



THE GLOBAL OCEAN GENOME: A “CATALOG” OF OCEAN LIFE

Elisa Laiolo^{1,2,3†}, Intikhab Alam^{3†}, Mahmut Uludag³, Tahira Jamil^{1,2,3}, Susana Agusti^{1,2}, Takashi Gojobori³, Silvia G. Acinas⁴, Josep M. Gasol⁴ and Carlos M. Duarte^{1,2,3}

¹Marine Science Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

²Red Sea Research Center (RSCR), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

³Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

⁴Department of Marine Biology and Oceanography, Institute of Marine Sciences, Spanish National Research Council (CSIC), Barcelona, Spain

[†]These authors contributed equally to this work and share first authorship

YOUNG REVIEWERS:



IGOR

AGE: 12



SEBASTIAN

AGE: 13

Life has been evolving in the oceans much longer than it has on land, resulting in highly diverse ocean organisms—particularly microbes like bacteria and archaea. Ocean microbes perform crucial functions that influence the health of the ocean and ultimately impact Earth’s climate. To understand the diversity and functions of marine organisms, scientists have used a powerful technique called metagenomics to study the DNA of *all* the organisms present in an ocean-water sample at once. In our research, we combined results from multiple ocean metagenomic studies, taken from various locations and depth zones across the world’s oceans, to produce a global ocean genome composed of 317.5 million groups of

similar genes—approximately half of which could be categorized by type of organism and function. This unprecedented amount of data has much to teach us about varied ocean habitats and can help scientists answer many questions about ocean organisms and their functions.

THE OCEAN—THE CRADLE OF LIFE

The ocean is the world's largest habitat. Oceans cover 71% of our planet's surface and contain an absolutely enormous volume of water—1.335 billion km³, about 97% of all the water on Earth! The ocean is also the cradle of life—Earth's first life forms appeared in this watery world about 3.9 billion years ago.

Many kinds of organisms live in the ocean, from plants to animals to bacteria to viruses. While you are probably most familiar with larger marine animals like whales, dolphins, and octopi, the ocean actually contains a large array of animal life. Earth's first life forms originated in the ocean and only later evolved to conquer land. The long evolutionary history of ocean-based life is illustrated by the fact that, of the 34 major known animal phyla, there is only one that exists exclusively on land (that's Onychophora, a kind of worm that has legs and a texture like velvet) [1]. It should not be surprising, given how long life has been evolving there, that the ocean contains a tremendous biodiversity—much of which remains unexplored.

All animals—those in the ocean and those on land—are classified as **eukaryotes**, which means that each of their cells has a distinct nucleus and other membrane-bound functional units called organelles. However, by far the most abundant organisms in the ocean are the **prokaryotes**, ocean microbes that are invisible to the naked eye. Prokaryotes, which are single-celled organisms without nuclei or specialized organelles, include two types of organisms: bacteria, which you have probably heard of, and archaea, which you might be less familiar with. While bacteria and archaea have many things in common, one important feature that sets these two groups apart is that the cell walls of many bacteria contain a substance called **peptidoglycan**, which the cell walls of archaea completely lack.

While we know that there are more than two million species of bacteria in the global ocean, we still do not know much about them or the other types of ocean microbes. These organisms are extremely difficult to study because more than 99% of them have not been grown in the lab!

WHY ARE OCEAN MICROBES IMPORTANT?

Individual species of bacteria and archaea are fascinatingly varied. Importantly, they have many unique ways of obtaining energy.

EUKARYOTE

An organism whose cells contain a nucleus and other membrane-bound functional units called organelles. Animals, plants, and fungi are examples of eukaryotes.

PROKARYOTE

A single-celled organism that does not have a distinct nucleus or organelles. Bacteria and archaea are prokaryotes.

PEPTIDOGLYCAN

A substance forming the cell walls of many bacteria. Peptidoglycans are not present in the cell walls of archaea.

METABOLISM

Chemical reactions within cells that change “food” into energy.

