論 文 の 内 容 の 要 旨

論文題目

A large-scale comparative genomic analysis to reveal adaptation strategies of marine Flavobacteriia

（大規模比較ゲノムから探る、海洋性フラボバクテリアの適応戦略）

氏　　名　　熊谷 洋平

This thesis consists of five chapters. The outline of each chapter is as follows.

Chapter 1. General Introduction

Due to the technological innovation of the genome sequencing that occurred in the 21st century, the sequencing cost of a whole human genome, which is about 100 million dollars in 2001, came to be about $1,000 in 2014. The innovation provides two new approachs for microbial ecology: metagenomics and genomics. Metagenomics enables us to understand whole microbial process occurring in environments and microbial genomics provides us great insight into ecology, physiology, and evolution of microbial strain. Based on these technological breakthroughs, traditional questions in microbial ecology such as “What kind of bacteria are in what kind of environments”, or “What kind of genes do this bacterial strain possesses?” ware rapidly solved. In the post “$1,000 genome” era, "How to handle obtained sequence information" becomes more important than "How to obtain sequence information" and many new bioinformatics techniques had appeared. Comparative genomics is one of the novel approaches to reveal the ecophysiology of prokaryotes and to predict a function of functional unknown genes.

One of the biggest findings of metagenomics is the discovery of the proteorhodopsin (PR) gene. PR acts as a light-driven proton pump and bacteria with PR can generate ATP from sunlight. Since the discovery of the PR gene, the broad distribution of rhodopsins in the environment has been clarified, and it is estimated that there are more microorganisms possessing rhodopsin than microorganisms possessing photosynthetic systems on the earth. PR genes show a particularly wide distribution among various rhodopsins, and it is estimated that up to 80% of oceanic prokaryotes have PR in the open ocean. Not only geographical distribution studies but also physiological studies of PR also showed the important role of PR for adaptation to sunlit environments. On the other hand, the growing

understanding of PR function provokes another fundamental question—if the possession of PR is so advantageous for surface-dwelling microbes, why are there so many PR-lacking (PR−) prokaryotes in the marine photic zone? Comparative genomics is a potentially useful approach for answering such questions because genomes fundamentally reflect microbial ecophysiology. That is, systematic differences between PR− and PR+ prokaryote genomes might provide clues about differences in the lifestyles of these microbes. In addition, the previous study showed that PR− Flavobacteriia have significantly larger genomes than PR+ Flavobacteriia, although the ecophysiological reasons for this phenomenon remain enigmatic. Resolving this mystery should provide great insight for us to understand the genomic evolution of marine Flavobacteriia.

In this thesis, I sequenced 21 marine Flavobacteriia genomes and constructed genomic dataset including 41 PR− and 35 PR+ marine Flavobacteriia (chapter 2) and using this genomic dataset, I did phylogenetic profiling analysis and statistically detected the genes with biased distribution in PR+ strains or PR– strains (chapter 3). To reveal the function of the gene showed highest biased distribution to PR, I conducted an experimental assay (chapter 4). Accumulating all result of my thesis, I will discuss the ecology, physiology, and evolution of ocean surface bacteria (chapter 5).

Chapter 2. Construction of genome dataset of marine Flavobacteriia

To conduct comparative genomics, fulfilling two prerequisite conditions are required for successfully discovering systematic differences between different types of genomes. First, to attenuate strain-specific signals and achieve sufficient statistical power, a sufficiently large number of genomes that were not strongly biased within a single type are required. Second, genomes that belong to each different type need to be phylogenetically dispersed. Otherwise, genomic differences due to phylogenetic constraints (i.e., an effect that phylogenetically closely related species tend to have similar genomes just because they share a common ancestor) as opposed to ecophysiological adaptations, will bias the analysis. In this chapter, I sequenced 21 genomes of marine Flavobacteriia including 9 PR− and 12 PR+ strains and obtained 4 complete and 17 high-quality draft genomes. With available genome data from NCBI database and newly sequenced genome, I constructed 76 marine Flavobacteriia genome including 41 PR− and 35 PR+ strains. The sampling points of strains included in the genome dataset cover a wide range such as Antarctica, Mediterranean Sea, Pacific Ocean, Japan Sea, but there was no difference in the geographical distribution of PR– bacteria and PR+ strains.

