental in analyzing bacterial diversity and identifying possible bioactive compounds produced by the isolated actinomycete strains [33].The study

of *Morchella* and *Pseudomonas*, the CAZy database was able to identify chitinases and other types of enzymes associated with the interaction between these species [34].

With regard to antimicrobial compounds, the pathogen-host interactions database (PHI-Base) [35], CARD [36] and NaPDos2 [37] databases have been effectively employed to identify promising new antimicrobial candidates. In a set of metagenomic data from various biomes, NaPDosS2 was used to classify more than 35,000 type I KS (domains or enzymes within the PKS that catalyze the condensation reaction crucial for the poliquetidic chain elongation in the course of polyketide synthesis) domains out of 137. This analysis allowed a very comprehensive identification, contributing to the visualization of the diversity of PKSs (enzymes that catalyze the biosynthesis of polyketides, key bioactive compounds occurring in antibiotics and other natural products), and their distribution in the different biomes studied [38].

In the context of anticancer drugs, the PubChem BioAssay [39], DrugBank [40], CanSAR [41] and PharmacoDB [42] databases have been used in several studies. The CanSar database for example was used in specific cancers of female tissues to analyze gene expression and methylation of the MAPK isoform p38β. It provided important datasets associated with breast, cervical, ovarian and uterine endometrial cancer, allowing the expression and methylation of the MAPK11 gene to be analyzed [43].

In the case of biosurfactants, the use of BioSurfDB [44] has been outstanding in identifying promising new candidates in metagenomic datasets. BioSurfDB played a key role in a study of 46 metagenome samples from 20 different biomes, enabling a comprehensive, large-scale assessment of genes involved in biodegradation and biosurfactant processes [45].

3 Methodology

Seven raw metagenomic datasets were obtained from the sequencing of samples from Aveiro salterns (Portugal), Rio Maior salterns (Portugal) and Peña Hueca hypersaline lagoon (Spain), and were used in this work to find novel molecules with interest for different industries and sectors.

Initially, the quality of the raw data obtained was determined using FastQC software [46]. This provides a careful assessment of the data, forming a series of graphs that allow low-quality readings, adapters and contaminations to be identified. This ensures the reliability of the data, since after corrections, only the high-quality sequences were used for the further steps of the bioinformatics pipeline.

After quality control, assembly was carried out testing two programs: MEGAHIT [47] and SPAdes [48], which reconstruct the sequencing reads into contigs. After assembly, to ensure the quality of the contigs obtained, an evaluation was carried out using the QUAST software [49], that analyzed metrics such as the total length of the assembly, the number of contigs and their average size (N50) that allow to infer the assembly quality. Next, the obtained contigs were grouped into "bins" (a set of contigs that are grouped according to similar characteristics) using the tools: CONCOCT [50], MaxBin2 [51] and MetaBAT2 [52]. The Das Tool [53] was then used to concatenate results from the three software and obtain the metagenome assembled genomes

(MAGs), being sets of genomes reconstructed from DNA sequencing of a complex environmental sample, with restrictions of completeness > 90% and redundancy <10%.

MAGs were then taxonomically annotated using the Kaiju software [54]. General MAGs functional annotation was done with the eggNOG software [55], which contains orthologous groups constructed using the Smith-Waterman alignment technique which, based on homology, provides information on functions. MAGs were also mapped against the databases mentioned in the "database" section to find potential new enzymes and biosurfactants.

4 Work plan

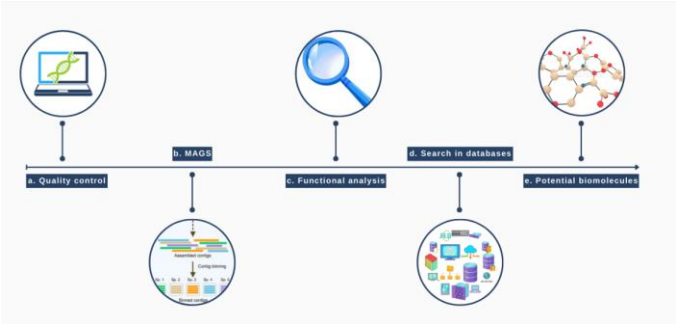

Fig. 1. A) Quality check and assembly of raw metagenomic sequencing data; b) Binning to obtain metagenome-assembled genomes (MAGs) using different software; c) Functional analysis of both clean reads and MAGs; d) Search for biomolecules of interest using specific databases; e) Check for novelty and potential of the identified biomolecules-coding sequences.

5 Results

5.1 General features of the metagenome

Metagenome sequencing using the Illumina NovaSeq6000 platform generated 88.8, 75.7, 67.3, 73.2, 67.8, 79.5, 105.3, 110.3, 98.9 and 134.3 million raw reads for the Aveiro D1, Aveiro G1, Aveiro G3, Aveiro H1, Aveiro H3 samples, Rio Maior A, Rio Maior C, Rio Maior D1, Rio Maior D2 and Peña Hueca PH, respectively, each with 150 bp (Table 1). Using MEGAHIT for de novo assembly, the following results were obtained: highest number of scaffolds above 5000 bp was obtained for Aveiro H3, while the lowest was recorded for Rio Maior C with only 27. On the other hand, if we look at the contig size, the largest obtained was for Aveiro G3 with 1,318,237 bp, and the smallest belonged to Aveiro D1 with a size of 168,745 bp. In addition, the average GC content varied from 61% as the maximum in Rio Maior D1 to the least of 53% in Aveiro H1 and H3.

Table 1 Summary of sequencing reads and assembly analyses.