Metabolism is the word for the energy-producing chemical reactions inside most cells. These reactions use “food” sources to produce the energy that cells need to grow, reproduce, and generally stay alive. Some bacteria and archaea use sunlight to produce energy, while others break down or transform compounds containing carbon, nitrogen, or sulfur—elements that can affect ocean chemistry and that also play important roles in Earth’s changing climate [2, 3].

Additionally, ocean microbes can be sources of molecules that play key roles in biotechnology. For example, a protein called Taq polymerase was isolated from bacteria growing in super-hot ocean water near hydrothermal vents, which are like deep-sea geysers. This protein is now an important part of a common laboratory test called PCR, which is used for many scientific and medical studies including the detection of the SARS-CoV-2 virus that causes COVID-19.

Just from these examples, you can see that, despite their tiny size, ocean microbes—particularly bacteria and archaea—are critically important ocean-dwellers. What is even more intriguing is that many ocean organisms and their associated functions have not yet been discovered. How can we get a picture of all the organisms that live in the ocean—from eukaryotes to the smallest and hardest-to-study prokaryotes? It would be great, would it not, if we could simply take samples of water or sediment from the ocean and quickly analyze them in the lab to find out what kinds of organisms are living there and which functions they perform? Well, thanks to a powerful new technique called metagenomics, researchers *can* do this!

STUDYING GENES TO UNDERSTAND ORGANISMS

Every living organism contains DNA, which holds the instructions—genes—that code for the structure and function of that organism. Collectively, all of an organism’s genes are called its **genome**.

One of the main ways that scientists study the genomes of organisms (a field called genomics) is by **sequencing** their DNA. Sequencing is a laboratory technique that allows scientists to “read” the DNA code. For many years, DNA sequencing was an error-prone, time-consuming process and scientists could only sequence short stretches of DNA in each experiment—it took an awfully long time to read the genome of an organism. For example, sequencing of the entire human genome, called the Human Genome Project, took 13 years of work by many scientists spread across 20 institutions worldwide [4].

However, in the past decades, some amazing technical advances in the field of DNA sequencing have not only made whole-genome sequencing faster, easier, and cheaper, but have enabled the field of **metagenomics** to explode. “Meta-” means “higher” or “beyond,” so,

GENOME

The collection of all the genes of an organism.

SEQUENCING

A laboratory technique that allows scientists to “read” the DNA code, either small segments or the entire genome of an organism.

METAGENOMICS

The study of the genomes of all the organisms in a sample, such as a sample of ocean water or those on the human skin.

metagenomics is “beyond” the study of one organism’s genome—it is the study of the *pooled* genetic information from *all* the organisms contained in a sample from the environment, like water or soil. Using metagenomics techniques, scientists can sequence *all* the DNA in a sample of ocean water to figure out the global ocean genome—all the genes from all the organisms in that sample. This information can tell scientists which types of organisms are present and what kinds of ecologically important functions they are performing in ocean habitats [5]!

THE POWER OF METAGENOMICS: WHO LIVES IN THE OCEAN AND WHAT ARE THEY DOING?

The first metagenomic study of the ocean was performed by the Sorcerer II Global Ocean Sampling Expedition, back in 2003–2004, which analyzed a marine plankton community. Over the past 10 years, other expeditions have occurred, including the TARA Ocean Expedition, which collected 243 samples from 68 sampling sites between 2009 and 2013, mostly from the upper ocean. Metagenomics analysis of these samples identified 33.3 million genes! (In contrast, the human genome contains approximately 30,000 genes—see [here](#) for more info.) Additional studies, including the Malaspina Expedition, the Ocean Sampling Day Program [6], and the Hawaii Ocean Time Series, have expanded on these results, sampling other oceans, specific marine habitats, and water from various depths.

It is important to recognize that the ocean environment is not the same from the surface to the ocean floor—we can divide it into four unique zones, as shown in [Figure 1A](#). In our research, we combined the vast amount of metagenome data from 2,112 ocean samples collected from multiple past studies ([Figures 1B, C](#)), to create a full “catalog” of the global ocean genome, called KMAP Global Ocean Gene Catalog 1.0. This catalog contains an amazing amount of information—317.5 million gene clusters (collections of similar genes grouped together according to function), approximately half of which are **annotated**, meaning they have been classified according to the type of organism they belong to and the functions of the proteins they code for.