Chapter 3. Phylogenetic profiling analysis of marine Flavobacteriia

In this chapter, I conducted a phylogenetic profiling analysis of PR− and PR+ marine Flavobacteriia. Phylogenetic profiling is a method of comparative genome analysis for detecting the relationship between genes by the similarity of phylogenetic distribution pattern of each gene. As a result, I found that 86 orthologous genes biased in PR– strains and 43 genes biased in PR+ strains. PR– strains tend to possess UV-absorbing pigment (aryl polyene (APE) or flexirubin-type pigment (FTP)) synthesis gene clusters. APE or FTP locate in outer membrane and known to function as "sunscreen". PR+ strains tend to possess photolyase gene cluster, a mechanism to deal with DNA UV damage. These results suggest that PR− and PR+ marine Flavobacteriia adopt contrasting strategies; the former produces APEs or FTPs to avoid UV damage, whereas the latter produces photolyase to repair DNA after UV damage. In accord with the analogy in which PR functions as microbial "solar panels", I propose a "solar-panel or parasol" hypothesis, in which APEs and FTPs are regarded as cellular "parasols". Another finding is about adaptation to anaerobic or microaerobic condition of PR– strains. PR– strains tend to possess several genes adaptive to anaerobic or microaerobic condition (e.g. genes involved in anaerobic N_2O respiration). Thus, the interiors of organic aggregates, which is known as typical niche of Flavobacteriia, are environments where facultative anaerobic PR− microbes may predominate.

Chapter 4. Functional analysis of DUF2237

In previous chapters, I did comparative genomic analysis and reveal the different adaptation strategies to light between PR+ and PR– Flavobacteriia. However, statistical correlation and anti-correlation do not necessarily indicate causation. In order to validate the results of comparative genomics, I conducted two experiments to examine physiological implications. To find out whether the genes with biased distribution are related to light utilization, I did knockout based experiment of DUF2237, a functionally unknown gene biased to PR+ bacteria. Because of the difficulty of genetic manipulation of Flavobacteriia, I used a model cyanobacteria *Synechocystis* sp. PCC 6803-P that possesses DUF2237 gene. By phototaxis assay, the DUF2237-knockout strain showed significantly less phototaxis movement to unidirectional white light than the wild-type strain. This result is consistent with the strong correlation between the presence of DUF2237 and phototrophy because phototaxis should be beneficial to organisms that utilize light.

Chapter 5. General Discussion

I started this study with the question, if PR is simply advantageous for marine surface-dwelling Flavobacteriia, why are there many bacteria lacking PR in the same group? By large-scale comparative genomics, I revealed that the possession of PR is not simply advantageous. PR+ strains have to possess DNA damage repair system, whereas PR– strains also show adaptation to light by acquiring light protection mechanism using sunscreening pigments. These findings provided a new perspective on the ecology of marine Flavobacteriia by considering light utilization. Traditionally, the research of oceanic Flavobacteriia mainly focused on their ability to metabolize high molecular weight compounds and their role as a degrader in the ocean, especially in marine particle. However, the interiors of macroscopic organic aggregates in the upper ocean are nutrient-rich conditions and likely decrease the advantage of possessing PR, thus the reason why Flavobacteriia possess PR is unclear. Based on my results, two contradictory features of Flavobacteriia, HMW compounds degrader and light utilization bacteria can be explained by the variety of adaptation strategies for light within Flavobacteriia. PR– strains may have more characteristics as HMW compounds degrader and PR+ may have more characteristics as light utilization bacteria. In methodological perspective, I proved the usefulness of phylogenetic profiling analysis in the field of microbial ecology. Phylogenetic profiling analysis is developed and applied to predict gene function by relationships of genes (e.g. formation of protein complex), so inverse correlation has not been analyzed. Detection of correlation and inverse correlation of genes by phylogenetic profiling analysis will be one of the important approaches in future microbial ecology studies since the genomic studies of prokaryotes are expected to proceed rapidly. The example of experimental validation of light-related function of DUF2237 shows that functional prediction of unknown genes by phylogenetic profiling analysis is useful in handling an enormous number of functional unknown genes possessed by environmental prokaryotes.