Analyzing metagenomic datasets from extreme environments to uncover biotechnologically valuable biomolecules

Ian Alves Machado[1], Sara Silverio[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 provides a comprehensive review of these microorganisms, with the main aim of understanding their biotechnological potential. For this approach, metagenomic datasets obtained from samples collected in hypersaline locations will be analyzed by different bioinformatics software aiming to reconstruct metagenome-assembled genomes (MAGs), that will then be explored in terms of functional annotation aiming to identify biomolecules of biotechnological interest, such as biosurfactants, enzymes, antimicrobial agents and anticancer agents.

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 will search for tools and databases that allow us to study and analyze biomolecules from microorganisms inhabiting hypersaline environments. Additionally, we will contribute to the enrichment of a database dedicated to biosurfactants.

2 State of the Art

2.1 Extreme environments and extremophiles

Extreme environments are classified as environments exposed to one or more extreme environmental parameter 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]. Because of their habitat, they developed biological mechanisms and adaptations to maintain their metabolic activity. Due to their unique adaptation characteristics, they produce a series of secondary metabolites that have a high biotechnological potential [3].

To emphasize the potential of extremophiles biomolecules, some examples of their reported applications will be provided. In the agricultural industry, for example, they act as biofertilizers, help in the cycling, fixation, solubilization of nutrients and mineralization, replacing current methods and techniques, promoting an increasingly clean and sustainable agriculture by replacing chemical products [4–7]. Bioenergy from the degradation of lignocellulosic biomass has the potential to be the main source of energy globally over the years. Its efficient conversion depends above all on the catalytic activity of carbohydrate-active enzymes. According to a study, there is a high abundance of *Alphaproteobacteria* and *Gammaproteobacteria*, thermophilic microorganisms that enhance the degradation of lignocellulose in samples from compost heaps containing this material [8]. In the health field, extremophiles are capable of producing a series of antibiotics, antitumor molecules and antifungals. They are well known for generating antimicrobial peptides such as halocins, which exhibit activities against a variety of microorganisms [4, 9]. In addition, they produce biosurfactants with a variety of therapeutic applications, acting to reduce microbial adhesion and colonization through their anti-adhesive and anti-biofilm properties [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. Different geomorphological factors explain the formation of these

environments, the vast majority of which are located in arid and semi-arid regions, such as the Great Salt Lake (Utah, USA), Salar of Atacama (Chile), Dead Sea (Israel) and Kati Thanda-Lake Eyre (South Australia, Australia), which favour hydrological imbalance caused by high evaporation rates that exceed precipitation, causing the accumulation of salts over time [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. Currently, a large number of enzymes in these deep regions have been identified. *Bacillus* sp. dsh19-1 and *Zunongwangia profunda* are adapted to extremely cold and saline conditions and are of industrial interest, being applied 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. Indeed, *Bacillus* sp., *Oceanobacillus* sp. and *Salinicoccus* sp. showed resistance to high concentrations of lead (4.1-7.2 mM) and nickel (3.6-4.1mM). In addition, *Oceanobacillus* sp. was shown to have the highest lead bioremediation capacity (98.8%), followed by *Bacillus* sp. (97.5%) and *Salinicoccus* sp. (92%). In relation to nickel, all the strains showed lower results when compared to lead, with *Bacillus* sp. (76%) having the highest removal capacity, followed by *Oceanobacillus* sp. (73.5%) and *Salinicoccus* sp. (71.7%) [20]. Halophilic microorganisms have been identified in the solar salt flats of Sfax, Tunisia, which have the ability to produce antimicrobial agents. Using specific selection criteria, microbial strains were isolated from both crystallizing and non-crystallizing ponds. The results obtained indicated that strains of *Salinarum* sp. showed antimicrobial activity, associated with the synthesis of a hydrophobic peptide called halocin [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. Currently, it is a molecule with great value in the market, mainly associated with cosmetics and health in the encapsulation of medicines, facilitating their administration [22]. Halophilic microorganisms isolated from salterns in Argentina were also shown to produce biosurfactants, that have applications in numerous industries ranging fromm food, to health to bioremediation [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 and evaluating the composition of microbial communities. It consists mainly of the extraction and analysis of the total DNA, without involving any selection, minimizing possible technical errors during the process, 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. The sequencing approach aims to identify different species in the sample by amplifying specific genomic regions to obtain detailed information on the taxonomic composition of the microbial community. On the other hand, the functional approach sequences all the genomic fragments present in the sample, allowing an assessment of the potential metabolic functions of microorganisms and how they interact with the ecosystem [27].

