CHAPTER 4 RESULTS AND DISCUSSIONS (I)

4.1 Coverage of cyanobacterial genomes with cyanoCOGs

Altogether, the process of cyanoCOGs construction started with 182,663 proteins encoded by 49 cyanobacterial genomes and ended with 160,409 of these proteins being included in 15,741 cyanoCOGs (the cyanoCOGs and accompanying materials are available in the supplement CD). With the growth of the genome collection and the various procedures for COG construction, the coverage of cyanobacterial genomes further increase. Figures 4.1 and 4.2 show that, on the average, the cyanoCOGs described here cover 87.8% of genes in a cyanobacterial genome as compared to 42% with previously release COGs which included total 69 genomes from the diverged organismal groups with only 4 of cyanobacterial genomes. For all cyanobacterial genome, this constructed cyanoCOGs showed the more coverage of the protein families than the COGs on the NCBI databases. The increasing of genes that exist in cyanoCOGs suggests that this constructed cyanoCOGs are more specific to the cyanobacterial species than previous NCBI COGs.

For finding the core and accessory gene sets of cyanobacteria, the distribution of number of species in cyanoCOGs was considered. As shows in figures 4.3 and 4.4, in the quantitative term, the cyanoCOGs with a large number of species (more than 44 genomes) are considered as the core gene sets, and the remainders are "shell" gene sets. More formally, assuming the distribution is described by an exponent, the best approximation was achieved with a sum of three exponential functions. The first exponent could be constructed to represent the conserved gene core (~915 cyanoCOGs), the second one describes the "shell" of moderately common genes (~4,860 cyanoCOGs), and the third one corresponds to the "ORFans" (~9,895 cyanoCOGs), which include the small number of (typically, but not necessarily, closely related) species. The cyanoCOGs that exist on only one species was determined as the recent paralog proteins.

For assigning the functional categories of the constructed cyanoCOGs, the similarity search against the previous version of NCBI COGs was performed. The whole COGs functional categories, that consist of the information storage and processing, the cellular processes and signaling, the metabolism, and the poorly characterized, are summarized their categories abbreviation in the table 4.1. Then, the best hit of NCBI COGs categories was assigned to the query cyanoCOGs. If the constructed cyanoCOGs was not found in COGs database, the X category was assigned to that query cyanoCOGs. figures 4.5 and 4.6 demonstrate the proportion of each cyanoCOGs categories for all cyanobacterial genomes, the X categories was found more than 40% in every species (except in TELOBP1 ~ 35%). Form this result; the small NCBI COGs database was taking into consideration, which means this database does not appropriate to determine the protein function for the cyanobacteria groups. On the other hand, the vast amount of unknown functionality genes or proteins could be concluded.

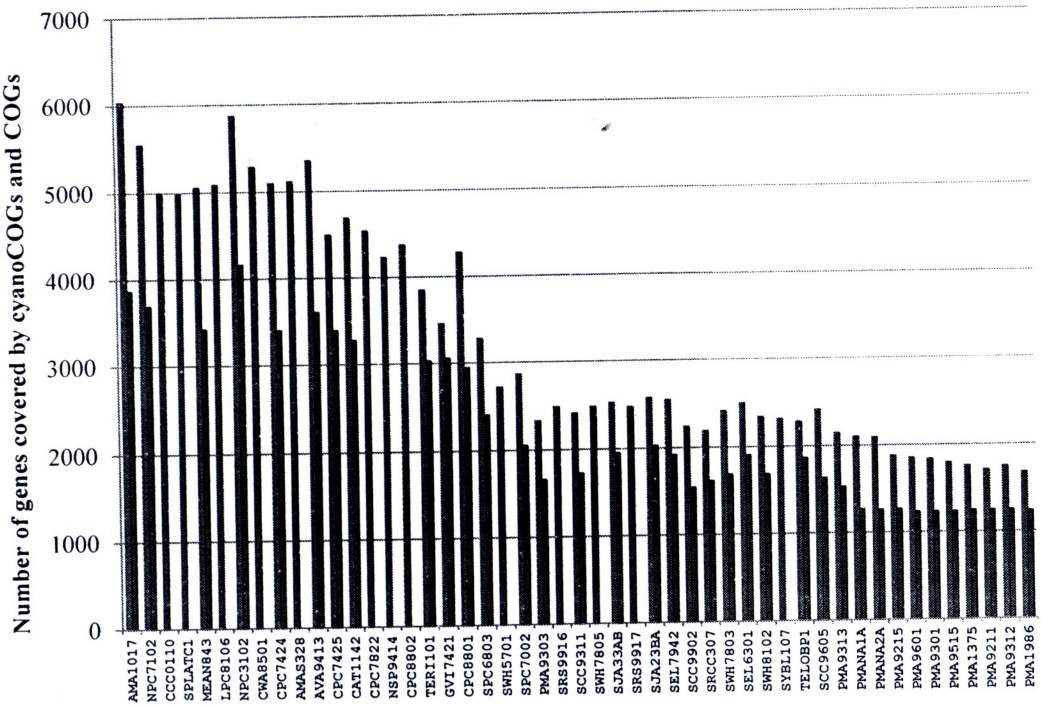

Figure 4.1 Coverage of cyanobacterial genomes with cyanoCOGs in green and NCBI COGs in red (Abbreviation as in Table 3.1).

