



2011

Young Researchers Conference on

# Evolutionary Genomics

<http://e-genomics.org>

1-2 August

National Center for Sciences Building

Tokyo, Japan

**Organizing Committee**

Atsushi Ogura, Chair, Ochanomizu University  
Yoshihito Niimura, Tokyo Medical and Dental University  
Kousuke Hanada, RIKEN Plant Science Center  
Takashi Makino, Tohoku University  
Takashi Gojobori, Honorary chair, National Institute of Genetics

**Sponsor**

National Institute of Genetics  
BGI  
Affymetrix Japan

**Contact**

contact@e-genomics.org

**Venue**

National Center for Sciences Building, 2nd floor  
Hitotsubashi 2-1-2, Chiyoda-ku, Zip101-0003, Tokyo, Japan



**中 会 議 場**

Conference Room

中規模な学術、講演会、説明会、レセプションなどにご利用いただけます。利用人数にあわせ、会場の規模を調整でき、様々なレイアウトでご利用いただけます。

【設備】(中会議場4のみ専用)

音響設備 / 無線マイク4本・ワイヤレスマイク4本

映像設備 / プロジェクター・スクリーン

同時通訳設備 / 4ヶ所指定同時通訳レシーバー200台

※移動式の音響機器、映像機器をご用意することができます。

【施設概要】中会議場1〜3:各64席、中会議場4:48席、最大240席利用可能



大学財務・経営センター総務部 経営支援課 会議室利用案内窓口

# Meeting Program

Monday, August 1, 2011

09:00-10:00 Registration

10:00- Opening remarks (Atsushi Ogura, Ochanomizu University)

10:05-12:30 Session 01 Session Chair: Takashi Makino (Tohoku University)

10:05-10:25 Dynamic evolution of the mechanisms of translation initiation

*So Nakagawa*

10:25-10:45 Evolution of transposable elements in Wolbachia bacterial endosymbionts

*Richard Cordaux*

10:45-11:05 Development of New Method for Constructing Ortholog Dataset Using Large Amount of Gene Data

*Tokumasa Horiike*

11:05-11:20 Break

11:20-11:40 Evolutionary and functional analysis of large-effect mutations

*Allan Moses*

11:40-12:00 Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes

*Yong Woo*

12:00-12:20 Measuring the evolutionary rate of protein-protein interaction

*Jianzhi Zhang*

12:20-12:40 Identification of dosage sensitive genes on the budding yeast genome

*Hisao Moriya*

14:00-16:30 Session 02 Session Chair: Kousuke Hanada (RIKEN)

14:00-14:20 Recurrent evolution of self-compatibility

*Kentaro Shimizu*

14:20-14:40 Cis-regulatory code of stress responsive transcription in Arabidopsis thaliana

*Shin-han Shiu*

14:40-15:00 Novel protein-coding genes in mammalian genomes

*Aoife Mclysaght*

15:00-15:20 When Time and Chance Happen to Worms: gene expression variation and the evolution of development

*Scott Rifkin*

15:00-15:35 Break

15:35-15:55 A Broad-ranging and Cost-efficient Inter-Species Microarray

*Jun Sese*

15:55-16:15 Navigating evolution-aware life science: our ancestral genome, 500 million years ago

*Shigehiro Kuraku*

16:15-16:35 Evolutionarily inflexible developmental programs?

~Transcriptome similarity supports vertebrate phylotypic hypothesis ~

*Naoki Irie*

15:00-16:30 Poster setup

17:00-19:00 Poster Session

Tuesday, August 2, 2011

09:30-12:00 Session 03 Session Chair: Yoshihito Niimura (Tokyo Medical and Dental University)

09:30-09:50 Higher evolutionary rate in genes with homopolymeric amino acid repeats that have nondisordered structure

*Jun Gojobori*

09:50-10:10 Comparative transcriptome and genome analysis in a chimpanzee trio

*Yasuhiro Go*

10:10-10:30 Understanding human brain evolution through transcriptome sequencing

*Philipp Khaitovich*

10:30-10:50 Pleistocene Affairs

*Johannes Krause*

10:50-11:05 Break

11:05-11:25 Evidence of Positive Selection Acting on Novel Primate Specific Transcripts Generated from Bidirectional Promoters

*Valer Gotea*

11:25-11:45 Effects of Natural Selection and Gene Conversion on the Evolution of Human Glycophorins Coding for MNS Blood Polymorphisms in Malaria-Endemic African Populations

*Wen-Ya Ko*

11:45-12:05 Evolutionary Genomic Medicine

*Sudhir Kumar*

12:00- 12:10 Closing remarks (Takashi Gojobori, National Institute of Genetics)

13:30- 15:30 Social Lunch

## Abstracts of Oral Presentations

01

Dynamic evolution of the mechanisms of translation initiation

So Nakagawa

National Institute of Genetics

Translation initiation is one of the most fundamental processes in the regulation of gene expression for all protein-coding genes in the genome of every species. In prokaryotes, it is generally believed that translation is initiated by the interaction between the Shine-Dalgarno (SD) sequence in the 5' untranslated region (UTR) of an mRNA and the anti-SD sequence in the 3' end of a 16S ribosomal RNA. However, there are two exceptional mechanisms, which do not require the SD sequence for translation initiation: one is mediated by a ribosomal protein S1 (RPS1) and the other used leaderless mRNA that lacks its 5' UTR. To understand the evolutionary changes of the mechanisms of translation initiation, we examined how universal the SD sequence is as an effective initiator for translation among prokaryotes. We identified the SD sequence in the genomes of 277 species (249 eubacteria and 28 archaeobacteria). We also devised an SD index that is a proportion of SD-containing genes in which the differences of GC contents are taken into account. We found that the SD indices varied among prokaryotic species, but were similar within each phylum. Although the anti-SD sequence is conserved among species, loss of the SD sequence seems to have occurred multiple times, independently, in different phyla. For those phyla, RPS1-mediated or leaderless mRNA-used mechanisms of translation initiation are considered to be working to a greater extent. Moreover, we also found that some species, such as Cyanobacteria, shows the nucleotide biases between upstream and downstream of the initiation codon, possibly suggesting the acquirement of new mechanisms of translation initiation. Our findings indicate that, despite the essentiality of translation initiation in gene expression, its mechanisms have dynamically changed during evolution.

Keywords:

## Evolution of transposable elements in Wolbachia bacterial endosymbionts

Richard Cordaux

Université de Poitiers – CNRS

The streamlined genomes of ancient obligate endosymbionts generally lack mobile genetic elements, which are otherwise relatively frequent in other prokaryotes. Yet, the genome of Wolbachia, one of the most abundant bacterial endosymbionts on Earth, is littered with transposable elements. Such paradox raises the question as to why there are so many transposable elements in the genome of this ancient endosymbiont. To address this question, we investigated the dynamics of various types of transposable elements (insertion sequences and group II introns) in Wolbachia genomes, using an evolutionary perspective. Our results indicate that several processes may explain the abundance of mobile DNA in Wolbachia, including recent activity, along with recurrent invasions through horizontal transfers and gene conversion. Overall, our fine-scale analysis of Wolbachia transposable elements provides insight into the micro-evolutionary processes that may explain the evolutionary interactions and dynamics of transposable elements in bacterial organisms.

Keyword:

## Development of New Method for Constructing Ortholog Dataset Using Large Amount of Gene Data

Tokumasa Horiike

Shizuoka University

Generally, ortholog databases such as COG (Cluster of Orthologous Groups) were utilized for prediction of protein functions. However they are not suitable for phylogenetic analysis because many paralogs and horizontally transferred genes are included in the estimated ortholog dataset. Therefore, we developed a new method to overcome the problems of constructing ortholog dataset that introduced the following ideas.