Lake Tyrrell, located in Melbourne, Australia, was subjected to successive sequential metagenomic analyses, where it was possible to identify new species of halophilic Archaea microorganisms, resulting in a new Nanohaloarchaea class [28–30]. Hypersaline soils from the Odiel salt marshes in Spain were subjected to metagenomic analysis in order to reveal valuable information about the local microbiota. A significant variety of microorganisms was found, in which the class Halobacteria and the presence of contigs related to Nanohaloarchaeota and Thaumarchaeota stood out [31]. The metagenomic study carried out in the Karak salt mine [32], Pakistan, highlights a diverse microbial community, including Euryarchaeota, Bacteroidetes, Proteobacteria, with the predominant genera being *Marinobacter* and *Salinibacter*.

2.5 Databases

Metagenomics generates a large quantity of data, 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 [33]. 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 [34], BRENDA [35], Expasy [36] and antiSMASH [37] 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 [38].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 [39].

With regard to antimicrobial compounds, the PHI-Base [40], CARD [41] and NaPDos2 [42] databases have been effectively employed to identify promising new antimicrobial candidates. The pathogen-host interactions database (PHI-Base) was used in a study to identify 112 hypervirulent mutations in 37 pathogen species. This analysis covered animal and plant pathogens and observed an increase in pathogen virulence, focusing mainly on trans-kingdom, molecular and cellular mutations [43]. In a set of metagenomic data from various biomes, NaPDosS2 was used to classify more than 35,000 type I KS domains out of 137. This analysis allowed a very comprehensive identification, contributing to the visualization of the diversity of PKSs, and their distribution in the different biomes studied [44].

In the context of anticancer drugs, the PubChem BioAssay [45], DrugBank [46], CanSAR [47] and PharmacoDB [48] databases have been used in several studies. The PubChem BioAssay database, for example, has identified nine potential inhibitors of active compounds in the avian infectious bronchitis virus (IBV) type 3c protease that can interact with the severe acute respiratory syndrome protease (SARS-CoV-2), four of which are covalent inhibitors and five non-covalent [49]. The CanSar database was used in specific cancers of female tissues to analyze gene expression and methylation of the MAPK isoform p38β. It provided important data sets associated with breast, cervical, ovarian and uterine endometrial cancer, allowing the expression and methylation of the MAPK11 gene to be analyzed [50].

In the case of biosurfactants, the use of BioSurfDB [51] 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. This enabled a better understanding of the distribution and diversity of these genes, contributing to the development of strategies associated with the manipulation of microbial consortia in areas with greater genetic variability in the production of biosurfactants [52].

3 Methodology

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

Initially, the quality of the raw data obtained will be determined using a software called FastQC [53]. 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 will be used for the further steps of the bioinformatics pipeline.

After quality control, assembly will be carried out testing two programs: MEGAHIT [54] and SPAdes [55], which will reconstruct the sequencing reads into contigs. After assembly, to ensure the quality of the contigs obtained, an evaluation will be carried out using the QUAST software [56], that will analyze 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 contigs obtained will be grouped into "bins" (a set of contigs that are grouped according to similar characteristics) using the tools: CONCOCT [57], MaxBin2 [58] and MetaBAT2 [59]. The Das Tool [60] will then be used to concatenate results from the three software and obtain metagenome assembled genomes (MAGs), being sets of genomes reconstructed from DNA sequencing of a complex environmental sample, with restrictions of completeness > 50% and redundancy <10%.

MAGs will then be taxonomically annotated using the Kaiju software [61]. General MAGs functional annotation will be done with the eggNOG software [62], which contains orthologous groups constructed using the Smith-Waterman alignment technique which, based on homology, provides information on functions. MAGs will also be mapped against the databases mentioned in the "database" section to find potential new enzymes, biosurfactants, antimicrobial and anticancer compounds. The novelty of the obtained sequences will be evaluated by BLAST analysis using the NCBI database.