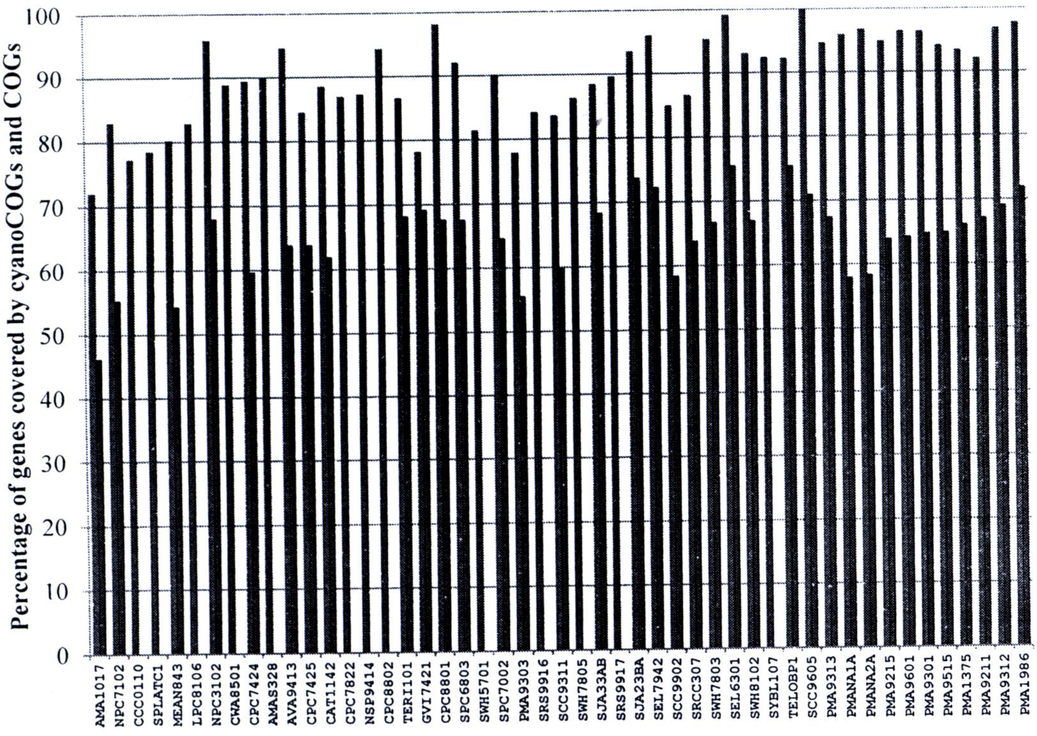

Figure 4.2 The percentage coverage of cyanobacterial genomes with cyanoCOGs in green and NCBI COGs in red (Abbreviation as in Table 3.1).

