論 文 の 内 容 の 要 旨

論文題目

Investigation of genome functions and their evolution using high-throughput sequencing
（超並列シーケンシングを用いたゲノム機能とその進化に関する研究）

氏　　名　　尾崎　遼

To date, more than 30,000 genomes have been sequenced since the publication of the first free-living organism genome. However, understanding of how genomes exert their functions is lagging behind. This is due to laboriousness of measuring dynamics of entities that link genomes and phenotypes (e.g., genes and proteins). Such information is required to investigate mechanisms of genome functions, and also should provide clues for understanding how genome functions have evolved. Recent advent of high-throughput sequencing enables us to obtain genome-wide information on not only DNA sequences but also gene expressions and DNA-binding protein occupancies more easily than ever before. Thus, by taking advantage of the high-throughput sequencing data, now we can investigate genome functions and their evolution.

In this thesis, I describe research on genome functions and their evolution using of high-throughput sequencing data in two topics.

First, I investigated contribution of CLOCK, a fundamental transcription factor in the mammalian circadian oscillator, to circadian rhythms in gene expression levels in a genome-wide manner. Circadian rhythms are oscillation with a period of approximately 24 hours in biochemistry, physiology, and behavior of organisms. The circadian rhythms are generated by circadian clock. Mammalian circadian clock have been well characterized: many key genes are identified and systems-biological approaches have been applied. In that, transcriptional and translational feedback loops (TTFLs) with dozens of genes are well modeled. Meanwhile, early microarray experiments found that many genes out of the TTLFs showed circadian expressions in cells and tissues, which raises a question: how these many genes are regulated on a genome-wide manner? To address this question, I analyzed high-throughput sequencing data including CLOCK ChIP-Seq, mRNA-Seq, and small RNA-Seq data from mouse liver. The new method to enumerate DNA-binding motifs from ChIP-Seq data was developed. Application of the method to the ChIP-Seq data revealed comprehensive set of CLOCK-binding motifs. In addition, I found that contribution of CLOCK to the transcriptome-wide circadian gene expressions as a direct transactivator was smaller than expected. Several plausible mechanisms of CLOCK indirectly regulating rhythmically expressed genes expressions were discussed.

Second, I show the possibility of positive selection on gene copy number variations by taking

advantage of the system of parallel evolution of three-spined sticklebacks. Positive selection is an evolutionary process by which an allele increases its frequency in a population. Copy number variations (CNVs) constitute a significant proportion of genomic diversity. In particular, gene copy number variations (GCNVs), which change the numbers of gene loci in genomes, can significantly alter gene functions and dosages, and thus can undergo positive selection. However, positive selection of CNVs has not been proved. One way of assessing the possibility is to search for increase or decrease of copy numbers in parallel evolution of freshwater groups of three-spined stickleback. Parallel evolution is the adaptive evolution in which the same genotypes or phenotypes are selected in different but related lineages, and provides strong evidence of positive selections. To address the possibility of positive selection on GCNVs using the system of parallel evolution of three-spined stickleback, I analyzed resequencing data of multiple individuals from freshwater and marine populations. A novel approach was devised to detect GCNVs under parallel selection from whole-genome sequencing data with low coverage by comparing two resequencing datasets. Application of the method to the resequencing data of sticklebacks revealed GCNVs that were likely under parallel selection. Many of the identified GCNVs showed increase in gene copy numbers in freshwater individuals, which is consistent with the notion that increase in gene copy number would facilitate adaptive evolution. These results suggest that contribution of GCNVs should be considered in studies on adaptive evolution.