

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS (I)

5.1 Conclusion

The constructed cyanoCOGs in this study was primarily used for classifying the cyanobacterial proteomes with the same biological function into clusters by the OrthoMCL method. The constructed cyanoCOGs were also used for detecting the phyletic pattern that represents the presence or absence of genes from each cyanobacterial genome. Besides, the ribosomal proteins that are embedded in the strictly core cyanoCOGs were used to reconstruct the evolutionary tree for this important domain of life. Both the phylogenetic tree and the phyletic pattern of each cyanoCOGs was used for uncovering the evolutionary scenario for each cyanoCOGs, which is explained by the gain and loss events of the orthologous genes along the cyanobacterial lineages by implementing the parsimonious evolutionary scenarios algorithm (Mirkin, *et al.*, 2003). Moreover, the ancestral gene sets for every cyanobacterial ancestral state were also revealed. In this study, the inferred gene sets of last common cyanobacterial ancestors have indicated that it is a photoautotroph and their genome contains all core cyanoCOGs and some accessory genes. Furthermore, the same analysis was emphasized on the photosynthesis gene sets in order to delineate the dynamic and evolution of those genes. Then, results illustrate that the most photosynthetic machinery genes of the modern cyanobacterial genomes were inherited from the LCCA, whereas some photosynthetic machinery genes were acquired from other sources, such as by horizontal gene transfer, genomic recombination, or gene duplication. Conversely, the gene loss events have extensively occurred, especially, in the cyanobacteria that inhabit stable environments, such as *Prochlorococcus*. The analysis of photosynthetic gene content has suggested that the extensive gene loss was found in the proteins, which are involved in the photosynthesis antenna, such as phycocyanin, allophycocyanin, phycoerythrin, and phycobilisome linker proteins.

5.2 Recommendations

During this evolutionary scenario, studies have been done. The results from this study could be used in various directions, from the cyanoCOG to the gain and loss patterns. These following recommendations are the examples of the further studies:

- 1) The cyanoCOGs has been constructed by using the completed genome sequences, and draft assembly genome sequences could be used to curate the prior annotated genes for those draft assembly genome sequences, and also could be used to infer the protein function for the hypothetical proteins. The phyletic pattern of cyanoCOGs and phylogenomics could be used for identifying the functional linkage of the same cyanoCOGs (Eisen, 1998). In this study, the unknown functional proteins have been found in the core cyanoCOGs, which means those proteins may have an important role for cyanobacteria. Those essential gene groups would be the next target genes for the experimental design and discovery of their functions.
- 2) Since there are some ambiguous functional assignments of the constructed cyanoCOGs in this study, human expert curation by checking the protein family of each cyanoCOGs one-by-one is needed in order to improve the accuracy and specificity check. For example, the same gene functions are clustered into the difference cyanoCOGs or vice versa.

- 3) For more systematic curation of the constructed cyanoCOGs and for gaining better insight into understanding the evolutionary scenarios of cyanobacterial lineages, the cyanoCOGs would be categorized according to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway databases and the parsimonious evolutionary scenarios algorithm should be modified and applied to every metabolic pathway (Wang, *et al.*, 2006). Then, the evolutionary study of the entire metabolic networks could be more representative of the relationships between the evolution of the cyanobacterial genome and their environmental factors, habitats, and ecological niches.
- 4) For gaining better understanding of the photosynthetic machinery evolution, additional analysis on the genomic context, and genome rearrangement of cyanobacteria and the other photoautotrophic organisms, such as the other photosynthetic bacteria and plants, should be taken into consideration. As well, cyanophages or non-photoautotrophic organisms that contain the photosynthetic apparatus could elaborate the evolutionary dynamics of these significant gene groups by using phylogenetic and phylogenomic analysis.
- 5) The biological explanation of all the events along evolutionary scenarios such as gene gain, gene loss, and horizontal gene transfer are required. For instance, why some genomes have lost nitrogen fixation genes or some photosynthetic genes, or which genes and biological pathways those genomes use or compensate for the missing genes, is still unclear.