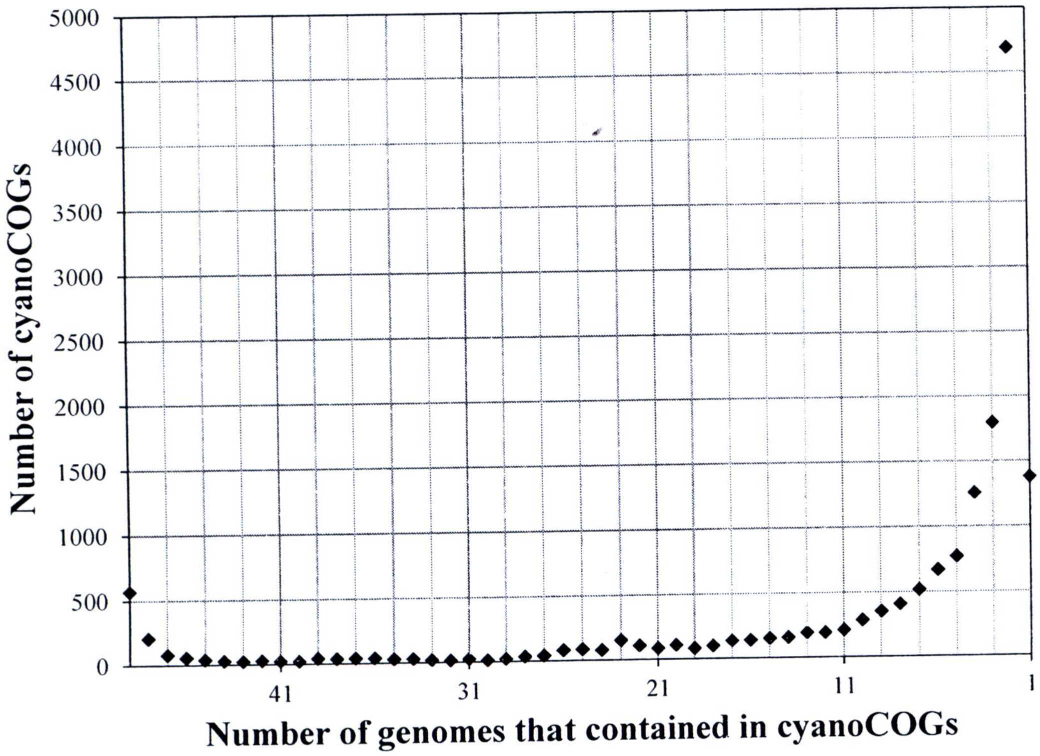

Figure 4.3 Distribution of the number of species in cyanoCOGs

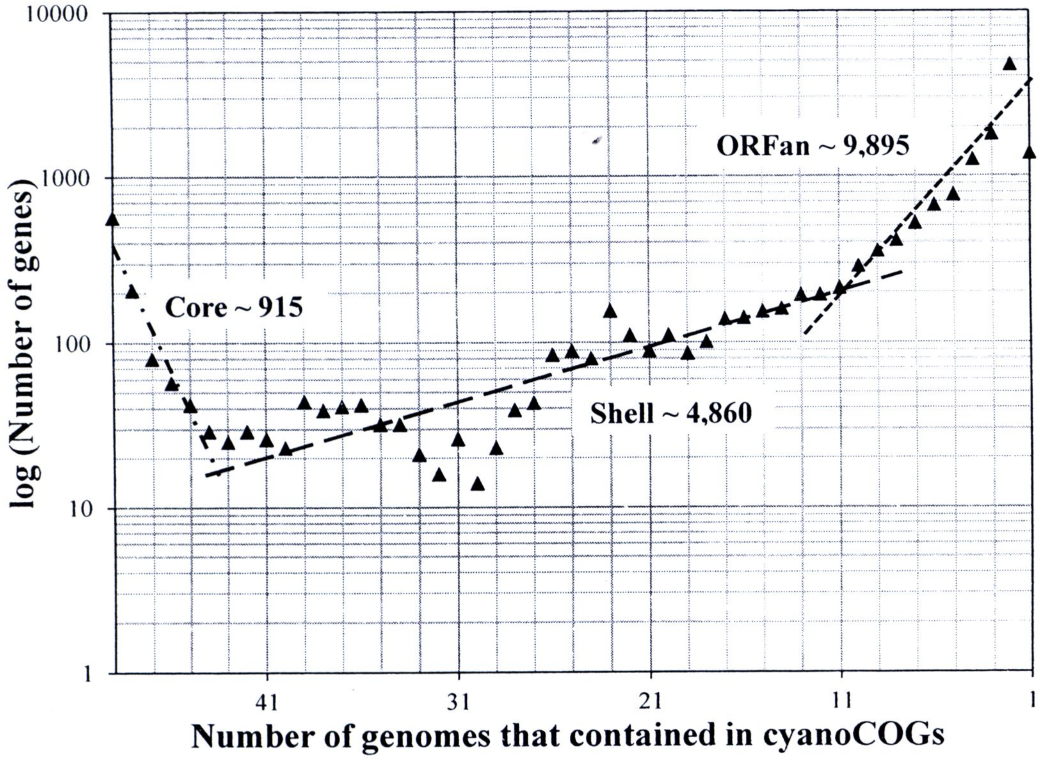

Figure 4.4 Distribution of the number of species in cyanoCOGs (a semi-logarithmic plot).

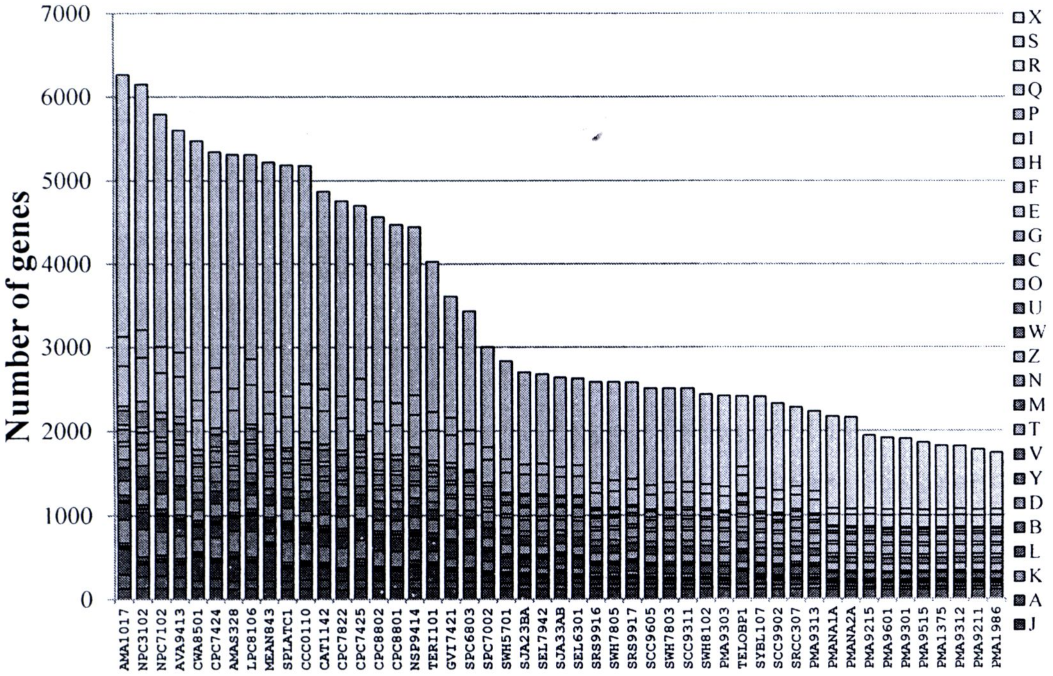

Figure 4.5 Functional breakdown of the entire set of cyanoCOGs of each species the function of cyanoCOGs (the genome abbreviations on the X-axis as shown in Table 3.1 and the abbreviation of COGs categories on the right Y-axis as shown in the Table 4.1).

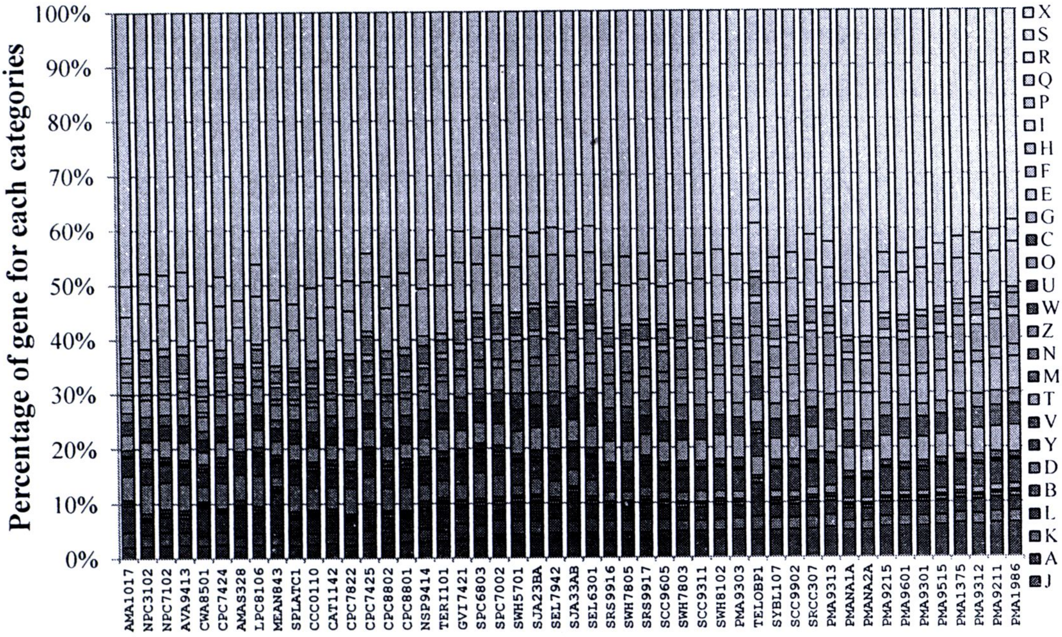

Figure 4.6 The percentage functional breakdown of the entire set of cyanoCOGs of each species (the genome abbreviations on the X-axis as shown in Table 3.1 and the abbreviation of COGs categories on the right Y-axis as shown in the Table 4.1).