ANNOTATED

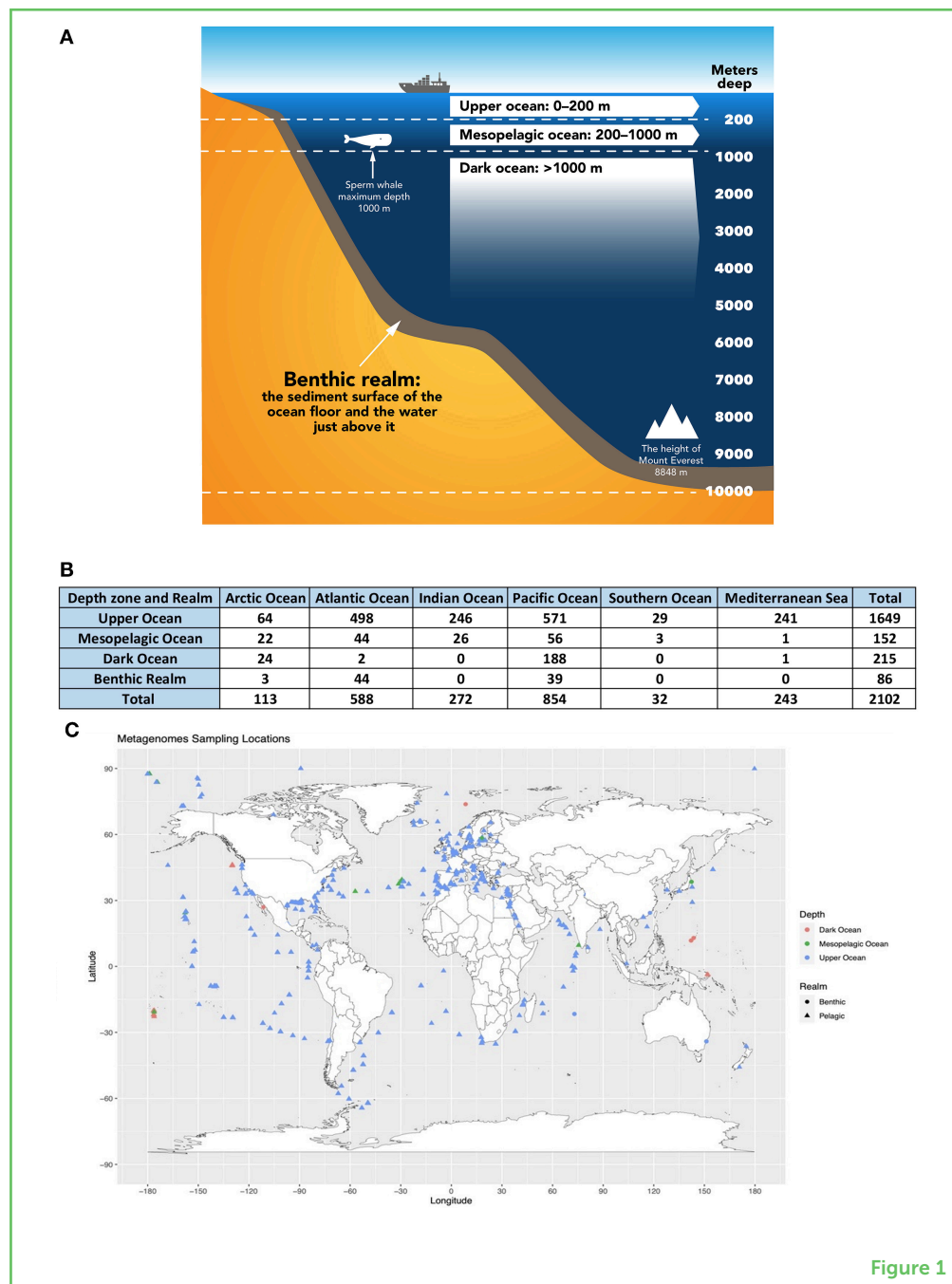
Classified according to the type of organism the genes belong to and the functions of the proteins they code for.

