The 50\textsuperscript{th} Anniversary of World Data Center for Microorganisms

“Dynamic Changes of Marine Microbial Communities over Time”

16 September, 2016

Grand Building, Beijing Friendship Hotel, Beijing

Takashi Gojobori

KAUST

(King Abdullah University of Science and Technology), Saudi Arabia

National Institute of Genetics, Japan

The University of Tokyo, Japan

Waseda University, Japan
Big Data

~ Key Point ~

Monitoring (time and space) is a Key!
~Marine Meta-Genomics~

[写真] Apollo 17号からみた地球 (NASA)
—A View of marine micro-organism ecosystem—

Tohoku sea coast ➔ One of species-richest spots

Foods chain

Large plant planktons ➔ Animal planktons ➔ Fish

Micro plant planktons

Soluble organiuics

Protozoa

Micro-organism loop

Bacteria

Distribution of Chlorophyll a
In Tohoku sea coast
Microorganisms play important roles in cycling of materials in marine environments.
Metagenome analysis will become a useful tool for monitoring of biological diversity.

Bacteria in the sea are abundant (ca. $10^9$ cells/L) and highly diverse. They grow faster and respond to environmental changes.
Environmental conditions fluctuate widely in coastal areas.

Temporal discharges

Phytoplankton blooming

Water currents
How and Why do microbial communities change?

OBJECTIVES

Clarify the spatial and temporal variation of microbial community at coastal areas in relation to physical and biological environmental factors.
Observation Points at the Sendai Bay and Nemuro Sea as a Control

Observation Points (Sendai Bay) A-line (Near Sea)
Sea Water Sampling Points in Kyushu for "Red Tide" Monitoring

Comparative Metagenomics for Detecting any changes in marine microorganism diversity

Periodical Sampling

Red Tide Occurrence Sea Area

Yatsushiro Bay

Red Tide Non-Occurrence Sea Area

ShiBushi Bay
―Sea water sample and DNA extraction―

Flow of DNA extraction

Sea water sample

Filter selection

Recovery of Filtered liquid

Filtering

DNA extraction

Sea water samples A, B Each 2 litres

Large planktons

10 µm Filter

DNA
A: 30 ng
B: 100 ng

2 µm Filter

DNA
A: 400 ng
B: 300 ng

0.2 µm Filter

DNA
A: 1500 ng
B: 400 ng

NGSSequencing
（約400種類の微生物ゲノムが含まれることがわかった）
Assembling analysis (Digital DNA Chip Analysis)

Sequence data → Fragmentation → Grouping → Assembling → Digital Probe → Digital Hybridization

Contig

Operational Taxonomic Unit (OTU) → Community Composition Data
3. NGS Data analysis

Bacterial community at Sendai bay in May, 2014

- Proteobacteria: 70.3%
- Bacteroidetes: 19.9%
- Actinobacteria: 7.5%

(National institute of genetics)
3. NGS Data analysis

Annual bacterial community at the P1 observation point

<table>
<thead>
<tr>
<th>Depth (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
</tbody>
</table>

Sea bottom

- Spring
- Summer
- Autumn
- Winter

- Actinobacteria <phylum>
- Aquificae <phylum>
- Bacteroidetes/Chlorobi group
- Chlamydiae/Verrucomicrobia group
- Chloroflexi <phylum>
- Cyanobacteria
- Deferrribacteres <phylum>
- Deinococcus-Thermus
- environmental samples <Bacteria>
- Fibrobacteres/ Acidobacteria group
- Firmicutes
- Fusobacteria
- Gemmatimonadetes
- Nitrospirae
- Planctomycetes
- Proteobacteria
- Spirochaetes
- Synergistetes
- Tenericutes
- Thermodesulfobacteria <phylum>
- Thermotogae <phylum>
- unclassified Bacteria
### Statistics of Samples  （As of 28 August, 2016）

<table>
<thead>
<tr>
<th>No. of Samples</th>
<th>No. of Microbial Fraction</th>
<th>No. of DNA Extraction</th>
<th>Nucleotide Sequence Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>139</td>
<td>573</td>
<td>573</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.6 billion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>330 billion</td>
</tr>
</tbody>
</table>