Parameters	Aveiro D1	Aveiro G1	Aveiro G3	Aveiro H1	Aveiro H3	Rio Maior A	Rio Maior C	Rio Maior D1	Rio Maior D2	Pena Hueca PH
Salinity (%)	41.1	6.2	6.2	9.9	9.9	30.0	32.2	44.5	44.5	26.0
Raw reads	88,892,162	75,725,214	67,313,292	73,288,356	67,808,662	79,514,900	105,343,554	110,356,200	98,914,676	134,313,112
Assembly length (bp)	888,000,690	1,119,366,983	935,059,259	1,138,817,834	1,033,283,183	403,336,330	465,419,789	1,278,113,235	946,503,416	621,296,442
Number of contigs (> 0 bp)	1,475,900	1,380,067	502,497	1,494,447	1,383,445	680,83	900,483	2,795,354	1,945,251	863,734
Largest contig (bp)	168,745	601,99	1,318,237	336,566	584,432	1,222,312	558,573	294,882	338,724	866,352
Average contig length (bp)	602	811	1,861	762	747	592	517	457	487	719
N50 (bp)	1,053	1,483	1,309	1,364	1,402	2,158	1365	820	900	2265
L50 (bp)	102,068	104,681	98,783	118,889	96,501	16612	31565	117195	86810	30832
N90 (bp)	553	587	579	582	580	558	539	522	529	595
L90 (bp)	395,759	502,497	437,365	522,101	453,338	111869	147298	427762	337634	188747
GC content (%)	60	55	54	53	53	58	60	61	59	58
Number of scaffolds (> 1000 bp)	113,091	199,455	159,722	201,54	170,863	44,116	48,306	73,67	69,062	92,435
Number of scaffolds (> 5000 bp)	122	360	216	333	372	141	27	145	145	296
Assembly length (> 1000 bp)	275,858,726	536,103,642	410,043,115	510,906,813	455,089,106	147,797,709	128,468,895	203,952,990	187,785,135	291,469,159
Assembly length (> 50000 bp)	9,275,667	35,354,235	30,244,697	27,099,863	34,731,370	13,960,722	4,133,597	12,896,924	12,668,975	33,053,563

5.2 Taxonomic composition of compost microbiota

Taxonomic classification of metagenomic samples was done using the totality of the clean reads, before assembly, using Kaiju software and the nr_euk database. Identification at domain-level (Fig. 2) demonstrate that the Aveiro location was clearly dominated by Bacteria (79.2%) followed by Archaea (17.1%), Eukaryotes (0.56%) and Viruses (2.6%). In Rio Maior samples, the Archaea domain predominates (77.2%), followed by Bacteria (22%), Eukaryotes (0.28%) and Viruses (0.19%). Finally, in the Peña Hueca location in terms of abundance, just like Rio Maior, the domain Archaea (69%) predominated, followed by Bacteria (31%), Viruses (0.6%) and finally Eukaryotes (0.7%). Taxonomic analysis also enabled identification far beyond the domain, reaching species, genera, families, orders and classes for all samples, with average values for all samples of 10,691 species; 3743.8 genera; 990.8 families; 466.3 orders and 184.9 classes.

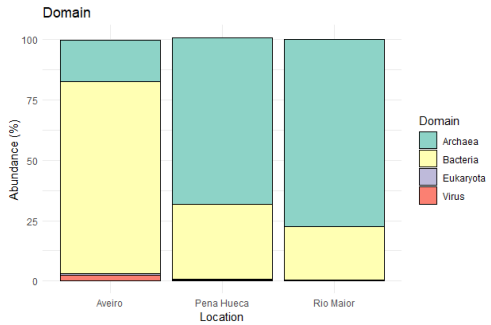

Fig. 2. Taxonomic overview of the samples Aveiro, Peña Hueca and Rio Maior. Taxonomic distribution of the metagenomes based on relative abundances of the clean metagenomic reads at domain level. The graph was constructed using the software 'R'.

In the taxonomic assessment at Class-level (Fig. 3), the Aveiro location was dominated by Gammaproteobacteria, followed by Alphaproteobacteria and Halobacteria. Unlike Aveiro, the Rio Maior location had the highest abundance relative to the Class Halobacteria, followed by Gammaproteobacteria and Balneolia. Finally, the Peña Hueca location, the Class Halobacteria, followed by Gammaproteobacteria and Nanohaloarchaea.

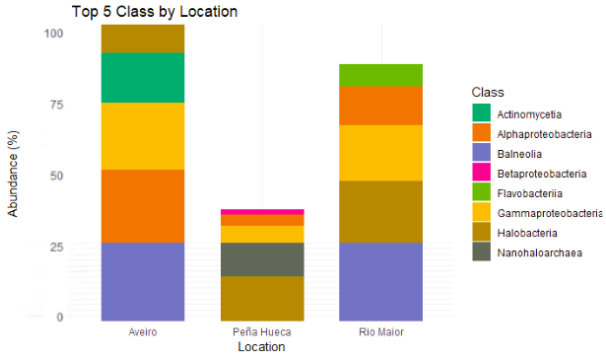

Fig. 3. Taxonomic overview of the samples Aveiro, Peña Hueca and Rio Maior. Taxonomic distribution of the metagenomes based on relative abundances of the clean metagenomic reads at Class level. The graph was constructed using the software 'R'.

At Order-level (Fig.4), a high predominance of Haloferacales was detected in all locations. In Peña Hueca, in addition to the Haloferacales, two other orders were extremely abundant, these being the Halobacteriales and Chromatiales. In Aveiro and Rio Maior, other Orders also stand out, although they are not very high, these being the Rhodobacterales in Aveiro and Halobacteriales in Rio Maior.

Fig. 4. Taxonomic overview of the samples Aveiro, Peña Hueca and Rio Maior. Taxonomic distribution of the metagenomes based on relative abundances of the clean metagenomic reads at Class Order. The graph was constructed using the software 'R'.

5.3 Identification and choice of bins

For subsequent analysis, the discovered contigs were grouped into "bins" according to their similarity, using the "binning" process. As previously mentioned, to evaluate the quality of the acquired bins the following parameters were used: "> 90% completeness and <10% redundancy" (Fig. 5), based on literature recommendation [56]. It was verified that a small quantity has the desired characteristics for future analyzes (represented in red). Furthermore, it was also possible to predict the proportion of bins that have the desired characteristics in each sample, where the samples "Aveiro H3", "Aveiro H1", "Aveiro G3", "Aveiro G1" and "Peña Hueca (PH)" stand out.