BACTERIA: THE OCEAN’S MAJOR PLAYERS

Focusing on the annotated gene clusters—those that we can link to a particular type of organism and for which we have functional information—we found that, across all ocean samples, 78.34% of the annotated genes were from bacteria, followed by eukaryotes (12.16%), archaea (6.05%), and viruses (3.44%) ([Figure 2A](#)). While genes from bacteria were dominant in every depth zone, genes from these four major types of ocean organisms were not evenly distributed throughout all zones—the benthic realm and the various depth

Figure 1

Samples compiled for our study. **(A)** Ocean zones and depths. **(B)** Locations and depth zones of the samples used for our research. You can see that certain oceans, like the Southern Ocean, and certain depth zones, like the dark ocean and the benthic realm, are underrepresented in these samples. **(C)** Locations from which the metagenomes used in our research were obtained across the benthic and pelagic realms. Depth zones are shown by the colors of the circles.



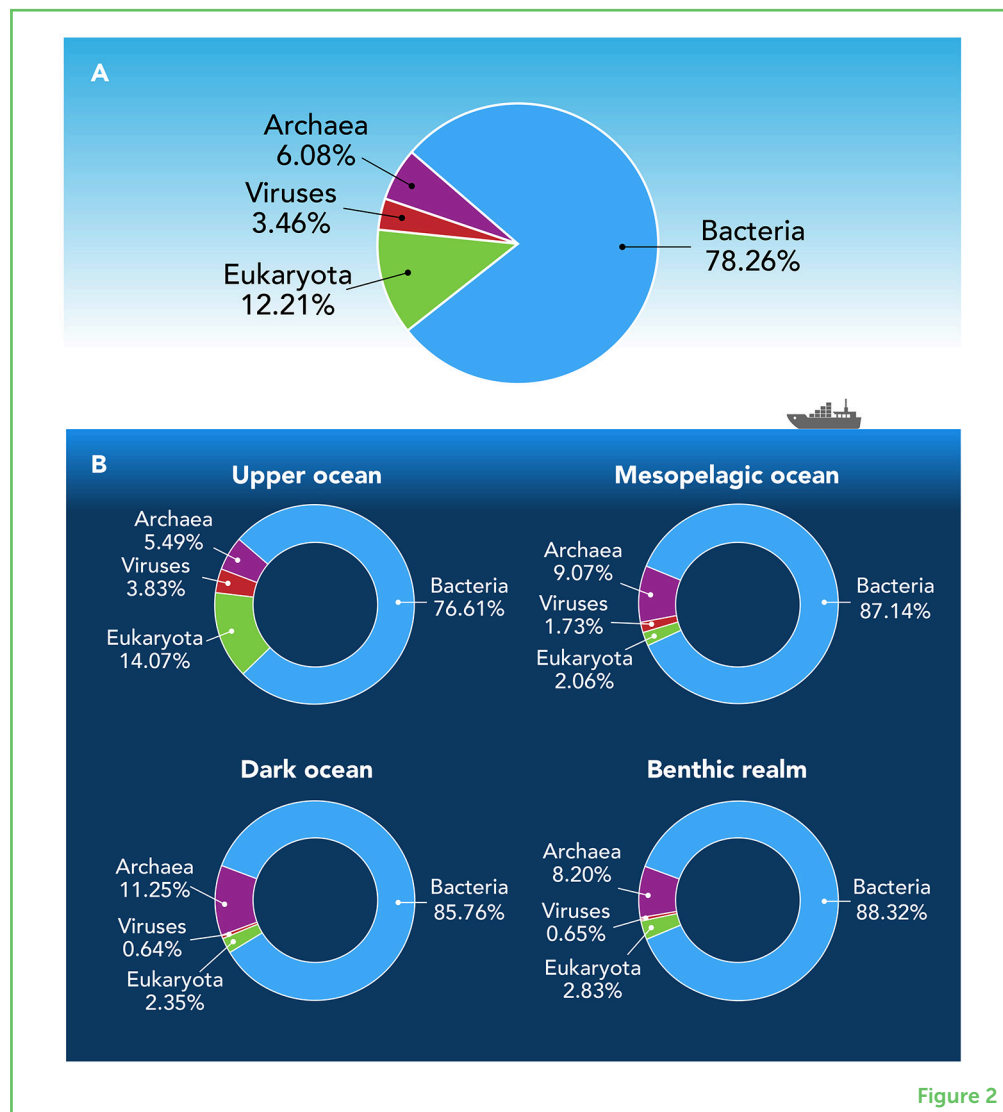
zones of the pelagic realm contained differing proportions of each organism type (Figure 2B). This is not surprising because conditions across depth zones vary—particularly in the amount of light that is available—creating a variety of unique ecological niches in which specific types of organisms can thrive.