We have another 500 billion nucleotide sequence data for microbes in the red tide preventive project, Japan (As of 31 March, 2016)
Marine Metagenome Database

Microbial Composition
Spatial Variation

Bacterial community profiling using the 16S rRNA gene PCR amplicons (Terminal–Restriction digestion Fragment Length Polymorphism)
Sampling Points
(Apr, Jul, Sep, Dec 2012)
April 2012

Estuary circulation

Oceanic water intrusion
Chl α

Phytoplankton Sedimentation

At the expiry of spring bloom

Oceanic water intrusion
Bacterial Community Composition

Principal Component 1 (32%)

Principal Component 2 (22%)
Oceanic water intrusion

Phytoplankton Sedimentation

Oceanic water intrusion
December 2012

Vertical mixing season

PC1 (35%)

PC2 (15%)

Temp

Chla
Environmental factors which showed a significant correlation with the principal component scores of the bacterial community compositions (p<0.05)

Spatial variation of a bacterial community was often coordinated with the hydrographic profiles

<table>
<thead>
<tr>
<th></th>
<th>April</th>
<th>Jun</th>
<th>September</th>
<th>December</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Salinity</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td><strong>Phytoplankton</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll a</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>Nutrients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrate</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Nitrite</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ammonia</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Phosphate</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Silicate</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>
Temporal variation

Shotgun metagenome
&
16S rRNA gene amplicon
Sampling points

C12

C05

Temp

Sal

Chl $\alpha$

Sampling

2012 Mar ~ 2014 Apr
contig_312525 Flavobacteriaceae

contig_772152 Alphaproteobacteria

Sampling time (2012 Mar - 2014 Apr)

Indicating a dynamics of a microorganism
Microbial community have an obvious seasonal variation.
Digital DNA chip analysis of shotgun metagenomes
Clustering using correlation coefficients

- September - January
- June - August
- February - May
Dec 1\textsuperscript{st}, 2013, Sendai Bay Locality of C5 0.2-0.8μm Fraction

DNA 0.2μm Markers

December of Sendai Bay Markers
Digital DNA Chip

Marine Genomic Identification of Seawater samples

Illustration cited: http://cdn.amanaimages.com/cen3tzG4fTr7Gtw1PoeRer/22451009380.jpg
**Sampling device with a monitoring system (Kyushu U.)**

自立式海洋微生物DNA採取装置

装置製作条件
表面から10mまでの柱状サンプル(2.5L程度)を1日に1回採取し、既定のフィルターでろ過を行う。
これを1週間行った後、サンプルとしてフィルター(7個)を回収する。
電源に関しては、今回は太陽電池等を用いないため動作に必要な電源を、鉛蓄電池で確保する。

卷取ドラム仕様

<table>
<thead>
<tr>
<th>駆動部</th>
<th>ギヤモータ(12V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>減速比</td>
<td>1/180</td>
</tr>
<tr>
<td>回転数</td>
<td>30.8rpm</td>
</tr>
<tr>
<td>トルク</td>
<td>35kgf+cm</td>
</tr>
<tr>
<td></td>
<td>さらに平歯車で1/6減速</td>
</tr>
<tr>
<td></td>
<td>ドラムの回転数 5.13rpm</td>
</tr>
<tr>
<td></td>
<td>トルク 2.1kgf+cm</td>
</tr>
<tr>
<td></td>
<td>ホースの移動速度 92mm/s</td>
</tr>
</tbody>
</table>
Summary

• The changes in the bacterial community spatially and temporally correspond to the hydrographic features of the environment.

• Spring phytoplankton bloom and collapse of stratification of seawater greatly affect the annual variation of the bacterial community.