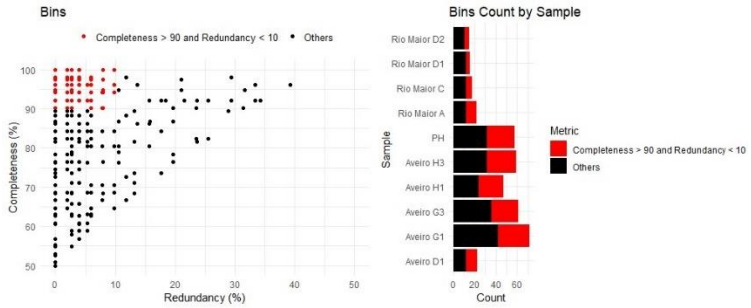

Fig. 5. Graphical representation of the quality and abundance of obtained bins. Color-coding for bins by criteria used, as described in literature, where red means the bins meet the criteria established and black means that they don't. Also, the percentage of bins per sample is depicted, giving an overview of the distribution of the bins within the different samples. The graph was constructed using the software 'R'.

5.4 Functional profiles and metabolic pathway annotation of the metagenomes

With the objective of analyzing the functional profiles of the metagenomes, using the DIAMOND algorithm, it was possible, after assembly, to identify regions capable of producing compounds of interest for subsequent analysis with databases. From the BioSurfDB database (Fig. 6, Fig. 7), it was possible to identify biosurfactants in different samples. There is a predominance of three samples from the same functional category, which is "Lipopeptide Biosynthesis". Within this category, the most abundant for all samples were "Putisolvin Biosynthesis", which is much superior to other products obtained in the study, followed by "Mycosubtilin Biosynthesis" from the same group. The "Glycolipid Biosynthesis" category was also highlighted in all samples, with higher prevalence of the sub-category "Biosynthesis of Trehalolipids". In terms of comparison between the samples, there are no major differences, with all following very similar patterns. Annotation using the antiSMASH database (Fig. 8) enabled the identification of secondary products. The prevalence of “Terpenes” is verified, which is very representative in all samples. The Aveiro location also presents a relatively high production of many other compounds of interest, with “NRPS”, “NRPS-like”, and “T3PKS” standing out. In the Peña Hueca and Rio Maior locations, the production of the remaining

compounds is not very high; however, “Lanthipeptide-class-II” and “RiPP-like” stand out in Peña Hueca, and “Lanthipeptide-class-II”, “Siderophore”, and “T3PKS” in Rio Maior. In terms of comparison between samples, Peña Hueca and Rio Maior show no major differences, following very similar patterns, while Aveiro deviates from these patterns and presents promising results in several products. Categorization of proteins using the CAZy database produced very striking results regarding the frequency of the different families of enzymes identified, with no overwhelming predominance of one of the biomolecules as identified in other databases (Fig. 9). Also, as typical of this analysis, some proteins were annotated to more than one CAZy class. Results showed that, in general, majority of proteins were annotated as "GT2", "GT4", "GH13" and “GH5” in Aveiro; in Peña Hueca, their frequency was generally lower, with "GT66" being the most frequent. In Rio Maior, the most represented family was "GT66", with a categorization of 5553 proteins.

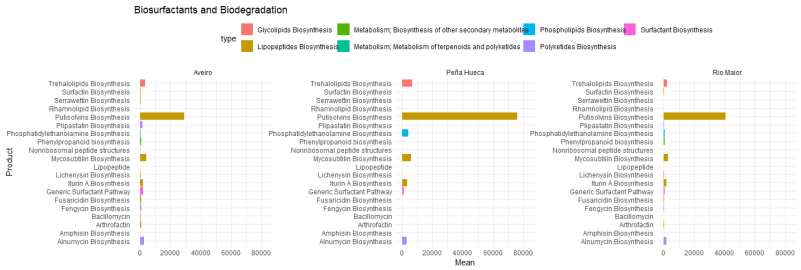

Fig. 6. Analysis of data using the Bio SurfDB database for Aveiro, Peña Hueca and Rio Maior samples. In this table, 'Mean' refers to the average number of reads mapped to each category within the database. The graph was constructed using the software 'R'.

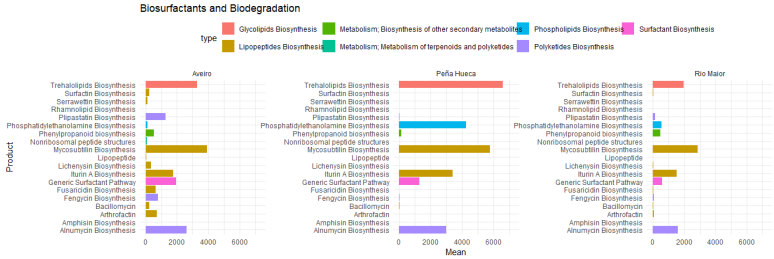

Fig. 7. Analysis of data using the BioSurfDB database for Aveiro, Peña Hueca and Rio Maior samples, excluding the “Putisolvins Biosynthesis” category, to highlight the other categories results. In this table, 'Mean' refers to the average number of reads mapped to each category within the database. The graph was constructed using the software 'R'.

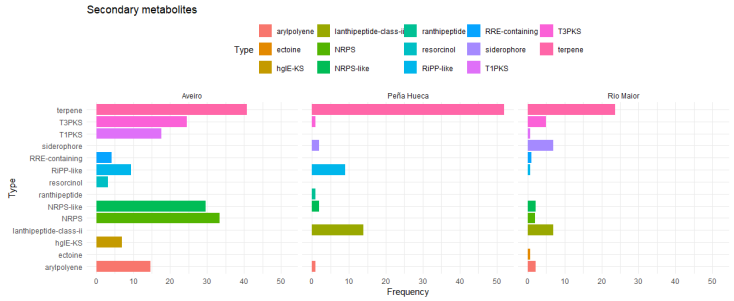

Fig. 8. Functional classification based on the antiSMASH database for Aveiro, Peña Hueca and Rio Maior samples, with a filter applied to select the top 10 entries per region. “Frequency” represents the quantity of compounds produced at each location. The 'Types' identified represent different enzyme categories, for example, NRPS (peptide synthesis), T1PKS (Type I polyketide synthesis), T3PKS (Type III Polyketide Synthases). The graph was constructed using the software 'R'.

