Bioinformatics Center
– Chemical Life Science –

https://cls.kuicr.kyoto-u.ac.jp/en/

Prof
OGATA, Hiroyuki
(D Sc)

Assist Prof
BLANC-MATHIEU, Romain
(D Sc)

Assist Prof
ENDO, Hisashi
(D Environmental Science)

Researcher (pt)
YAMAMOTO, Rumiko

Students
YOSHIKAWA, Genki (D3)
NISHIYAMA, Hiroki (D3)
PRODINGER, Florian (D3)
KANEKO, Hiroto (D1)
MENG, Lingjie (D1)
XIA, Jun (D1)
YOSHIDA, Koki (M2)
BAN, Hiroki (M1)
GARCIA, Kimberly Gacayon (RS)
ZHANG, Ruixuan (RS)
SUN, Weiyou (RS)
FANG, Yue (RS)
MIYAZAKI, Urara (UG)
HASHIMOTO, Kentaro (UG)

Guest Res Assoc
GRIMSLEY, Nigel (Ph D) Sorbonne University, France, 21–26 July
DELMONT, Tom (Ph D) Centre National de la Recherche Scientifique, France, 31 August–24 September

Scope of Research

We are interested in understanding the functioning and evolution of biological systems at varying scales from tiny microbes up to the Earth's environment, by leveraging rapidly accumulating big data in life science and bioinformatics approaches. We currently focus on 1) the evolution of viruses and their links to the origin of life, 2) microbial ecology in different ecosystems, and 3) the development of bioinformatics methods and biological knowledge resources for biomedical and industrial applications. To fuel these research activities, we take part in environmental sampling campaigns such as *Tara* Oceans. Our resources and developed tools are accessible through GenomeNet (www.genome.jp) to scientific communities and the public.

KEYWORDS
GenomeNet
Bioinformatics
Environmental Genomics
Virology
Molecular Evolution

Selected Publications

Aramaki, T.; Blanc-Mathieu, R.; Endo, H.; Ohkubo, K.; Kanehisa, M.; Goto, S.; Ogata, H., KofamKOALA: KEGG Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold, *Bioinformatics*, **btz859,** doi: 10.1093/bioinformatics/btz859 (2019).

Ibarbalz, F. M.; Henry, N.; Brandão, M. C.; Martini, S.; Busseni, G.; Byrne, H.; Coelho, L. P.; Endo, H.; Gasol, J. M.; Gregory, A. C.; Mahé, F.; Rigonato, J.; Royo-Llonch, M.; Salazar, G.; Sanz-Sáez, I.; Scalco, E.; Soviadan, D.; Zayed, A. A.; Zingone, A.; Labadie, K.; Ferland, J.; Marec, C.; Kandels, S.; Picheral, M.; Dimier, C.; Poulain, J.; Pisarev, S.; Carmichael, M.; Pesant, S.; Tara Oceans Coordinators; Babin, M.; Boss, E.; Iudicone, D.; Jaillon, O.; Acinas, S. G.; Ogata, H.; Pelletier, E.; Stemmann, L.; Sullivan, M. B.; Sunagawa, S.; Bopp, L.; de Vargas, C.; Karp-Boss, L.; Wincker, P.; Lombard, F.; Bowler, C.; Zinger, L., Global Trends in Marine Plankton Diversity Across Kingdoms of Life, *Cell*, **179,** 1084-1097 (2019).

Endo, H.; Suzuki, K., Spatial Variations in Community Structure of Haptophytes Across the Kuroshio front in the Tokara Strait, *Nagai, T., Saito, H., Suzuki, K., Takahashi, M.(eds.), Kuroshio Current: Physical, Biogeochemical and Ecosystem Dynamics, AGU Geophysical Monograph Series, AGU-Wiley*, 207-221 (2019).

Li, Y.; Endo, H.; Gotoh, Y.; Watai, H.; Ogawa, N.; Blanc-Mathieu, R.; Yoshida, T.; Ogata, H., The Earth is Small for "leviathans": Long Distance Dispersal of Giant Viruses Across Aquatic Environments, *Microbes Environ.*, **34,** 334-339 (2019).

Yoshikawa, G.; Blanc-Mathieu, R.; Song, C.; Kayama, Y.; Mochizuki, T.; Murata, K.; Ogata, H.; Takemura, M., Medusavirus, A Novel Large DNA Virus Discovered from Hot Spring Water, *J. Virol.*, **93,** e02130-18 (2019).

Comparing Two Megaviridae Communities Using Meta-barcode During a Red Tide Period in an Enclosed Bay

"Megaviridae" is a proposed family of giant viruses infecting unicellular eukaryotes. They are very abundant and ubiquitous in the sea water and affect marine microbial community by their lytic infection cycle. However, only a few Megaviridae members have been explored in the marine environments, and their ecological roles in marine ecosystems remains unknow.