- 1) HGT Filter: Deletion of genes that are derived from horizontal gene transfer from initial sequence dataset.
- 2) Out-paralog filter: Deletion of out-paralogs from result data of BLAST using the similarity score. Remaining out-paralogs are deleted in the following steps.
- 3) Tree split: This program split a tree into two to remove the out-paralogs from the candidates of ortholog dataset with the information of phylogenetic tree's topology, such as monophyletic or polyphyletic states in the group and species level.
- 4) Changing threshold: This program cut the candidate of ortholog group (including out-paralog) into two ortholog groups by the difference of evolutionary distance among true ortholog members and that among out-paralogs.

We compared the ortholog dataset made by this method and ortholog dataset without the key ideas of this method. The result suggest that this method has some tolerance for Out-paralogs and HGTs.

Keywords:

04

## Evolutionary and functional analysis of large-effect mutations

Alan Moses

University of Toronto

In the search for genetic variation that underlies phenotypic and functional diversity, amino acid differences in proteins have usually been the focus. However, low-cost high-throughput sequencing technologies have led to the identification of large amounts of genetic variation of different types, such as non-sense mutations, gene copy number variations, and insertions and deletions in both protein coding and non-coding DNA. We have been exploring the functional consequences of these types of mutations using evolutionary genomics methods. By quantifying the evolutionary constraint on the regions in which these mutations fall, we show that as has been observed for SNPs, indels and stop codons affecting more conserved regions segregate at lower allele frequencies in the population, consistent with their effects being more deleterious. In addition, by systematically phenotyping budding yeast, we have begun relating these mutations to phenotypic consequences. Our analysis indicates that deleterious large-effect mutations segregating at low frequency have a large functional and phenotypic impact on natural populations.

Keywords:

05

## Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes

Yong Woo

The University of Chicago

A balance between gene expression stability and evolvability is essential for the long-term maintenance of a living system. We studied whether the genetic and epigenetic properties of the promoter affect gene expression variability. We hypothesized that upstream distance and orientation (head-to-head or head-to-tail) are important for the promoter architecture and gene expression variability. We found that in budding yeast genes with a short upstream distance tend to have low gene expression variability, and their promoter is flanked by strongly positioned nucleosomes and tends to have low nucleosome occupancy. These observations suggest that in vivo positioning of the flanking nucleosomes facilitates stable nucleosome depletion at the core promoter region and enhances gene expression stability. Head-to-head genes have, on average, lower gene expression variability, greater nucleosome depletion at the core promoter region, and more strongly positioned nucleosomes that flank the core promoter than do head-to-tail genes. These observations hold for diverse eukaryotes. In complex organisms such as mammals, only a small fraction of head-to-tail genes have retained a short upstream distance, probably because the promoter may not be flanked by a strongly positioned nucleosome on the upstream side.

Keywords:

06

## Measuring the evolutionary rate of protein-protein interaction

Jianzhi Zhang

University of Michigan

Despite our extensive knowledge about the rate of protein sequence evolution for thousands of genes in hundreds of species, the corresponding rate of protein function evolution is virtually unknown, especially at the genomic scale. This lack of knowledge is primarily due to the huge diversity in protein function and the consequent difficulty in gauging and comparing rates of protein function evolution. Nevertheless, most proteins function through interacting with other proteins, and protein-protein interaction (PPI) can be tested by standard assays. Thus, the rate of protein function evolution may be measured by the rate of PPI evolution. Here we experimentally examine 87 potential interactions between *Kluyveromyces waltii* proteins whose one-to-one orthologs in the related budding yeast *Saccharomyces cerevisiae* have been reported to interact. Combining our results with available data from other eukaryotes, we estimate that the evolutionary rate of protein interaction is 0.26 per PPI per billion years, three orders of magnitude lower than the rate of protein sequence evolution measured by the number of amino acid substitutions per protein per year. The extremely slow evolution of protein molecular function may account for the remarkable conservation of life at molecular and cellular levels and allow studying the mechanistic basis of human disease in much simpler organisms.

Keywords: protein interaction, yeast two hybrid, evolutionary rate, yeast, protein function

07

## Identification of dosage sensitive genes on the budding yeast genome

Hisao Moriya

Okayama University

Expression level of each gene within a cell should be optimized so that the cellular system can perform its function precisely and effectively. On the other hand, the expression level should have some permissible range so that the cellular system has robustness against perturbations in gene expression. We have previously developed a genetic method designated genetic Tug-Of-War (gTOW), by which permissible limit gene-overexpression of a target gene can be measured as the copy number (Moriya et al, 2006).

In this study, we have cloned >95% of genes in the budding yeast *Saccharomyces cerevisiae* genome into the plasmid for gTOW, and measured the upper limit copy numbers of them. As a result, we have isolated about 100 genes with the upper limits <10 copies. Gene ontology analysis revealed that genes involved in cytoskeletal organization and intracellular transport are significantly enriched. The reason for the very low limits seemed to be generally explained by the dosage imbalance between partner genes, as we have found in the cell cycle regulatory system (Kaizu et al., 2010). We suggest that these dosage balances might function to secure genome integrity as a “genome skeleton”.

Keywords:

## Recurrent evolution of self-compatibility

Kentaro Shimizu  
University of Zurich

Ever since Darwin's pioneering research, the evolutionary of self-fertilization (selfing) by the loss of self-incompatibility (SI) has been regarded as one of the most prevalent events in flowering plants. We have studied two genera *Arabidopsis* and *Cardamine*, in which the transition occurred many times independently. Evolutionary theory predicts that mutations disabling male function have the highest rate of increase in frequency from positive selection, because male mutations would increase through pollen as well as through seeds. Despite many studies on the genetic basis of loss of SI in the predominantly selfing plant *Arabidopsis thaliana*, it remained unknown whether selfing arose through mutations in the female specificity gene (SRK), male specificity gene (SCR/SP11), or modifier genes and whether any of them rose to high frequency across large geographic regions. We revealed that a disruptive 213-bp inversion in the SCR gene (or its derivative deletions) is found in 95% of European accessions, which contrasts with the genome-wide pattern of polymorphism in European *A. thaliana*. When the inversion was restored in transgenic plants, ancestral SI was revived. These results suggest that a disruptive mutation in the male gene was advantageous in overcoming mate limitation, possibly associated with historical population expansion(s) from glacial refugia.

Keywords:

## Cis-regulatory code of stress responsive transcription in *Arabidopsis thaliana*

Shin-Han Shiu  
Michigan State University

Environmental stress leads to dramatic transcriptional re-programming central to plant survival. At the cis-regulation level, substantial knowledge has accumulated on how a few plant cis-regulatory elements (CREs) function in stress regulation but many more CREs remain to be discovered. In addition, the plant stress cis-regulatory code, i.e. how CREs work independently and/or in concert to specify stress responsive expression, is mostly unknown. Using multiple stress gene expression data, we identified a large number of putative CREs (pCREs) in *Arabidopsis thaliana* with characteristics of authentic cis-elements. In addition, using these pCREs, we uncovered cis-regulatory codes specifying how the presence and absence of these motifs and their combinatorial relationships can be used to predict stress responsive expression. Expression prediction models based on pCRE combinations perform significantly better than those based simply on pCRE presence and absence. Furthermore, instead of a few master regulatory rules for each stress condition, many rules were discovered, and each appears to control only a small subset of stress responsive genes. This study contributes significantly to a better understanding of plant stress cis-regulatory logic and provides prioritized targets for further experimentation.

Keywords: Stress, Transcription, Regulation, Plan

## Novel protein-coding genes in mammalian genomes

Aoife McLysaght  
Trinity College Dublin

François Jacob famously declared that the “probability that a functional protein would appear de novo by random association of amino acids is practically zero”. However, within the last few years convincing examples of de novo origination of protein-coding genes have been uncovered by meticulous analysis of gene and genome sequences of flies, yeasts and mammals. Using extremely strict criteria, so as to avoid false positives, we searched for de novo genes within the Great Ape lineage leading to human, and also within murid genomes. We uncover new examples of de novo genes, several with a role in disease including cancers.