MICROBE METABOLISM CAN INFLUENCE EARTH'S CLIMATE

Microbial metabolism keeps the oceans healthy by controlling the flow of nutrients and energy. As we mentioned earlier,

Figure 2

(A) Overall abundance of genes from the four major types of organisms across all ocean samples. (B) The abundance of genes from each type of organism differs between depth zones of the pelagic realm (upper, mesopelagic, and dark ocean) and the benthic realm. This tells us that each region of the ocean forms a unique habitat with specific growth conditions, in which certain organisms can thrive.

**Figure 2**

some microbial metabolic processes influence the cycling of elements—like carbon—that can influence Earth’s climate. Given the importance of microbial metabolism, we took an in-depth look at the metabolism-related genes in our catalog. In samples from the benthic realm, 41.3% of the annotated gene clusters were involved in metabolic processes, compared to just 25% of the gene clusters from the pelagic communities. This tells us that there is a lot going on in this commonly neglected and understudied region of the ocean!

When we sorted the annotated metabolism-related genes found across all ocean samples by the substances being metabolized, we found that that 50% of those genes took part in some form of carbon-related metabolism (Figure 3). Carbon-containing molecules, like carbon dioxide (CO₂) and methane (CH₄) are **greenhouse gases** that contribute to **global warming**. Some bacteria and algae can use photosynthesis to turn CO₂ into carbohydrates, with help from sunlight, thus removing CO₂ from the atmosphere. But photosynthesis is not the only pathway that ocean organisms use to metabolize

GREENHOUSE GASES

Gases in the Earth’s atmosphere that trap the sun’s heat, preventing it from escaping back into space.