Table 4.1 The NCBI COGs functional categories (Tatusov, *et al.*, 2001).

Information storage and processing	
J	Translation, ribosomal structure and biogenesis
A	RNA processing and modification
K	Transcription
L	Replication, recombination and repair
B	Chromatin structure and dynamics
Cellular processes and signaling	
D	Cell cycle control, cell division, chromosome partitioning
Y	Nuclear structure
V	Defense mechanisms
T	Signal transduction mechanisms
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
Z	Cytoskeleton
W	Extracellular structures
U	Intracellular trafficking, secretion, and vesicular transport
O	Posttranslational modification, protein turnover, chaperones
Metabolism	
C	Energy production and conversion
G	Carbohydrate transport and metabolism
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport and catabolism
Poorly characterized	
R	General function prediction only
S	Function unknown

4.2 Phyletic pattern

The notion of a phyletic pattern, which is the pattern of presence-absence of a cyanoCOGs in the analyzed set of species, has been developed in the original COGs study. Subsequently, phyletic pattern has been extensively employed for both functional prediction and starting material for evolutionary reconstruction. Figure 4.7 shows that the distribution of phyletic patterns in the new set of cyanoCOGs. The decay of the curve is remarkable steep, which is a substantial majority of the patterns (5,013 of 15,741) are unique, that is, represented by one cyanoCOGs only. Examination of the list of top 25 widespread cyanoCOGs is particularly instructive (as shown in table 4.2). In this list, 15 patterns are "trivial", which is represented in multiple species of a compact monophyletic group, such as *Nostocales* or *Oscillatioriales*. The one exception are the "all" pattern which describes the strictly defined core of 570 cyanobacterial genes represented in all species, and the other nine exception are the paralog pattern which describe the gene that occur in the unique genome. The most "non-trivial" pattern is the one that includes in cyanoCOGs that represent in two species which are *A. maxima* and *S. platensis* (760 cyanoCOGs).

Phyletic pattern of cyanobacteria was also used for reconstruction phylogenetic network by using the SplitTree 4.8 program (Huson and Bryant, 2006), then, the result shows in Figure 4.8. The phylogenetic network represent how close between each genome, which mean the closely related genome are closely link together. The individual branch length of the phylogenetic network represents how much individual genes are. On the other hand, the share branch length from two or more genomes shows the share properties between genomes, which represented by the orthologous gene from those genomes. This phylogenetic network can separate the cyanobacterial into several groups, such as, the *Nostocales*, *Oscillatoriales*, *Choococcales* and picocyanobacteria. This reconstructed phylogenetic network shown the same topology with the previous phylogenetic tree that using the conserved protein families across cyanobacteria genome (Swingley, *et al.*, 2008). This phylogenetic network can separate a marine picocyanobacteria from another cyanobacterial group. The prior study on the marine cyanobacteria showed the similar topology of phylogenetic networks with this study (Dufresne, et al., 2008). The organisms which have the same environmental niche are grouped together. For instance, both of SEL7942 and SEL6301 are *Synechococcus* spp., but they are not grouped with another *Synechococcus* spp. groups. Because they are the fresh water *Synechococcus* spp. the other are marine *Synechococcus* spp. On the other hand, both of SJA33AB and SJA23BA also separate to other *Synechococcus* groups, because these two organisms live in the extreme environment (Yellowstone).

