NIG International Symposium 2017
Commemorating the 30th Anniversary of DDBJ

Life, Environment, and Evolution Revealed by Genomes

May 27(Sat) - May 29(Mon), 2017
Mishima Citizens Cultual Hall, Mishima, Shizuoka

News & Topics

2017/07/04	Videos were released on TOGO TV.
2017/05/28	Due to flight trouble of airline, the lecture of Janet Thornton, schedulled May 28, was canceled.
2017/04/26	Oral ・ Poster Speaker was decided.
2017/04/20	Oral and poster abstract submission deadline is extended to April 23rd 17:00 JST.
2017/03/15	Website was updated (program details, speakers' profile, etc)
2016/11/30	Abstract submission and Registration are opened!!
2016/11/30	Our official website has been opened!!

Introduction

Recent developments of life sciences owe much to the open, free availability of genomic information. The genome information is the lifeline of life sciences, as are electricity and gas for households. DNA Data Bank of Japan (DDBJ) at the National Institute of Genetics has been the worldwide provider of the lifeline. Together with GenBank at the National Institute of Biotechnology Information (NCBI) and ENA at the European Bioinformatics Institute (EBI), DDBJ has been managing the International Nucleotide Sequence Database (INSD) that celebrates its 30th anniversary.

Thirty years ago, few of us could enjoy the internet, and the delivery of international mail required a week. What the founders of INSD dreamed of 30 years ago has become reality. Now we organize a symposium focusing not only on cutting-edge science but also on the history of DDBJ. I hope this symposium serves as a primer to invite you into the world of science.

Chairman, NIG International symposium 2017 Steering Committee
Masanori Arita
Yasukazu Nakamura

大学共同利用機関法人 情報・システム研究機構
国立遺伝学研究所   DDBJ   30th SINCE 1987

せせらぎと緑と元気あふれる協働のまち
三島市   日本学術会議 SCIENCE COUNCIL OF JAPAN   富士山麓先端健康産業集積プロジェクト ファルマバレープロジェクト

NIG International Symposium 2017
Commemorating the 30th Anniversary of DDBJ

**Life, Environment, and Evolution
Revealed by Genomes**

May 27(Sat) - May 29(Mon), 2017
Mishima Citizens Cultual Hall, Mishima, Shizuoka

Home   Outline   Program   Submission   Registration   Sponsors   Access

jp en

Outeline

Date	2017/5/27（Sat） ～ 29（Mon）
Venue	MISHIMA Citizens Cultural Hall
Language	5/27-28 Japanese(incl. translation), 5/29 English
Admission	Free
Registration	Mandatory
Organizer	Research Organization of Information and Systems National Institute of Genetics
Support	Mishima city Science Council of Japan Pharma Valley Center

NIG International Symposium 2017
Commemorating the 30th Anniversary of DDBJ

Life, Environment, and Evolution Revealed by Genomes

May 27 (Sat) - May 29 (Mon), 2017
Mishima Citizens Cultual Hall, Mishima, Shizuoka

Home   Outline   Program   Submission   Registration   Sponsors   Access

Program

Notice and Apology

Due to airline flight trouble, the lecture of Janet Thornton, scheduled May 28 afternoon, was canceled.
Please note that the timetable of May 28 afternoon might be changed, because of this.
Thank you for your understanding.

May 27th（Sat）

All-in-one Database Tutorial 2017 (in Japanese)

Databases

National Bioscience Database Center（NBDC）

Database Center for Life Science（DBCLS）

Protein Data Bank Japan（PDBj）

DNA Data Bank of Japan（DDBJ）

(Part 1)

| 13:00 - 13:10 | Opening | **Opening remarks** |
| | | Isao Katsura (Director-General, National Institute of Genetics) |

| 13:10 - 13:40 | Lecture | **EBI Lecture** |
| | | Guy Cochrane
  (European Bioinformatics Institute) |

| 13:40 - 13:50 | break | |

| 13:50 - 14:20 | Lecture | **Genbank Lecture** |
| | | Ilene Mizrachi
  (National Center for Biotechnology Information) |

| 14:20 - 14:40 | break | |

(Part 2)

Time	Type	Details
14:40 - 15:05	Lecture	**NBDC Lecture** Mari Minowa (National Bioscience Database Center)
15:05 - 15:30	Lecture	**DBCLS Lecture** Hiromasa Ono (Databas Center for Life Science)
15:30 - 15:50	break	
15:50 - 16:15	Lecture	**PDBj Lecture** Hirohumi Suzuki (Institute for Protain Research, Osaka University)
16:15 - 16:40	Lecture	**DDBJ Lecture** Masanori Arita (National Institute of Genetics)