Keywords:

## When Time and Chance Happen to Worms: gene expression variation and the evolution of development

Scott Rifkin  
University of California, San Diego

The mapping between genotype and phenotype is a critical but troublesome component of evolutionary theory. It can be extraordinarily complex, and the same mutation can have different phenotypic effects depending upon genetic background or environmental conditions. However, even genetically identical organisms in homogeneous environments vary, suggesting that randomness in developmental processes such as gene expression may also generate diversity. This stochasticity threatens to further complicate our understanding of how allelic effects arise. I will discuss the consequences and relevance of gene expression variability for the operation and evolution of genetic networks in multicellular organisms, using the intestinal specification network in *Caenorhabditis elegans* and some of its close relatives as an example.

Keywords

## A Broad-ranging and Cost-efficient Inter-Species Microarray

Jun Sese

Dept. of Computer Science, Tokyo Institute of Technology

After the success of comparative genomic studies, comparative transcriptomics has been attracting considerable attention. However, comparative transcriptomics in non-model organisms is hindered by the shortage of easy-to-use transcriptome tools, such as genome sequence and microarray, that are specifically prepared for non-model organisms. This study facilitates observations of the comprehensive gene expression observation of non-model organism.

In this study, we report a novel tool, the inter-species array, which can measure the gene expression levels of multiple species at once. An inter-species array is cost efficient, because it operates simply by changing the probes on an Agilent customized commercial array, and uses conventional facilities and protocols. To design the necessary probes, we investigated the effects of mutations on expression levels by using systematically mutated probes.

To evaluate the strategy for designing probes, we generated an array for humans, rats and mice that comprised 6683 genes. The number of genes is larger than that of previous arrays. We measured the expression of genes in human and rat astrocytes and in rat and mouse cortex. The experimental results obtained using our array show a high correlation in gene expressions between orthologous genes.

Keywords:

## Navigating evolution-aware life science: our ancestral genome, 500 million years ago

Shigehiro Kuraku

University of Konstanz

Complex life is achieved by elaborate networks composed of dozens of thousands of proteins encoded by 'genes' and non-protein-coding elements, each of which has a long or short history since its origin. One of the tasks of evolutionary biology is to provide non-evolutionary research fields with time scales and explanations for complex biological systems. I have analyzed molecular phylogeny and embryonic expression patterns of a variety of protein-coding genes in non-model vertebrate species, such as jawless fishes. In fact, this avenue of studies on non-model vertebrates has shed light on fine-scale pictures of a number of ancient events that have shaped the genomes of well-studied species, including us humans. First, the major event which introduced a high level of redundancy in our genome, known as 'two-round whole genome duplications (2R-WGD)', has been revealed to date back to the era before the radiation of all extant vertebrates, including jawless fishes. Second, it has been shown likely that the ancestral vertebrate genome contained additional protein-coding genes (designated 'cryptic pan-vertebrate genes') that were successively lost in the course of evolution and have not been recognized previously. Evolutionary genomics can now play an indispensable role in bridging separate research fields focusing on different species, genes and biological processes.

Keywords: whole genome duplication, jawless fish, gene loss, life science, human genome

14

Evolutionarily inflexible developmental programs? ~Transcriptome similarity supports vertebrate phylotypic hypothesis ~

Naoki Irie

RIKEN, CDB

One of the biggest questions in Evolutionary Developmental biology is that how animal developmental systems accomplished the evolution of extensive morphological diversity, while meeting requirements of creating a viable organism every generation. A long-standing controversy concerning this question is that which embryonic period (earliest or organogenesis period) has been conserved the most during evolution, and has been imposed constraints on morphological evolution. Originally proposed by Ernst Haeckel, former hypothesis insist that the earliest embryo is the stage of conservation, whilst the latter hypothesis (phylotypic stage hypothesis or the developmental hourglass model) emphasizes conservation of mid-embryonic or organogenesis stages. In order to investigate which developmental stage is conserved the most among vertebrate embryos, here we collected and identified the transcriptomes of early to late embryos of mouse, chicken, xenopus, and zebrafish, and further evaluated their transcriptome similarities as an index for evolutionary distance. Our results indicated that so-called pharyngular stage is the stage of conservation, while early and later stages are rather diverged. We will discuss how developmental hourglass model is related to the concept of phylotype, and vertebrate's basic body plan.

Keywords: Evolution, Development, Transcriptome comparison, Vertebrates

15

Higher evolutionary rate in genes with homopolymeric amino acid repeats that have nondisordered structure

Jun Gojobori

The Graduate University for Advanced Studies

In the human genome, about 650 genes are known to have repeats of single amino acids. This kind of repeat are also called as homopolymeric amino acid repeats. By investigating the evolutionary conservativeness among mammals, we classified the repeats into three categories: those whose length is conserved among mammals (CM), those whose length differs among nonprimate mammals but is conserved among primates (CP), and those whose length differs among primates (VP). The 3D structure of these amino acid repeats is considered to be intrinsically disordered. As expected, many repeats are predicted to have disordered structure. On the other hand, 13% and 25% of the repeats for categories CM and VP are predicted to have nondisordered structure, respectively. The genes with nondisordered repeats showed a higher Ka/Ks ratio than the genes with disordered repeats. These results indicate that amino acid substitution rates have been elevated in the genes with nondisordered repeats.

Keywords: Homopolymeric amino acid repeats, protein structure, intrinsically disordered region, mammal, primate

## Comparative transcriptome and genome analysis in a chimpanzee trio

Yasuhiro Go

Primate Research Institute, Kyoto University

Transcriptome can be defined as the total set of transcripts in a given cell or a population of cells (e.g., tissue). Using RNA-seq technology, which sequences the transcripts with massively parallel DNA sequencing, we can see the dynamics of transcripts among different type of cells (e.g., cancer cells vs. normal cells) or different organisms (e.g., human brain vs. chimpanzee brain). Here we used this RNA-seq for transcriptome analysis of leukocyte in a chimpanzee trio (father-mother-child). We obtained 3~5 Gb from each individual by Illumina-GAI sequencer and found about a hundred thousand SNPs against a reference chimpanzee genome sequence. Moreover we found ten thousands and several thousands of transcribed SNPs (tSNPs) and several thousands of transcribed INDELs (insertions and deletions) on average among trio, and revealed that tSNPs are more enriched on 3'-UTR than CDS or 5'-UTR, suggesting a major role of tSNPs as a gene expression regulator. Moreover, NGS enables us to quantify a level of gene expression in an allele-specific way, and it is thus possible to estimate relative contribution of cis and trans mutations for gene expression divergence, and more importantly, to detect candidate imprinting genes straightforwardly. At present, we have thousands of such tSNPs in the data. We recently obtained whole genome sequencing data from a same trio, covering more than 30X of the chimpanzee genome. Integrating chimpanzee transcriptome and genome data as well as conjunction with human trio data could help us to get a comprehensive picture for evolutionary dynamics of transcriptome and genome both intra- (trio) and inter- (humans and chimpanzees) species levels.

Keywords: chimpanzee, comparative transcriptome, comparative genome, regulation of gene expression

## Understanding human brain evolution through transcriptome sequencing

Philipp Khaitovich

PICB

Transcription is the first step connecting genetic information with an organism's phenotype. While gene expression changes in the human brain has been characterized extensively, our knowledge about changes in transcript structure, as well as mechanisms of transcriptional regulation is limited. Here, we use high-throughput transcriptome sequencing (RNA-Seq) to characterize the human brain transcriptome at different life stages: from newborns to old age. We investigate changes in transcript splicing, as well as use of alternative transcription start and termination sites at different age. Further, we link these changes to splicing regulation. Finally, we compare human splicing patterns to those in chimpanzees and macaques and identify human-specific splicing changes.