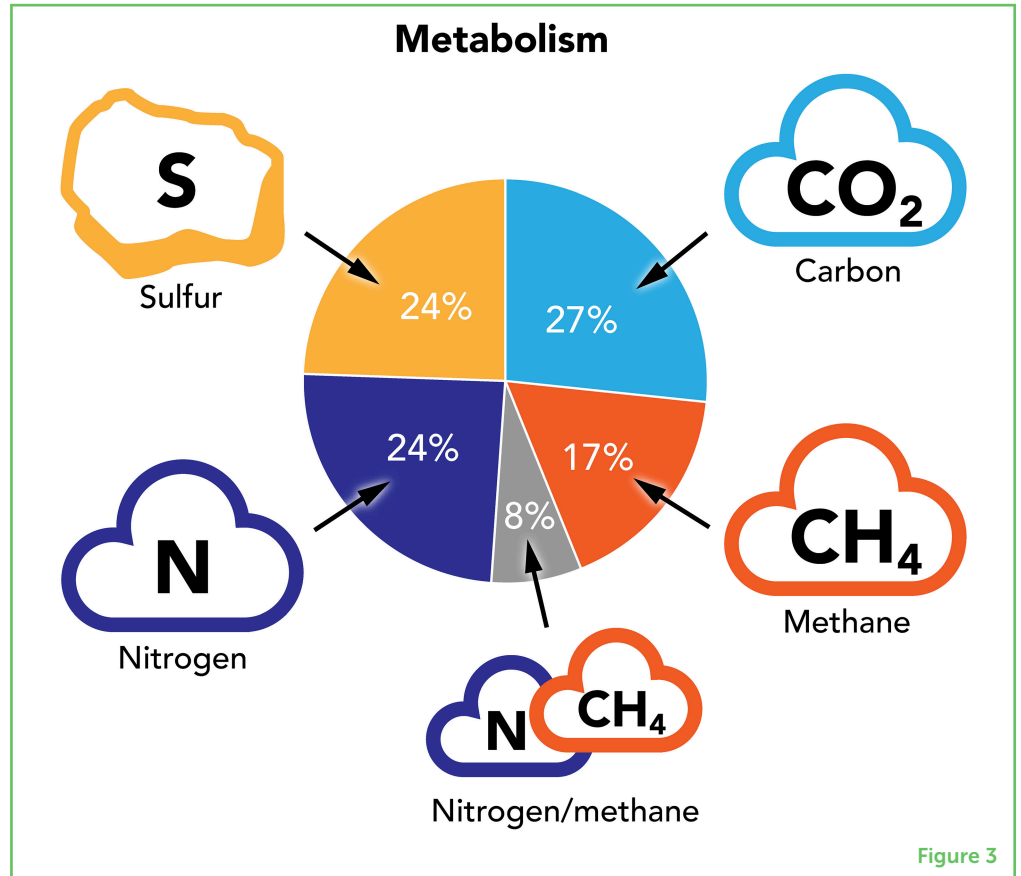
GLOBAL WARMING

The rising of Earth’s average temperature caused by greenhouse gases in the atmosphere.

carbon! Methane-metabolizing pathways, for example, do not require light and can function in the deep ocean and benthic realm, where light does not penetrate. Some of these pathways have been discovered quite recently and scientists still have much to learn about them.

Figure 3

Across all ocean samples, more than 50% of the annotated metabolism genes identified in our metagenomic analysis were related to some form of carbon metabolism (CO_2 or CH_4). Genes involved in the metabolism of nitrogen and sulfur were also present.



METAGENOMICS: THE FUTURE OF OCEAN RESEARCH?

The KMAP Ocean Gene Catalog 1.0 consists of approximately 163 million annotated gene clusters, providing us with an extraordinary amount of information about the types of organisms living at various depths of the ocean and the functions they perform. The global ocean genome is much more than just a simple catalog of the many organisms living there and their functions—it can deepen our understanding of the ocean and provide us with a “baseline” from which to evaluate the effects of human disturbances that can upset the ocean’s delicate balance, including pollution, overfishing, and ocean mining.

While sequencing and studying the global ocean genome has been occurring for some time, two factors have limited the size of those studies: technology and cost. Recent breakthroughs in sequencing technologies have resulted in systems that are faster, more

accurate, and portable [7]. Portability is especially important to marine scientists because portable systems can be used onboard oceangoing research vessels. These advances in sequencing technology have also dramatically reduced its cost. In the past, scientists had to ask smaller questions and carefully choose which sequencing experiments to perform because this research was extremely expensive. Today, modern, inexpensive sequencing technology allows relatively simple collection of genome data on an unprecedented scale. These data can help scientists answer questions about the vast array of ecosystem services that ocean organisms provide to keep our entire planet—and us—healthy.

Do *you* have questions about ocean organisms and their functions? Maybe future metagenomic studies of the global ocean genome will provide answers!

ACKNOWLEDGMENTS

This research was supported by King Abdullah University of Science and Technology through competitive center funding provided to the Computational Bioscience Research Center. Co-written by Susan Debad Ph.D., graduate of the University of Massachusetts Graduate School of Biomedical Sciences (USA) and scientific writer/editor at SJD Consulting, LLC.

ORIGINAL SOURCE ARTICLE

Laiolo, E., Alam, I., Uludag, M., Jamil, T., Agusti, S., Gojobori, T., et al. 2024. Metagenomic probing toward an atlas of the taxonomic and metabolic foundations of the global ocean genome. *Front. Sci.* 1:1038696. doi: 10.3389/fsci.2023.1038696

REFERENCES

1. Jaume, D., and Duarte, C. M. 2006. *General Aspects Concerning Marine and Terrestrial Biodiversity. The Exploration of Marine Biodiversity—Scientific and Technological Challenges*. Bilbao: Fundación BBVA.
2. Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74. doi: 10.1126/science.1093857
3. Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. 2007. The sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical pacific. *PLoS Biol.* 5:e77. doi: 10.1371/journal.pbio.0050077
4. Wikipedia. 2022. *Human Genome Project*. Available online at: https://en.wikipedia.org/wiki/Human_Genome_Project (accessed January 18, 2023).