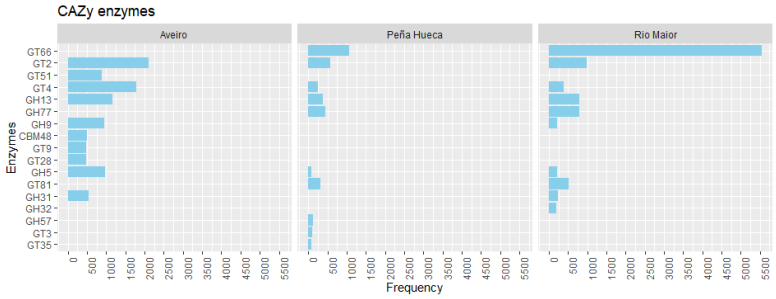

Fig. 9. Functional classification based on CAZy database for Aveiro, Peña Hueca and Rio Maior samples, with a filter applied to select the top 10 entries per region. “Frequency” represents the quantity of categories of enzymes produced at each location. The designations of the families obtained: GH (Glycoside Hydrolase Family), GT (Glycosyltransferases Family), CBM (Carbohydrate-Binding Module Family). The graph was constructed using the software 'R'.

6 Discussion

Our results obtained at domain-level (Fig. 2) are in accordance with the literature. The Archaea domain, along with the Bacteria domain, prevails in halophilic environments. In Aveiro, a higher abundance of Bacteria was observed, likely due to the lower mean salinity of the collected samples, in comparison with Rio Maior and PH. However, it is important to note that other domains can also be found in these areas, albeit in smaller numbers, indicating that the ability to adapt to high levels of salt is not exclusive to the Archaea and Bacteria domains [57]. In fact, at class-level (Fig. 3), samples are practically dominated by Halobacteria and Gammaproteobacteria. Halobacteria are an extremely interesting group of microorganisms within the Archaea domain, their

prominent presence can be attributed to their need for high salt concentrations for their growth [58]. In addition, halobacteria produce a significant amount of carotenoids. These have antioxidant properties that protect against oxidative stress and give the cells a characteristic colour, contributing to the absorption of light energy and providing photoprotection against intense solar radiation [59]. The Gammaproteobacteria belong to the domain Bacteria, according to studies they are classified as chitinotrophic, which may explain their prevalence compared to other groups, as chitin is abundant and is used as a growth substrate [60]. Regarding the Order (Fig. 4), it can be seen that, as in the literature, the predominance of Halobacteriales is common, being prominent within the class Halobacteria, making up this group of extremophile archaea, which are still poorly known and classified [61]. Their rise in these environments can be explained by their osmoadaptation strategies and optimal growth in salinities above 2 M NaCl [62].

Analyzing the BioSurfDB results (Fig. 6, Fig. 7), the predominance of putisolvins in all samples can be explained according to the literature by the regulation of the two-component system GacA/GacS, with GacS being primarily responsible for generating responses to environmental stimuli. Reduced temperatures and high salt concentrations contribute to the increased expression of GacA and GacS, influencing the biosynthesis of these substances. From a functional perspective, due to their amphipathic nature, putisolvins reduce surface tension at oil/water and air/water interfaces, reduce biofilm formation, disperse crystals of aromatic compounds such as naphthalene and phenanthrene, and promote bacterial surface motility [63]. Studies indicate that their production mainly occurs at the end of the exponential growth phase, causing a reduction in the medium's surface tension by approximately 40% and inhibiting biofilm formation in various contexts, especially on medical surfaces, without exhibiting phytotoxic effects [64]. Belonging to another group of products, trehalolipids stood out in all three samples. According to the literature, they play important roles due to their high structural diversity, such as: reducing surface tension, possessing immunomodulatory and antitumor activities, and facilitating the solubilization of hydrophobic compounds. Unfortunately, clinical use is still unlikely due to the high cytotoxicity of trehalolipids; however, trehalolipids produced by *Rhodococcus* exhibit reduced cytotoxicity, enabling medical use [65].

Analysis of the AntiSMASH data (Fig. 8) shows that all the samples contain terpenoids, which are considered to be the largest class of secondary metabolites [66]. They are generally associated with playing essential roles in plants, such as direct and indirect defense against pathogens and herbivores. From a functional point of view, a study revealed remarkable antibacterial properties [67]. In addition, the literature proves how extremely useful terpenoids are. In cosmetics, they protect the skin against ultraviolet rays and ageing, mainly due to their antioxidant effects [68]. In flavors and fragrances, several studies report their widespread use as essential oils, being applied in the food, beverage and perfume industries [69–71].

Analyzing the CAZy database results, we can see the prevalence of five main families, namely “GT2”, “GT4”, “GH13” “GH5” and “GT66”. The GT2 family is involved in synthesizing β-glucans with a wide range of properties that are important for biotechnological applications [72]. This family encodes glucose transporters that are

tissue-specific and insulin-sensitive, playing an essential role in the metabolic regulation of glucose in various tissues in response to systemic insulin levels [73]. The GT4 family of glycosyltransferases, which is abundant in *Lactiplantibacillus plantarum* 84-3, isolated from dairy samples, have a considerable impact on carbohydrate metabolism, benefitting the degradation of resistant starches and butanoate metabolism [74]. The GH13 family, particularly α-amylase, is studied because of its great importance in industrial applications. Their importance in hydrolyzing α-1,4-glycosidic bonds has therefore been prominent in several industrial sectors such as in the food industry, biofuels, and detergents [75]. According to the literature, the GH5 family, such as the GH5 xylanase of *Penicillium funiculosum*, has activity for degrading lignocellulosic residues, including the ones from sugarcane bagasse, and this enzyme is of interest for second-generation biofuel production [76]. An additional example is Bpman5 enzyme of the family GH5, of which studies have reported resistance toward different ions, organic solvents, and detergents [77]. The GT66 family includes proteins with glycosyltransferase activity, which are crucial in the N-glycosylation of proteins. These proteins represent the catalytic subunits of oligosaccharyltransferases [78].

