Initiation of a comparative metagenomic study of the Red Sea and Pacific Ocean marine microbiomes

Shugo Watabe1, Kazuho Ikeo2, Takahisa Mori3, Katsumiho Mineta4, Intikhab Alam5, Rintanis Kodzius6, John A. C. Archer7, Vladimir B. Bajic8, Takashi Gojobori9

1 Department of Marine Biosciences, Kaisei University, Kanagawa, Japan
2 Division of Biotechnology, Hokkaido University, Hokkaido, Japan
3 Division of Biotechnology and Biometric and Bioinformatics, Hokkaido University, Hokkaido, Japan
4 Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology

Introduction

The marine microbiome is a fundamental component of the biosphere. Its bacteria are abundant and play critical roles within the ocean environment. The majority of this important group of bacteria are genetically uncharacterized. Relatively few species have been studied in the laboratory. However, by applying metagenomic analyses to marine microbial populations, genomic ‘snapshots’ may be taken and from appropriate time series experiments their dynamics established. As a key component of the CBRC Centre Research Program (2014-2020), we are initiating a comparative study of the Red Sea and North Eastern Japanese coastal and bay complexes (Figure 1). These environments differ in physical characteristics significantly. The Red Sea exhibits consistently high salinity, temperature and insolation characteristics (Figure 2), whereas the Japanese waters are less saline, cooler and receive lower insolation. Here, we present initial data and analytical pipelines for Phase 1 of our collaborative research program.

Study areas

The Japanese study area is located on the north east Pacific side of Japan. The water offshore area of north-east Japan is hydrographically divided into three areas: the Kurishio water (KY) area, the Oyashio water (OW) area, and the Kurishio-Oyashio mixed water region (KOMWR). These areas comprised both deep ocean (Sendsi), as well as more shallow inland areas (Seto Inland Sea). By comparison with the Red Sea, these areas exhibit a more constant moderate salinity (30-34 p.s.u.). However, these areas experience a wide range of temperature and insolation variations. August temperatures in the north-east Pacific, surrounding Japan, range from 26°C to 29°C and February temperatures range from 16°C to 23°C generating a broad variation in physical characteristics. The warm Kurishio ocean current heats the south coast of Honshu, Kyushu, Shikoku. In the east of Honshu, East China Sea, and the Sea of Japan Chubu, water temperature varies greatly in the east-west direction relative to Japan. Sea surface temperature in August rises to 25°C or more in the south of the southern coast of Honshu. In addition, due to the influence of cold water Oyashio current, sea surface temperature in the east of Honshu is lower than that of the Sea of Japan side of the Tsushima Current. In the vicinity of the Kuril Islands, because of mixing with deep cold water below sea level by the Kuroshio, water temperature is low (Figure 2a).

By contrast to the Japanese seas, the Red Sea is a desert-bound ocean whose oligotrophic waters receive nutrients only seasonally from the Indian Ocean. Very high surface evaporation in the northern Red Sea, estimated at 2.0 m yr$^{-1}$, has established a salinity gradient ranging from 34.0 p.s.u. in the south to 41.0 p.s.u. in the north, as compared to 33 p.s.u. to 34 p.s.u. in the Japanese test areas. Airing from this, the usual decreases in water temperature with depth is not seen in the Red Sea; below the strongly heated 28°C to 32°C surface layer, a uniform temperature of 21°C is maintained through the water-column thus forming the warmest and most saline deep-water body on Earth (Figure 2b).

Sampling regimens and data generation

Microbial populations will be sampled using boat-based repeat sampling of the euphotic zone along a series of transects in Japanese coastal and Red Sea areas (Figures 3a and 3b) and supported by chemical and weather data for each sample. Longitudinal surveys in year 1 will establish a baseline for further more detailed sampling and analysis in subsequent years, allowing us to focus on key environmental indicator groups/metabolism. RNA sequencing will be carried out using Illumina Hi-Seq technology, which generates very high coverage data, albeit at a reduced sequence length of <105 bp. To support metagenomic assembly, longer read data will be generated using Illumina Mi-Seq technology. Should more advanced sequencing technology arise, these will be incorporated into the primary data generation pipeline. For selected sample types and locations in later phases of the research, transcriptomic metagenome data will be obtained to capture highly expressed genes.

Analytical pipeline

We use two complementary computational analytical pipelines. During the course of the collaboration, annotations resulting from the use of these pipelines will be integrated to provide comprehensive information for modelling microbial population diversity and metabolic capabilities of microbial communities. The workflows, presented in Figure 4, describe the data analysis pipeline and datasets (NCBI, ESpiDOG, InterProScan, MEGAN, EC, TCDB and SILVA) interrogated to generate a consistent framework for comparison. These uniformly annotated data will form a common marine microbial comparative dataset that captures microbial taxonomic diversity, as well as, metabolic and transport functions of microbial populations from these environments.

Acknowledgements

This program is generously supported by the KAUST Center Competitive Centre Funding program 2014-2020 awarded to the CBRC consortium.