Keywords: human, chimpanzee, brain, RNA-seq, splicing"

## Pleistocene Affairs

Johannes Krause

Tübingen University, Germany

A genetic comparison between modern humans and their extinct relatives could both address the relationship between us and them and offer the possibility to identify genetic changes that happened specifically on the human lineage. Using a combination of high-throughput DNA sequencing technologies and multiple improvements in ancient DNA retrieval, library construction and targeted library enrichments, the Neandertal Genome Analysis Consortium has recently completed a first version of the Neandertal genome as well as a genome sequence of an extinct hominin discovered in the Altai mountains in southern Siberia named Denisovan. The analysis of both the Neandertal and Denisovan genome revealed evidence of gene flow between certain modern human populations and both extinct hominins. From the analysis of the data we were furthermore able to draw conclusion about diversity within and among the extinct hominins and by scanning the human genome for regions of positive selection using the Neandertal and Denisovan genome, we identified several strong candidate genes involved in diet, cognitive traits, and skeletal morphology that were potentially selected on the modern human lineage.

Keywords: Ancient DNA, Hominin, Neandertal, Denisovan, Next Generation Sequencing

## Evidence of Positive Selection Acting on Novel Primate Specific Transcripts Generated from Bidirectional Promoters

Valer Gotea

National Human Genome Research Institute

The evolution of novel characters implies the emergence of genes with novel functions, which was mostly studied in genes that were subjected to various forms of duplication. Here I focus on a different class of genes that are capable of providing new functions, namely lineage-specific novel transcripts (NT) that are not the result of gene duplication. More specifically, I analyze a collection of 1,162 non-coding, non-conserved human-specific NTs that flank conserved genes in a head-to-head fashion. While their origin from bidirectional promoters is supported by significant correlation of expression to that of their protein-coding neighbors, NTs appear to be randomly distributed across the genome. Additionally, their novel character is supported by their high transposable element (TE) content and the splice sites usage pattern. Many of these transcripts can be viewed as transcriptional noise, but evidence of functional importance, in the form of positive selection, can be found in few transcripts. A novel test that takes advantage of the high TE content of NTs highlights less than 5% of NTs as having significantly elevated rates of human-specific substitution rates (after correction for local mutation bias). In a few specific cases, the mutation pattern can be linked (and be validated experimentally) to the increased strength of neighboring splice sites that resulted in the recruitment of novel exons, thus supporting the gain of function hypothesis. Based on these findings, I will outline a lineage-independent mechanism that is an important evolutionary force by continuously generating genuinely novel transcripts.

## Effects of Natural Selection and Gene Conversion on the Evolution of Human Glycophorins Coding for MNS Blood Polymorphisms in Malaria-Endemic African Populations

Wen-Ya Ko

Center for Evolutionary Medicine & Informatics @ Biodesign Institute, Arizona State University, USA

Malaria has been a very strong selection pressure in recent human evolution, particularly in Africa. Of the one million deaths per year due to malaria, more than 90% are in sub-Saharan Africa, a region with high levels of genetic variation and population substructure. However, there have been few studies of nucleotide variation at genetic loci that are relevant to malaria susceptibility across geographically and genetically diverse ethnic groups in Africa. Invasion of erythrocytes by *Plasmodium falciparum* parasites is central to the pathology of malaria. Glycophorin A (GYPA) and B (GYPB), which determine MN and Ss blood types, are two major receptors that are expressed on erythrocyte surfaces and interact with parasite ligands. We analyzed nucleotide diversity of the glycophorin gene family in 15 African populations with different levels of malaria exposure. High levels of nucleotide diversity and gene conversion were found at these genes. We observed divergent patterns of genetic variation between these duplicated genes and between different extracellular domains of GYPA. Specifically, we identified fixed adaptive changes at exons 3–4 of GYPA. By contrast, we observed an allele frequency spectrum skewed toward a significant excess of intermediate-frequency alleles at GYPA exon 2 in many populations; the degree of spectrum distortion is correlated with malaria exposure, possibly because of the joint effects of gene conversion and balancing selection. We also identified a haplotype causing three amino acid changes in the extracellular domain of glycophorin B. This haplotype might have evolved adaptively in five populations with high exposure to malaria.

## Evolutionary Genomic Medicine

Sudhir Kumar

Center for Evolutionary Medicine & Informatics @ Biodesign Institute, Arizona State University, USA

Modern technologies have made the sequencing of personal genomes routine and revealed millions of variants per genome. What do these variants foretell about an individual's predisposition to diseases? The experimental technologies and resources required to carry out such evaluations are not yet available, even for single nucleotide variants (SNVs) of protein coding DNA that constitutes only 1 – 3% of the genome. Nature is the greatest experimenter. By the process of natural selection, Nature evaluates new mutations and existing variations in the diversity of species. Outcomes of these experiments are revealed by multispecies genome comparisons. Here I will discuss studies from our group (and others) that have evaluated the relationship of evolutionary information gleaned from these comparisons with characteristics and in silico functional diagnoses of SNVs associated with monogenic, complex, and somatic-cell diseases. I conclude that the patterns of long-term evolutionary conservation and permissible sequence diversity constitute two lenses of an evolutionary telescope to investigate human variation.

Keywords: Evolution, Medicine, Polymorphisms, Disease, Evolutionary rate

# Abstracts of Poster Presentations

P-01

Identification and characterization of protochordate promoters

Mayu Fushimi<sup>1</sup> and Kohji Okamura<sup>2</sup>

1 Department of Biology, Faculty of Science, Ochanomizu University

2 Centre for Informational Biology, Ochanomizu University

DNA methylation is an ancient process conducted by conserved methyltransferase families. However, its biological functions are highly diverged and still elusive in some groups of organisms such as protochordates. While plants and fungi use DNA methylation to defend against transposable elements, mammals employ it for the sake of stage-, tissue-, or cell type-specific regulation of gene expression. It is most likely that the gene regulation by DNA methylation is also seen in all kinds of vertebrate animals, but not in other animals. How did the earliest vertebrates exploit DNA methylation for their unique gene regulation? To address this issue, we scrutinized RNA-seq data of a protochordate, *Ciona intestinalis*, not only to identify but also to characterize its promoter sequences. By comparing them to those of vertebrates', we can surmise what had happened in the promoters of the primordial vertebrates.

Keywords: promoter, transcription start site, CpG island, vertebrate, *Ciona intestinalis*

P-02

## Birth and death of genes linked to chromosomal inversion

Yoshikazu Furuta

Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo

Genome rearrangements such as inversion, duplication and gene conversion are important for adaptive evolution. Availability of closely-related complete genome sequences now helps to detect these and novel mechanisms of genome rearrangements. A Gram-negative human stomach pathogen *Helicobacter pylori* is known for its plastic genome and geographical differentiation. We sequenced 4 Japanese *H. pylori* strains and compared them with other 6 *H. pylori* strains, including European, West African and Amerind. We found copy number changes specific to East Asian strains in genes of outer membrane protein families. Some of these changes were mapped at an endpoint of a large genome inversion. Their sequence analysis suggested occurrence of DNA Duplication Associated with Inversion (DDAI), a novel mechanism of DNA duplication. All the large inversions found between these 10 *H. pylori* genomes were analyzed and their mechanism were classified into 4 types: (i) DDAI, (ii) homologous recombination at long inverted repeats, (iii) recombination at short inverted repeats, and (iv) inversion adjacent to a mobile element. Recognition of these inversion modes allowed reconstruction of synteny evolution in this species. This result may serve as a paradigm in analyzing long and short-term genome evolution in various organisms and in cancer cells through extensive DNA sequencing.

Keywords: genome comparison, gene duplication, genome rearrangement, replicative inversion, structural variation

P-03

## Important roles of coding small open reading frames (sORFs) hidden in plant genomes

Kousuke Hanada, Mieko Higuchi, Masanori Okamoto, Minami Shimizu, Kazuo Shinozaki, Motoaki Seki and Minami Matsui  
RIKEN, Plant Science Center

It is revealed that peptides translated from small open reading frames (sORF) play essential roles in multicellular organisms. However, small coding sequences tend not be identified in their genomes so that we may miss essential mechanisms at molecular level. Here, we identified novel 7901 sORFs with high coding potential hiding in a representative model plant species (*A. thaliana*). Toward functional analysis of these coding sORFs, we generated the expression atlas in 16 organs and 17 environmental conditions by our designed array. We found that 4664 coding sORFs were expressed in at least one experimental condition. In addition, the evolutionary conservation of each coding sORF was examined in 16 land plant species. Out of 7901 coding sORFs with homologous sequences to at least a plant species, 6516 are subject to purifying selection. There are 7321 coding sORFs with the evidence of transcription and/or purifying selection. We manually chose 472 coding sORFs, and generated overexpression mutants for each of 472 coding sORFs. Approximately ten percentage (48/472) of chosen coding sORFs induced visible phenotypic effects. Taken together, our finding suggests that a large number of coding sORFs hidden in plant genomes are associated with morphogenesis.

