Genetic mechanism underlying the parallel evolution of lip thickness in east African cichlids

One of the most interesting phenomena describing the evolution of east African cichlids is the occurrence of morphological parallelisms within the context of adaptive radiation. Understanding the molecular mechanism underlying morphological parallelism may be of primary importance in recent evolutionary biology. We are now focusing on the fleshy lip, which was evolved independently in each Lake, as a textbook example of parallel evolution. We constructed a hybrid cross of two Lake Victoria cichlid species, which are distinct in the degree of lip thickness. By using this hybrid cross, we performed a Quantitative Trait Locus (QTL) analysis to identify the genomic region(s) that affects the lip thickness. As a result, we detected one major QTL in which the candidate gene encoding one of the major component extracellular matrix proteins exist. We expect that the quantity of the deposition of this protein affect to the lip thickness.

Whole-genome Japanese Reference Panel and future directions

Tokyo University Tohoku Medical Megabank Organization (ToMMo) has constructed the whole-genome reference panel with 1,070 healthy individuals (1KJPN) who were attending to the prospective cohort project. The 1KJPN contains 21.2 million single nucleotide variants (SNVs) more than half of which are novel. The panel also contains millions insertions and deletions with the similar novelty rate. These genetic information will be an important infrastructure to explore the traits and disease especially for Japanese population and will accelerate the personalized preventive medicine and personalized healthcare in Japan. We have also created the Japanese population scale copy number profiles in genic regions and the medicine and personalized healthcare in Japan. We have also created the panel also contains millions insertions and deletions with the similar novelty rate. These genetic information will be an important infrastructure to explore the traits and disease especially for Japanese population and will accelerate the personalized preventive medicine and personalized healthcare in Japan. We have also created the genome of the archaean Thermoplasma acidophilum. Here, we provide evidence that TacXerA is a critical recombinase that resolves chromosome dimers in T. acidophilum. ChIP-Seq analysis with anti-TacXerA antibody identified two TacXerA binding sites (named as dif1 and dif2) on the chromosome; this finding was confirmed by real-time qPCR. Consistently, in vitro experiments revealed that a dimer plasmid containing dif2 was resolved by purified TacXerA into two monomer plasmids. The catalytically important nucleotide sequences were identified by an in vitro inter-molecule recombination assay with plasmids containing various mutations in the dif2 sequence. These studies strongly suggest that TacXerA is indeed a functional Xer recombinase involved in dimer resolution of the chromosome at the dif2 site.

Molecular analysis of the site-specific recombination mediated by XerA from Thermoplasma acidophilum

Xer-mediated site-specific recombination plays a critical role in the segregation of circular chromosomes after DNA replication. It has been shown that two tyrosine-recombinases, XerC and XerD, form a complex that resolves chromosome dimers in Escherichia coli. In contrast, only a single orf (referred to as TacXerA) sharing similarity with the common Xer recombinases has been found in the genome of the archaean Thermoplasma acidophilum. Here, we provide evidence that TacXerA is a critical recombinase that resolves chromosome dimers in T. acidophilum. ChIP-Seq analysis with anti-TacXerA antibody identified two TacXerA binding sites (named as dif1 and dif2) on the chromosome; this finding was confirmed by real-time qPCR. Consistently, in vitro experiments revealed that a dimer plasmid containing dif2 was resolved by purified TacXerA into two monomer plasmids. The catalytically important nucleotide sequences were identified by an in vitro inter-molecule recombination assay with plasmids containing various mutations in the dif2 sequence. These studies strongly suggest that TacXerA is indeed a functional Xer recombinase involved in dimer resolution of the chromosome at the dif2 site.

General sessions (ISYR-1 – ISYR-6, 1A-01 – 3F-09)

Fission yeast CENP-T nclosseuome promote the isochromosome formation in centromere

Histone-like complexes in combination with other CENP proteins and localizes to centromere, mediating the attachment to microtubules. However, little is known about its effect on gross chromosomal rearrangement (GCR) that occurs in centromere. Here, we found in fission yeast that CENP-T determines the modes of GCR in centromere: isochromosome formation and translocation. The M447T mutation in the histone-fold domain impaired the binding of CENP-T to CENP-W, and decreased its association to centromere. To our surprise, the CENP-T mutation decreased the spontaneous rate of isochromosome formation, while it increased that of translocation that occurs between different centromeres, suggesting a preference for the isochromosome formation. The CENP-T mutation increased the rate of gene conversion in centromere, and caused the focus formation of the Rad52 recombinase in centromere. Our data suggest that CENP-T complexes promote the isochromosome formation through the regulation of spontaneous recombination between centromere repeats.
Genome-wide identification and classification of G protein-coupled receptors in fish genomes based on conserved domain signatures

G protein-coupled receptors (GPCRs) prevalently exist in eukaryotes and responsible for sensing different kinds of stimuli via the interaction with specific signaling molecules. They are well characterized in important model organisms but poorly characterized in recently sequenced fish genomes.

We present a genome-wide prediction and collection of members in 10 GPCR families in 31 fish genomes using a bioinformatics approach based on the known protein domain signatures. We found that certain fish genomes possess much higher numbers of members in certain GPCR families compared to other genomes. The finding implies that some fish require many GPCR genes in these families for their lifestyle.

We also conducted a detailed analysis of visual opsin genes. By comparing phylogeny and composition of neighboring genes, we found several new gene duplication events that have not been reported previously and linked them to the genome duplications specific in some fish families.

Generally speaking, our results reflected the importance of some GPCR families in certain fish species and can serve as a good starting point for future studies of specific GPCR-mediated signaling pathways.

Brain gene expression profiling in the fire ant, Solenopsis invicta: division of labor and colony form

Differential gene expression in the brain may underlie some behavioral differences among individuals. In eusocial insects, the behavior is different amongst castes: queens reproduce and workers work. While, in the fire ant, Solenopsis invicta, workers from single (monogyne) versus multiple (polygyne) queen colonies behave differently. We aim to identify the brain genes that are differentially expressed depending on their caste or colony form. To this end we have dissected brains from virgin queens and workers at two stages: pupae and adults. Four biological replicates of each sample were obtained. With respect to colony form, we have isolated 7 biological replicates of RNA from the brains of adult workers from monogyne and polygyne colonies. These samples are subjected to RNA-sequencing. A preliminary analysis comparing the two castes (queens vs. workers) revealed more than 16,000 (17% of total) and 14,000 (15% of total) transcripts differentially expressed in the adult and pupal stages, respectively. We found about 2% of those differently expressed transcripts are transposon transcripts which may contribute to the behavior of worker and virgin queens.

Study on the origin of fibromelanosis using genetic comparison between Indonesian Cemani chicken and other domesticated chickens

Fibromelanosis (Fm) phenotype, the hyperpigmentation observed in the skin and internal organs in the two chicken breeds, Cemani and Silky has been caused by the duplication of a genomic segment including EDN3. At EDN3 that is responsible for the melanocyte movement, a pair of particular haplotypes was always found in the two breeds. The haplotypes were diverged ca.1.2±0.4 mya, suggesting that the presence of the haplotypes far before the domestication of chicken. The investigation of the 417kb sequence flanking to EDN3 revealed the hitchhiking effect has acted on the duplicated region.

Accumulation of genetic incompatibilities across cycles of isolation and migration

Speciation can be viewed as the establishment of Reproductive Isolation (RI) between two groups of organisms. Although RI comes into various forms, it is genetic incompatibility that once evolved, the speciation process is irreversible. We focus on the evolution of a Dobzhansky-Muller incompatibility (DMI) between populations in one migration cycle model increases the likelihood of fixation of incompatible alleles for both simple and complex DMI genetic architecture compared with CM model. We use a simulation method and find that MIM cycle model increases the rate of speciation compared with the CM model. Our results demonstrate that isolation periods help the DMI-type RI to evolve, which is critical to the ultimate fate of incipient specie.
Estimation of allele frequency of pathological variants based on whole-genome sequencing of 1070 Japanese individuals

Tohoku University Tohoku Medical Megabank Organization (ToMMo) have sequenced whole genomes of 1,070 cohort participants, and constructed the whole-genome reference panel (1KJPN). We started partial public release of whole-genome Japanese reference panel, and opened a website “integrative Japanese Genome Variation Database (iJGVDB) http://ijgvd.megabank.tohoku.ac.jp/”. Variants of 1KJPN were annotated with biological and medical information, and overlaps between the SNVs and known pathological SNVs from the Human Gene Mutation Database (HGMD) were identified. We calculated individual variant load for disease-causing mutations and stop-gained (nonsense) variants for 1KJPN individuals. The estimation with high-confidence SNV set showed that on average one individual had 11.2 disease-causing variants (9.6 as heterozygous and 1.6 as homozygous) and 53.7 stop-gained variants (41.6 as heterozygous and 12.1 and homozygous). These estimates were very similar with those in other East Asian populations, and we did not find any signature of strong population bottleneck in history of Japanese population.

Spread of reduced activity of STX promoter throughout Great Journey

STX is an enzyme responsible for the synthesis of polysialic acids on neural cell adhesion molecule, and plays an important role in human brain function. Variants of the STX gene have shown association with various mental disorders. Here we focus on the four SNP combinations (SNP types) composing of three SNPs in the STX promoter region and one (“CGC”) of the four shows significantly lower promoter activity than others. Population genetic analysis of haplotype sequences of the 10 kb region surrounding the three SNPs from 63 human individual samples from ethnic groups reveal that the ancestor of all the types emerged about 0.58 MYA, and the CGC type expanded about 0.1 MYA, coincidentally with the time of African exodus. Analysis using SNP data from 1000 genome project showed that this CGC type is prevalent in Asia, and shows high homozygosity in a region extending for 18 kb around CGC SNPs with sharp boundaries, suggesting a possible signature of natural selection. Interestingly, some particular haplotypes, belonging to non-CGC types, show also relatively high frequency but low diversity in each population and raise a possibility of some global selection on this promoter region.

Transcriptome analysis of early embryos of interspecific avian hybrids

Interspecific F1 hybrids can be obtained from crosses between male chickens (Gallus gallus domesticus) and female Japanese quails (Coturnix japonica) using artificial insemination; however, most of hybrid embryos die before hatching. Our previous study showed that developmental arrest occurs frequently at the pre-primitive streak stage in F1 hybrid embryos. To study the molecular basis of this hybrid incompatibility, we performed mRNA-sequencing analysis for blastoderms at stage X and stage XIII-XIV of chickens, Japanese quails, and their F1 hybrids and then compared expression levels of 11,000 genes at each stage among chickens, quails, and F1 hybrids and between two stages in each species and F1 hybrids. In 31 genes, expression levels increased from stage X to stage XIII-XIV in parental species but decreased or did not change significantly in both chicken and quail alleles in F1 hybrids. Two primitive streak formation-related genes were included in the 31 genes, suggesting that the defect in primitive streak formation possibly cause the early developmental arrest of F1 hybrid embryos.