May 28th（Sun）

Lectures / Poster Session (translation available)

Time	Type	Details
10:00 - 11:00	Lecture	**Microbiome analysis of the human gut and the ocean** **Peer Bork** Group Leader of EMBL, Heidelberg, Germany
11:00 - 12:00	Lecture	**Evolution as an integrative theory of biology** **Mariko Hasegawa** President, National University SOKENDAI
12:00 - 13:00	lunch break	
13:00 - 14:00	Lecture	**Big data in medical research** **Janet Thornton** Former director of EMBL-EBI, Hinxton, UK
14:00 - 15:00	Lecture	**The human genome project: its history and impact** **Yoshiyuki Sakaki** Emeritus Professor, The University of Tokyo The President, Shizuoka Futaba Senior/Junior High School
15:00 - 15:15	Lightning Talks	
15:15 - 16:30	Poster session	
16:30 - 16:45	Commendation	**Poster review** **Isao Katsura** Director-General, National Institute of Genetics
16:45 - 17:45	Lecture	**Database development for the life sciences in Japan: status quo and challenges** **Toshihisa Takagi** Professor,The University of Tokyo/Director of DDBJ Center, National Institute of Genetics

Oral Sessions (English only)

Session1: Systems Biology and Informatics (9:30 - 10:30)

9:30 - 10:10	Invited lecture	**Diagnosing un-occurred diseases by dynamic network biomarkers** **-- detecting the tipping points of biological processes by omics data** **Luonan Chen** Excutive Director, Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, China
10:10 - 10:30	Oral presentation	**Folding transition" in the sequence subspace around a protein family** Akira Kinjo (Institute for Protein Researc, Osaka University)
10:30 - 10:50	Break	

Session2: Transdisciplinary and Omics (10:50 - 12:10)

10:50- 11:30	Invited lecture	**No end in sight for gene function discovery: How long will it take to understand the human genome and what can be learned from the GAA1/GPAA1 function discovery story?** **Frank Eisenhaber** Executive Director, Bioinformatics Institute,A*Star Singapore
11:30 - 11:50	Oral presentation	**Classification of alkaloid compounds based on subring skeleton (SRS)** Ryohei Eguchi (Graduate School of Information Science, Nara Institute of Science and Technology)
11:50 - 12:10	Oral presentation	**LC-HRMS metabolomics method with high specificity for metabolite identification using all ion fragmentation (AIF)** Romanas Chaleckis (Gunma University Initiative for Advanced Research)
12:10 - 13:30	Lunch break	

Session3: Biology and Database (13:30 - 14:40)

13:30 - 14:00	Invited lecture	**From database to database integration** **Susumu Goto** Professor, DBCLS
14:00 - 14:20	Oral presentation	**Single-cell enhancer RNA analysis in mouse embryonic stem cells** Haruka Ozaki (Advanced Center for Computing and Communication, RIKEN)
14:20 - 14:40	Oral presentation	**gEVE, genome-scale endogenous viral element database, and its applications** So Nakagawa (Micro / Nano Technology Center, Tokai University)
14:40 - 15:00	Break	

Session4: Genetics and Genomics (15:00 - 16:10)

15:00 - 15:30	Invited lecture	**Toward creating future medicine based on genome information**
		Tadashi Imanishi
		Professor, Tokai University School of Medicine

| 15:30 - 15:50 | Oral presentation | **Global deceleration of gene evolution following recent genome hybridizations in fungi** |
| | | Wataru Iwasaki (Graduate School of Science, The University of Tokyo) |

| 15:50 - 16:10 | Oral presentation | **A new protocol for constructing the ortholog table in microbial genome database for comparative analysis (MBGD)** |
| | | Ikuo Uchiyama (National Institute for Basic Biology) |

Introduction of Speakers

Peer Bork
Group Leader of EMBL, Heidelberg

Group Leader and senior scientist at EMBL and is known as the most cited European researcher in Molecular Biology and Genetics. He received numerous awards for this contribution in science, one of which was Nature Awards for Mentoring in Science (2008) and Felix Burda award for medicine and science (2016).