Keywords: Denovo genes, Plant, Comparative genomics, Transcriptome, Overexpression

P-04

## Difference in gene duplicability causes the difference in overall structure of protein-protein interaction networks among eukaryotes

Takeshi Hase

Dept. of Bioinformatics, Tokyo Medical and Dental University; The Systems Biology Institute

Protein-protein interaction networks (PINs) were suggested to be disassortative networks. In such networks, interactions between high- and low-degree nodes are favored while hub-hub interactions are suppressed. It was postulated that a disassortative structure minimizes unfavorable cross-talks between different hub-centric functional modules and was positively selected in evolution. However, in this study, we investigated several PINs from various eukaryotes (e.g., yeast, worm, fly, human, and malaria parasite) and showed that disassortative structures are not common features among eukaryotes. By simulation studies on the basis of a duplication-divergence model, we demonstrated that a preferential duplication of low- and high-degree nodes can generate disassortative and non-disassortative networks, respectively. From these results, we hypothesized that the difference in degree dependence on gene duplications accounts for the difference in assortativity of PINs among species. By comparing 55 proteomes in eukaryotes, we found that proteins with lower degrees showed higher duplicabilities for PINs with a disassortative structure, while opposite trend was observed for PIN without disassortative structure. These observations supported our hypothesis. Therefore, disassortative structures observed in PINs are merely a byproduct of preferential duplications of low-degree genes, which may be caused by an organism's living environment.

Keywords: protein-protein interaction networks, gene duplication, eukaryotes

P-05

## Molecular Evolution of the Bitter Taste Receptor Gene Family in Three Chimpanzee Subspecies

Takashi Hayakawa(1, Tohru Sugawara(1, Yasuhiro Go(1, Toshifumi Udon(2, Hirohisa Hirai(1 and Hiroo Imai(1

1)Primate Research Institute, Kyoto University, 2)Chimpanzee Sanctuary Uto

In mammals, bitter taste sense is mediated by TAS2R proteins which belong to a large family of G-protein coupled receptors. Previous study revealed that the TAS2R gene family in Western chimpanzees (*Pan troglodytes verus*) had high intrasubspecific nucleotide diversity comparative to autosomal noncoding regions. It was suggested that the haplotypes which might respond different repertoires of ligands had been evolutionally maintained with weak balancing selection. However, it is unclear whether the TAS2Rs has been adaptively evolved to environments, such as vegetation. We determined sequences of all 28 TAS2Rs in 10 Eastern chimpanzees (*P. t. schweinfurthii*) and two Central chimpanzees (*P. t. troglodytes*) and analyzed relationship between intrasubspecific diversity and intersubspecific divergence of them together with sequences of 46 Western chimpanzees which were reported previously. Like Western chimpanzees, Eastern and Central chimpanzees had comparatively high intrasubspecific diversity in the TAS2R family. Besides, the intersubspecific divergence is a little higher than the intrasubspecific diversity and most of haplotypes were not shared between subspecies. Therefore, compositions of the TAS2R haplotypes in each subspecies substantially seemed to differ. These results allow us to imagine that the TAS2R family in each chimpanzee subspecies might independently adapt to each environment, maintaining the subspecies-specific haplotype diversity.

Keywords: chimpanzee, subspecies, TAS2R family, adaptive evolution

P-06

## A speciation gene for left-right reversal in snails results in anti-predator adaptation

Masaki Hosoi

NCB Naturalis

How speciation genes can spread in a population is poorly understood. In land snails, a single gene for left-right reversal could be responsible for instant speciation, because dextral and sinistral snails have difficulty in mating. However, the traditional two-locus speciation model predicts that a mating disadvantage for the reversal should counteract this speciation. In this study, we show that specialized snake predation of the dextral majority drives prey speciation by reversal. Our experiments demonstrate that sinistral Satsuma snails (*Stylommatophora*: *Camaenidae*) survive predation by *Pareas iwasakii* (*Colubroidea*: *Pareatidae*). Worldwide biogeography reveals that stylommatophoran snail speciation by reversal has been accelerated in the range of pareatid snakes, especially in snails that gain stronger anti-snake defense and reproductive isolation from dextrals by sinistrality. Molecular phylogeny of Satsuma snails further provides intriguing evidence of repetitive speciation under snake predation. Our study demonstrates that a speciation gene can be fixed in populations by positive pleiotropic effects on survival.

Keywords: Speciation, Coevolution, Biogeography, Animal Behavior, Asymmetry

P-07

## Diversification of Activation Domain of Vertebrate Nr2e1

Ai Kamijo, Kei Yura, Atsushi Ogura

Ochanomizu University

A common ancestor of the current animal phyla appeared by the Cambrian explosion. In “light switch theory” hypothesis, the Cambrian explosion is presumably caused by the acquisition of eyes. The theory tells that predator-prey relationships changed dramatically due to the acquisition of eyesight by the predators, and that preys started to evolve in various forms by selection pressure. If the theory is correct, genes related to eye formation have been drastically changed in the era of the Cambrian explosion.

Nr2e1 is one of the Eye Field Transcription Factors, and is known to function in morphogenesis of eyes. We therefore searched for nr2e1 in all the available genome sequences of animal phyla and examined phylogenetic relationships to estimate the evolutionary process of nr2e1. Based on the phylogenetic tree, possible ancestral sequences and the protein three-dimensional structures of nr2e1 are deduced to find functionally critical changes in amino acid residues. From these predicted structures, the DNA-binding domain was found better conserved than the activation domain. This result raised the possibilities that the protein that binds to the activation domain of nr2e1 coevolves along with the changes of the activation domains.

Keywords: eye, evolution, eye field transcription factor

P-08

## The Differentiation of Sex Chromosomes in Eutherians and Marsupials

Yukako Katsura  
SOKENDAI

Mammalian sex chromosomes originated from pairs of autosomes, and homologous genes on sex chromosomes (gametologs) differentiated as a result of arrest of recombination between the chromosomes. In eutherians, this differentiation took place in a stepwise fashion. It is believed that the first two steps were generated in the ancestor of Theria (eutherians and marsupials). In marsupials, however, such differentiation has not been investigated yet. In this study, orthologous pairs of eutherian gametologs were identified in opossums. Phylogenetic analysis and the estimated divergence time of these gametologs revealed that they had differentiated simultaneously in the therian ancestor. A possible gene conversion between X and Y chromosomes was observed in eutherians. Moreover, to understand the mode of functional divergence of gametologs, a ratio of nonsynonymous to synonymous substitutions on a branch leading to each X and Y gametologs was examined. Based on the results, we concluded that 1) at least eight genes differentiated simultaneously in the therian ancestor, but gene conversion in eutherians reduced the nucleotide divergence between some gametologs and 2) some Y gametologs differentiated to attain a new function, such as sex determination or spermatogenesis, whereas some gene might have maintained its current function as it was on an autosome.