Keyword(s): Hybrid inviability mRNA sequencing Aves
Hybrid sterility and functional diversification of a chromatin-binding protein in threespine sticklebacks

Hybrid sterility is one of the major reproductive barriers in speciation. However, genes causing hybrid sterility have not been identified in natural vertebrate. In present study, we screened genes causing hybrid sterility in Japanese three-spine stickleback species pairs, Gasterosteus aculeatus and Gasterosteus nipponicus, which recently diverged but are reproductively isolated even in sympatry. Hybrid male sterility was reported between the two species. In a QTL region that we previously found associated with hybrid male sterility, we found a gene that is rapidly evolving, encodes a chromatin-binding protein, and is expressed in the testis. Rapid evolution and chromatin-binding are common features of genes causing hybrid sterility in Drosophila and mouse. To identify functional differences of the candidate gene between the two species, we conducted histone binding assays in vitro and cell localization assays in a zebrafish cell line. The in vitro assays indicated that G. aculeatus protein bound to a modified histone characteristic of heterochromatin, but G. nipponicus protein did not bind to that. The cell assays indicated that the candidate gene showed a different localization pattern in the nuclei between proteins of the two species. These results suggested that the function of the candidate gene is different between the recently diverged species. We are now investigating the roles of this gene in hybrid male sterility using CRISPR/Cas9 genome editing technology.

Analysis of genetic diversity of black woodpecker based on mitochondrial DNA

To estimate the genetic diversity of the black woodpecker Dryocopus martius, nucleotide sequences of the control region of mitochondrial DNA were analyzed using skeletal muscle, feathers, or intestinal mucosal cells from feces as the source of total genomic DNA. The total length of the control region was 1,187 or 1,188 bp. Eleven haplotypes were detected in 7 and 15 samples obtained from Hokkaido and Honshu, respectively, in Japan. Furthermore, nucleotide diversity (π; Nei 1987) and haplotypic diversity (h; Wenink et al. 1993) were calculated. The π values were 0.282 and 0.216, and the h values were 0.952 and 0.647 in the Hokkaido and Honshu samples, respectively. Despite the larger sample size from Honshu, both the π and h values in the Honshu samples were obviously lower than those in the Hokkaido samples. In addition, the h in the Honshu samples was nearly equal to that in the samples from the Northern Europe population of the pink-footed goose (0.645 ± 0.038; Ruokonen et al. 2005). These findings indicate that inbreeding depression would be a concern for this species population in Honshu.
The utility of genetic difference considering insertion and deletion

The Kimura two-parameter model (K2P) is one of the most widely used substitution models for estimating genetic differences (generally called evolutionary distances). Nucleotide changes seen during the evolutionary process include substitution, insertions and deletions. Although K2P is appropriate in some applications of nucleotide substitutions, it is desirable for evolutionary models of molecular sequences to include insertions and deletions in addition to substitutions. Therefore we extended K2P by considering gaps (insertions and/or deletions) and introduced a measure for estimating genetic difference between two nucleotide sequences in terms of nucleotide changes that have occurred during the evolutionary process. For all gene sequences of the mitochondrial genome in many animal species including humans, we constructed phylogenetic trees by calculating genetic differences in our model and K2P, and then evaluated the performance of the two genetic differences by comparing the topological distances of each tree from the tree based on accurate stratigraphic constraints. Our results indicate that how important it is to take into account the evolution by insertions and deletions.

Analysis of non-coding region proximal to the alpha operon in Bacillus subtilis

In bacteria, ribosomal protein genes often constitute an operon, and their expression is regulated through an autogenous feedback mechanism. The alpha operon of Escherichia coli consists of the genes coding for ribosomal proteins and alpha subunit of RNA polymerase (S13-S11-S4-α-L17), and its expression is regulated by S4 protein. On the other hand, the alpha operon of Bacillus subtilis includes about 150 bp non-coding region instead of the gene coding for S4 (S13-S11-α-L17). Therefore, it is assumed that regulation mechanisms of expression of the alpha operon in B. subtilis is different from that in E. coli. From our analysis, it was suggested that non-coding region located upstream of rpoA was involved in regulation of alpha operon in B. subtilis. We further analyzed the role of this non-coding region in its regulation.
Phenotypic research of *Bacillus subtilis* strain possessing minimum number of sigma factors

Sigma factor, which is involved in the initiation of transcription of bacteria, is one of subunits of RNA polymerase. *Bacillus subtilis* possesses 19 kinds of sigma factors, and function of individual sigma factor has been studied from the view of stress response and spore development, however, comprehensive understanding of whole sigma factors multiplied is not elucidated yet. In our former studies, these sigma factors have been inactivated simultaneously one by one, ultimately the strain, SigA only, in which all sigma factors other than the essential sigma factor, SigA, were inactivated, was obtained. The SigA only strain showed the representative phenotypes observed as lysis of colonies being transparency and rapid decrease in the number of the viable cells in colonies when growing on LB agar medium. Currently, we have succeeded in acquisition of the suppressor strains, which restored rapid decrease in the number of the viable cells. From analysis of these suppressor strains, we are trying to find factors responsible for the phenotypes of the SigA only strain.

Analysis of transcriptional regulation of *sigI* in *Bacillus subtilis*

In *Bacillus subtilis*, cell membrane contains 10% of glycolipids, which are synthesized by UgtP, and they play important roles as well as phospholipids in vivo. *ugtP* disruptant cells are bent and distended, and the activity of SigI, which is regulated by its cognate membrane protein RsgI, is activated in these cells. These findings suggest that glycolipids are required for cell shape maintenance. In this study, we focused on the *sigI* promoter activity in *ugtP* disruptant cells for elucidation of transcriptional regulation mechanism of *sigI* from SigI activation in these cells as a clue. We constructed *lacZ* transcriptional fusions with a series of *sigI* promoters which have deletions or mutations. In *ugtP* disruptant cells, the *sigI* promoter activity caused by both SigI and WalR, which is response regulator cognate with WalK sensor kinase of the essential two component system in *B. subtilis*, were activated. This suggests that glycolipids are involved in both activities of SigI and WalKR two component system in *B. subtilis*. 

Reconstitution of cyanobacterial RNA polymerase in *Bacillus subtilis* cell

To investigate the features of cyanobacterial RNA polymerase (RNAP), we planned to reconstitute cyanobacterial RNAP in *Bacillus subtilis* cells. The genes coding for *B. subtilis* sigma factors were eliminated from the genome and *rpoD1* (major sigma factor) of cyanobacteria was introduced into *B. subtilis*. Growth of the strain was retarded comparing with that of wild type. This suggested that the RpoD1 inhibited the formation of holo-RNAP with the competition with host's sigma factors or disturbed transcription of *B. subtilis* genome. The entire genes of cyanobacterial holo-RNAP were integrated and their transcription was induced. The expression of RpoA and RpoC2 was detected from the Western blotting analysis. Nevertheless, the obvious effect of the introduced genes on the phenotypic aspects of the strain was not observed. By the inactivation of the expression of host's RNAP, further analysis of complementation and function of heterologous RNAP are under consideration.
Difference in dominated mitochondrial CO1 haplotypes was observed in Bombyx mandarina populations from Hokuriku to Kyushu and its surrounding islands.

According to law concerning the conservation and sustainable use of biological diversity through regulations on the use of living modified organisms, we have continued to collect the information on genetic diversity maintained in the wild mulberry silkmoth, Bombyx mandarina populations in order to assess the effect of transgenic domesticate silkmoth on biodiversity of B. mandarina.

As a result, more than 5000 mitochondrial CO1 sequences were determined. In this study, we used CO1 sequences from one population from Toyama, 2 from Ishikawa, 4 from Fukui, 1 from Shiga, 2 from Kyoto, 2 from Hyogo and one from Tottori to assess differential occurrences in CO1 haplotypes among populations. At the 325th site, T was almost fixed in populations west of Fukui Takahama. The border can be drawn except for Takahama. In contrast, C was much frequently found in populations east of Fukui Takahama. The border is divided in the Wakasa area of Fukui prefecture. In addition, we analyzed the Osaka Toyono and Nara populations and found C was abundant at the 325 site. Incorporating Kyushu and its surrounding island populations into analysis, we surveyed geographical change of occurrences in CO1 haplotypes of B. mandarina.

Molecular phylogenetic study of the subgenus Sophophora (Diptera: Drosophilidae) and its related taxa based on 50 nuclear gene loci

The subgenus Sophophora is the largest groups in the genus Drosophila and comprised of 348 species including Drosophila melanogaster. Recent molecular studies suggested that this subgenus contained the genus Lordiphosa and Hirtodrosophila duncani. In addition, Drosophila "longicurra", the undescribed species from Malaysia, probably belongs to Sophophora. To reveal the phylogenetic relationships in the subgenus Sophophora and its related taxa, 52 species were analyzed using the nucleotide sequences of 51 nuclear gene loci. In the present study, the phylogenetic tree showed the high node support values at each branch, and Lordiphosa formed the sister clade of the willistoni and saltans species groups. However, H. duncani was placed at the basal branch of the subgenus Sophophora. Drosophila "longicurra" and its sibling was placed between the (melanogaster + obscura) and the (willistoni + saltans) clades, and recognized as the new species group. In the melanogaster species group, D. majtoi, endemic to Philippines and had never studies from molecular point of view, was relatively close to the ananassae subgroup, but recognized as the new species subgroup. The suzukii subgroup was divided into four groups including two newly established species subgroup, named as the lucipennis and uninspectata.

Sequence diversity of Drosomycin genes within Drosophila lutescens

Antimicrobial peptides (AMPs) are major innate immune mechanisms in insects. Seven AMPs have been identified in Drosophila melanogaster. Among them, Drosomycin is known to be a major AMP having seven copies (Drs, Dro2, Dro3, Dro4, Dro5, Dro6, and Drs-I) in the genome. Previous studies showed that there are variations in copy number and in expression pattern in Drosomycin genes among several species closely related to D. melanogaster. These variations should have evolved in a short evolutionary time by repeated gene duplications. In this study, to infer the mechanism of the gene duplications occurred at a high frequency, we compared the nucleotide sequences among sixteen strains of D. lutescens collected in the TMU campus, since this species was known to have many Drosomycin gene copies in the previous studies. Comparing the sequences of the region including the protein coding regions of Dro2, Dro3, Dro4 and Dro5, we found many insertions and deletions among these sequences and their breakpoints tend to be in AT-rich sites. We also found that there are variations in the copy number of Drosomycin genes within species as well as among species.
Feminization and masculinization have accelerated pseudogenization on neo-sex chromosomes in *Drosophila miranda*

Many Y chromosomes are degenerated after their origination. However, early evolutionary processes of sex chromosomes remain obscure. I have therefore investigated the degeneration processes of the neo-sex chromosomes that emerged in *Drosophila miranda* about 1 million years ago. I found that pseudogenization is accelerated in not only neo-Y but also neo-X after their emergence. Extensive comparisons of gene expression also revealed that the pseudogenes on neo-X and neo-Y tend to have possessed male- and female-biased functions, respectively, when they were located on the ancestral autosomes. Therefore, feminization and masculinization may have accelerated pseudogenization on both neo-X and neo-Y.