Microbiome analysis of the human gut and the ocean

The human gut microbiome can now be readily studied using metagenomics (Qin et al., Nature 2010) and harbours more than 1000 species that are associated with important functions, but also with more than 30 human diseases. I will illustrate the diagnostic potential of microbial marker species in disease states using colon cancer as an example, but will also point to challenges in the interpretation of such complex data. For example, drug intake can confound signals in disease associations as drugs often perturb the gut microbiota (e.g. Forslund et al., Nature 2015). For understanding other perturbations such as those caused by faecal microbiota transplantation, we need to analyse strain populations (Li et al., Science 2016), e.g. by utilising single nucleotide variation, SNVs (Schloissnig et al. 2013). SNVs analysis also enables to trace the fate of maternal strains after birth or to identify specific commensal or pathogenic strains in stool samples. Human pathogens are mostly coming from the environment and thus it is crucial to study biodiversity of microbes at planetary scale to understand their biogeography und functional potential. The feasibility of such a global approach is illustrated by the TARA oceans project surveying the microbial diversity of this vast ecosystem by studying plankton from 35000 samples collected from all major ocean regions (Bork et al., Science 2015 and refs therein).

Mariko Hasegawa
President, National University SOKENDAI

Evolution as an integrative theory of biology

Biological phenomena are consisted of several hierarchical levels such as genes, cells, organs, individuals, populations, communities, etc. Each level has its own research questions and methods. However, evolution is the key concept for the integrated understanding of biology as a whole. On the other hand, Niko Tinbergen, who was the Nobel laureate in 1973, proposed that there should be four different approaches to questions of "why" in biology: first is the mechanism of the trait (proximate cause), second is the function of the trait (ultimate cause), third is the developmental pathway (ontogeny), and the last is the evolutionary origin (phylogeny). Those are independent to each other, but again, the knowledge obtained from each approach can be connected to make a big picture only through the theory of evolution.

Evolution means the changes of traits over generation. In more strict sense, it is the changes of gene frequencies over generation. This does not mean that we cannot think of evolution when the genetic basis of the trait is unknown. There are many ways to do evolutionary analyses on phenotypic traits, and by doing so, we can have various insights and novel questions on the genetic basis of the traits.

What is the unit of selection and adaptation? This is a rather complicated question. Before 1970s, many arguments were implicitly based on group benefit, such as the one with "for the good of the species".

However, those were criticized as group selection fallacy, and fundamental idea was changed toward gene selection. Now it has been understood that selection and adaptation involve many types of conflict such as between genes within the same body, between the male and the female, between an individual and the community it belongs.

Nowadays, the theory of evolution has been extended to apply to human cultural development, in which the theory of cultural evolution is on the making. This area of study is very interesting and promising.

Janet Thornton

Former director of EMBL-EBI, Hinxton, UK

Senior scientist at EBI and was a director (from October 2001 to June 2015) playing a key role in the ELIXIR project. Formerly she led a group at University College London and started the CATH classification of protein structures with Christine Orengo. She was Dame Commander of the Order of the British Empire (DBE) in the 2012 for her significant contribution in bioinformatics.

Big data in medical research

Our understanding of biology at the molecular level has changed radically in the last 50 years. We have seen the emergence of 'BIG DATA' especially in genomics and transcriptomics in the research laboratory. For the first time the challenge is to interpret these data rather than in the experimental hurdles. Translating this knowledge into the clinic is a formidable task.

I will describe how the data landscape in biology has changed, drawing on my experiences at the European Bioinformatics Institute, which provides many of the core biomolecular data resources used worldwide. The changes in the medical data landscape are still emerging and have a long way to go, but the prospect of 'combining' the molecular data with the medical phenotypic data offers great hope for improved diagnoses and therapies for the future. The 100,000 genomes project, funded by the UK government through the NHS, provides a stimulus to bring this new technology into the clinic.

I will describe two vignettes of how the basic knowledge can help in the clinic. The first in 'Rare Diseases' shows how we can use the 3D structures of proteins to understand and identify variants which cause developmental problems in children. The second shows how we can use transcriptome data to identify drugs, which may have an impact on Ageing.

Ultimately it is essential to share our knowledge without boundaries to improve health globally. This will require an extensive computational infrastructure for biomedical data and close collaboration between computational specialists, biologists, data scientists and clinicians. At EMBL-EBI we have built a large infrastructure for biological data, in collaboration with partners around the world. Together with others we have built the ELIXIR consortium across Europe to share bioinformatics data, tools, infrastructure and expertise. The medical data will require a much larger, more diverse infrastructure – this is just beginning.