5. Escobar-Zepeda, A., Godoy-Lozano, E. E., Raggi, L., Segovia, L., Merino, E., Gutiérrez-Rios, R. M., et al. 2018. Analysis of sequencing strategies and tools for taxonomic annotation: defining standards for progressive metagenomics. *Sci. Rep.* 8:12034. doi: 10.1038/s41598-018-30515-5
6. Kopf, A., Bicak, M., Kottmann, R., Schnetzer, J., Kostadinov, I., Lehmann, K., et al. 2015. The ocean sampling day consortium. *GigaSci.* 4:27. doi: 10.1186/s13742-015-0066-5
7. Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., et al. 2018. Evaluation of oxford nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Sci. Rep.* 8:10931. doi: 10.1038/s41598-018-29334-5

SUBMITTED: 23 September 2022; **ACCEPTED:** 06 January 2023;

PUBLISHED ONLINE: 16 January 2024.

EDITOR: [Pedro Morais](#), Florida International University, United States

SCIENCE MENTORS: [Gustavo Fonseca](#) and [Luisa I. Falcon](#)

CITATION: Laiolo E, Alam I, Uludag M, Jamil T, Agusti S, Gojobori T, Acinas SG, Gasol JM and Duarte CM (2024) The Global Ocean Genome: A "Catalog" of Ocean Life. *Front. Young Minds* 11:1052361. doi: 10.3389/frym.2023.1052361

CONFLICT OF INTEREST: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

COPYRIGHT © 2024 Laiolo, Alam, Uludag, Jamil, Agusti, Gojobori, Acinas, Gasol and Duarte. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

YOUNG REVIEWERS

IGOR, AGE: 12

I live in Brazil on the beach. I like playing with Lego, traveling, surfing, and I love to read. In the future, I will become an architect and I love science. In my house I have a huge yard with a tree house.



SEBASTIAN, AGE: 13

I like sports, reading, math, physics, and all things space!



AUTHORS



ELISA LAIOLO

Elisa is a Ph.D. student at King Abdullah University of Science and Technology (KAUST), under the supervision of Prof. Carlos M. Duarte. She earned her B.Sc. in biological sciences at Marche Polytechnic University, Ancona, Italy in 2018, and her M.Sc. in marine biology at the same university in 2020. During her B.Sc. and M.Sc. studies, Elisa started using metagenomics to investigate the patterns of microbial diversity in a broad range of aquatic environments and to understand the functions of those microbes, both in terms of their metabolism and the substances they produce. Microbial ecology has always been Elisa's main field of interest and her current studies focus on using metagenomics to study the functions of microbial communities in the sediment, with particular attention to coastal blue carbon habitats—ecosystems that can store carbon within their plants and sediments.

*elisa.laiolo@kaust.edu.sa



INTIKHAB ALAM

Intikhab has been a computational biologist at the Computational Bioscience Research Center (CBRC) in King Abdullah University of Science and Technology since 2010. He completed his Ph.D. in computational biology (bioinformatics) from the University of Bielefeld, Germany in 2004, and earned his Master's of Philosophy degree in molecular biology from the Center of Excellence in Molecular Biology, University of Punjab, Pakistan. Intikhab develops ways to assign taxonomic and functional information to genes from new genomes or metagenomes, to compare them with reference databases using supercomputers such as Shaheen and its associated computer clusters. He is developing metagenomic gene catalogs from publicly available metagenomics data.



MAHMUT ULUDAG

Mahmut Uludag is a software developer at King Abdullah University of Science and Technology. He is interested in improving accessibility and interpretation of research results using open-source software technologies, such as the Solr search platform and the Weka machine-learning library.