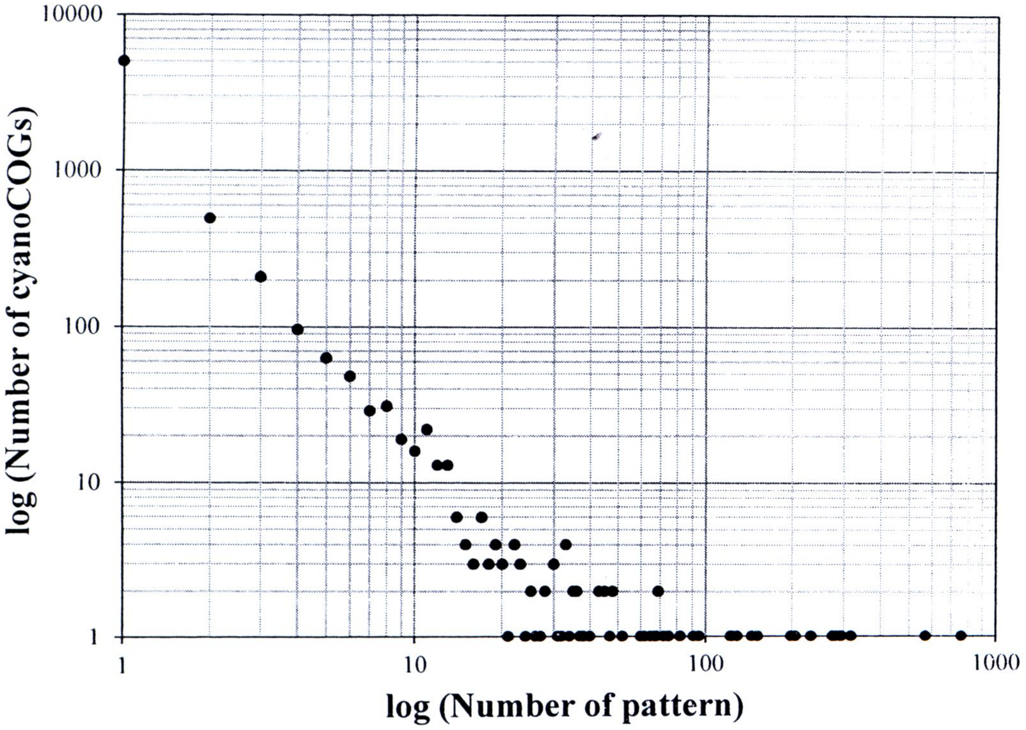

Figure 4.7 Distribution of phyletic patterns by the number of cyanoCOGs. This figure shows 5,013 unique patterns for cyanoCOGs and the most 25 frequent phyletic patterns have shown the Table 4.2.

Table 4.2 The 25 most frequent phyletic patterns in the cyanoCOGs.

Phyletic pattern§	Number of cyanoCOGs	Comment
0100000000000000000000000000000000000010000000000	760	Arthospira spp.
1111111111111111111111111111111111111111111111111	570	All species
0000000000000000000000000001100000000000000000000	318	
0000000000000000000000000000000000110000000000000	296	
0000000011000000000000000000000000000000000000000	293	
1000000000000000000000000000000000000000000000000	283	AMA1017 paralog
0010000000000001000000000000000000000000000000000	276	
0000000000000000000000000000000000110000000000000000	233	Synecchococcus elongatus
0000010100000000000000000000000000000000000000000	206	
0001100000000000000000000000000000000000000000000	199	
0010000000000011000000000000000000000000000000000	152	
0010000000000011100000000000000000000000000000000	145	Nostocales
0000000000000100000000000000000000000000000000000	129	MEAN843 paralog
0100000000001000000000000000000000000010000000000	124	
0000000000000000000000101000000000000000000000000	122	
0000000000001000000000000000000000000000000000000	95	LPC8106 paralog
0000000000010000000000000000000000000000000000000	91	GVI7421 paralog
0000000000000000000000000000000000000010000000000	82	SPLATC1 paralog
0010000000000010000000000000000000000000000000000	75	
1111111011111111111111111111111111111111111111111	72	
0000000000000010000000000000000000000000000000000	69	NPC3102 paralog
0000000000000000001011010110000000000000000000000	69	
0000000000000011000000000000000000000000000000000	68	
0000000000000000011111111111111100000001111111100	67	
0100000000001000000000000000000000000010000000001	65	Oscillatoriales
0100000000000000000000000000000000000000000000000	62	AMAS328 paralog
0000000000000000000000000000000000000000000000001	60	TERI101 paralog
0000001000000000000000000000000000000000000000000	52	CPC7425 paralog
1000001000000000000000000000000000000000000000000	48	
0001100000100000000000000000000000000000000000000	48	

§The phyletic pattern is the pattern that indicated absent (0) or present (1) for each genome. Each position of phyletic pattern are represent the genome of AMA1017, AMAS328, AVA9413, CAT1142, CCC110, CPC7424, CPC7425, CPC7822, CPC8801, CPC8802, CWA8501, GVI7421, LPC8106, MEAN843, NPC7102, NPU3102, NSP9441, PMA1375, PMA1986, PMA9211, PMA9215, PMA9301, PMA9303, PMA9312, PMA9313, PMA9315, PMA9601, PMANA1A, PMANA2A, SCC9311, SCC9502, SCC9605, SEL6301, SEL7942, SJA23BA, SJA33AB, SPC6803, SPC7002, SPLAC1, SRCC307, SRS9916, SRS9917, SWH5701, SWH7803, SWH7805, SWH8102, SYBL107, TELOBP1, and TERI101, respectively (Abbreviation as in Table 3.1).