Yoshiyuki Sakaki

Emeritus Professor, The University of Tokyo
The President, Shizuoka Futaba Senior/Junior High School

The human genome project: its history and impact

Discovery of the double-stranded structure of DNA by Watson and Crick revealed that genetic information is coded by using four letters (bases), A,G, T and C of DNA. This opened a door of molecular biology and eventually enabled us to decode the whole human genome of three Giga bases under the Human Genome Project (HGP). HGP is, however, not a simple extension nor expansion of traditional molecular biology. HGP introduced various new approaches and challenges to biology and medicine, among which following three innovative approaches/challenges should be noted. The first, the concept of engineering was introduced to biology through the automation of DNA sequencing technology, in which Japanese scientists played leading roles. The second, the concept of "open innovation" was introduced. The international Human Genome Sequencing Consortium set the so called "Bermuda Principle" of immediate release of the sequence data to public databases without any condition, which successfully provided a powerful platform to accelerate biological and medical researches. The third, a new academic field called "bioinformatics" has been established through HGP. Bioinformatics coupled with high speed DNA sequencer and high performance computer enabled us to develop a new style of biology called "data-driven biology", in which DNA databases play an integral role.

The completion of the Human Genome provided us a gold standard to open a new era of human biology and medicine. The complete genome sequence, coupled with revolutionary progress of new sequencing technologies, enabled us to carry out sequence-based large scale population study to identify a large number of disease related genes, which has brought about a new stream of genome-based medicine called personalized medicine. As well, the large scale sequence data of various ethnic populations provided deep insight into the human evolution. Several other recent progress of human biology and medicine including "molecular epi-genomics" and "human microbiome" will be discussed.

Toshihisa Takagi

Tokyo University professor ／国立遺伝学研究所 DDBJ center センター長

Database development for the life sciences in Japan: status quo and challenges

30 years have passed since the birth of DDBJ (DNA Data Bank of Japan).

DDBJ is a member of INSDC (International Nucleotide Sequence Database Collaboration), and was established to cooperatively develop and maintain the international nucleotide sequence archive. The past 30 years have witnessed many events that significantly changed the life sciences, such as the determination of the human genome sequence. It was also the last 30 years that the importance and necessity of databases and bioinformatics have been acknowledged.

High-performance genome sequencing devices have been developed one after another, genome sequences of a vast number of organisms are determined day by day, and furthermore, data other than genomes have been acccumulated inexpensively, such as gene expressions, proteins, and metabolites, (e.g. omics data). Not only data but also knowledge has exploded. A data driven approach became the mainstream of life sciences that shares and integrates those big data in the form of a database and analyzes them using artificial intelligence. Databases now become essential for life science research. In addition, data sharing through databases is also very important from the standpoint of open science, which has become a global trend in recent years.

In this lecture, I will outline the history of database development in the field of life sciences in Japan, and introduce its current situation and challenges, mainly focusing on DDBJ. Open science is one important issue here. For database development, it is essential to manage supercomputers to store and analyze data, and raise human resources that are responsible for data construction and analysis. I will mention these issues too.

Luonan Chen

Excutive Director, Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, China

Diagnosing un-occurred diseases by dynamic network biomarkers
-- detecting the tipping points of biological processes by omics data

Considerable evidence suggests that during the progression of complex diseases, the deteriorations are not necessarily smooth but are abrupt, and may cause a critical transition from one state to another at a tipping point. Here, we develop a model-free method to detect early-warning signals of such critical transitions (or un-occurred diseases), even with only a small number of samples. Specifically, we theoretically derive an index based on a dynamical network biomarker (DNB) that serves as a general early-warning signal indicating an imminent sudden deterioration before the critical transition occurs. Based on theoretical analyses, we show that predicting a sudden transition from small samples is achievable provided that there are a large number of measurements for each sample, e.g., high-throughput data. We employ gene expression data of three diseases to demonstrate the effectiveness of our method. The relevance of DNBs with the diseases was also validated by related experimental data (e.g., liver cancer, lung injury, influenza, type-2 diabetes) and functional analysis. DNB can also be used for the analysis of nonlinear biological processes, e.g., cell differentiation process.

Frank Eisenhaber

Executive Director, Bioinformatics Institute,A*Star Singapore

No end in sight for gene function discovery: How long will it take to understand the human genome and what can be learned from theGAA1/GPAA1 function discovery story?