4 Work plan

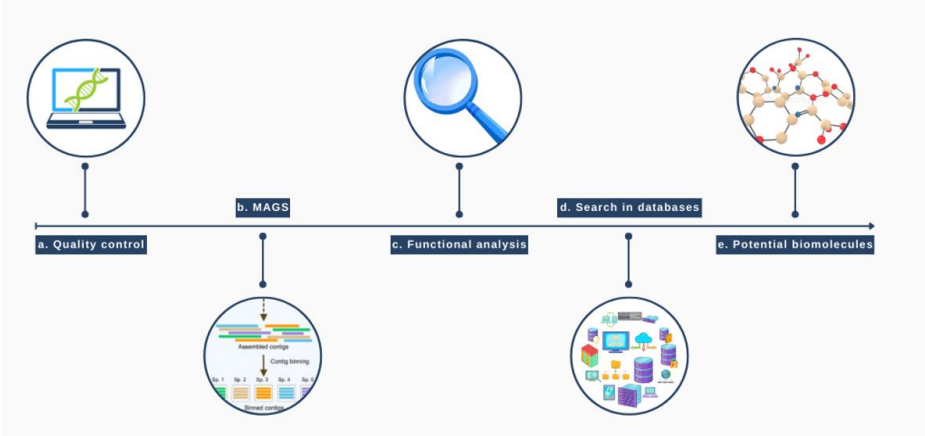

Fig. 1. a) Quality check and assembly of raw metagenomic sequencing data; b) Binning to obtain metagenome-assembled genomes (MAGs) using different software; c) Functional analysis of both clean reads and MAGs; d) Search for biomolecules of interest

using specific databases; e) Check for novelty and potential of the identified biomolecules-coding sequences.

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. Ventosa, A., de la Haba, R.R., Sánchez-Porro, C., Papke, R.T.: Microbial diversity of hypersaline environments: a metagenomic approach. Curr Opin Microbiol. 25, 80–87 (2015). https://doi.org/10.1016/j.mib.2015.05.002.
29. Emerson, J.B., Andrade, K., Thomas, B.C., Norman, A., Allen, E.E., Heidelberg, K.B., Banfield, J.F.: Virus-Host and CRISPR Dynamics in Archaea-Dominated Hypersaline Lake Tyrrell, Victoria, Australia. Archaea. 2013, 1–12 (2013). https://doi.org/10.1155/2013/370871.
30. Narasingarao, P., Podell, S., Ugalde, J.A., Brochier-Armanet, C., Emerson, J.B., Brocks, J.J., Heidelberg, K.B., Banfield, J.F., Allen, E.E.: *De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. ISME J. 6, 81–93 (2012). https://doi.org/10.1038/ismej.2011.78.
31. Vera-Gargallo, B., Ventosa, A.: Metagenomic Insights into the Phylogenetic and Metabolic Diversity of the Prokaryotic Community Dwelling in Hypersaline Soils from the Odiel Saltmarshes (SW Spain). Genes (Basel). 9, 152 (2018). https://doi.org/10.3390/genes9030152.
32. Cycil, L.M., DasSarma, S., Pecher, W., McDonald, R., AbdulSalam, M., Hasan, F.: Metagenomic Insights Into the Diversity of Halophilic Microorganisms Indigenous to the Karak Salt Mine, Pakistan. Front Microbiol. 11, (2020). https://doi.org/10.3389/fmicb.2020.01567.
33. 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.
34. 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.
35. 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.

36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.
43. Brown, N.A., Urban, M., Hammond-Kosack, K.E.: The trans-kingdom identification of negative regulators of pathogen hypervirulence. FEMS Microbiol Rev. 40, 19–40 (2016). https://doi.org/10.1093/femsre/fuv042.

44. 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.
45. 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.
46. 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.
47. 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.
48. 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.
49. Soulère, L., Barbier, T., Queneau, Y.: In Silico Identification of Potential Inhibitors of the SARS-CoV-2 Main Protease among a PubChem Database of Avian Infectious Bronchitis Virus 3CLPro Inhibitors. Biomolecules. 13, 956 (2023). https://doi.org/10.3390/biom13060956.
50. 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.
51. 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.
52. 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.
53. Andrews, S.: FastQC A Quality Control tool for High Throughput Sequence Data. Babraham.ac.uk.

54. 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.
55. 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.
56. 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.
57. 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.
58. 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.
59. 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.
60. 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.
61. 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.
62. 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.