Future work will be dedicated to find which of the identified sequences/molecules are novel and promising to further explore, and on using additional databases to explore the presence of antimicrobial and anticancer compounds on these metagenomic datasets.

7 Conclusion

Extreme environments, such as hypersaline locations, harbor tremendous biodiversity with regard to extremophiles. These organisms developed unique adaptations to live within such conditions and, due to such an adaptation, produce a wide diversity of biotechnologically important compounds in different industries, such as pharmaceuticals, environmental, and food.

The saltworks of Aveiro and Rio Maior in Portugal, and the hypersaline lagoon of Peña Hueca, in Spain, were studied to search for new biomolecules of industrial interest. The corresponding procedures involved the analysis of the metagenome, that is, the assembly of contigs, binning, and functional annotation, by using specific databases to identify enzymes and biosurfactants.

Results from the study depicted a high degree of taxonomic diversity present in the samples under study. The domains Bacteria and Archaea were observed to be the most predominant. Moreover, a large number of contigs were grouped into high-quality bins. Functional annotation identified the presence of biosurfactants, secondary metabolites, enzymes, and other relevant compounds. The results depended on different locations, with the Aveiro sample showing a greater diversity of identified compounds.

This study thus underscored the significance of metagenomics and its application to the extreme environments in the search for new biomolecules that could lead to new technologies and therapies in pharmaceutical, environmental, and food industries.

References

1. Giordano, D.: Bioactive Molecules from Extreme Environments. Mar Drugs. 18, 640 (2020). https://doi.org/10.3390/md18120640.
2. Poli, A., Finore, I., Romano, I., Gioiello, A., Lama, L., Nicolaus, B.: Microbial Diversity in Extreme Marine Habitats and Their Biomolecules. Microorganisms. 5, 25 (2017). https://doi.org/10.3390/microorganisms5020025.
3. Counts, J.A., Zeldes, B.M., Lee, L.L., Straub, C.T., Adams, M.W.W., Kelly, R.M.: Physiological, metabolic and biotechnological features of extremely thermophilic microorganisms. WIREs Systems Biology and Medicine. 9, (2017). https://doi.org/10.1002/wsbm.1377.
4. Kochhar, N., I.K, K., Shrivastava, S., Ghosh, A., Rawat, V.S., Sodhi, K.K., Kumar, M.: Perspectives on the microorganism of extreme environments and their applications. Curr Res Microb Sci. 3, 100134 (2022). https://doi.org/10.1016/j.crmicr.2022.100134.
5. Ajar Nath, Y.: Biodiversity and bioprospecting of extremophilic microbiomes for agro-environmental sustainability. J Appl Biol Biotechnol. (2021). https://doi.org/10.7324/JABB.2021.9301.
6. Tiwari, S., Prasad, V., Lata, C.: Bacillus: Plant Growth Promoting Bacteria for Sustainable Agriculture and Environment. In: New and Future Developments in Microbial Biotechnology and Bioengineering. pp. 43–55. Elsevier (2019). https://doi.org/10.1016/B978-0-444-64191-5.00003-1.
7. Biodiversity and biotechnological applications of halophilic microbes for sustainable agriculture. Journal of Applied Biology & Biotechnolog. (2018). https://doi.org/10.7324/JABB.2018.60109.
8. Santos-Pereira, C., Sousa, J., Costa, Â.M.A., Santos, A.O., Rito, T., Soares, P., Franco-Duarte, R., Silvério, S.C., Rodrigues, L.R.: Functional and sequence-based metagenomics to uncover carbohydrate-degrading enzymes from composting samples. Appl Microbiol Biotechnol. 107, 5379–5401 (2023). https://doi.org/10.1007/s00253-023-12627-9.
9. Coker, J.A.: Extremophiles and biotechnology: current uses and prospects. F1000Res. 5, 396 (2016). https://doi.org/10.12688/f1000research.7432.1.
10. Schultz, J., Rosado, A.S.: Extreme environments: a source of biosurfactants for biotechnological applications. Extremophiles. 24, 189–206 (2020). https://doi.org/10.1007/s00792-019-01151-2.
11. GudiÃ±a, E.J., Fernandes, E.C., Rodrigues, A.I., Teixeira, J.A., Rodrigues, L.R.: Biosurfactant production by Bacillus subtilis using corn steep liquor as culture medium. Front Microbiol. 6, (2015). https://doi.org/10.3389/fmicb.2015.00059.
12. Balan, S.S., Kumar, C.G., Jayalakshmi, S.: Aneurinifactin, a new lipopeptide biosurfactant produced by a marine Aneurinibacillus aneurinilyticus SBP-11 isolated from Gulf of Mannar: Purification, characterization and its biological evaluation. Microbiol Res. 194, 1–9 (2017). https://doi.org/10.1016/j.micres.2016.10.005.