Using Pooled RNA-seq, detecting adaptive evolution for cold temperature of *D. albomicans*

Pooled RNA-seq is a useful method to obtain SNP information in exons and estimate gene expression levels at a low cost. It can be applied to non-model organisms for which genome sequence data is unavailable. In the last few decades, *D. albomicans* has expanded the distribution to the temperate zone. In our previous studies, it has been shown that the temperate population has a stronger cold tolerance than the tropical population. Therefore, in this study, we tried to detect the genes responsible for the stronger cold tolerance in the temperate population from their digestive organs, as they feed and breed on fermented fruits and sap, where a variety of bacteria and fungi propagate. *Drosophila* flies defend themselves from invading microorganisms with innate immune system. In our previous studies, we found that *Drosophila virilis*, which feeds on slime fluxes and decaying parts of trees, is more resistant to *Penicillium* fungus infection than *D. melanogaster*, which feeds on fermented fruits. To clarify the immune mechanism responsible for the difference in the antifungal resistance, we compared the expression patterns of immune-related genes in gut and in fat body between the two species in response to the infection of *Penicillium* fungus. We found that antimicrobial peptides and lysozymes synergistically act against the infected fungi in *D. virilis* unlike the antifungal immune response in the fat body of *D. melanogaster*. These results indicate that the immune system has been substantially differentiated between *D. melanogaster* and *D. virilis*.
Genome history of people on Japanese Archipelago

Ainus, Mainlanders, and Okinawans live on Japanese Archipelago. We have been conducting genome-wide SNP data analyses of these people. We found that Ainus and Okinawans are tightly clustered followed by Mainlanders and Koreans in phylogenetic tree of East Asians and Ainus are genetically different from Mainlanders living in Tohoku. Neighbor-net network based on Fst distances among seven geographical area populations shows that Tohoku shares a short split with Okinawa, suggesting genetic affinity between Emishi people of Tohoku and ancestral Okinawans. We recently examined Izumo population in Shimane Prefecture using both Affymetrix 6.0 and Japonica Array. Izumo people were genetically more apart from Koreans than Japanese Mainlanders in Tokyo. This suggests existence of some heterogeneous migrants to Japanese Archipelago Mainland. We thus propose new hypothesis, “inner dual structure”. Two phases of migrations are assumed to form current Japanese Mainlanders: phase 1 migrants arrived at Archipelago in late Jomon period, while phase 2 migrants came to Northern Kyushu and later spread eastward, and they brought paddy field rice agriculture.

Biological processes enriched in gene with long poly-Q may be pathomechanism. Confirmed that long poly-Qs are distinct from other repeats in between length polymorphism and poly-Q expansion disease, and causing loci in healthy subjects. Thus, this study suggests association poly-Qs causing expansion diseases are longer than non-disease- is mainly molecular binding function. This study also shows that the pathomechanism of neurodegeneration observed in repeat expansion diseases. To obtain a clue about this at the whole gene level, we tried to extract functional feature of poly-glutamine (poly-Q) repeats with length polymorphism using data of short tandem repeats collected from the human gene databases, H-InvDB and VarySysDB [Pre- sented at the 85th meeting]. All disease-causing poly-Q had length polymorphism that was found in one fifth of total genes containing poly-Q. Genes with polymorphic poly-Q exhibited significant enrichment of biological processes of GO terms regarding apoptosis and nervous system development, but not in those with monomorphic poly-Q. However, no significant difference in molecular function was found between genes with and without length polymorphism, which is mainly molecular binding function. This study also shows that the poly-Q causing expansion diseases are longer than non-disease-causing loci in healthy subjects. Thus, this study suggests association between length polymorphism and poly-Q expansion disease, and confirmed that long poly-Qs are distinct from other repeats in pathomechanism.

Patterns of Measures of Divergence Between Populations

FST was introduced by Wright (1951) and it is used as a convenient measurement to detect genetic differentiation. While the concept of FST is widely accepted, different authors proposed different methods of calculation and there is some disagreement among estimates. To evaluate concordance and discordance among measurements, we compared the distribution and tempo- ral change of those measures.

In this study, we conducted coalescent simulation under a simple divergence model and calculated several FST-related measures from simulated data. As genetic markers, DNA sequences, SNPs and microsatellites were considered. We found differences among estimates depending on the definition and assumed marker types even when data are generated under the same population model. It was also found FST calculated from DNA sequence data has narrower distribution than those from SNPs and microsatellites. We also evaluated the effect of ascertainment on the distribution of FST when SNP typing data were used. Three types of SNPs data were considered; random SNPs, SNPs ascertained in a small discovery panel and highly heterozygous SNPs. Distribution was shifted toward higher values when the ascertainment SNPs were used, especially when SNPs were selected based on their heterozygos- ity. Therefore, careful attention should be paid when different FST measures are compared.
Experimental evolutionary study based on big data of mouse germline mutations from ENU mutagenesis

Due to extremely low mutation rate, it is virtually impossible to observe “on-going” evolutionary process in the mammalian genome. Typical evolutionary studies are based on “backward” analysis starting from extant data. But there are little means to directly prove a model by “forward” analysis: i.e. it is extremely difficult to conduct experiments to elucidate evolutionary mechanisms.

One of the few ideas to conduct an “experimental evolution” study is to accelerate the evolutionary rate by a feasible mutagen. N-ethyl-N-nitrosourea (ENU) is a widely-used mutagen. In case of mice, ENU can induce one de novo mutation in ~700 kb per gamete. The ENU-mutagenesis has been devised as a powerful tool for the forward genetics. However, the countless raw mutations, which are completely trackable in the genealogy, can be used as excellent on-going evolutionary data by virtue of their broad spectra.

By using a total of 3,426 ENU-induced de novo mutations in mouse germ cells, we analyzed mutation patterns and found bidirectional mutation pressure in the mammalian genome, which is consistent with Sueoka’s mutation pressure theory. This finding may shed light on the enigmatic isochore evolution.

Identification of genetic loci associated with tame behavior using selective breeding and genome-wide analysis in mice

Tame behavior is one of the major elements in domestication and defined as increased interaction of animals with human. To identify genes associated with active tameness which is defined as contacting human hand (contacting), we performed selective breeding for contacting using wild-derived heterogeneous stock (WHS). WHS is a mixed population derived from 8 wild mouse strains originated in various geographic regions. After the selection of 8 generations, we obtained over 50K SNPs and behavioral data and performed three types of analysis, (1) identification of selected SNP by using simulation based on non-selection model, (2) identification of selected region by using statistic which is a population differentiation score between selection and control population (XP-CLR), and (3) identification of SNP associated with contacting by using Genome-wide association studies (GWAS). We identified 32 genetic regions as a candidate region associated with tameness. One genetic locus on chromosome 10 were exceeded the genome-wide threshold by XP-CLR and GWAS, suggested that the locus should be associated with contacting.
Pathways including Xrs2-FHA–dependent Tel1 activation

The MRX complex ensures NHEJ fidelity through multiple


Arachis

Tetrasomic recombination in allotetraploid Arachis

Cultivated peanut (Arachis hypogaea) is an allotetraploid (AABB) with 2n = 4x = 40. It has been assumed for peanut that homoeologous recombinations between the A and B genomes are suppressed while homologous recombinations are occurred as diploid-like disomic manners. In this study, we found the homoeologous recombination is frequent in progeny of a cross between cultivated peanut and an induced allotetraploid derived from probable ancestral diploid species. We suggest a novel finding would have a significant impact on the genome and partly disomic and partly tetrasomic genetic behavior. This finding would have a significant impact on the genome and partly disomic and partly tetrasomic genetic behavior. This finding would have a significant impact on the genome and partly disomic and partly tetrasomic genetic behavior. This finding would have a significant impact on the genome and partly disomic and partly tetrasomic genetic behavior.
Sensitivity of mgs1-18 rad18

A genetic screen for genes that suppress the temperature evisiae recombination, and centromere architecture. of cellular processes, including DNA replication, homologous

identified 7 genes or elements, which are involved in a variety

by using a yeast genomic library. This screen successfully

that suppress the temperature sensitivity of mgs1

Previous studies have shown that

polymerase/exonuclease reactions by archaeal replicative DNA

Effects of the PCNA-DNA interactions on the DNA polymerase/exonuclease reactions by archaeal replicative DNA polymerase

DNA replication is accomplished by the orchestration of the several enzymes. This complicated process is mainly coordinated by DNA clamp (Proliferating Cell Nuclear Antigen; PCNA). PCNA stabilizes and accelerates the reaction of replicative DNA polymerase by tethering the polymerase with the substrate DNA. Replicative DNA polymerase comprises two independent domains, Polymerase (DNA-synthesis) and Exonuclease (DNA-correction (degradation)) domains, and switches the DNA interacting domain by altering the relative arrangement to PCNA molecule. In the two DNA polymerase/PCNA/DNA complex structures in polymerase and exonuclease modes, PCNA seems to hold the DNA substrate, whereas the DNA polymerase dynamically alters its relative arrangement to the others. In this study, we constructed several alanine-substituted mutants of the basic residues inside the PCNA ring presumably responsible for the DNA interaction, and examined the polymerase and exonuclease reactions performed by DNA polymerase with PCNA. As a result, the alanine-substituted mutants of PCNA exhibited the substantial enhancement only in the exonuclease reaction.

A genetic screen for genes that suppress the temperature sensitivity of mgs1-18 rad18α cells in Saccharomyces cerevisiae

Progression of the replication fork is often impeded by DNA lesions caused by exogenous or endogenous DNA-damaging agents. If these stalled forks are bypassed incorrectly, it can lead to genome instability, which associated with tumorigenesis in humans. Stalled replication forks activate the DNA damage tolerance pathway including Rad18 ubiquitin ligase, which allows replication to resume without removing the replication-blocking lesion through its role in PCNA ubiquitination. Saccharomyces cerevisiae Mgs1, a member of the AAA⁺ class ATPase family, is highly conserved from bacteria to human. Previous studies have shown that mgs1Δα is synthetically lethal with rad18α, indicating that at least one of the two gene products is required for cell survival. To better understand this essential function, we carried out a multicopy suppressor screen for genes that suppress the temperature sensitivity of mgs1-18 rad18α cells by using a yeast genomic library. This screen successfully identified 7 genes or elements, which are involved in a variety of cellular processes, including DNA replication, homologous recombination, and centromere architecture.
Roles of Drosophila orthologs of human male-infertility genes in spermatogenesis

Spermatogenesis is a complex process requiring many genes. About 700 genes are currently known to be involved in this process of Drosophila; however, it is very likely that many still remain to be discovered. Based on the Gendoo disease database, 405 genes are suspected to cause human male infertility and 107 of them have Drosophila orthologs. In this study, with an aim to identify new spermatogenesis genes in Drosophila, we investigated the roles of 103 Drosophila orthologs of human male-infertility genes in spermatogenesis by using four GAL4 drivers and 156 RNAi lines. From this screening, we found that RNAi knockdown of 74 genes significantly decreased male fertility with at least one GAL4 driver, out of which 58 are newly identified genes involved in male fertility.

