論文内容の要旨

論文題目　Reconstruction of Evolutionary History of Genomes Using Abundant Sequence Data

（大量のゲノム配列情報を用いたゲノム進化過程の再構築）

氏　　名　　岩　崎　　渉

More than 10 years have passed since the first free-living organism genome was sequenced. Since then, the pace of the genome sequencing projects has been getting faster and faster, and the number of the sequenced genomes is soon reaching the 1,000 mark. This fact would mean that, by taking advantage of the abundant genomic data from the wide range of organisms, now we can reconstruct the evolutionary history of the genomes and investigate how the extant species' genomes have been organized through evolution until today.

In this thesis, first, I present a novel, effective and efficient method for estimating the whole gene sets of the ancient species by utilizing abundant genomic data from hundreds of species.

To reconstruct plausible evolutionary history, rates of gene gain/loss should be estimated by considering the high level of heterogeneity: for example, genome duplication and parasitization, respectively, result in high rates of gene gain and loss. However, gene-content evolution reconstruction methods that consider this heterogeneity and that are both effective in estimating the rates of gene gain and loss and sufficiently efficient to analyze abundant genomic data had not been developed. The present method comprises analytically integrable modeling of gene-content evolution, analytical formulation of expectation-maximization, and efficient calculation of marginal likelihood using an inside-outside-like algorithm. Simulation tests on the scale of hundreds of genomes showed that both the gene gain/loss rates and evolutionary history were effectively estimated within a few days of computational time. Subsequently,

this algorithm was applied to an actual data set of nearly 200 genomes to reconstruct the heterogeneous gene-content evolution across the three domains of life. The reconstructed history, which contained several features consistent with biological observations, showed that the trends of gene-content evolution were not only drastically different between prokaryotes and eukaryotes, but were highly variable within each form of life. The results suggest that heterogeneity should be considered in studies of the evolution of gene content, genomes, and biological systems.

Second, I show the results of the application of the present method to the evolution of metabolic pathways of prokaryotes.

The evolutionary history of biological pathways is of general interest especially in this post-genomic era, because it may provide clues for understanding how complex systems encoded on genomes have been organized. To explain how pathways can evolve *de novo*, some noteworthy models have been proposed until today. However, direct reconstruction of pathway evolutionary history both on a genomic scale and at the depth of the tree of life has suffered from artificial effects in estimating the gene content of ancestral species. Thus I applied the present algorithm that effectively reconstructs gene-content evolution without these artificial effects to this problem, to provide a definitive version of the pathway evolutionary history. The carefully reconstructed history, which was based on the metabolic pathways of 160 prokaryotic species, confirmed that pathways have grown actively and not just through the random acquisition of individual genes. The pathway acquisition took place in short time, probably eliminating the difficulty in holding functionless genes during the course of the pathway evolution. This rapid evolution was due to massive horizontal gene transfers as gene groups, a part of which was probably operon transfers, which would convey existing pathways but not be able to generate novel pathways. To this end, I analyzed how these pathways originally appeared in the ancient era, and found that the ancient acquisition of pathways occurred significantly contemporaneously across distant phylogenetic clades. As a possible model that explains this observation, I propose that novel pathway evolution may be facilitated by bidirectional horizontal gene transfers in prokaryotic communities, which model would complement the existing pathway-evolution models.