

CHAPTER 1 INTRODUCTION (I)

1.1 Background and Rationale

Cyanobacteria have diverse morphological, biochemical, and physiological properties due to a wide range of their genomic content, which have enabled this ancient group of micro-organisms to settle and persist in a wide range of habitats. They also significantly contribute to global primary production (Hess, 2004), of which the diazotrophic taxa are central to global nitrogen cycle (Zehr, *et al.*, 2001). Marine cyanobacteria are continually important sources of structurally diverse bioactive secondary metabolites, as an attractive source of natural products for anticancer or neurotoxin acting agents (Tan, 2007). These diverse capabilities of cyanobacteria are driven by the evolution and adaptation to environmental niches. All modern cyanobacteria are descended from the common ancestor, Pre-cyanobacteria, which is generally accepted as the progenitors of oxygenic photosynthesis. Consequently, the photosynthetic capabilities of cyanobacteria and other modern photosynthetic organisms were laterally and horizontally transferred from the Pre-cyanobacteria. During the evolution of cyanobacteria, an environmental pressure plays an important role for the adaptation of cyanobacteria by acquiring and/or losing genomic contents (Mulikidjanian, *et al.*, 2006). Therefore, the evolutionary study is a key step to delineating a relationship between environmental driven forces and the genome dynamics.

An organismal evolution can be traced by its genome sequences with the advancement of comparative genomic approach. Evolutionary scenarios in the compact organismal groups have been done, for example, in the α -proteobacteria (Boussau, *et al.*, 2004), *Thermus-Deinococcus* group (Omelchenko, *et al.*, 2005), *Cyanobacteria* (Mulikidjanian, *et al.*, 2006), *Streptococcus* (Marri, *et al.*, 2006), *Lactobacillales* (Makarova, *et al.*, 2006; Makarova and Koonin, 2007), Archaea (Makarova, *et al.*, 2007), and the *Prochlorococcus* (Kettler, *et al.*, 2007). In the case of all these analyses, a considerably high resolution of the homologous relationship in the overall orthologous group of protein sets has been achieved. Both the phylogenetic content and evolutionary scenarios—which are explained by the gene gain and loss events along the lineages of each organismal group—have been revealed to provide a better understanding of how living organisms could survive and inhabit specific environmental niches.

The prior evolutionary study of cyanobacteria has revealed the evolution of the photosynthetic genes in these distinctive photosynthetic organisms (Mulikidjanian, *et al.*, 2006). Pro-cyanobacteria, which are hypothesized as the common ancestor of modern descendent cyanobacteria, have been postulated as a photosynthetic ancestor of all modern photosynthetic organisms. Also, the cyanobacterial evolutionary study has been done in the genus of *Prochlorococcus*, which can uncover the environmental factors driving the gained and lost events on their genomic evolution (Kettler, *et al.*, 2007).

Currently, the cyanobacterial genomic data are enormously increasing every day. However, the mechanism and tempo of the cyanobacterial genome evolution driven by their habitats and environment factors remain ambiguous. The availability of these cyanobacteria genomic data provides ample opportunity to uncover their complexities and diversities through the genome evolutionary study. Thus, in this thesis, the complete and draft assembly cyanobacterial genomic sequences were used to delineate their evolutionary scenarios, such as loss and gain events, by using the parsimonious

evolutionary algorithm. The genome evolutionary scenarios could indicate the relationship between particular biological properties, habitats, and adaptation to the specific environmental niches of modern descendant cyanobacteria. Furthermore, this algorithm was also applied to study and explain the evolutionary scenario of the photosynthetic apparatus of cyanobacteria.

1.2 Objectives

- 3) To study the evolutionary scenarios, gene gain, and loss events in cyanobacteria lineages by comparative genomic approach.
- 4) To better understand the evolutionary scenarios of photosynthesis gene groups and explain how they evolved along the environmental niches.

1.3 Scope of Work

In order to achieve the objectives mentioned above, the scope of work was as follows:

- 7) For finding the orthologous gene sets in cyanobacteria, the construction of cyanobacterial Cluster of Orthologous Groups of proteins (cyanoCOGs) was performed using Makov Clustering algorithm (MCL) and OrthoMCL tool.
- 8) For determining phyletic patterns from cyanoCOGs, the detection of presence and absence of genes in each cyanoCOGs was performed by using the Python scripts.
- 9) For assigning the cyanobacterial evolutionary tree, the concatenated ribosomal proteins that are contained in the core cyanoCOGs genes was done by using Neighbor Joining (NJ) methods.
- 10) For determining gene gain and loss events along the cyanobacterial lineages, the parsimonious evolutionary scenarios algorithm was implemented by using the Python scripts.
- 11) For tracing the evolutionary scenarios of cyanobacterial lineages, the phyletic pattern of all cyanoCOGs and the reconstructed phylogenetic tree were used with the implemented parsimonious evolutionary scenarios program.
- 12) For understanding the evolution of the photosynthetic gene sets, the implement parsimonious evolutionary scenarios program was applied to these gene sets.

1.4 Expected Outputs

- 4) The database of orthologous groups of proteins in cyanobacterial lineage or cyanoCOGs.
- 5) New tools for identifying evolutionary scenarios along the cyanobacterial lineages.
- 6) The insight for understanding the evolution of cyanobacteria genome context and the evolution of photosynthetic apparatus.