Elongation of synaptonemal complex in meiosis regulates crossover numbers between homologs dependent on chromosome-size

During meiosis, at least one crossover (CO) between homologous chromosomes must be assured on each homolog pair to ensure proper segregations in meiosis I. For this, meiosis-specific chromosome structure synaptonemal complex (SC) component ZMM (Zip-Mer-Msh) complexes play an important role for CO control as well as SC elongation between homologs. However, the function of SC elongation in the CO control is still unknown. Recently it was reported that a novel SC component, Gmc2-Ecm11 complex, facilitates SC elongation via Zip1 polymerization. Then, we focused on gmc2 or ecm11 mutations as a convenient tool to examine the function of SC elongation. To investigate the role of SC elongation in CO regulation, we analyzed genetic CO frequency in the gmc2 null mutant. While we observed reduction of CO frequency in short chromosomes, we observed increase of CO frequency in long chromosome. This result indicates that there would be a specific control mechanism to vary CO formation dependent on chromosome-size through SC elongation. Furthermore, cytological analysis indicates that this chromosome-size-dependent control of CO formation might occur in the middle stage of meiotic recombination.
X chromosome inactivation in the mouse embryos deficient for SmcHD1

In mammals, one of the two X chromosomes of female cells is inactivated in early embryogenesis for dosage compensation between the sexes. It is thought that X chromosome inactivation is established and maintained by epigenetic modifications of DNA and histones. Homozygous mutation of SmcHD1, a member of SMC family proteins, has been shown to cause derepression of X-inactivated genes in post implantation female mouse embryos and their subsequent lethality at the mid-gestation stage, suggesting a role of SmcHD1 in the maintenance of X inactivation. A recent study showed that the Barr body in human cultured cells becomes decondensed upon depletion of SMCHD1 although the X-inactivated state is not affected. In this study, we find that X-inactivated genes become sporadically reactivated later on in the embryo deficient for SmcHD1, but are stably silenced in immortalized embryonic fibroblasts, in which X inactivation has been fully established and maintained, in the absence of SmcHD1. Taken all together, we suggest that SmcHD1 would be rather required for the process of the establishment of X inactivation, whose defects would not allow the maintenance of X-inactivated state later on.

The role of DNA methylation in Epigenetic barrier between Naïve and Primed state

Mouse ESC and EpiSC are at distinct pluripotent states representing different developmental stages, and they are classified as naïve (mESC) and primed (EpiSC) type stem cells. These two stem cells are known to have different DNA methylation patterns and recruitments of transcription factor (TF), and such differences are thought to contribute to form an epigenetic barrier between naïve and primed state. To ask if DNA methylation is needed to establish the barrier, we used Dnmt3a/3b knockout (DKO) ESCs that lack de novo DNA methylation and induced differentiation of the ESC to EpiSC. Intriguingly, DKO ESC could differentiate into EpiSC, which showed characteristics similar to wild-type EpiSCs. We examined recruitments of OCT3/4 and SOX2 to several loci by ChIP and found that although the DKO-derived EpiSCs have hypomethylated genome, TF recruitments can occur properly at all the gene loci examined. Our study suggests that recruitments of OCT3/4 or SOX2 can occur independently from DNA methylation and may provide a new framework to understand the epigenetic barrier that distinguishes two states of mammalian pluripotent stem cells.

Differential expression of Xist in the mouse preimplantation embryo

The dosage difference of X-linked genes between the sexes in mammals is compensated for by genetically inactivating one or the other X chromosomes in XX females. A noncoding RNA transcribed from the Xist gene at the onset of X-inactivation coats the X chromosome in cis and induces chromosome-wide heterochromatinization. We have created a new Xist allele (Xist<sup>CAG</sup>) driven by a constitutively active CAG promoter. The paternal transmission of Xist<sup>CAG</sup> resulted in the preferential inactivation of the targeted paternal X (Xp) not only in the extraembryonic, but also embryonic lineage, whereas maternal transmission ended with embryonic lethality at the early postimplantation stage with a phenotype that resembled mutant embryos carrying a maternal deficiency in Tsix in both sexes. Interestingly, we found that the upregulation of Xist<sup>CAG</sup> in preimplantation embryos temporarily differed depending on its parental origin with the paternal allele starting at the 8-cell stages, whereas the maternal one at the blastocyst stage. These findings may indicate that the maternal Xist allele may manifest a chromatin structure inaccessible by transcription factors relative to the paternal allele.
The prediction of the effect of CpG around transcription start site on sex-biased gene expression

It is expected that epigenetic control has a critical role for sex differentiation, since male and female use almost identical genome. In teleost fish, it is shown that transcriptional control depending on DNA methylation is important for sex differentiation. In this study, to predict influence of DNA methylation for sex differentiation, we investigated the relationship between CpG distribution around transcription start site (TSS) and male to female gene expression ratio of guppy (*Poecilia reticulata*) that has remarkable sexual dimorphism. As a result of this analysis, we found that sex-biased genes and non-biased genes had different value of CpG Observed/Expected (CpGO/E) in downstream of TSS, and furthermore sex-biased genes with low CpG O/E had large biased expression. This study provides an evidence that CpGs in TSS downstream region has important roles in transcriptional controls.

DNA methylation of the germ cells in Japanese hagfish, *Eptatretus burgeri*

It is generally believed that genomic DNA is conserved in all cells of multicellular organisms throughout their development and differentiation. The various species of Japanese hagfish containing Eptatretus burgeri (Cyclostomata, Vertebrata), are known to be eliminate a fraction of their chromosomes during early embryogenesis. To elucidate how the cells distinguish the chromosome which is eliminate or not, we try to examine the DNA methylation patterns in the germ cells. As a result, 5-methylcytosine and 5-hydroxymethylcytosine, which is the marker of cytosine demethylation, were equally distributed on the all chromosomes in spermatogonia, whereas 5-methylsytosine was exclusively decreased on eliminated chromosome in spermatocytes. Since 5-hydroxymethylcytosine in spermatocyte was observed on all chromosomes, we concluded that methylation and demethylation activity is equal in spermatogonia and demethylation was promoted on eliminated chromosomes in spermatocytes. These results suggested that DNA cytosine methylation may act as a epigenetic marker to decide the eliminated chromosomes in early embryogenesis.

Diversity of P element piRNAs production and mRNA expression in natural populations of *Drosophila melanogaster*

In *Drosophila melanogaster*, germline abnormalities caused by transposition of P element, is known as P-M hybrid dysgenesis. P element mobilization is prevented by P element piRNAs in crossing between P strains. 60kDa products and KP polypeptides also suppress P activity. However, in natural populations, the suppression by these components is complicated, especially the repression by piRNAs is largely unknown. Therefore, we compared suppression by P element piRNAs and mRNA in natural populations, using 7 lines of Q strain and 2 lines of M’ strain with RNA-seq and qRT-PCR. In embryos, we found significant differences in piRNAs production and mRNA expression among the lines. These two values were negatively correlated. In addition, the weak susceptibility of M’ strain may associated with low level of piRNAs production in germline and high level of KP element mRNA in embryos. In conclusion, we firstly proposed that the quantity of piRNAs was one of the essential factor influencing phenotypes in P-M system of natural strains, and that Q strains showed diverse expression of silencing components.

HDAC regulates a novel mechanism for plant drought tolerance

Histone modification is important information for post-translational regulation in eukaryotes. To survive under environmental stress, plants use shrewd strategies. Histone modifications regulate gene activity in response to abiotic stresses in plants. *Arabidopsis* histone deacetylase 6, HDA6 has multi functions in control of genome maintenance, environmental stress responses, plant development and hormone signaling pathways. Here we show that HDA6 regulates acetate fermentation pathway involved in drought stress tolerance. Furthermore, our data indicates that connaturally plants have the dexterous system using acetic acid biosynthetic pathway to tolerate water deficiencty. Acetic acid production is induced via glycolysis by metabolic flow change especially under drought condition. HDA6 directly binds and regulates acetic acid biosynthesis genes in Arabidopsis. By the treatment with appropriate concentration of acetic acid, monocots and dicots enhance drought tolerance.
The analysis of suppression mechanism in *AtRE1*, a copia-type retrotransposon found in *Arabidopsis thaliana*

The expression of retrotransposon is activated by some stress, although it is usually suppressed by epigenetic silencing mechanisms. We aim to analyze these activation and suppression mechanisms using *AtRE1*, a copia-type retrotransposon found in *Arabidopsis thaliana*. Thus, we investigated the transcriptional quantity of RNA and the promoter activity by means of Real time PCR technology and GUS staining assays in associated mutant plants. Experimental results up to now suggest that the genotoxic stress, UV irradiation and under methylation of DNAs activate the transcription of *AtRE1*, however the transcription of sense RNA in mutant plants associated with RdDM has not been found. Methylation of DNA is a common epigenetic signaling tool that cells use to lock genes in the "off" position. In order to examine the effects of inhibiting methylation on *AtRE1* expression, a experiment using 5-azacytidine was performed in *Col* Wild-type seeds. The experimental results show that when seeds were transferred from 5-aza included medium to drug-free medium, the drug induced transcriptional level was decreased as time went by.

Imprinted genes show differential expression which primarily from either maternally or paternally inherited alleles. In plants, genomic imprinting primarily occurs in the endosperm of developing seeds. The purpose of this study is to identify the candidate imprinted genes in *Brassica rapa*.

Searching candidate imprinted genes in *Brassica rapa*

In this study, we assayed genome-wide gene expression pattern by performing high-throughput sequencing of RNA derived from endosperm of developing seeds. The same development stage of seeds in *Arabidopsis thaliana*.

Sensitivity of *Escherichia coli* lactose repressor to the inducer molecule

SrS RNA, also called tmRNA, plays a crucial role in bacterial major ribosome rescue system, translation. Among the phenotypes sssA-deficient *Escherichia coli* cells show is delayed response in lactose operon induction. Lactose repressor is one of the natural targets of translation and its C-terminally deleted version is produced in sssA-deficient *E. coli* cells. We have previously shown that those truncated lactose repressor show reduced sensitivity to the inducer and proposed this is causative of the delayed response. Here we performed calorimetric analysis and found that C-terminal truncation reduced the affinity of inducer molecule to lactose repressor, supporting our model.
Analysis of the activation mechanism of an extracytoplasmic function σ factor σV in the Bacillus subtilis cells lacking glucolipids

The Bacillus subtilis cells lacking glucolipids by disruption of the ugpP gene show aberrant morphology and induction of the three extracytoplasmic function (ECF) σ factors, σM, σV, and σX. These ECF σ factors are regulated directly by their respective cognate transmembrane anti-σ factors, which sequester the σ factors to the membrane. The σV is induced by lysozyme. When B. subtilis cells are challenged by lysozyme, the anti-σV are cleaved by regulated intramembrane proteolysis (RIP). However, we could not observe the proteolysis of the anti-σV in the cells lacking glucolipids. Moreover, deletion of the rnsP gene encoding the second site cleavage enzyme of the anti-σV abolished activation of σV by lysozyme, but not activation by the lack of glucolipids. These results suggest that the lack of glucolipids induces the σV by another mechanism. The lack of glucolipids may lead to the conformation change of the anti-σV.