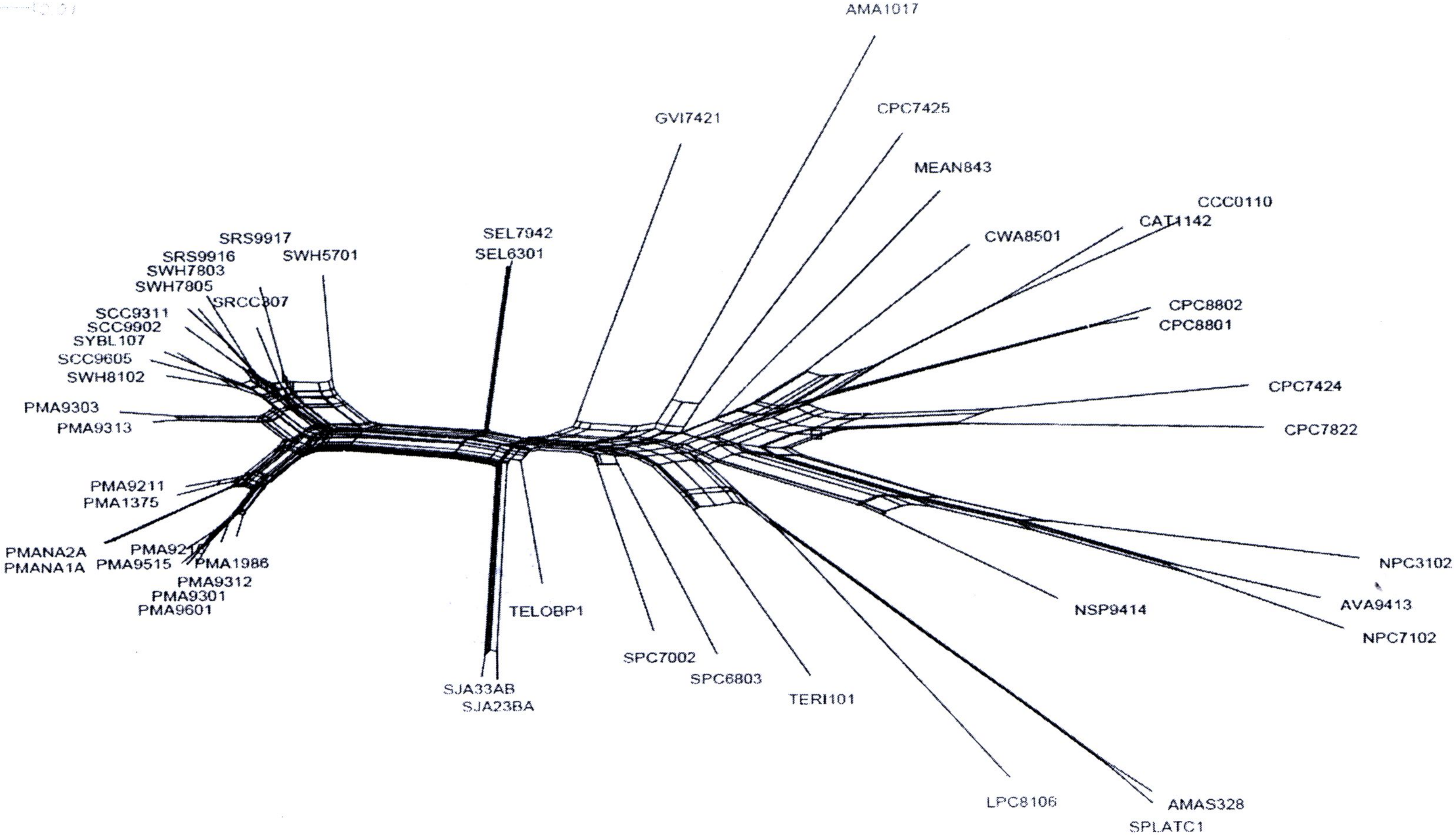

Figure 4.8 The phylogenetic networks, which is explained how closely related between each cyanobacterial genome is, are constructed by using gene-content phyletic patterns and the SplitTree 4.8 program (Huson and Bryant, 2006) (the genome abbreviation as shown in Table 3.1)

4.3 Phylogenomic of cyanobacteria

Tredditionally, the phylogenetic tree buildings base on a small subunit of ribosomal RNA (16S rRNA), which is the most popular molecular marker in cyanobacteria also (Honda, *et al.*, 1999). However, it is not sufficient for study at a sub-generic level because it is highly conserved among closely related species and strain. Then, there are several studies purposed to use another molecular marker such as using another proteins, using the consensus phylogenetic tree that reconstructed from several protiens, using the genome context, or using the concatenated ribosomal proteins in order to identify the phylogenetic relationships in the sub-generic level (Luo, *et al.*, 2008; Han, *et al.*, 2009). In this study, the concatenated ribosomal protein genes, including 20 large and 16 small subunits were used to reconstruct the evolutionary tree for the cyanobacteria. The inferred phylogenetic tree that is shown in the figure 4.9 was reconstructed by using the neighbor joining methods. The 1,000 bootstrapping was performed in order to evaluate the robustness of the inferred phylogenetic tree and result the high number of bootstrapping (more than 60 for every node). Then, the reconstructed evolutionary tree can also separate the cyanobacterial in to the specific clade, such as *Nostocales*, *Oscillatioriales*, *Chroococcales* or picocyanobacteria. The reconstructed phylogenetic tree from concatenated ribosomal protein genes is equivalent to the previous16s ribosomal RNA tree and paleontology study (Tomitani, *et al.*, 2006). In addition, this result also equivalent to the evolutionary tree that reconstruct by using the conserved protein families across the cyanobacterial species (Swingley, *et al.*, 2008).

