Bioinformatics Center
– Chemical Life Science –

http://cls.kuicr.kyoto-u.ac.jp/

Prof
OGATA, Hiroyuki
(D Sc)

Assoc Prof
GOTO, Susumu
(D Eng)

Program-Specific Assist Professors

TOKIMATSU, Toshiaki (D Agr)
KOTERA, Masaaki (D Sc)

Program-Specific Researchers

MORIYA, Yuki
NAKAGAWA, Zenichi

Researcher(pt)

SHIMIZU, Yugo

Students

MIZUTANI, Sayaka (D3)
FUJITA, Megumi (D3)
NISHIMURA, Yosuke (D3)
MIHARA, Tomoko (D3)

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 integrated bioinformatics approaches. We currently focus on 1) the evolution of giant DNA viruses and their links to the origin of life, 2) the ecological and evolutionary significance of microorganisms in the oceanic ecosystem, and 3) the development of new bioinformatics methods and biological knowledge resources for biomedical and industrial applications. To fuel these research activities, we take part in several environmental sampling campaigns such as the *Tara* Oceans project. Our resource and methodological developments are accessible from the GenomeNet service to scientific communities and public.

KEYWORDS

GenomeNet
Bioinformatics
(Meta)genomics
Evolutionary Biology
Pharmacoinformatics

Selected Publications

von Dassow, P.; John, U.; Ogata, H.; Probert, I.; Bendif, E. M.; Kegel, J. U.; Audic, S.; Wincker, P.; Da Silva, C.; Claverie, J.-M.; Doney, S.; Glover, D. M.; Flores, D. M.; Herrera, Y.; Lescot, M.; Garet-Delmas, M.-J.; de Vargas, C., Life Cycle Modification in Open Oceans Accounts for Genome Variability in a Cosmopolitan Phytoplankton, *ISME J.*, doi: 10.1038/ismej.2014.221. (2014).

Wilson, W. H.; Gilg, I. C.; Duarte, A.; Ogata, H., Development of DNA Mismatch Repair Gene, MutS, as a Diagnostic Marker for Detection and Phylogenetic Analysis of Algal Megaviruses, *Virology*, doi, 10.1016/j.virol.2014.07.001. (2014).

Kotera, M.; Tabei, Y.; Yamanishi, Y.; Muto, A.; Moriya, Y.; Tokimatsu, T.; Goto, S., Metabolome-scale Prediction of Intermediate Compounds in Multistep Metabolic Pathways with a Recursive Supervised Approach, *Bioinformatics*, **30,** i165-i174 (2014).

Yamanishi, Y.; Kotera, M.; Moriya, Y.; Sawada, R.; Kanehisa, M.; Goto, S., DINIES: Drug-target Interaction Network Inference Engine Based on Supervised Analysis, *Nucleic Acids Res.*, **42,** W39-W45 (2014).

Mizutani, S.; Noro, Y.; Kotera, M.; Goto, S., Pharmacoepidemiological Characterization of Drug-induced Adverse Reaction Clusters towards Understanding of Their Mechanisms, *Comput. Biol. Chem.*, **50,** 50-59 (2014).

Functional and Evolutionary Analysis of Intragenic miRNAs in Animals

MicroRNAs (miRNAs) are short single-stranded non-coding RNAs present in diverse organisms and suppress the expression of several target genes by binding to the 3'UTR of their mRNAs. They are key regulators in gene expression networks and have influence on various genetic pathway and pathologic states in multiple diseases. Expanding repertoire of miRNAs and variation of their target sites in animal evolution are associated with major body-plan innovations. However, individual functions of most miRNAs still remain unclear.