Targeting preference and expression profile of human micro-RNAs

MicroRNAs (miRNAs) are short noncoding RNAs that modulate the transcriptome mainly through post-transcriptional repression. To know miRNAs' targeting tendency, we examined expression profiles of miRNAs and mRNAs in representative normal human organs using publicly available RNA-seq data sets. The results showed that, in each of the organs examined, highly expressed miRNAs tend to avoid targeting highly expressed genes, while lowly expressed miRNAs tend to have target sites in highly expressed mRNAs. These targeting features were distinct for miRNAs of old origin.

Evolution and diversity of receptor as the sensor of the environment and essential protein for survive

Sensory receptor genes recognize various ligands in their environments and thus are assumed to be genetically diverged compared to house-keeping (or vital) genes that function to maintain life or reproduction. To understand the relationship between ability to bind various ligands and the genetic diversity among receptors, I classified human genes into two groups: environmental sensory genes (E-genes) and vital genes (V-genes). I collected 11 E-genes and 24 V-genes from human and mouse genomes. Then, I conducted analyses of their evolutionary rates (dNdS) and the extent of genetic diversity within the global human population. So far, averages of evolutionary rates show no significant difference between two groups. However, the average proportion of variable amino acid sites in human E-genes is significantly higher than that of V-genes. This observation suggests either the relaxation of functional constraint for E-genes or positive selection for increasing diversity of E-genes in humans.

Time-dependent directional change in oligonucleotide composition found by big data analyses on zoonotic RNA viruses

Viruses always pose significant threats to public health, as exemplified by the ebolavirus threat highlighted recently. To face world-wide serious threats caused by infectious viruses (e.g. Ebola, influenza, and MARS), we should innovate advanced technologies including big data analysis, e.g. a large-scale word count of oligonucleotides. Importantly, the above-mentioned zoonotic RNA viruses have a very high evolutionary rate. By using BLSOM for oligonucleotide composition in influenza virus genomes, we revealed directional changes of oligonucleotide composition that have occurred during human-to-human transmission after invasion from nonhuman sources [1,2]. By conducting a large-scale word count on oligonucleotide compositions, we have found the directional sequence changes of Ebola, MARS and influenza virus genomes that should become important for designing diagnostic PCR-primer and therapeutic oligonucleotides. 1) Iwasaki et al. (2011) DNA Res. 18, 125-136. 2) Iwasaki et al. (2013) Novel bioinformatics strategies for prediction of directional sequence changes in influenza virus genomes and for surveillance of potentially hazardous strains, BMC Infect Dis. 21, 386.
Molecular evolutionary characteristics of mitogenome in Ctenophora

To date, genome data were published in only 2 species of Cydippida and Lobata, Ctenophora last year, because DNA analysis of Ctenophora is extremely difficult. It was reported that many gene loss/gain were recognized between genomes of the 2 species and their circular mitogenomes are extremely small, 10 ~ 11 kbp in size. In this study, we analyzed 15 species collected in Japan to clarify molecular evolutionary characteristics of mitogenome in Ctenophora. We extracted DNA from mitochondrial fraction of living individuals, and inferred their mitogenome structures and phylogenetic relationships among the species by using 85 orthologous genes of transcription, translation, and replication systems and mitochondrial IRs. We also estimated the molecular phylogenetic trees of Ctenophora using 85 orthologous genes. Our results were as follows: 1) it was strongly suggested that Ctenophora was the earliest lineage within Metazoa. 2) Porifera-Placozoa clade made a cluster with Ctenophora. The linear mitochondrial genome (mitogenome) of Medusozoa has characteristics unique among those in multicellular animals, and inferred their mitogenome structures and phylogenetic relationships among the species based on partial mitochondrial gene sequences. The analyses showed that gene rearrangements occurred even between sibling species in the same genera. It is presumed from the phylogenetic relationships that the gene rearrangements occurred during short evolutionary time.

Evolution of TRP gene copy number in Echinoderms

TRP (Transient Receptor Potential) gene super family is ion channels with six-transmembrane domain. Some of them are temperature sensing channels and called as thermotriangles. The copy number of TRP in the genome differs from species to species and even an orthologous pair of thermotriangles from closely related species shows different range of temperature for their sensing. In this study, we investigated the number of TRP gene copy in two Echinoderms, sea urchin (Hemicentrotus pulcherrimus) and starfish (Patiria miniata). We extracted RNAs from larvae of H. pulcherrimus and P. miniata and conducted RNA-seq. Using amino acid sequences of known TRP genes, we performed hmm search to identify candidate TRP genes in these two species from their transcriptomes. It turned out that H. pulcherrimus has at least one TRPV, four TRPMs and seven TRPAs. It also turned out that P. miniata has at least TRPV, six TRPMs and two TRPAs. From phylogenetic tree, we assume that one copy of TRPA of H. pulcherrimus belongs to TRPAI clade while the other copy belongs to Basal clade.
Phylogenetic relationship of coelacanths, lungfishes, and tetrapods

Lobe-finned fishes are close relatives of tetrapods and coelacanths and lungfishes are the two extant lineages in this group of fishes. Despite the importance in understanding the origin of tetrapods the phylogenetic relationship among the coelacanths, lungfishes, and tetrapods has been controversial in morphological studies and in studies with molecular data. Recently, owing to the genome sequencing of coelacanth, two studies [Amemiya et al. 2013; Liang et al. 2013] reconstructed the sister relationship of lungfishes and tetrapods with high statistical support. In this study we analyzed the data used in these studies and our newly collected data set. In our result, the lungfish-tetrapod sister relationship was strongly supported when cartilaginous fishes were used as outgroup, as was done by the previous studies, whether or not ray-finned fishes are included. However, when only ray-finned fishes as outgroup, all the three possible relationships of the three taxonomic groups were generated depending on the phylogeny construction methods, substitution models and data sets used, and the statistical supports were mostly low.

Molecular Evolution of ATP Synthase in Notothenioid Fish

Fish belonging to the perciform suborder notothenioidi have adapted to cold waters of the Southern Ocean by producing AFGP (antifreeze glycopeptide). And significant amino acid changes in ATP8/6 proteins has been observed (Papetti, 2007). To understand the evolution of ATP synthase in notothenioids, we analyzed all subunits of ATP synthase encoded in nuclear and mitochondrial genome. In most subunits, the ratios of the rates of nonsynonymous substitutions to synonymous substitutions (dN/dS) is higher in the branches of notothenioids than the other branches. The statistical test implemented in CODEML showed some subunits encoded in nuclear genome contained sites under positive selection. In addition, we found parallel amino acid changes in two α subunits, whose duplication took place before the divergence of protacanthopterygiids and neoteleostei. The number of parallel amino acid changes observed is significantly more than the random expectation. These results have suggested that adaptive evolution occurred in some subunits of ATP synthase, while the functional constraints on most subunits of ATP synthase were relaxed in notothenioid lineages.

Analysis of population structure of Cryptomeria japonica based on amplicon sequenced data

Cryptomeria japonica is a coniferous tree species and broad natural distribution in the Japanese archipelago. There are two varieties distinguished by their morphological traits: the ura-sugi variety (C. japonica var. radicans) has slender branchlets with soft leaves and is mainly distributed on the Japan Sea side, while the omote-sugi variety (C. japonica) has rough branchlets with hard leaves and is distributed on the Pacific Ocean side. Because each variety consists of two groups of populations, C. japonica comprises 4 groups of populations. In this study, 94 samples from 4 natural populations (Ajigasawa, Bijodaira, Shimowada, Yakushima), each representing one of the 4 groups, were used to determine sequences at 142 nuclear genes using next generation sequencing. The total length of the analyzed sequences was 76,757bp, and we found 2,053 segregating sites. The average nucleotide diversity was 0.0036 and higher than those obtained previously. The average values of Tajima’s D were different among populations, indicating their different demographic histories. The FsT values and the UPGMA tree suggested that the Yakushima population first diverged from the other populations.

An introgressive hybrid between wild Vigna species found in a limestone rock mountain

V. exilis lives only in steep-edged limestone hills in Southeast Asia. In contrast, V. umbellata is found in various places such as mountain ranges and lowlands except the limestone environments. As such, it was surprising to find a population of V. umbellata-like plants in a limestone hill. However, we found some V. exilis-like phenotypes in the population, and because of its habitat, we suspected that it is a hybrid between V. umbellata and V. exilis. The analysis using chloroplast DNA and nuclear DNA clearly revealed it is a hybrid. The identical chloroplast DNA sequences between V. umbellata and the hybrid indicated the first step of the hybrid formation was a pollination of V. exilis pollen to V. umbellata. In addition, according to the dominance of V. umbellata-type alleles at nuclear SSR loci, the hybrid had probably undergone at least one round of backcross by V. umbellata. The SSR analysis also indicated the hybrid has not undergone ploidy change, since most of the loci were fixed with single alleles.

As such, this is the first case of homoplod hybrid in genus Vigna, and one of the few cases in the angiosperms.
Divergence of RNA editing in chloroplast genome among Arabidopsis species

RNA editing is found in organelle genomes in higher plants that exchange cytosine to uracil after transcription. We surveyed divergence of RNA editing among Arabidopsis species. Two of 34 previously known RNA editing sites in *A. thaliana* were lost in *A. lyrata* ssp. *lyrata*. In *A. lyrata* ssp. *lyrata*, there are 5 novel RNA editing sites compared to *A. thaliana*. Among the 8 variations of RNA editing sites, five losses in *A. thaliana* might have occurred by nucleotide substitution in chloroplast genome sequences. Other three changes could be caused by mutations in nuclear genome, possibly by loss of function in PPR genes. A editing site (89th site in *rpl23*) loss in *A. lyrata* was found to be occurred in OTP80.

Characterisation of major histocompatibility complex class II (MHCII) in Japanese frogs

MHCII molecules recognize and present extracellular pathogens and there is substantial interest in understanding the role of MHCII and other immune genes involved in immunity against chytridiomycosis, a serious fungal disease of amphibians caused by the chytrid fungus. In Japan, most endemic frogs are not infected with chytrid, suggesting that they are resistant against the fungus, with MHCII in the immune system or antimicrobial peptides in the skin playing a role. We examine such immune genes in native Japanese frogs including the Japanese brown frog (*Rana japonica*). Effects of nutrient on cellular ROS level and oxidative DNA damage in *Escherichia coli*

The maintenance of genetic information is important for organisms to conserve species. However, spontaneous mutation occurs at very low frequency by genetic mistransmission, being involved in evolution, genetic disease, and carcinogenesis. In previous studies on mutagenesis using *Escherichia coli*, bacterial cells logarithmically dividing in rich media were used for the investigation. However, such laboratory conditions are quite different from actual growth environment for *E. coli*. In most of the case, the environment is poor in nutrient and the cells grow much slower. The nutrient condition affects the metabolism including energy production and protein synthesis. A major cause of spontaneous mutation is oxidative DNA damage caused by reactive oxygen species (ROS) produced in oxygen respiration, in particular hydroxyl radical. For example, hydroxyl radical generates 8-oxoG by oxidizing guanine in DNA, and if unreppaired the 8-oxoG induces G:C to T:A base substitution. In this study, we examined effects of nutrient on ROS-mediated mutagenesis in *E.coli* cells. So far, we obtained evidence that nutrient does affect 8-oxoG-mediated mutagenesis and cellular level of ROS. We are now examining a possibility that nutrient changes intracellular free Fe²⁺ concentration.
Identification of novel oxidative stress response gene in *Saccharomyces cerevisiae*

Reactive oxygen species (ROS) can damage lipids, protein, and DNA. For isolation of novel oxidative stress response gene, we constructed gene disruption strains in *Saccharomyces cerevisiae*, and selected by H$_2$O$_2$ sensitivity. Mutant strain showed sensitivity to not only H$_2$O$_2$, but also to alkylating agent (MMS). Furthermore, the mutant showed higher spontaneous mutation frequency and higher intracellular ROS level after H$_2$O$_2$ treatment. Increased ROS sensitivity and ROS level was canceled in ogg1 double deletion mutant, but MMS sensitivity was not complemented. These results implies the existence of ROS detoxify pathway induced by oxidative DNA damage repair.