• The most dominant operational taxonomic unit is an oceanic bacterial taxon, and its abundance varies irrespective of environmental factors, suggesting that seawater intrusion from outside of the bay may strongly influence composition of the bacterial community in the bay.
Comparative META-genomics among different marine resources

Red Sea

Sea of Japan/Pacific Ocean

Environmental differences
Diversity of micro organisms
Red Sea

- The Red Sea is the world's northernmost tropical sea, positioned between Africa and Asia.
- The Red Sea is 360 km at its widest point and 1,930 km in length.
- Molecular and genomic-basis analyses are limited.
Red Sea (Temperature and Salinity)

From World Ocean Atlas 2009
Objective

• To understand the microbial diversity and its functions in the community over time by the time-series sampling

• To see the dynamics of microbial communities (1) between the Red Sea and the other locations such as in Japan, and (2) between seasons

• To find the novel genes for the industrial application
Overview of entire project

Samples / Extraction / Sequencing

In silico Analysis

Bacterial Metabolite Functions

- Sea Water Samples
- DNA Extraction
- DNA Sequencing
- Database Search
- Information Integration
  - Biological Network

- Functional Prediction
- Genomic Mining

Database construction

- Bacterial Fermentation
- Metabolite database Screening
  - Compound Evaluation
- Genome Modification
  - Biosynthetic pathway Analysis

Judgement of Usefulness

Outputs

- Novel Drugs
- Health care product
- Animal Feed
- Energy source

Triangle Value

Value / Unit

High

Low

Small

Large

Quantity

Overview of entire project
Sampling location in the Red sea

- Every month from Nov 2014
- 1 site (ca 25m depth)
- Seawater from 3 depths (surface, 10m and 20m)
- Sediment
- Chemicals Properties
From Sample collection to NGS

Seawater and sediment collection at the site in the Red sea

Filtration of seawater

DNA extraction by MO BIO kit

NGS library construction by Nextera XT

MiSeq
Outline of analytical pipeline

For multiple comparison, we first created a gene catalog based on all samples. Then, each read was mapped onto this gene catalog.

DMAP (Alam et al., unpublished).
INDIGO (Alam et al., 2013 PlosOne).

Methods
Data set

• Monthly samples (Feb, Mar, May, Jun, Jul, Aug, Sep, Oct, Nov 2015: 9 months)
• Seawater from surface (1m) and 10m depths
• Sediment (Mar 2015)
• Filtration (0.22 and 0.8μm)
• Whole genome shotgun NGS
Overview: Taxonomic composition of the Red sea Metagenomes

Total reads from 16 monthly samples: 67,531,568
Total number of contigs with genes as a gene catalog: 640,286

Dominant category is Bacteria. Ratio of virus is relatively high.
Seasonal difference of taxonomic composition

Results

There are seasonal differences.
Family composition is biased

Only 6 families are dominant in all the samples

(1m – surface water)
Family composition is biased

(10m)
Phylum composition of sea water and sediment samples

Results

- **Sea water (20m)**
  - Proteobacteria: 75.4%
  - Cyanobacteria
  - Euryarchaeota
  - Chloroflexi
  - Aquificae
  - Chlamydiae
  - Acidobacteria

- **Sediment (25m)**
  - Proteobacteria: 63%
  - Actinobacteria
  - Firmicutes
  - Verrucomicrobia
  - Lentisphaerae
  - Deferribacteres
  - Fusobacteria
  - Chlorobi
  - Bacteroidetes
  - Synergistetes
  - Planctomycetes
  - Tenericutes
  - Spirochaetes
  - OD1
  - Thermodesulfovibrio
Class composition within proteobacteria

Results

Class composition within proteobacteria:

Sea water (20m):
- Alpha: 85%
- Gamma: 14%
- Delta: 0%
- Epsilon: 0%

Sediment (25m):
- Alpha: 13%
- Gamma: 46%
- Delta: 40%
- Epsilon: 1%
Functional annotation

The composition of Functional annotation looks resemble among monthly samples.
Isolation of the useful enzymes from the Red sea

(1) Conventional screening for **Cellulases**

- Sampling inoculation: Seawater Sediment Plankton seaweed
- Inoculation on Nutrient media and Incubation at 30 °C for 48 hr
- Get single colony
- Transfer to CMC media and incubate at 30 °C for 48 hr
- Screening by staining with Congo Red or Gram’s Iodine
Isolation of the useful enzymes from the Red sea