Keywords: Sex chromosome, Mammals, Gene conversion, Evolutionary strata

P-09

## Comparison of network structures of functional modules between the yeast and the mycoplasma protein interaction networks

Masataka Kikuchi, Soichi Ogishima, Naoki Masuda, Yoshihito Niimura,  
Hiroshi Tanaka  
Tokyo Medical and Dental University

In the protein interaction network (PIN), remarkable structures have been uncovered: scale-free, small-world and modular structures. Modular structure is the structure that proteins having similar functions interact with each other and form the functional module. Entities (proteins) composing functional modules have been evolutionarily studied, however, it is unclear how PIN structure evolve with functional modules. Here, we compared network structures of functional modules between the yeast and the mycoplasma PINs. First we determined functional modules whose proteins show high clustering and functional coefficients, and found that unexpected characteristic of functional modules in connection degrees for both yeast and mycoplasma PINs: proteins having middle connection degree interact with each other in the yeast PIN, whereas those having high connection degree do in the mycoplasma PIN. Indeed, those proteins constitute the same protein complexes (e.g., RNA polymerase, proteasome, anaphase-promoting complex). By comparing the ratio of the centrality of functional module proteins to that of other proteins between two species, we found that functional modules in the mycoplasma PIN centralize, whereas those in the yeast PIN do not. Our results suggest that the yeast PINs decentralized and diverged compared to the mycoplasma PINs, which may reflect expansion of functions from prokaryote to eukaryote.

Keywords: Modular structure, Functional module, Protein interaction network

P-10

## Ancient Domestication of Tyrosine Recombinase-encoding Crypton Family of DNA Transposons

Kenji Kojima

Genetic Information Research Institute

Domestication of transposable elements (TEs) led to evolutionary breakthroughs including the origin of telomerase and vertebrate acquired immune system. These breakthroughs were accomplished by adaptation of molecular functions essential for TEs, such as reverse transcription, DNA cutting and ligation or DNA-binding. Crypton represents a unique class of TEs using tyrosine recombinase (YR) to cut and rejoin the recombining DNA molecules. Here we report Cryptons from animals, fungi, and stramenopiles, as well as genes derived from Crypton domestication events. Human genome harbors 6 genes containing DUF3504 domain, which are conserved among jawed vertebrates. We found that DUF3504 domain was derived from Crypton YR. ZMYM2, 3, and 4 genes were duplicated through 2 rounds of genome duplication in early vertebrates, and are orthologs of the WOC gene, which is conserved among bilaterians. The domestication of Crypton leading to origin of the WOC gene over 910 million years ago, is the second oldest TE domestication event known to date. Many of DUF3504 genes are transcriptional regulators and the acquisition of DUF3504 domain could have added new regulatory pathways via protein-DNA or protein-protein interactions.

Keywords: transposon, domestication, Crypton, tyrosine recombinase,

P-11

## Bioinformatic analysis of variation in the 16S rRNA gene in search for universal primers in metagenomics

Kyungtaek Lim

The University of Tokyo

Background: Many universal primers for conserved regions on the 16S ribosomal RNA (rRNA) gene have been used to obtain partial or nearly full-length sequences of 16S rRNA genes from metagenome. However, it was hardly evaluated whether those primer sequences allow unbiased representation of bacterial population.

Purpose: We aimed to determine suitable primers for the 16S rRNA gene by bioinformatic analyses of all the available complete genome sequences of eubacteria.

Method: We retrieved full-length 16S rRNA gene sequences from completely sequenced eubacterial genomes, after filtering out the partial gene sequences. One reference 16S rRNA gene sequence was selected from each strain to construct a reference 16S rRNA library for sorting out minimal number of variations in each primer-binding site.

Major findings: Loss of the anti-Shine-Dalgarno (anti-SD) sequence in 12 bacterial strains disqualified a known universal primer-binding site (C10:1522-1541). We also identified many rearranged forms of the gene, such as deletion, substitution, transposon insertion, inversion, and partial duplication. Universality of many of the other primer-binding sites was also challenged by the high nucleotide level variations. Only 5 primer-binding sites (C4:515-531, C4:517-533, C7:1061-1074, C9: 1390-1407 and C10:1492-1507) showed sufficient conservation for design of universal primers. We designed primer formulas for those conserved sites to cover all the sequences in the reference 16S rRNA library.

Significance: The universal primer sequences we recommended are good subjects for experimental evaluation.

Keywords: 16S rRNA gene, Genome rearrangement, Intragenomic variation, Universal primer, Anti-Shine-Dalgarno sequence

P-12

## Effective Gene Collection from the Metatranscriptome of Marine Microorganisms

Mengjie Lin, Atsushi Ogura  
Ochanomizu University

Metatranscriptomic studies have not been popular due to problems with reliability, repeatability, redundancy and cost performance. Here, we propose a normalized metatranscriptomic method that is suitable for the collection of genes from samples as a platform for comparative transcriptomics.

We constructed two libraries, a non-normalized and a normalized library, from samples of marine microorganisms taken from Hiroshima bay during the day. We sequenced 0.6M reads for each sample on a Roche GS FLX, and obtained 0.2M genes after quality control and assembly. A comparison of the two libraries showed that the number of unique genes was larger in the normalized library. Functional analysis of genes revealed that ribosomal RNA genes and chloroplast genes were dominant in both libraries. Taxonomic distribution analysis of the libraries suggests that Stramenopiles form a major taxon that includes diatoms. These results are consistent with previous information that the diatom is a dominant species in this area and that genes related to photosynthesis might be active in the daytime.

Normalization of marine metatranscriptome could be useful in increasing the number of genes collected, and in reducing redundancies among highly expressed genes. Gene collection through the normalization method was effective in providing a foundation for comparative transcriptomics.

Keywords: metatranscriptome, marine microorganisms

P-13

## Habitat variability correlates with duplicate content of Drosophila genomes

Takashi Makino and Masakado Kawata  
Tohoku university

The factors limiting the habitat range of species are crucial in understanding their biodiversity and response to environmental change. Yet the genetic and genomic architectures that produce genetic variation to enable environmental adaptation have remained poorly understood. Here we show that the proportion of duplicated genes (PD) in the whole genomes of fully sequenced *Drosophila* species is significantly correlated with environmental variability within the habitats measured by the climatic tolerance and habitat diversity. Furthermore, species with a low PD tend to lose the duplicated genes owing to their faster evolution. These results indicate that the rapid relaxation of functional constraints on duplicated genes resulted in a low PD for species with lower climatic tolerance, and suggest that the maintenance of duplicated genes gives organisms an ecological advantage during evolution. We therefore propose that the PD in a genome is related to adaptation to environmental variation.

Keywords: habitat diversity, climatic tolerance, habitat distributions, evolvability, adaptation, duplicated genes

P-14

Designing microarray: chordate *Ciona intestinalis*, and its gene expression analysis of light-induced spawning

Hiroshi Matsumae

Tokyo Medical and Dental University

Microarray of *Ciona intestinalis* had already designed several years ago. However, gene models and functional annotations of *Ciona.intestinalis* and microarray technology have been updated in these days. So we developed new microarray for *Ciona intestinalis*. I will present one example how to design custom-made microarrays for animals which genomic resources are available by comparing two designed arrays. Moreover, I will show you the result of gene expression analysis focused on light-induced spawning of adult *Ciona intestinalis* using by the new tool.

Keywords: chordate, gene expression analysis, microarray, *Ciona intestinalis*, reproduction

P-15

Paralogous Conserved Non-coding Sequences Derived from the Two-round Whole Genome Duplications

Masatoshi Matsunami and Naruya Saitou

Department of Genetics, School of Life Science, The Graduate University for Advanced Studies (SOKENDAI), Japan

The evidence of the two-round whole genome duplications (2R WGD) occurred in the ancestral vertebrate genome is the existence of paralogous conserved synteny blocks within each vertebrate genome. These synteny blocks are bearing paralogs and conserved non-coding sequences (CNSs) that are candidate of cis-regulatory elements. The CNSs might play important roles in the vertebrate evolution. However, the relation between paralogs and paralogous CNSs is still unclear. Thus, detecting paralogous CNSs and inferring their relation is important to understand evolution after the 2R WGD.

Orthologous CNSs of each synteny block were detected by using previously described synteny data, and 7,924 CNSs are conserved among 8 vertebrate species. These orthologous CNSs were compared each other to detect paralogous CNSs. From these analysis, we found that there is weak correlation between the distribution of paralogous CNS and the distribution of paralogous genes. This result suggests that paralogous CNSs might contribute to paralogous synteny retention.