Studies on the de-ubiquitination mechanism for the xeroderma pigmentosum group C protein in UV damage response

The xeroderma pigmentosum (XP) is a rare genetic disorder characterized by a skin cancer susceptibility, which is more than 1,000 times higher than the general population. Most XP patients harbor mutations in genes involved in nucleotide excision repair (NER), which removes a wide variety of DNA damage including UV-induced cyclobutane pyrimidine dimer (CPD) and 6-4 photoproduct (6-4PP). The XPC protein plays a key role in recognizing DNA damage for global genome NER. Previously, we reported that XPC undergoes mostly reversible ubiquitination mediated by the CRL4$^{DDB2}$ E3 ligase upon UV irradiation. However, the biological significance of this ubiquitin turnover on XPC has remained unclear. Here we identified a candidate enzyme that is responsible for de-ubiquitination of XPC by a siRNA library screen targeting 118 human de-ubiquitinating enzymes (DUBs). Expression of the mutant DUB lacking its catalytic activity showed a defect in removal of UV-induced CPDs from the global genome, whereas 6-4PPs were repaired normally. Taken together, our results uncover possible roles and molecular mechanisms of XPC de-ubiquitination in cellular responses to UV-induced DNA damage.

The effects of AP sites repair enzymes on *Ciona intestinalis* embryogenesis

In this study, we focused on Apurinic/apyrimidinic (AP) sites, which induce genomic instability if left unrepaired. AP sites are produced continuously in living cells even under normal conditions and AP site repair system has been thought to be important especially during the early embryogenesis. In spite of the significance of AP sites repair, it is not entirely clear that how AP sites are repaired in vivo and how AP sites accumulation impact on living body. Therefore, we have been investigating the AP sites repair mechanism during the early developmental stage by using *Ciona intestinalis*, which is a model organism in developmental biology. In this presentation, we report the in vitro analysis of *Ciona intestinalis* AP endonucleases, which are the key enzymes in repairing AP sites, and the effects of *Ciona intestinalis* embryo when inhibiting the AP endonucleases.
The Analysis of Functional Relationship between Oxidation Resistance 1 (OXR1) Gene and Oxidative Defense Genes in *C. elegans*

The OXR1 gene was identified as a human gene that contribute to defense against oxidative stress, by using the ability to suppress the spontaneous mutator phenotype of *E. coli* *mutH* *nth* strain. The OXR1 gene is highly conserved among eukaryotes, but its function still remains uncertain. In this study, we tested whether the OXR1 gene has DNA repair activity toward oxidative base damage. We purified the human OXR1 protein, and tested its binding and cleavage activity to some double-stranded oligonucleotides containing an oxidative base damage. The human OXR1 protein binds to and cuts those containing 8-oxoG and Tg, but does not bind to those containing AP site mimic and normal bases only. Next, we examined *in vivo* relationship between the OXR1 gene and oxidative DNA repair genes, using *C. elegans*. The OXR1 mutation recovered the developmental delay phenotype of *exo-3* strain, which is a class of genes the product of which mutation recovered the developmental delay phenotype of OXR1 gene and oxidative DNA repair genes, using *C. elegans*. The OXR1 mutation recovered the developmental delay phenotype of *exo-3* strain, which is a class of genes the product of which functions as AP endonuclease. These results suggest that the OXR1 gene defends DNA against oxidative stress by repairing oxidative base damages as DNA glycosylase.

The role of mismatch repair system in the integrity of germline genome

As well as DNA polymerase fidelity, mismatch repair (MMR) is greatly contributing to the accurate DNA replication. To clarify the role of MMR on mammalian germline mutation, we analyzed *de novo* germline mutations with use of parents-child trio sample of *Msh2* deficient mice by whole exome sequencing. We found elevated frequency of *de novo* mutation, which include base substitutions, indels at dinucleotide repeat sequence. The spectrum of observed mutations in this study is consistent with it of somatic mutations that has been reported in previous studies. These results indicated that *Msh2* is required for the stable transmission of genetic information to the next generation.

The role in BER of *DIMT1L* and *TFB1M*, human homologs of KsgA, DNA glycosylase of *E.coli*

Oxidative stress damages DNA and induces mutation and stop of DNA replication, leading to apoptosis or cancer. Base Excision Repair (BER) is one of the systems which prevent such aberrations. DNA glycosylase cuts damaged base and makes AP site in the first step of BER. Subsequently, some of them work as AP lyase which cleaves DNA strand at AP site. KsgA was originally identified as 16S methyltransferase in *E.coli*. Previously, we found that KsgA have DNA glycosylase activity and AP lyase activity which removes 5-hydroxycytosine and cytosine opposite to oxidized thymine. KsgA have 2 human homologs, *DIMT1L* and *TFB1M*. To reveal whether these homologs are multifunctional, we investigated the presence of DNA glycosylase activity of *DIMT1L* and *TFB1M*, revealed that they have DNA glycosylase activity which excises 5-hydroxycytosine and cytosine paired with oxidized thymine last year. Additionally, it turned out that *DIMT1L* and *TFB1M* have AP lyase activity this year. At present, we study involvement of *TFB1M* in BER *in vivo*.
Mathematical model which can estimate the mutation frequency induced by radiation, considering dose rate explicitly

We propose a mathematical model by which we can estimate the damage of living object caused by artificial radiation exposure. In our theory, the dependence of mutation frequencies on the dose rate is critically important to predict both the time course and the stationary effect of the DNA mutation in cell cycles. Our theory shows the saturation of mutation frequencies, which marks a substantial difference from existing theories based on the total dose. The mutation frequency increases linearly in time, first, i.e., it depends linearly on the total dose. Then it starts to bend and approaches a stationary value, which is determined by the dose rate. In this way, our model can reproduce not only the linear behavior found in high dose-rate experiments, but also the deviation from it which is found in low dose-rate experiments. Importantly, we have derived a scaling function from our rate equations that predicts a universal feature in the mutation frequency of living organisms. In this study, we have analyzed the experimental data of five living organisms; mouse, drosophila, chrysanthemum, maize, and tradescantia. Despite the difference between animal and plant, all these data reasonably fall on a single line for our scaling function.
Analyses of spatiotemporal regulatory mechanisms of mes-endoderm segregation in zebrafish

Three primary germ layers (endoderm, mesoderm, and ectoderm) are established from multi-potent cells by activation of the germ layer specific genes in vertebrates. In zebrafish, it is thought that blastoderm cells that receive large amount of Nodal ligand differentiate into endoderm cells, and the Nodal signaling from yolk syncytial layer (YSL) and enveloping layer (EVL) plays critical roles in the endoderm specific gene activation. However, spatiotemporal patterning of endoderm differentiation still remains to be elucidated. We found that endoderm cells are emerged from the most marginal cells in the blastoderm adjacent to YSL and EVL in zebrafish embryos. In this study, we are now focusing on the functions of extra-embryonic cells (YSL and EVL) in endoderm cell differentiation.

Identification of genes related to the light-induced degradation of Chlamydomonas clock protein ROC15

In the unicellular green alga Chlamydomonas reinhardtii, RHYTHM OF CHLOROPLAST (ROC15), one of the clock proteins, is rapidly degraded after light exposure and this process is thought to be related to the phase-resetting mechanism of circadian clock (Niwa et al, PNAS, 2013, 110, 13666-13671). This process is sensitive over a wide range of wavelengths, especially red and violet light. However, since there are no known photoreceptors showing such absorption spectrum in Chlamydomonas, unknown light-signaling pathways may involved in this process. In this study, we tried to identify components of this process by a forward genetic analysis using a ROC15-Luciferase fusion protein as a reporter for the light-induced degradation of ROC15. We screened 14,200 clones from an insertional mutant library and isolated 10 mutants. One of them showed a defect in a red/violet-light-specific manner. We identified candidate genes in several isolated mutants. Most of them were previously unknown genes. These genes provide important clues to understanding the molecular mechanisms of this process.

An attempt to replace major membrane lipids by monoglucoyl(diacylglycerol) in Escherichia coli

Escherichia coli membranes consist of three kinds of major membrane phospholipids: phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and cardiolipin (CL). Previous studies reported that an E.coli mutant lacking PE due to inactivation of the pssA gene is viable in the presence of high concentration of divalent cations. Besides, an E.coli mutant lacking both PG and CL due to inactivation of the pgsA gene is viable in the absence of the major outer membrane lipoprotein. These study suggest that each major membrane lipids are not necessarily essential for viability of E.coli. Therefore we attempted to replace major membrane lipids by foreign lipid monoglucoyl(diacylglycerol) (MGlcDAG) and an acidic biosynthetic precursor phosphatidic acid (PA). It was reported that introduction of the MGlcDAG relieved divalent cation dependence to some extent in the null pssA mutant and that PA was accumulated in the pgsA null mutant. We introduce the Acholeplasma laidlawii mgs gene and the cdsA9 mutation which cause accumulation of PA into the pgsA null, pssA expression repressible mutant. Contrary to our expectation the mutant lacking all major membrane lipids was not viable.
Rodent phenotypic diversity: insights from conserved non-coding evolution

Rodent is the most phenotypically diverse and ecologically successful mammalian order. To understand what is responsible for these features, we analyzed the conserved non-coding sequence (CNS) evolution. The evolutionary and population genetics analyses suggest that CNSs are not just mutational cold spots, but are under purifying selection. The analyses of the genomic location of the CNSs show that they cluster around genes involved in development, transcription and nervous system. Analyzing the expression and ChiP-Seq data, we show that CNSs are mostly regulatory elements associated with conserved gene expression. We compared CNS evolution in rodents with three other mammalian orders and found that rodents have the least number. The low number of CNSs in rodents is due to higher rate of loss and lower rate of gain. The low number of CNSs in rodents suggests that rodent regulatory elements are not as conserved as in other mammalian orders. Our results suggest that rodent diversity may be attributable to higher regulatory turnover rate reflected in the lower number of rodent CNSs.