In animal genomes, miRNAs are enriched in protein coding regions (genes), several of which are evolutionary conserved. The overlapping miRNA and gene are called as an intragenic miRNA and its host gene, respectively, and not a small number of them are transcribed as the same transcription unit, which is also indicated by a correlation of their expression levels. Thus, we hypothesized that each intragenic miRNA has a functional association with its host gene. In this study, we first collected miRNA data from a database and predicted their targets, and then performed statistical tests on two types of regulation model (Figure 1); *A. direct regulation model*, where a miRNA regulates its host gene directly, and *B. functional regulation model*, where a miRNA regulates a set of genes with the related function to its host gene. As a result, 34.9% of intragenic miRNAs in animals turned out to follow these two models.

Emergence of widely conserved miRNAs and their conserved target sites is important events in animal evolution. We detected 12 intragenic miRNA families that are specifically conserved among mammals to reveal evolutionary origin of conserved target sites (on-going, Figure 2).

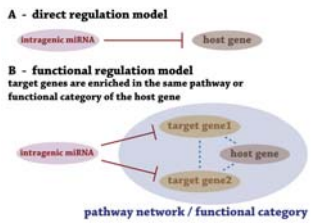

Figure 1. Two regulation models of intragenic miRNAs.

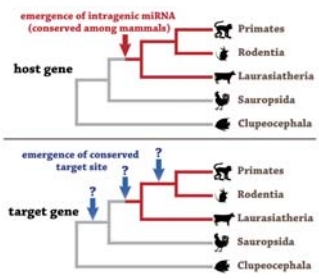

Figure 2. Emergence period of conserved target site of intragenic miRNAs that are specifically conserved among mammals.

Phylogenomics of Virus-encoded Amino Acid Metabolic Enzymes

Viruses do not encode their own translation machinery and thus depend on their hosts regarding the metabolism for protein synthesis. However, certain viruses are known to encode a few amino acid (AA) biosynthetic enzymes in their genomes. Curiously, these viral enzymes correspond to only a part of entire AA metabolic pathways, while many other enzymes are likely to be required (and supplied by their hosts) for a complete metabolism. To better understand why these viruses encode such specific AA enzymatic reactions and to clarify their evolutionary origins, we systematically identified virus-encoded AA enzymes in complete viral genomes and performed comparative genomics and phylogenetic analyses.

We searched complete viral genomes (200 bp to 2.5 Mbp) for AA metabolic genes using KAAS (KEGG Automatic Annotation Server) with cellular homologs as queries and identified 147 viral genes (32 KEGG orthologs involved in nine KEGG AA pathways) in 69 viral genomes. These viral genomes were relatively large (32 kbp to 2.5 Mbp), and the encoded viral enzymes were more widely distributed across different cellular organisms than other AA enzymes (p<0.007). Phylogenetic analyses of one of the viral ortholog groups indicate that viral homologs are distributed in two clades, being relatively similar to eukaryotic homologs than to prokaryotic ones with no clear phylogenetic affinity to the homologs from their hosts (Figure 3). We are currently undertaking network analyses and metagenomic (virome) data mining to investigate the role of these viral enzymes and to characterize the extent of this phenomenon in natural samples, eventually to get better insight into viral strategies in the control of amino acid metabolisms for their survival.

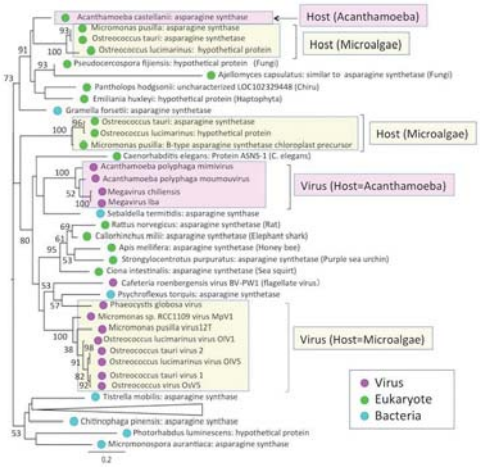

Figure 3. Maximum likelihood phylogenetic tree of viral asparagine synthases and their homologs. Thirteen viruses infecting eukaryotes including Mimivirus encode an asparagine synthase. Bootstrap values greater than 50% are shown along the branches.