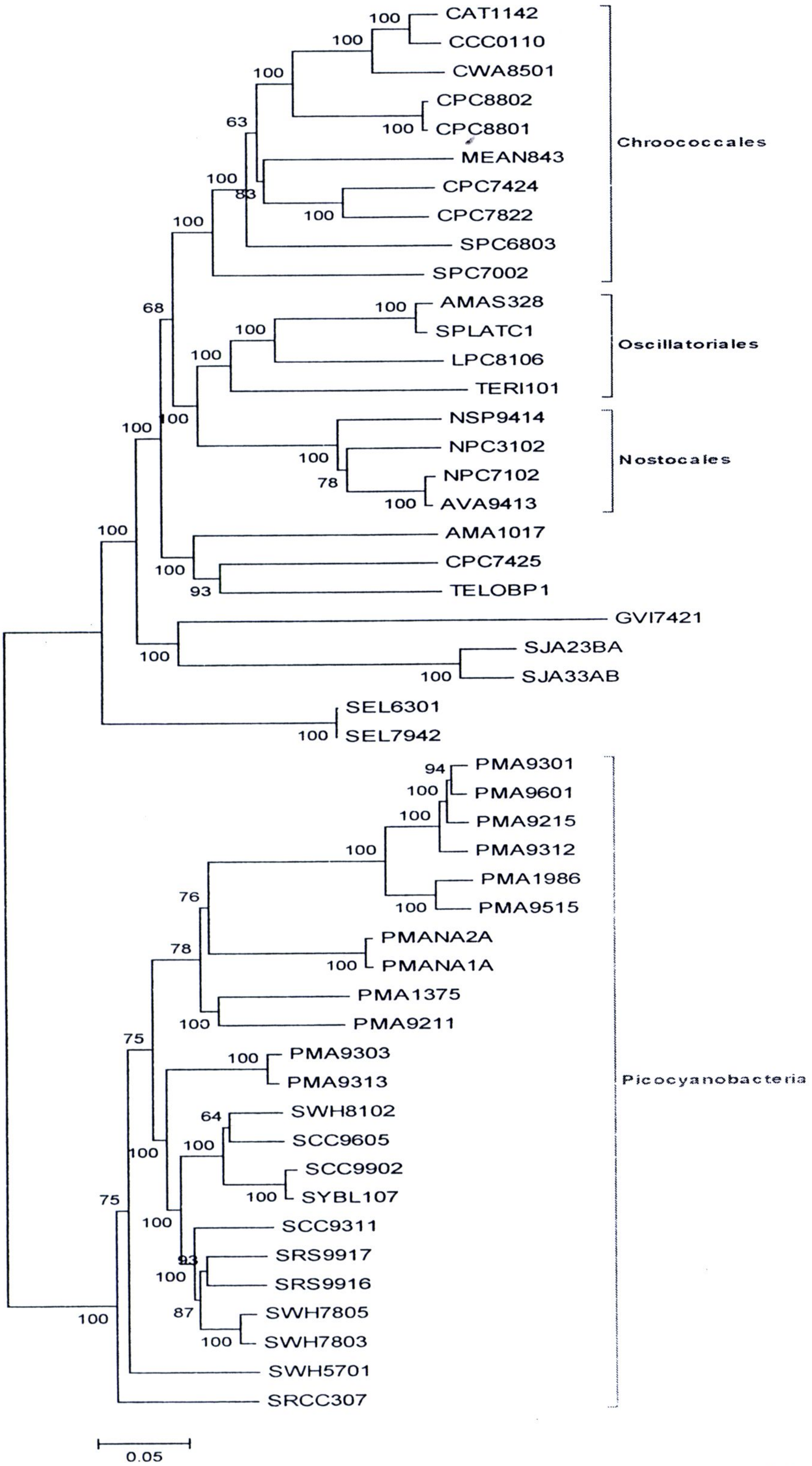

Figure 4.9 The evolutionary tree of cyanobacteria genome reconstructed by using concatenated ribosomal proteins with Neighbor Joining method (Abbreviation as in table 3.1). The number at each node represents the Bootstrapping value.

4.4 Evolutionary scenarios in cyanobacteria lineage

For the evolution for cyanobacterial lineage, the biological properties and environmental niches of ancestral state for each cyanobacterial groups in the figure 4.10 as represented into the "LCCA", "A", "B", "C", "D", "E", "F", and "G" group were described in the table 4.3. The prior cyanoCOGs and evolutionary tree were used for tracing the evolutionary scenarios in the cyanobacterial lineage by applying the parsimonious evolutionary scenario algorithm (Mirkin, et al., 2003). The number of gene loss and gain along each branch of the tree was shown in figure 4.11. The LCCA is conservatively estimated to contain 2,468 genes compared to 3,128, 2,722, and 2,055 genes of the last common ancestor of *Nostocales*, *Oscillatoriales*, and picocyanobacteria, respectively. When comparing with 181 signature genes in cyanobacteria that have been reported before (Martin, *et al*., 2003), the LCCA shown a large amount of genes for its cellular processing. However, the 181 signature genes are the shared orthologous gene in the cyanobacterial species that cannot found in other organismal groups. Moreover, there has extensively lost in the marine picocyanobacteria clades, and extensively gained in the fresh water cyanobacteria, such as the cyanobacteria that habits in the extreme environmental condition, *Nostocales*, *Oscillatoriales*, and *Chroococcales*, the number of gene gain and loss was shown in the Figure 4.11. For assigning the biological properties of every cyanobacterial ancestral states, the analysis of the gene sets of cyanoCOGs by study their functional breakdown with the cyanoCOGs categories was performed. The amount and the proportion of the metabolism gene groups of the cyanoCOGs for each ancestral genome represented in the figure 4.12.