Functional analysis of uORF44-derived protein in mouse

It is widely accepted that human genes are monocistronic and thus that ~60,000 small ORFs in the 5’ untranslated regions (upstream ORFs or uORFs) encode no proteins underlying physiological functions. Recently, development and prevalence of ribosome profiling methodology allow us to witness that many more uORFs are capable of translation than previously believed, but still any physiologically functional proteins encoded in the uORFs has yet to be reported. Here, we present the potentially first case: a protein encoded in a uORF (named uORF44p). In this presentation, we will discuss up-to-date results of phenotypic analyses on uORF44 knockout mice we created.
Adaptive significance and molecular mechanisms of intra-specific body color variation in *D. melanogaster*

Body pigmentation is polymorphic in *D. melanogaster*, and studies have suggested that it is an adaptive trait. We measured desiccation resistance of four wild-derived inbred strains from the *Drosophila* Genetic Reference Panel (DGRP). The darkest strain was most vulnerable to desiccation. Because the expression level of *ebony* in the darkest strain during the period right after eclosion was the lowest among the four strains, we investigated the relationship between its expression level and desiccation resistance. RNAi knockdown of *ebony* was conducted by utilizing GAL4-UAS system. We measured desiccation resistance of whole-body knockdown and control individuals at 25°C in ~2% relative humidity. The survival time (h) was shorter and the dehydration speed (mg/h) was higher in the knockdown individuals compared to the control. A partial knockdown of *ebony* in the epidermis by *pnr*-GAL4 resulted in a slightly higher dehydration speed than the control. Whereas, knocking down this gene in trachea did not give any difference. Our result suggested that the dark phenotype due to the reduced expression of *ebony* in the epidermis is associated with the reduced desiccation resistance.

Involvement of the perinuclear anchorage of DNA double-strand break in damage-induced sister chromatid cohesion

DNA double-strand breaks (DSBs) are the most serious DNA damage for the cells. In *Saccharomyces cerevisiae*, it is known that persistent DSBs are anchored to nuclear pore complex (NPC) and the nuclear membrane protein Mps3. However, molecular mechanisms and physiological significance of the perinuclear anchoring of DSBs had been unknown. Recently, we reported that the perinuclear anchorage of DSBs requires the activity of the SWR1 and INO80 chromatin remodeling complexes. We also showed that the frequency of inappropriate DSB repair caused by unequal sister chromatid recombination (uSCR) was increased by *nup120* delta *mps3* delta65-145 double mutation. According to this result, we suppose a possibility that the perinuclear anchorage of DSBs is involved in the damage-induced sister chromatid cohesion, which is required for efficient DSB repair by homologous recombination. We are analyzing the involvement of SWR1/INO80 chromatin remodeling complexes, NPC and Mps3 in damage-induced cohesion using their mutant strains.

Attempts to generate artificial ring chromosomes in rice

In *Arabidopsis thaliana*, we have successfully generated 2.85- and 0.98-Mb sized artificial ring chromosomes using the Cre/LoxP and Ac/Ds system. Despite the ring structure, they are stable during mitotic divisions and are transmissible to the next generation through meiosis. To confirm the stability of ring minichromosomes, we attempted to generate artificial ring chromosomes in rice. Although a similar strategy is also applicable to rice, more simplified methods are needed to be developed, because rice takes more time to mature than *A. thaliana*. In the present study, a new binary vector, pDLHC in which a Cre recombinase gene was incorporated with an artificial intron, was constructed. The pDLHC construct was effective in the induction of deletions between two *LoxPs* at T2 generation of ‘Nihon-bare’ expressing Ac transposase. No stable artificial ring chromosomes have been found, possibly because no centromere DNA sequences had been involved in the circular DNA molecules that was generated by Cre/LoxP system. This suggests that pDLHC has a potential for efficient generation of artificial ring chromosomes in rice and other plant species.
Chromosome mapping by flow cytometry

Flow cytometry enables chromosomes to be sorted into different groups based on their characteristics, such as their size and specific repetitive sequences. Despite remarkable recent progress in the analysis of plant genome organization and chromosome structure, there is a need for easy methods of mapping specific DNA sequences to specific chromosomes. Flow-sorted chromosomes, which can be collected in a vial or on a microscope slide, can be analyzed by PCR to tell whether or not they harbor specific DNA sequences, and, at the same time, can be identified by fluorescence in situ hybridization (FISH) using chromosome-specific repetitive sequences. A slight modification of the procedures of flow cytometry causes the hyperexpansion of chromosomes in wheat, about eight times larger than the size of chromosomes from squashed preparations: The longest chromosomes exceed 100 micrometer. On hyperexpanded chromosomes, FISH of repetitive sequences, such as 45S rDNA, 3S rDNA and centromeric repeats, has been proved to be practicable, and single-copy FISH would be probably easier. Such FISH-based mapping of single-copy sequences would clarify the structure of chromosomes in more detail.

Inference of the evolutionary process of microchromosomes in vertebrates from comparative gene mapping for *Polypterus*

The karyotypes of non-avian reptiles and birds generally have a large number of morphologically indistinguishable microchromosomes. In contrast, the majority of amphibians and all mammalian and teleost fish species have no microchromosomes. To delineate the evolutionary process of microchromosomes in vertebrates, we constructed a chromosome map of a ray-finned fish, *Polypterus senegalus*, which is positioned as the most basal group of the extant fish, with more than 150 functional genes, and compared it with the genome and/or chromosome maps of several model species of vertebrates. Five linkage groups of chicken macrochromosomes and 19 groups of chicken microchromosomes, which are conserved in non-avian reptiles and Xenopus tropicalis but not in the teleost fish and human, was also highly conserved in *Polypterus*. Our present finding provides the possibility that the ancestral tetrapods had several pairs of macrochromosomes and many microchromosomes, each of which corresponded to the chicken chromosomes. The disappearance of microchromosomes in amphibians and mammals was caused by repeated fusions of microchromosomes that occurred independently in each lineage.

Whole genome sequencing of Asian pit viper, Habu, *Protobothrops flavoviridis* and whole gene finding

Snake venoms are the promising resource for the pharmaceutical discovery. Towards the complete understanding of the snake proteins including all venomic proteins and endogenous venom inhibitors, we conducted the whole genome sequencing of Japanese endemic pit viper, Habu snake, *Protobothrops flavoviridis*. By FACS analysis, we estimated the genome size of Habu to be 1.8 Gb. We assembled 98 Gb of shotgun sequence reads into 1,924,902 contigs. By adding Illumina mate-pair reads with five different insert sizes, we finally obtained 84,502 scaffolds with the N50 length of 467 kb. By ab initio gene prediction using RNA-seq data as exon/intron hints, we identified 25,132 protein-coding genes in total. Through the homology search against the public databases, we identified 20,540 genes with functional predictions. Through the keyword search of venom/toxin related terms against the annotations, we identified 329 genes as venom-related genes. We also identified >1,000 candidate genes for inhibitors of venomic proteins.
Re-assessment of reference sequences of whole genome DNA
Whole human genome sequences have been open to public; however, they were re-sequenced by using long-read NGS in 2015. Not only uncompleted regions but also whole structures including insertions, deletions, inversions, translocations, duplications, etc. were updated. The mouse genomic sequences have been considered to be more precise than human’s, since they are from C57BL/6J inbred genomic DNA. Fairfield et al. (2015) reported that only -50% of Mendelian traits in the mouse were identified by NGS, the efficiency of which was very equivalent to those in human. It implies that the mouse reference sequences also have some unexpected issues in addition to those found in human sequences. We also found peculiar findings during the SNV identification in C57BL/6J background by using NGS technologies. The re-assessment of reference genomic sequences seems to be urgent and necessary.

Functional enrichment analysis of human transcriptional target genes and its application to their prediction
An examination of transcriptional cascades is important to understand the regulatory mechanisms of gene expression. Functional enrichments of transcriptional target genes have been utilized to understand the functions of transcription factors and cascades in a cell. To promote the genome-wide analysis, I predicted transcriptional target genes using open chromatin regions of human immune cells, as well as known transcription factor binding sequences (TFBS). Of the 10 annotation databases of gene functions and pathways, 8 annotation databases showed larger numbers of functional enrichments in transcriptional target genes than those identified by gene expression information alone in the three cell types. Interestingly, from the comparison with transcriptional target genes including randomly selected genes, native predicted transcriptional target genes showed the most functional enrichments. The effects of gene expression levels, enhancer regions, and CTCF on the functional enrichments of transcriptional target genes were also examined. These analyses would be useful to improve the methods and conditions for prediction of transcriptional target genes from enhancer regions.

Systematic analysis of mutation distribution in three dimensional protein structures identifies cancer driver genes
Protein tertiary structure determines molecular function, interaction, and stability of the protein, therefore distribution of mutation in the tertiary structure can facilitate to identify new driver genes in cancer. To analyze mutation distribution in protein tertiary structures, we applied a novel three dimensional permutation test. We analyzed somatic mutation datasets of 21 types of cancers obtained by the TCGA project. Of the 3,608 genes that had ≥23 mutations in the regions with tertiary structure data, 98 genes showed significant skew. Known tumor suppressors and oncogenes were significantly enriched in these identified cancer gene sets. Twenty-three genes, including TP53, PIK3CA, PTEN, and PARG, were detected in multiple cancers. Candidate genes with significant 3D skew of the mutation positions included tumor suppressor genes, oncogenes, kinases, apoptosis related genes, RNA splicing factor, miRNA processing factor, and transcription factors. Our study suggests that systematic analysis of mutation distribution in the tertiary protein structure can help identify cancer driver genes, and contribute to the functional interpretation of the role of the mutations.

A computer simulation study on the power of haplotype-based GWAS
Since more than a million single-nucleotide polymorphisms (SNPs) are analyzed in any given genome-wide association study (GWAS), performing multiple comparisons can be problematic. To cope with multiple-comparison problems in GWAS, haplotype-based algorithms were developed to correct for multiple comparisons at single SNP loci in linkage disequilibrium (Misawa et al. 2008 J. Hum. Genet. 53: 789-). To tackle the same problem, FAIS was developed (Llinares-López et al. 2015, Bioinformatics 31: i240-). FAIS automatically finds intervals of SNPs in the genome that are jointly associated with the phenotype. It also solves the statistical problem of multiple hypothesis testing.

We conducted computer simulations to estimate statistical power of ParaHaplo and FAIS. DNA sequences were generated by computer simulations. Causative SNPs were randomly chosen, then phenotypes were determined by using given penetrance. Based on the phenotypes, GWAS were conducted by using ParaHaplo and FAIS. Computer simulations indicate that FAIS tends to detect more causative SNPs than ParaHaplo.
Do melanocytes contribute to the postnatal development of the mouse blood vessels?

Melanocytes are derived from the vertebrate embryo-specific neural crest. These cells migrate to and settle in various organs, including not only the skin but also extracutaneous locations such as the choroid of eye, inner ear, heart, etc. Skin melanocytes produce melanin and primarily determine the skin color. How do melanocytes, which also produce melanin, function in the extracutaneous locations, where only dimly light may illuminate them? In order to elucidate whether strial melanocytes contribute to the structure of their habitats, we used a melanocyte-deficient mouse mutant strain, Mitf<sup>mi-bw</sup>. The stria vascularis normally develop rich capillary networks surrounded by melanocytes. Our study suggests that melanocytes may contribute to the normal blood vessel vasculature in the stria vascularis.