Keywords: Genome duplication, Non-coding region, Vertebrates

P-16

## Diversity of Olfactory Receptor Gene Repertoires among 38 Mammals

Yoshihito Niimura (1), Atsushi Matsui (2)

(1) Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University

(2) Primate Research Institute, Kyoto University

Olfaction, the sense of smell, is essential for the survival of animals. Odor molecules in the environment are detected by olfactory receptors (ORs) encoded by a large multigene family. Previous studies using the whole genome sequences showed that the numbers of OR genes are significantly smaller in higher primates including humans (<400) than most of other mammals examined (e.g., mice and rats have >1,000). To further investigate the diversity of OR gene repertoires among mammals, I extensively identified the OR genes from the draft genome sequences of 38 diverse mammals. Because the genome sequences of many species are at low coverage (<2x), I estimated the number of OR genes in the entire genome by using TraceArchive database. The results demonstrated that the estimated numbers of functional OR genes are extremely variable, ranging from only ~10 in dolphins to ~2,000 in elephants. Moreover, the number of functional OR genes and the fractions of pseudogenes are not correlated to each other, suggesting that the fraction of OR pseudogenes is a poor indicator to the olfactory ability of an animal. These results supports the hypothesis that OR gene families dynamically changed during evolution depending on each organism's living environment.

Keywords: Olfactory receptor, Multigene family, Gene duplication, Mammalian evolution,  
Comparative genomics

P-17

## Cultural Domestication of Chicken

\*Atsushi Ogura, \*Lin Mengjie, ΔTomoyoshi Komiyama

\*Ochanomizu University, ΔTokai University

Chickens were first domesticated from Red jungle fowl (*Gallus gallus gallus*) around in Southeast Asia. Domesticated chickens are currently used for various purposes. Since a long time ago, chickens may have fascinated the ancient human, not only because of the usefulness as potential foods but also because of their colorful appearance and characteristic songs. Fighting cocks, Shamo and long-crowing chicken, Naganakidori are typical examples of chickens that have been bred for purposes other than foods. The Shamo has much larger and more aggressive than other domesticated chickens. Naganakidori have been bred for the purposes of entertainment and appreciation. Therefore, the question arises as to what sides in these chickens are under selective pressures. We examined genes that are responsible for aggressiveness and crowing behavior by utilizing aCGH designed from chicken genome with seven samples of cultural domesticated chicken. From these results, we found that a number of genes are mutated in cultural domesticated chickens, and those mutations related to cultural domestication are different from those of broiler and layer.

P-18

## Molecular Evolutionary Analysis of Retrogenes in Green Algae.

Kazutaka Takeshita, Mathieu Deblieck, Dorota Buczek, Kanako O. Koyanagi, Hidemi Watanabe, Wojciech Makalowski

Graduate School of Information Science and Technology, Hokkaido University

The genomes of two green algae, one from unicellular green alga *Chlamydomonas reinhardtii* and the other from multicellular *Volvox carteri*, revealed both genomes, in spite of their fundamental differences in organismal complexity and life history, have surprising similarities at the molecular level. This indicated that the root of multicellularity, at least in multicellular green algae, can be associated with lineage specific modification of protein coding genes shared between two algae rather than acquisitions of new protein coding genes in the multicellular lineage. As a potential evolutionary driving force for the multicellularity in green algae, we focused on gene duplications by retrotransposition and its resultant retrogenes. We found that the amount of retrogenes in *Volvox* was significantly bigger than those in *Chlamydomonas* and have conducted molecular evolutionary analysis of retrogenes in both species. We show some interesting retrogenes in *Volvox* and discuss about retrogenes' contribution to the multicellularity in green algae.

Keywords: retrogene, retrotransposition, green algae, multicellularity

P-19

## Evolution of gene regulatory networks by fluctuating selection and intrinsic constraints

Masaki E. Tsuda, Masakado Kawata  
RIKEN Advanced Science Institute

Various characteristics of complex gene regulatory networks (GRNs) have been discovered during the last decade, e.g., redundancy, exponential indegree distributions, scale-free outdegree distributions, mutational robustness, and evolvability. However, it is not well understood whether these characteristics are the direct products of selection or those of other evolutionary forces such as mutational biases and biophysical constraints. To elucidate the causal factors that promoted the evolution of complex GRNs, we examined the effect of fluctuating environmental selection and some intrinsic constraining factors on GRN evolution by using an individual-based model. We found that the evolution of complex GRNs is remarkably promoted by fixation of beneficial gene duplications under unpredictably fluctuating environmental conditions and that some internal factors inherent in organisms, such as mutational bias, gene expression costs, and constraints on expression dynamics, are also important for the evolution of GRNs. The results demonstrated that various biological properties observed in GRNs could evolve as a result of not only adaptation to unpredictable environmental changes but also non-adaptive processes owing to the properties of the organisms themselves.

Keywords: gene regulatory network, gene duplication, genome architecture, evolution, fluctuating environment

P-20

## Lineage-specific Changes Within Genome Components Of Molluscs Revealed By Partial Genome Shotgun Sequencing

Masa-aki Yoshida, Atsushi Ogura  
Ochanomizu University

The cephalopods (squids and octopuses) provide an useful model to examine lineage specific genomic changes causing nervous system evolution of molluscs in which cephalopods developed complex eye structure and large-sized brain. We have projected the pygmy squid genome sequencing project because the pygmy squid have the smallest genome (2.1Gb) in the cephalopods. For the first attempt, we conducted comparative genome structure analyses of three molluscs, pygmy squid, nautilus and scallops using partial genome shotgun sequencing. We tried to find whole genome duplication that might produced large amount of novel functional genes and to extract lineage specific repetitive element (RE) that might caused genome size expansion. We, first, observed large-scale gene duplications in all the three molluscs. We, then, identified common and lineage-specific REs including 38,258 novel REs as well as RE frequencies by homology-based RE detection and de novo RE detection. Nautilus has the largest genome among three species, but also the smallest proportion of REs, indicating that the increase in genome size resulted not from RE expansion. Squid, on the other hand, has the largest proportion of REs but smaller genome size than nautilus even though they have complicated eye and brain.

Keywords: Genome sequencing, NGS, whole genome duplication, repetitive elements, cephalopods

P-21

## A codon substitution model that incorporates the effect of the GC contents, the gene density and the density of CpG islands of human chromosomes

Kazuharu Misawa  
RIKEN

### Abstract

Background: Developing a model for codon substitutions is essential for the analyses of protein sequences. Recent studies on the mutation rates in the non-coding regions have shown that CpG mutation rates in the human genome are negatively correlated to the local GC content and to the densities of functional elements. This study aimed at understanding the effect of genomic features, namely, GC content, gene density, and frequency of CpG islands, on the rates of codon substitution in human chromosomes.

Results: Codon substitution rates of CpG to TpG mutations, TpG to CpG mutations, and non-CpG transitions and transversions in humans were estimated by comparing the coding regions of thousands of human and chimpanzee genes and inferring their ancestral sequences by using macaque genes as the outgroup. Since the genomic features are depending on each other, partial regression coefficients of these features were obtained.

Conclusion: The substitution rates of codons depend on gene densities of the chromosomes. Transcription-associated mutation is one such pressure. On the basis of these results, a model of codon substitutions that incorporates the effect of genomic features on codon substitution in human chromosomes was developed.