13. Sayed, A.M., Hassan, M.H.A., Alhadrami, H.A., Hassan, H.M., Goodfellow, M., Rateb, M.E.: Extreme environments: microbiology leading to specialized metabolites. J Appl Microbiol. 128, 630–657 (2020). https://doi.org/10.1111/jam.14386.
14. Saccò, M., White, N.E., Harrod, C., Salazar, G., Aguilar, P., Cubillos, C.F., Meredith, K., Baxter, B.K., Oren, A., Anufriieva, E., Shadrin, N., Marambio-Alfaro, Y., Bravo-Naranjo, V., Allentoft, M.E.: Salt to conserve: a review on the ecology and preservation of hypersaline ecosystems. Biological Reviews. 96, 2828–2850 (2021). https://doi.org/10.1111/brv.12780.
15. Mokashe, N., Chaudhari, B., Patil, U.: Operative utility of salt-stable proteases of halophilic and halotolerant bacteria in the biotechnology sector. Int J Biol Macromol. 117, 493–522 (2018). https://doi.org/10.1016/j.ijbiomac.2018.05.217.
16. Yun, J.-H., Ohki, M., Park, J.-H., Ishimoto, N., Sato-Tomita, A., Lee, W., Jin, Z., Tame, J.R.H., Shibayama, N., Park, S.-Y., Lee, W.: Pumping mechanism of NM-R3, a light-driven bacterial chloride importer in the rhodopsin family. Sci Adv. 6, (2020). https://doi.org/10.1126/sciadv.aay2042.
17. Zajc, J., Kogej, T., Galinski, E.A., Ramos, J., Gunde-Cimerman, N.: Osmoadaptation Strategy of the Most Halophilic Fungus, Wallemia ichthyophaga, Growing Optimally at Salinities above 15% NaCl. Appl Environ Microbiol. 80, 247–256 (2014). https://doi.org/10.1128/AEM.02702-13.
18. Martínez, G.M., Pire, C., Martínez-Espinosa, R.M.: Hypersaline environments as natural sources of microbes with potential applications in biotechnology: The case of solar evaporation systems to produce salt in Alicante County (Spain). Curr Res Microb Sci. 3, 100136 (2022). https://doi.org/10.1016/j.crmicr.2022.100136.
19. Jin, M., Gai, Y., Guo, X., Hou, Y., Zeng, R.: Properties and Applications of Extremozymes from Deep-Sea Extremophilic Microorganisms: A Mini Review. Mar Drugs. 17, 656 (2019). https://doi.org/10.3390/md17120656.
20. Diba, H., Cohan, R.A., Salimian, M., Mirjani, R., Soleimani, M., Khodabakhsh, F.: Isolation and characterization of halophilic bacteria with the ability of heavy metal bioremediation and nanoparticle synthesis from Khara salt lake in Iran. Arch Microbiol. 203, 3893–3903 (2021). https://doi.org/10.1007/s00203-021-02380-w.
21. Ghanmi, F., Carré-Mlouka, A., Vandervennet, M., Boujelben, I., Frikha, D., Ayadi, H., Peduzzi, J., Rebuffat, S., Maalej, S.: Antagonistic interactions and production of halocin antimicrobial peptides among extremely halophilic prokaryotes isolated from the solar saltern of Sfax, Tunisia. Extremophiles. 20, 363–374 (2016). https://doi.org/10.1007/s00792-016-0827-9.
22. Serrano, S., Mendo, S., Caetano, T.: Haloarchaea have a high genomic diversity for the biosynthesis of carotenoids of biotechnological interest. Res Microbiol. 173, 103919 (2022). https://doi.org/10.1016/j.resmic.2021.103919.
23. Nercessian, D., Di Meglio, L., De Castro, R., Paggi, R.: Exploring the multiple biotechnological potential of halophilic microorganisms isolated from two

Argentinean salterns. Extremophiles. 19, 1133–1143 (2015). https://doi.org/10.1007/s00792-015-0785-7.
24. Amann, R.: Who is out there? Microbial Aspects of Biodiversity. Syst Appl Microbiol. 23, 1–8 (2000). https://doi.org/10.1016/S0723-2020(00)80039-9.
25. Lewin, A., Wentzel, A., Valla, S.: Metagenomics of microbial life in extreme temperature environments. Curr Opin Biotechnol. 24, 516–525 (2013). https://doi.org/10.1016/j.copbio.2012.10.012.
26. Sleator, R.D., Shortall, C., Hill, C.: Metagenomics. Lett Appl Microbiol. 47, 361–366 (2008). https://doi.org/10.1111/j.1472-765X.2008.02444.x.
27. Dindhoria, K., Manyapu, V., Ali, A., Kumar, R.: Unveiling the role of emerging metagenomics for the examination of hypersaline environments. Biotechnol Genet Eng Rev. 1–39 (2023). https://doi.org/10.1080/02648725.2023.2197717.
28. Ju, F., Zhang, T.: Experimental Design and Bioinformatics Analysis for the Application of Metagenomics in Environmental Sciences and Biotechnology. Environ Sci Technol. 49, 12628–12640 (2015). https://doi.org/10.1021/acs.est.5b03719.
29. Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., Terrapon, N.: The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 50, D571–D577 (2022). https://doi.org/10.1093/nar/gkab1045.
30. Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., Schomburg, D.: BRENDA, the ELIXIR core data resource in 2021: new developments and updates. Nucleic Acids Res. 49, D498–D508 (2021). https://doi.org/10.1093/nar/gkaa1025.
31. Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., Durinx, C.: Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. Nucleic Acids Res. 49, W216–W227 (2021). https://doi.org/10.1093/nar/gkab225.
32. Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R.: antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39, W339–W346 (2011). https://doi.org/10.1093/nar/gkr466.
33. Pipite, A., Siro, G., Subramani, R., Srinivasan, S.: Microbiological analysis, antimicrobial activity, heavy-metals content and physico-chemical properties of Fijian mud pool samples. Science of The Total Environment. 854, 158725 (2023). https://doi.org/10.1016/j.scitotenv.2022.158725.
34. Cailleau, G., Hanson, B.T., Cravero, M., Zhioua, S., Hilpish, P., Ruiz, C., Robinson, A.J., Kelliher, J.M., Morales, D., Gallegos-Graves, L.V., Bonito, G., Chain, P.S.G., Bindschedler, S., Junier, P.: Associated bacterial communities, confrontation studies, and comparative genomics reveal important interactions between Morchella with Pseudomonas spp. Frontiers in Fungal Biology. 4, (2023). https://doi.org/10.3389/ffunb.2023.1285531.
35. Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., Sahu, J., Iyer, S.V., Khamari, L., De Silva, N., Martinez, M.C., Pedro, H.,

Yates, A.D., Hammond-Kosack, K.E.: PHI-base in 2022: a multi-species phenotype database for Pathogen–Host Interactions. Nucleic Acids Res. 50, D837–D847 (2022). https://doi.org/10.1093/nar/gkab1037.

36. Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A., Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., Baker, S.J.C., Dave, M., McCarthy, M.C., Mukiri, K.M., Nasir, J.A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., Kwong, M., Liang, Z.C., Niu, K.C., Shan, P., Yang, J.Y.J., Gray, K.L., Hoad, G.R., Jia, B., Bhando, T., Carfrae, L.A., Farha, M.A., French, S., Gordzevich, R., Rachwalski, K., Tu, M.M., Bordeleau, E., Dooley, D., Griffiths, E., Zubyk, H.L., Brown, E.D., Maguire, F., Beiko, R.G., Hsiao, W.W.L., Brinkman, F.S.L., Van Domselaar, G., McArthur, A.G.: CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 51, D690–D699 (2023). https://doi.org/10.1093/nar/gkac920.
37. Klau, L.J., Podell, S., Creamer, K.E., Demko, A.M., Singh, H.W., Allen, E.E., Moore, B.S., Ziemert, N., Letzel, A.C., Jensen, P.R.: The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. Journal of Biological Chemistry. 298, 102480 (2022). https://doi.org/10.1016/j.jbc.2022.102480.
38. Singh, H.W., Creamer, K.E., Chase, A.B., Klau, L.J., Podell, S., Jensen, P.R.: Metagenomic data reveals type I polyketide synthase distributions across biomes. mSystems. 8, (2023). https://doi.org/10.1128/msystems.00012-23.
39. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., Zaslavsky, L., Zhang, J., Bolton, E.E.: PubChem 2023 update. Nucleic Acids Res. 51, D1373–D1380 (2023). https://doi.org/10.1093/nar/gkac956.
40. Knox, C., Wilson, M., Klinger, C.M., Franklin, M., Oler, E., Wilson, A., Pon, A., Cox, J., Chin, N.E. (Lucy), Strawbridge, S.A., Garcia-Patino, M., Kruger, R., Sivakumaran, A., Sanford, S., Doshi, R., Khetarpal, N., Fatokun, O., Doucet, D., Zubkowski, A., Rayat, D.Y., Jackson, H., Harford, K., Anjum, A., Zakir, M., Wang, F., Tian, S., Lee, B., Liigand, J., Peters, H., Wang, R.Q. (Rachel), Nguyen, T., So, D., Sharp, M., da Silva, R., Gabriel, C., Scantlebury, J., Jasinski, M., Ackerman, D., Jewison, T., Sajed, T., Gautam, V., Wishart, D.S.: DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Res. 52, D1265–D1275 (2024). https://doi.org/10.1093/nar/gkad976.
41. di Micco, P., Antolin, A.A., Mitsopoulos, C., Villasclaras-Fernandez, E., Sanfelice, D., Dolciami, D., Ramagiri, P., Mica, I.L., Tym, J.E., Gingrich, P.W., Hu, H., Workman, P., Al-Lazikani, B.: canSAR: update to the cancer translational research and drug discovery knowledgebase. Nucleic Acids Res. 51, D1212–D1219 (2023). https://doi.org/10.1093/nar/gkac1004.
42. Smirnov, P., Kofia, V., Maru, A., Freeman, M., Ho, C., El-Hachem, N., Adam, G.-A., Ba-alawi, W., Safikhani, Z., Haibe-Kains, B.: PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. Nucleic Acids Res. 46, D994–D1002 (2018). https://doi.org/10.1093/nar/gkx911.

43. Katopodis, P., Kerslake, R., Zikopoulos, A., Beri, N., Anikin, V.: p38β - MAPK11 and its role in female cancers. J Ovarian Res. 14, 84 (2021). https://doi.org/10.1186/s13048-021-00834-9.
44. Oliveira, J.S., Araújo, W., Lopes Sales, A.I., Brito Guerra, A. de, Silva Araújo, S.C. da, de Vasconcelos, A.T.R., Agnez-Lima, L.F., Freitas, A.T.: BioSurfDB: knowledge and algorithms to support biosurfactants and biodegradation studies. Database. 2015, (2015). https://doi.org/10.1093/database/bav033.
45. Oliveira, J.S., Araújo, W.J., Figueiredo, R.M., Silva-Portela, R.C.B., de Brito Guerra, A., da Silva Araújo, S.C., Minnicelli, C., Carlos, A.C., de Vasconcelos, A.T.R., Freitas, A.T., Agnez-Lima, L.F.: Biogeographical distribution analysis of hydrocarbon degrading and biosurfactant producing genes suggests that near-equatorial biomes have higher abundance of genes with potential for bioremediation. BMC Microbiol. 17, 168 (2017). https://doi.org/10.1186/s12866-017-1077-4.
46. Andrews, S.: FastQC A Quality Control tool for High Throughput Sequence Data. Babraham.ac.uk.
47. Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W.: MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. Bioinformatics. 31, 1674–1676 (2015). https://doi.org/10.1093/bioinformatics/btv033.
48. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A.: SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 19, 455–477 (2012). https://doi.org/10.1089/cmb.2012.0021.
49. Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G.: QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29, 1072–1075 (2013). https://doi.org/10.1093/bioinformatics/btt086.
50. Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C.: Binning metagenomic contigs by coverage and composition. Nat Methods. 11, 1144–1146 (2014). https://doi.org/10.1038/nmeth.3103.
51. Wu, Y.-W., Simmons, B.A., Singer, S.W.: MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 32, 605–607 (2016). https://doi.org/10.1093/bioinformatics/btv638.
52. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z.: MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 7, e7359 (2019). https://doi.org/10.7717/peerj.7359.
53. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., Banfield, J.F.: Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 3, 836–843 (2018). https://doi.org/10.1038/s41564-018-0171-1.