(2) High throughput screening for Lipases

Hosokawa et al. (2014) doi:10.1016/j.bios.2014.08.059
Hybrid Pipe Line of Genomics

- Bio Samples
- DNA Extraction
- DNA Sequencing
- Database Search
- Functional Prediction
- Genomic Mining
- Information Integration
- Biological Network
- Pathway Analysis
- Genome Modification
- Bacterial Metabolite Functions
- Metabolite database Screening
- Compound Evaluation
- Bacterial Fermentation
- Yes
- Judgement of Usefulness
- No
- Yes
- Judgement of Usefulness

Overview of entire pipeline

Database construction

In silico Analysis

Quantitative Analysis

Triangle Value

- Novel Drugs
- Health care product
- Animal Feed
- Energy source

Value / Unit

- Low
- High

Quantity

- Small
- Large

Output

in National Institute of Genetics
Summary

• Since 2014, monthly samplings of the Red Sea metagenomes were initiated for a reference.
• The majority of organisms in the Red Sea metagenome was bacteria, and virus was observed as a second major group.
• We successfully observed that taxonomic composition was changed over seasons.
• At the family level, only 6 families are dominant in all the samples.
• Hierarchal clustering of GO/KEGG top categories showed no clear differences among the samples, suggesting functional constraint at the community level.
• Comparative metagenomics will be a key to understand the microbial diversity.
Conclusion and Summary
Information Explosion in Life Science

Beyond the 4th Paradigm proposed by Jim Gray

1st Paradigm: Experiments

2nd Paradigm: Theory

3rd Paradigm: Simulation

4th Paradigm: Data-driven Scientific Discovery

Information Explosion

Scientific Innovation and its Application to the Society

5th Paradigm: Data-driven Scientific Innovation

Information Explosion in Life Science

Genome Information Society

2010 2020 2030 2040
~Vision~
Proposal of a view of the new society by innovation~
Genetic Diversity!
Biome
Marine Monitoring System by use of Meta-Genomics

Satellite

Public Database

DDBJ
NCBI
EMBL

Marine Environmental Information

Meta-Genome Ship

Ministries, Universities, Institutes
Summary

“Genome Information-oriented Society”

〜 g−Society 〜
DNA for World Peace!

DNA FOR PEACE

This ad was created on behalf of the UN, by Art Director Max Belanger, Copywriter/Skip, Inc., Turner Advertising Inc.
Acknowledgements

- CBRC, KAUST
  Katsuhiko Mineta
  and other Gojobori`s lab members
  Vlad Bajic and John Archer
- Shugo Watabe (Kitasato University)
  - CREST Project, Japan-
- Haruko Takeyama (Waseda University)
- National Institute of Genetics (NIG), Japan
- FRA (Fishery Research Agency), Japan
- Yoshizumi Ishino (Kyushu University)
Acknowledgements

KAUST
- Takashi Gojobori
- John Archer
- Vladimir Bajic
- Yoshimoto Saito
- Intikhab Alam
- Hayedeh Behzad
- Mohamed Alarawi
- Hajime Ohyanagi
- Amani Al-ma'abadi
- Siham Fatani
- Magbubah Essack
- Hiroshi Shimizu

Kitasato University
- Shugo Watabe
- Kazuho Ikeo
- Toshiaki Kudo
- Takehiko Ogata
- Shigeru Sato
- Atsushi Kobiyama
- Md Shaheed Reza
- Yuichiro Yamada
- Daisuke Ikeda
- Nanami Mizusawa
- Yuri Ikeda
- Yumiko Kawai

Iwate Fisheries Technology Center
- Kimiaki Naiki
- Shinnosuke Kaga
- Yoshimasa Kaga
- Shiho Watanabe

Waseda University
- Haruko Takeyama
- Masato Hosokawa

and CBRC CCF team

جامعة الملك عبد الله للعلوم والتقنية
King Abdullah University of Science and Technology

www.kaust.edu.sa
Thank you!