TAHIRA JAMIL

Tahira Jamil is a researcher at King Abdullah University of Science and Technology (KAUST). She has a passion for data science and analyzing big data. While working in Amsterdam, she developed and implemented the code for Bayesian models for qualitative data in open-source statistical software packages. At KAUST, she has worked on various projects analyzing environmental and biological data. She analyzed the effect of multiple stressors on the Red Sea coral reef ecosystem. She uses marine heat waves as a novel approach for detecting coral reef zones that are susceptible to bleaching, using the Red Sea as a model system. She has worked on sequencing data from genomes, collaborated on the global biosphere gene catalog, and performed metagenomic analysis on SARS-CoV-2, plastic-degrading enzymes, and antibiotic-resistance genes.



SUSANA AGUSTI

Susana Agusti is a full professor of marine science at King Abdullah University of Science and Technology, in the Red Sea Research Center (Saudi Arabia). She was formerly a professor at UWA Oceans Institute and School of Plant Biology and the Spanish National Research Council, and an adjunct professor at the University of Tromsø (Norway). She holds Bachelor's and Ph.D. degrees from the Autonomous University of Madrid. She has participated in oceanographic expeditions in the polar Arctic and Southern (Antarctic) Oceans, as well as the Atlantic, Indian, and Pacific Oceans. She also studies the Red Sea, one of the warmest seas on earth. Prof. Agusti studies photosynthetic plankton and the metabolic balance of the oceans, with the goal of evaluating the impact of climate change on the oceans. She also studies the sensitivity of marine organisms to increased water temperature and other stressors, such as UV-B radiation, persistent pollutants, and combinations of stressors.



TAKASHI GOJOBORI

Professor Gojobori is a distinguished professor at King Abdullah University of Science and Technology. His research interests range from comparative genomics and gene expression of neural cells in sensory organs of various organisms to the marine metagenomics of a diversity of microorganisms. He aims to elucidate the evolutionary origin of the genome and how it can be used for the development of bioenergy. He holds a number of prestigious academy memberships.



SILVIA G. ACINAS

Silvia G. Acinas is a research scientist at the Institut de Ciències del Mar in Barcelona. She loves working with the smallest living microbes on Earth, the prokaryotes (bacteria and archaea). She earned her Ph.D. in Alicante, Spain, looking at the diversity of bacteria and archaea in marine ecosystems. After her Ph.D., Silvia moved to the US for 7 years, five of which were spent working at the Massachusetts Institute of Technology, where she learnt about how different microbial genomes can be, even between closely related species. This spurred her fascination with the evolutionary mechanisms driving speciation in microbes. She is now working on understanding the evolution and genetic and functional diversity of bacteria in polar marine ecosystems, which are threatened by climate change.



JOSEP M. GASOL

Josep M. Gasol is a research professor at the Institut de Ciències del Mar in Barcelona. He became fascinated by the microbial inhabitants of aquatic environments when studying biology, and started working on a tiny (20 m across!) red lake in Catalonia. This was a bit too small, so he switched to many lakes around Montreal and southern Quebec during his postdoctoral work. When even those became too small, he decided that the ocean was largest ecosystem on Earth and became fascinated with ocean microbes. He still likes freshwater environments, and has also worked on mountain and saline lakes, small and large, in South America, Africa, and the Pyrenees. Pep is now interested in groundwater at the land-ocean interface. His ocean work combines research on a long-term coastal site (the Blanes Bay Microbial Observatory) and large oceanographic cruises (such as the Malaspina), and his experimental approaches range from observational to experimental. He also likes watching birds—but not doing research on them!

**CARLOS M. DUARTE**

Carlos is a distinguished professor at King Abdullah University of Science and Technology. After nearly four decades documenting how human pressures impact marine organisms, Carlos hopes to spearhead a global scientific effort to rebuild the abundance of marine life. His research is characterized by independence, creativity, serendipity, and linking of multiple disciplines, as well as the capacity to organize and collaborate with large, interdisciplinary teams. He loves dogs and enjoys being out on the open sea, reading, swimming, snorkeling, walking, and playing with his grandson, Oliver.