Keywords: Genome sequencing, Mutation rate, codon substitution, male-driven evolution, CpG hypermutability, methylation

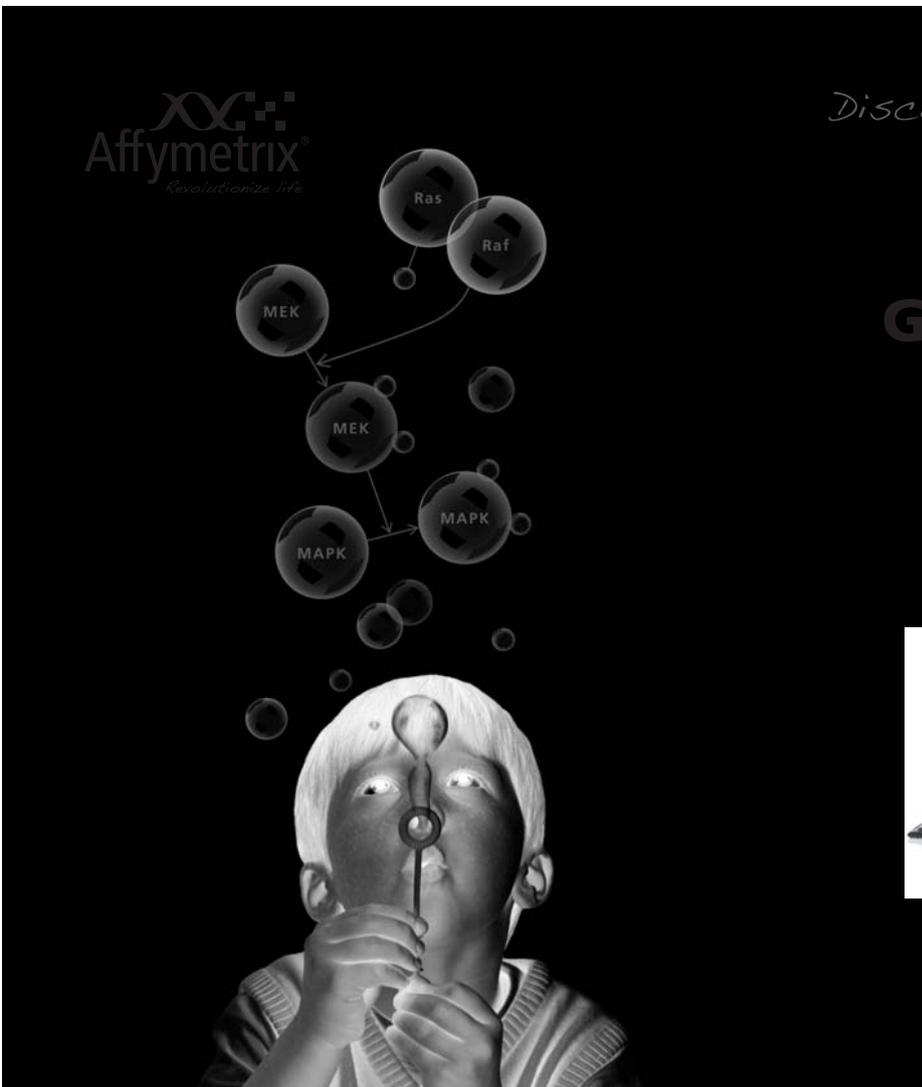
SRA Hacks: better ways to use public data of next generation sequencing.

Tazro Ohta

Research Organization of Information and Systems, Database Center for Life Science

Next generation sequencer (NGS) has now become one of the most important method for genetics. Sequence Read Archive (SRA) is a public database maintained by NCBI, EBI, and DDBJ to archive sequence data generated by NGS, but large amount of sequence files and complicated metadata files which describe the background of the data make the database too complicated to use. Here, we introduce our web services, Survey of Read Archives (SRAs: <http://sra.dbcls.jp/>) and Kusarinoko (<http://g86.dbcls.jp/kusarinoko/>). SRAs offers statistics of SRA to grasp what it contains, and the interface to search those data by the information of the published article. This enables to find the reliable data and it also provides more background information of the data which is needed for analysis. Then, we developed Kusarinoko, which is the interface to show integrated metadata and information from published article, for example, experimental design, sequencer platform, and library construction. These services will help researchers who want to use public NGS data for their research activity, and we will keep improving them by integrating other databases and web services.

Keywords: SRA Hacks: better ways to use public data of next generation sequencing.



*Discover how easy  
microarrays can be*

マイクロアレイを見近なソリューションに

## GeneAtlas® システム

6,300,000円(税別)

革新的な低価格と低ランニングコスト

簡単操作と半自動化でハンドリング時間は最小限  
アカデミア向けには1年間解析ソフトのライセンス付与

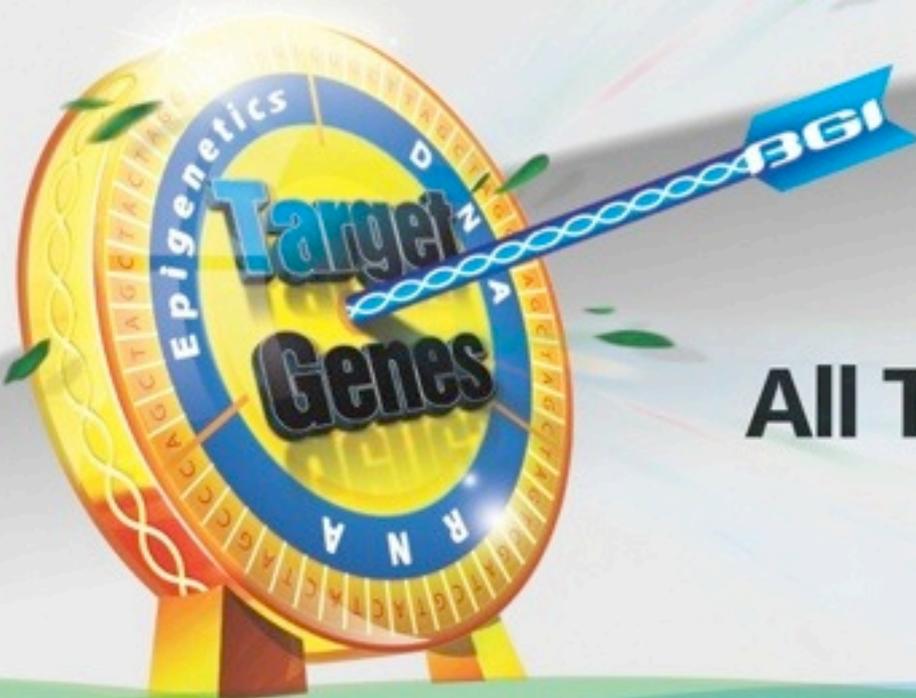
省スペース設計で場所をとりません

- 〈構成品〉 ・イメージャー
- ・自動洗浄・染色装置
- ・ハイブリダイゼーション装置
- ・コントロールPC
- ・バーコードリーダー



GeneAtlas® について、もっと知りたい方は資料をご請求ください。  
salesjapan@affymetrix.com

アフィメトリクス・ジャパン株式会社  
〒105-0013 東京都港区浜松町1-24-8 ORIX浜松町ビル7F  
TEL:03-6430-4020 FAX:03-6430-4021  
URL:www.affymetrix.com/jp/



# All Targeted Genes

## Genomics Solutions

Whole Genome Sequencing  
Exome Sequencing  
Target Region Sequencing  
High-throughput Genotyping  
Metagenomics

RNA-Seq (Transcriptome/Quantification)  
Small RNA Sequencing  
Non-coding RNA Sequencing  
Degradome Sequencing

ChIP Sequencing  
Reduced Representation Bisulfite Sequencing  
Whole Genome Bisulfite Sequencing  
MeDIP Sequencing

## Promotion for *de novo* Sequencing

\*High accuracy

\*High productivity

\*High efficiency

\*Low cost

\*Short turnaround time

Now the cost for *de novo* sequencing can be as low as **20,000 USD!**

Until 15<sup>th</sup> September 2011, Japan only (not including tax)

\*Note:

### This service includes:

1. Sequencing data: 60 folds (60Gb clean data)
2. Free data filtering

1. This is only applicable to the genome of which the genome size is 1G, the heterozygosity is less than 0.5% and GC content is between 35% and 65%.
2. We can also construct 200bp, 500bp, 800bp short-insert libraries and 2Kb, 5Kb, 10Kb and 20Kb long-insert libraries and provide assembly, annotation and other bioinformatics analysis, of which the cost is not included in the 20,000USD. Please contact us about the details.