An alternative pluripotent state confers interspecies chimaeric competency

Pluripotency, the ability to generate any cell type of the body, is an evanescent attribute of embryonic cells. Transitory pluripotent cells can be captured at different time points during embryogenesis and maintained as embryonic stem cells or epiblast stem cells in culture. Since ontogenesis is a dynamic process in both space and time, it seems counterintuitive that these two temporal states represent the full spectrum of organismal pluripotency. Here we show that by modulating culture parameters, a stem-cell type with unique spatial characteristics and distinct molecular and functional features, designated as region-selective pluripotent stem cells (rsPSCs), can be efficiently obtained from mouse embryos and primate pluripotent stem cells, including humans. The ease of culturing and editing the genome of human rsPSCs offers advantages for regenerative medicine applications. The unique ability of human rsPSCs to generate post-implantation interspecies chimaeric embryos may facilitate our understanding of early human development and evolution.

Comparative proteomics database for gene spectral analysis

To address genetic background specific to species is a big challenge in evolutionary developmental biology. For this purpose, we are developing a comparative omix database designed especially for the clarification of species specific or overrepresented genes in terms of gene sequences and corresponding ontology. We designate gene spectrum for the classified frequency of gene types in a genome so that it should represent the quantitative genome identity in terms of gene ontological context - i.e. the ensemble of biological process, cellular component and molecular function.
Regulatory mechanisms of the plasticity of cell fate decision during early embryonic development

In vertebrates, three germ layers (endoderm, mesoderm, and ectoderm) are established before gastrulation, and it is well known that Nodal signaling is essential for mesendoderm induction. Although the competency of different germ layers is thought to be lost after the specification of germ layer, epigenetic regulation of the competency for mesendodermal gene expression in ectoderm cells remains to be elucidated. In this study, we will present the regulation of mesendoderm competency in ectodermal cells by using zebrafish embryos.

Genetics terminology in the highschool "biology" textbook: current status and problems

In accordance with the new education guidelines (2012) issued by the Japanese MEXT (Ministry of Education, Culture, Science and Technology), all the textbooks for "Science" or "Biology" in junior high schools and high schools were renewed. Traditionally, the technical terms in textbooks had been based on "The Japanese Scientific Words (Genetics) " compiled by the MEXT(1993). In the renewed textbooks, however, there are so many words which were not contained in the "The Japanese Scientific Words (Genetics)", mainly because of the recent rapid advancement in the fields of life science. Here a discussion was given on how much range of biological science fields should be taught in the middle course of education. Also introduced was the current activities of the editorial committee for “the Japanese Glossary of Genetics” which was established by the GSJ in 2009.

A bioinformatics strategy to identify horizontal gene transfer on the basis of Self-Compressing BLSOM (Batch-Learning Self-Organizing Map)

With remarkable increase in genomic sequence data, novel tools are needed for comprehensive analyses of big sequence data available. We previously developed a BLSOM, which can cluster genomic fragment sequences according to phylotype solely dependent on oligonucleotide composition, and applied to genome studies. In addition, we have recently developed a Self-Compressing BLSOM (SC-BLSOM) for computation time reduction, which allows us comprehensive but convenient analysis of big sequence data.

We used SC-BLSOM to analyze the horizontal gene transfer (HGT) between Arthropods known to be pathogen-transmitting vector and their symbionts such as protozoa and bacteria. We could detect candidates of transferred genes and assign to their potential origins. This method is useful for revealing whole picture of HGT process for both hosts and their symbionts, and therefore, is suitable for efficient knowledge discovery from big sequence data.

Function prediction of poorly characterized proteins on the basis of oligopeptide composition similarity

As the result of extensive decoding of genome sequences, a large number of proteins, whose function cannot be identified by homology search of amino acid sequences, has been accumulated progressively and thus remains of no use in science and industry. A method for predicting their functions that does not depend on the sequence homology search is in need.

Here, we developed a new method to predict protein function on the basis of similarity in oligopeptide composition in protein candidates. Oligopeptides are compositional elements of a protein and are involved in formation of its functional motifs and higher-order structural parts. Using this method, the proteins, whose functions have not yet been known because of lack of significant global homology to function-known proteins with sequence homology search, could be related to function-known proteins. Therefore, this method is useful to predict function of a vast amount of poorly characterized proteins.
Spatio-temporal imaging of the Gli transcription factors: establishment of Gli3-deficient cell lines by CRISPR/Cas system

Gli transcription factors are indispensable mediators of Hedgehog signaling. Gli proteins are thought to move to specific locations inside the cells depending on upstream signal. However, the regulatory mechanism of Gli activity is still unclear. To elucidate the relationship between subcellular localization and activity of Gli, we challenge spatio-temporal imaging of the Gli transcription factors.

In this study, we knocked down the Gli3 gene in 3T3 cell by CRISPR/Cas system. We established cell lines with frame-shift mutations in the second and third exons, which encode N-terminal part of Gli3. To investigate expression of the Gli3 protein in these mutant cell lines, we performed Western analysis with an anti-Gli3 antibody. As a result, we observed the Gli3 protein in the mutant cell lines at a slightly lower molecular weight than that in 3T3. This observation suggested that nonsense mutation near 5' region in the Gli3 ORF allowed reinitiation of translation from the second inframe ATG codon. Thus, expression of the protein product of interest needs to be carefully analyzed, even if a frame-shift mutation is introduced to the gene by the gene-editing technology.
Transposable elements acting as cis-regulatory elements in mammary gland development

Mammary gland is a mammalian-specific organ that is required to feed their young offspring. To reveal the evolutionary origin of the genetic elements responsible for mammary gland development, I reanalyzed published ChIP-seq data of various transcription factors such as ERalpha, FoxA1, GATA3 and AP2gamma which are responsible for mammary gland development, and compared the distribution of their binding sites and transposable elements (TEs). I found that the location of binding sites are biased within several families SINEs, LINEs and LTR-retrotransposons. Particularly, it was revealed that the consensus sequences of some of the TE families contain the binding motifs of the transcription factors. In addition, I found that two waves of retrotransposition bursts produced the major TE-derived functional elements during mammalian evolution. These facts suggest that a part of TE families had spread the binding motifs of the transcription factors via retrotransposition, and that they contributed to increase the functional elements involved in the mammary gland evolution.

Analysis of DNA methylaiton of the newly inserted LINE elements in zebrafish

Long interspersed elements (LINEs) are mobile genetic elements that exist in most eukaryotic genomes. LINEs mobilize and amplify their own copies by a mechanism called retrotransposition in which the LINE RNA is reverse-transcribed into DNA and then integrated into the host genome. Integrations of LINEs cause insertional mutations that are sometimes deleterious to the hosts. Thus, new LINE retrotransposition events can be suppressed by epigenetic regulations, such as DNA methylation. However, little is known about the epigenetic regulation of LINE retrotransposition. To elucidate epigenetic regulation of LINE retrotransposition, we developed a system to detect the de novo zebrafish LINE insertions in vivo, and analyzed DNA methylation status at the site of newly inserted LINEs. From the analysis, I discuss the relation between LINE retrotransposition and DNA methylation.

The active MITE mPing produces transcriptional variants post-transcriptionally in the rice genome

Transposable elements (TEs) are well-known to disrupt the function of genes. In addition, recent studies have reported that TEs also contribute to the regulation of genes both at transcriptional and post-transcriptional levels. Miniature Ping (mPing) is an active miniature inverted-repeat TE discovered in the rice genome, and its insertion renders adjacent genes stress-inducible. Little is known, however, about the effects of mPing on the post-transcriptional regulation of genes. In this study, we performed 3’-RACE to investigate the structure of transcripts derived from mPing-inserted mutant allele and demonstrated that mPing is creating new transcript variants by inducing various alternative splicing events. mPing is actively transposing in japonica landrace EG4 under natural growth conditions, and its copy number reaches approximately 500 copies. To investigate whether these mPing insertions affect post-transcriptional regulation of genes, we also performed RNA-seq. Consequently, we identified that several mPing insertions in intron induce various alternative splicing events.
Evolution of the Au SINE retrotransposon in plant genomes

Many TE families show surprisingly high levels of similarity between distantly related species. This high similarity, coupled with a patchy distribution, has often been attributed to frequent horizontal transfers of TEs between species, even though the mechanistic basis tends to be speculative. Here, we studied the evolution of the Au SINE family, in which high similarity between distantly related plant species has been reported. We find that the evolution of the Au SINE can be readily explained by vertical transmission where the Au SINE was present in the common ancestor of all angiosperms and has been retained in some lineages while lost in others. Based on our results, we propose a model of TE evolution that may explain the patchy distribution and high similarity between distantly related species without having to invoke hypothetical scenarios of rampant horizontal transfers or exaptations.

Chromosome substitution strains of medaka fish

Chromosome substitution strains (CSSs), also known as consomic strains, are powerful tools for assigning the polygenes that control quantitative complex traits to specific chromosomes. Three years ago we started to establish full set of medaka fish CSSs using HNI-II (Oryzias sakaizumii; wild type body color) as the donor and Hd-rR (O. latipes; orange-red males and white females) as the host, and to date about a half of the full set strains has been successfully completed. For these strains, each chromosomes of the HNI-II strain is individually transferred into the Hd-rR background. Because two major loci, r on Chr1 (sex chromosome) and b on Chr12 (autosome), control the body color, males and females were orange-red in CSS1, brown males and blue females in CSS12, and orange-red males and white females in other strains. Viability was variable among strains: five normal (Chr11, 12, 13, 15), five significantly low (Chr2, 4, 14, 16, 20), two extremely low (Chr7, 21), and lethal for Chr23. The locus responsible for the lethality was mapped at the terminal of Chr23. Genome-wide linkage analysis successfully detected a locus on Hd-rR Chr20 that interact with HNI Chr23.

Functional analyses of vertebrate ultraconserved elements in humanized fruit flies

Comparative genomics has identified a large number of non-coding segments that have been highly conserved over hundreds of millions of years of vertebrate evolution. To examine their functions in vivo, we generated over 300 Drosophila transgenic lines, each containing a unique human conserved non-coding segment inserted upstream of a core promoter fused to a GAL4 gene, and then assessed their enhancer activities. We observed reporter GFP expression in one or more tissues of embryos and larvae in all 44 segments tested. Therefore, it is evident that human conserved segments can function as enhancer in Drosophila. Moreover, our assay system is more powerful and effective than the previously described mouse assay, which detected reporter gene expression only in a half of tested segments. In addition to segment-specific expression, all segments commonly induced GFP expression in mushroom body of the central brain and medulla neurons of the optic lobe. This result may imply that human conserved segments share some unknown enhancer functions despite lacking sequence homology. We will soon make the newly developed lines available to the community through KYOTO stock center (DGRC).