54. Menzel, P., Ng, K.L., Krogh, A.: Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 7, 11257 (2016). https://doi.org/10.1038/ncomms11257.
55. Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., Bork, P.: eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. 36, D250–D254 (2007). https://doi.org/10.1093/nar/gkm796.
56. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., DuLong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., Segata, N.: Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell. 176, 649-662.e20 (2019). https://doi.org/10.1016/j.cell.2019.01.001.
57. Oren, A.: Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Syst. 4, 2 (2008). https://doi.org/10.1186/1746-1448-4-2.
58. Kamekura, M.: Diversity of extremely halophilic bacteria. Extremophiles. 2, 289–295 (1998). https://doi.org/10.1007/s007920050071.
59. Rodrigo-Baños, M., Garbayo, I., Vílchez, C., Bonete, M., Martínez-Espinosa, R.: Carotenoids from Haloarchaea and Their Potential in Biotechnology. Mar Drugs. 13, 5508–5532 (2015). https://doi.org/10.3390/md13095508.
60. Sorokin, D.Y., Kolganova, T. V.: Bacterial chitin utilization at halophilic conditions. Extremophiles. 18, 243–248 (2014). https://doi.org/10.1007/s00792-013-0611-z.
61. Bachran, M., Kluge, S., Lopez-Fernandez, M., Cherkouk, A.: Microbial Diversity in an Arid, Naturally Saline Environment. Microb Ecol. 78, 494–505 (2019). https://doi.org/10.1007/s00248-018-1301-2.
62. Pal, S., Biswas, R., Misra, A., Sar, A., Banerjee, S., Mukherjee, P., Dam, B.: Poorly known microbial taxa dominate the microbiome of hypersaline Sambhar Lake salterns in India. Extremophiles. 24, 875–885 (2020). https://doi.org/10.1007/s00792-020-01201-0.
63. Dubern, J.-F., Bloemberg, G. V.: Influence of environmental conditions on putisolvins I and II production in *Pseudomonas putida* strain PCL1445. FEMS Microbiol Lett. 263, 169–175 (2006). https://doi.org/10.1111/j.1574-6968.2006.00406.x.
64. Kuiper, I., Lagendijk, E.L., Pickford, R., Derrick, J.P., Lamers, G.E.M., Thomas-Oates, J.E., Lugtenberg, B.J.J., Bloemberg, G. V.: Characterization of two *Pseudomonas putida* lipopeptide biosurfactants, putisolvin I and II, which inhibit biofilm formation and break down existing biofilms. Mol Microbiol. 51, 97–113 (2004). https://doi.org/10.1046/j.1365-2958.2003.03751.x.
65. Kuyukina, M.S., Ivshina, I.B., Baeva, T.A., Kochina, O.A., Gein, S. V., Chereshnev, V.A.: Trehalolipid biosurfactants from nonpathogenic Rhodococcus actinobacteria with diverse immunomodulatory activities. N Biotechnol. 32, 559–568 (2015). https://doi.org/10.1016/j.nbt.2015.03.006.

66. Oldfield, E., Lin, F.: Terpene Biosynthesis: Modularity Rules. Angewandte Chemie International Edition. 51, 1124–1137 (2012). https://doi.org/10.1002/anie.201103110.
67. Zhao, Y., Hu, T., Liu, R., Hao, Z., Liang, G., Li, G.: Biochemical Characterization and Function of Eight Microbial Type Terpene Synthases from Lycophyte Selaginella moellendorffii. Int J Mol Sci. 22, 605 (2021). https://doi.org/10.3390/ijms22020605.
68. Lee, C.: Fifty years of research and development of cosmeceuticals: a contemporary review. J Cosmet Dermatol. 15, 527–539 (2016). https://doi.org/10.1111/jocd.12261.
69. Bohlmann, J., Keeling, C.I.: Terpenoid biomaterials. The Plant Journal. 54, 656–669 (2008). https://doi.org/10.1111/j.1365-313X.2008.03449.x.
70. Hausch, B.J., Lorjaroenphon, Y., Cadwallader, K.R.: Flavor Chemistry of Lemon-Lime Carbonated Beverages. J Agric Food Chem. 63, 112–119 (2015). https://doi.org/10.1021/jf504852z.
71. Schwieterman, M.L., Colquhoun, T.A., Jaworski, E.A., Bartoshuk, L.M., Gilbert, J.L., Tieman, D.M., Odabasi, A.Z., Moskowitz, H.R., Folta, K.M., Klee, H.J., Sims, C.A., Whitaker, V.M., Clark, D.G.: Strawberry Flavor: Diverse Chemical Compositions, a Seasonal Influence, and Effects on Sensory Perception. PLoS One. 9, e88446 (2014). https://doi.org/10.1371/journal.pone.0088446.
72. Oehme, D.P., Shafee, T., Downton, M.T., Bacic, A., Doblin, M.S.: Differences in protein structural regions that impact functional specificity in GT2 family β-glucan synthases. PLoS One. 14, e0224442 (2019). https://doi.org/10.1371/journal.pone.0224442.
73. Lane, M.D., Flores-Riveros, J.R., Hresko, R.C., Kaestner, K.H., Liao, K., Janicot, M., Hoffman, R.D., McLenithan, J.C., Kastelic, T., Christy, R.J.: Insulin-Receptor Tyrosine Kinase and Glucose Transport. Diabetes Care. 13, 565–575 (1990). https://doi.org/10.2337/diacare.13.6.565.
74. Liang, T., Jiang, T., Liang, Z., Zhang, N., Dong, B., Wu, Q., Gu, B.: Carbohydrate-active enzyme profiles of Lactiplantibacillus plantarum strain 84-3 contribute to flavor formation in fermented dairy and vegetable products. Food Chem X. 20, 101036 (2023). https://doi.org/10.1016/j.fochx.2023.101036.
75. Janeček, Š., Svensson, B., MacGregor, E.A.: α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. Cellular and Molecular Life Sciences. 71, 1149–1170 (2014). https://doi.org/10.1007/s00018-013-1388-z.
76. Ogunyewo, O.A., Okereke, O.E., Kumar, S., Yazdani, S.S.: Characterization of a GH5 endoxylanase from Penicillium funiculosum and its synergism with GH16 endo-1,3(4)-glucanase in saccharification of sugarcane bagasse. Sci Rep. 12, 17219 (2022). https://doi.org/10.1038/s41598-022-21529-1.
77. Zang, H., Xie, S., Wu, H., Wang, W., Shao, X., Wu, L., Rajer, F.U., Gao, X.: A novel thermostable GH5_7 β-mannanase from Bacillus pumilus GBSW19 and its application in manno-oligosaccharides (MOS) production. Enzyme Microb Technol. 78, 1–9 (2015). https://doi.org/10.1016/j.enzmictec.2015.06.007.

78. Liang, D.-M., Liu, J.-H., Wu, H., Wang, B.-B., Zhu, H.-J., Qiao, J.-J.: Glycosyltransferases: mechanisms and applications in natural product development. Chem Soc Rev. 44, 8350–8374 (2015). https://doi.org/10.1039/C5CS00600G.