To study the differences of Megaviridae community in two different kind of water, surface sea water samples (0.22–3 μm) were collected from an enclosed bay (Uranouchi Inret, Kochi) during a red tide period. A set of degenerated primers (called "MEGAPRIMER"), which target family B DNA polymerase genes of Megaviridae, was used to assess the composition of Megaviridae community. Bioinformatic analysis was applied after sequencing. As a result, hundreds of operational taxonomic units (OTUs) were identified in each sample. When we compared the Megaviridae structures between within and without red-tide waters, there was a clear difference between these two samples in a community variation analysis. The difference was also visible in dominant OTU proportions (Figure 1) between the two types of samples. Phylogenetic analysis based on a maximum-likelihood method showed that almost all of the OTUs belong to unknown Megaviridae branches (Figure 1). Future study will mainly focus on the host-virus interactions and the relationship between viral community and environmental factors, that may reveal ecological functions of Megaviridae in a coastal ecosystem.

An Assessment of NCLDV Host Prediction Based on Co-occurrence Analysis

Nucleocytoplasmic large DNA viruses (NCLDVs) form a monophyletic group of viruses and our understanding of these virus-host systems is limited. Co-culture with their host is the "Gold Standard" for the identification of virus-host relationship but limited by our ability to cultivate microorganisms (thus hosts). Therefore, consideration should be put into cultivation-independent approaches.

In our research, we used deeply sequenced metagenomic and amplicon data generated by the *Tara* Oceans project to examine if the use of co-occurrence approaches can provide virus-host relationship for this group of viruses. Samples were collected from the surface (SRF) and deep chlorophyll maximum (DCM) layers of global oceans. The abundance matrices of NCLDVs and eukaryotes were constructed from pico-size fraction (0.22–1.6 or 0.22–3.0 μm) and four larger size fractions (0.8–5, 3–20_5–20, 20–180 and 180–2000 μm), respectively. Co-occurrence analysis was carried out by Flashweave. As a result, we got a large number of associations between NCLDVs and eukaryotes (Figure 2). However, because of the deficiency of information about known NCLDVs, it will be necessary to quantitatively evaluate the FlashWeave host-virus relationship predictions. Furthermore, the prediction of the NCLDV-host pairs by abundance-based methods may be improved by combining with other data, such as phylogenetic relationships among various viruses. These bioinformatic tools eventually will help us to understand the natural diversity, life cycle, interactions and co-evolution of NCLDVs and their host.

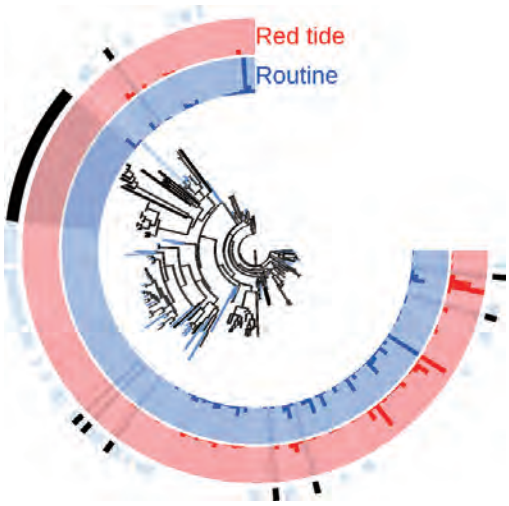

Figure 1. Phylogenetic tree of Megaviridae OTUs (97% identity) based on ML method. Relative abundances of Megaviridae in routine and red-tide samples were also shown in blue and red colors, respectively. Reference sequences-black; other pelagic sea water sequences-light blue.

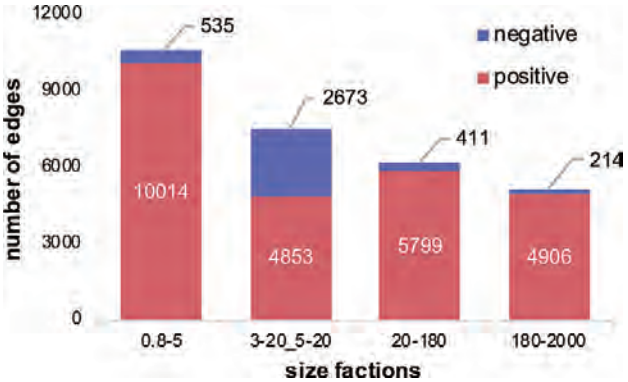

Figure 2. Number of edges between NCLDVs and eukaryotes in inferred network by size fraction. Red: edges with positive weight; Blue: edges with negative weight.