Despite dramatic technical progress in genome and transcriptome sequencing, the ability to link changes in human sequences with phenotypic outcomes is severely limited. After the enthusiastic era of first full genome sequencing that started with a few bacteria and yeast in the middle of the nineties and culminated in the first human genome draft, the expectations with regards to cures of not yet treatable diseases or to new biotechnologies have not been fulfilled even nearly to the extent as the original hype might have promised. Whereas the impact is dramatic in cases where biomolecular mechanisms are known, little progress even over several decades in the future should be expected where this is not the case. Except for few cases of clear statistical links between certain genomic aberrations and phenotypic properties, genotypes can be linked to phenotypes only via the explanation level of biomolecular mechanisms the knowledge of which is currently fragmentary at best. It is most urgent to discover or augment function description for > 10000 non- or poorly characterized human genes. The key to understanding biomolecular sequences is via function prediction from protein sequences. The plethora of methods for structure and function prediction from protein sequence integrated in BII's ANNOTATOR environment is reviewed. A new software GUI, the human protein mutation viewer, is useful for mapping mutations onto sequence architectural features and for delineating possible mechanistic implications. Examples of function discovery for GPI lipid anchor biosynthesis pathway's

transamidase subunit genes pig-T and Gpaa1 as well as some others are presented.

[1] Eisenhaber F, J. Bioinform Comput Biol. 10, 1271001 (2012)
[2] Eisenhaber B, et al. Cell Cycle. 13, 1912-7 (2014).
[3] Kulemzina I, et al. Mol Cell. 2016 Sep 15;63(6):1044-5
[4] Lua WH et al. Cell Cycle. 2017 Mar 4;16(5):457-467

Susumu Goto

DBCLS Professor

From database to database integration

In the big data era for life science, it is crucial for researchers to be able to access up-to-date and easy-to-use databases. Over 1,500 databases are catalogued in the online Nucleic Acids Research (NAR) database issue web site, and it is still not an easy task to find an appropriate database for each researcher in terms of both freshness and usefulness. While developing and maintaining each useful database is of course very important, integration of the databases is also indispensable for easier understanding and interpreting of the experimental results.

Databases are classified using several criteria such as data types and source information used in the category list of NAR web site. Another classification can be archival data repositories such as DDBJ and knowledgebases such as KEGG. KEGG can be considered as an integrated database as well by curating information from several data sources using its own ontology, which provides unique biological contents in an integrated and uniform way. It also serves as a backend data source for several bioinformatics analysis tools including functional annotation system for omics data.

To further integrate databases for more useful applications, technologies to support interoperability of databases distributed in the internet will be necessary. Semantic web is one such technology exploited by National Bioscience Database Center, JST and Database Center for Life Science, ROIS. Most data are represented in resource description framework (RDF) format with properly designed ontologies to follow FAIR principle to make data findable, accessible, interoperable and reusable. DBCLS has been developing technologies to integrate databases using RDF, more broadly the linked open data concept, and harnessing the community to provide, integrate and utilize the biological data in an integrated way. I have recently moved to DBCLS from KEGG group, so I would like to talk about the two approaches and discuss their differences.

Tadashi Imanishi

Tokai University School of Medicine professor

Toward creating future medicine based on genome information

As a consequence of rapidly advanced DNA sequencing technology, genome information is being used widely in various areas of biomedical research. On the other hand, the application of genome information to practical medical services is still in its infancy. We are thus engaged in the research and development of genome sequencing and bioinformatics, with the aim of creating future medicine through clinical applications of genome information.

Firstly, we developed a computational system for predicting disease risks based on personal genome information. By collecting information of single nucleotide polymorphisms from literatures of genome-wide association studies, we compiled a database of genetic risk factors of various diseases, and released it as VaDE (http://bmi-tokai.jp/VaDE/). Then, using the VaDE database as a reference, we developed a computational tool for calculating the risks of various diseases from personal genome information. This enabled us to calculate the personal risks (odds ratios) of various diseases systematically. Though we still need to objectively evaluate its prediction accuracy, the system could specify an apparent high-risk group for some of the diseases. We suppose that the system will provide an indispensable technology to realize a scheme of preventive medicine in the future.

Secondly, we developed a genome analyzing system for the rapid diagnosis of infectious diseases. Using a portable, single-molecule DNA sequencer and high-spec laptop computers, we could set up a system that can in principle identify bacterial species in about one hour from a DNA sample of bacterial infection. This technology will replace the current methods of diagnosis for bacterial infections in the near future.

By utilizing genome information databases as a key technology, we will continue to develop the medical services of the future.