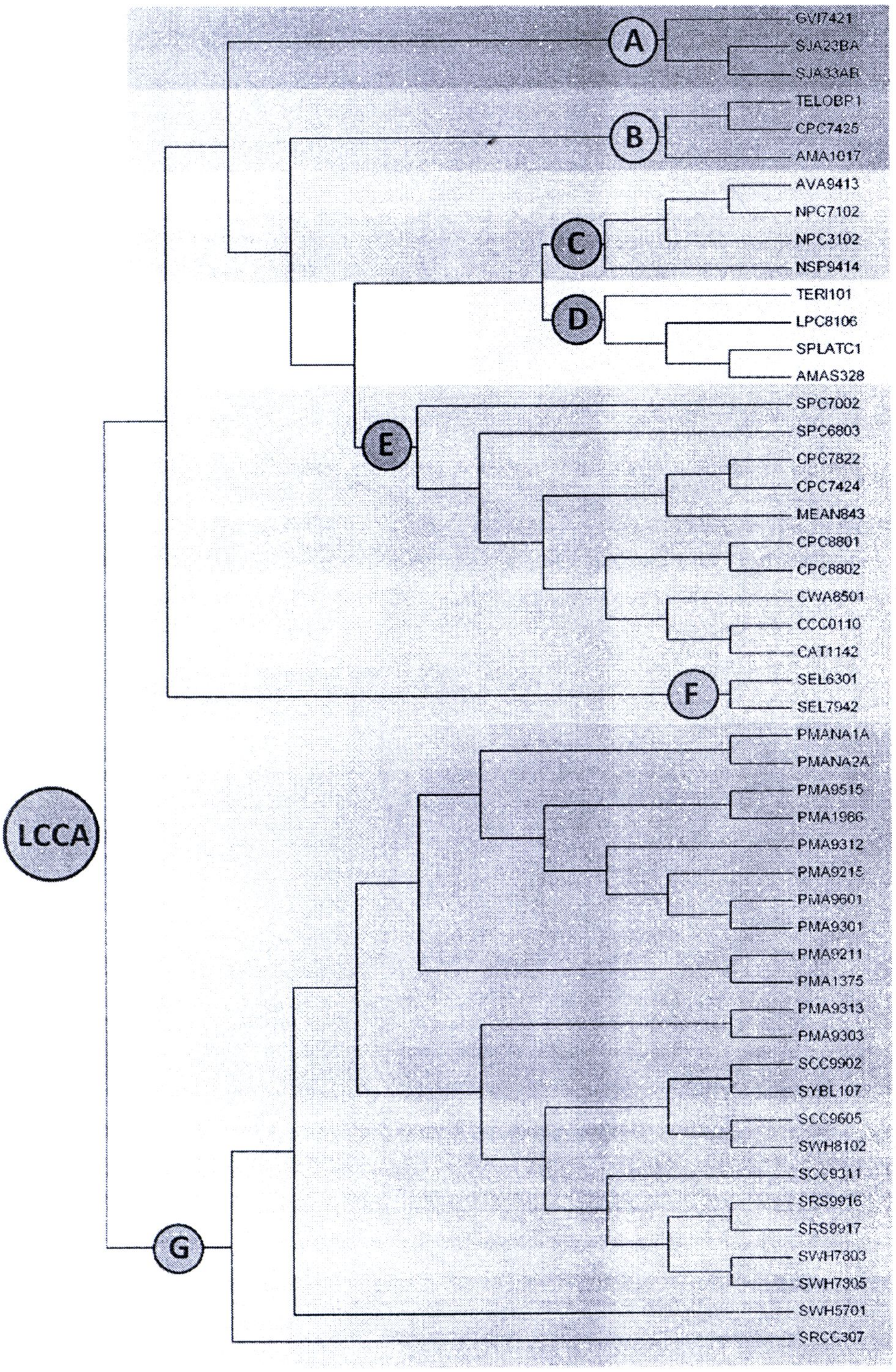

Figure 4.10 The cyanobacterial ancestral form represent by "LCCA", "A", "B", "C", "D", "E", "F", and "G" , the description of their share biological property of each group was described in the Table 4.3 (Abbreviation as in Table 3.1).

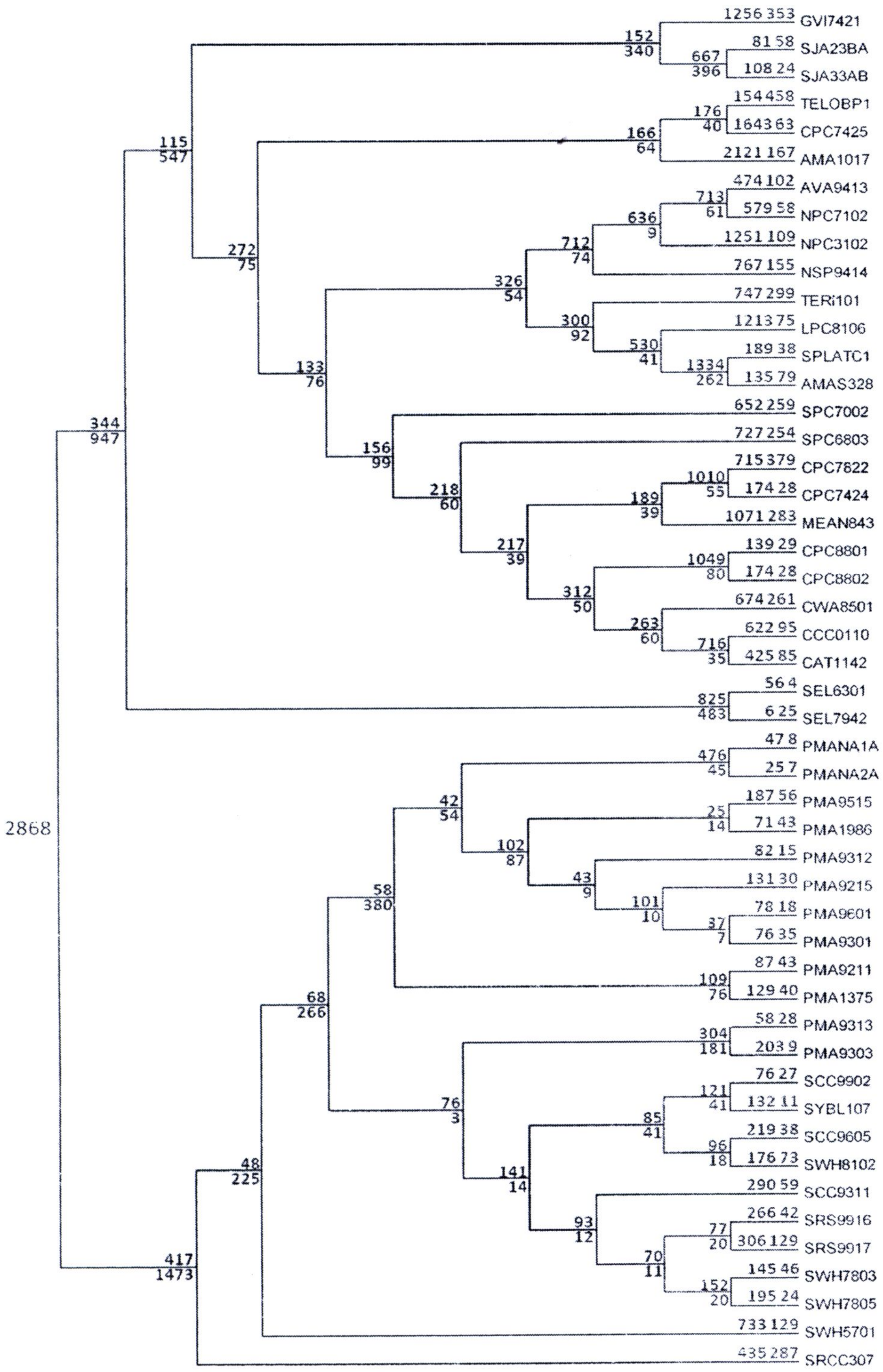

Figure 4.11 The summary of gene gain and loss along the cyanobacterial lineages. Each branch is labeled by 2 numbers: blue, the number of gained along the branch; red, the number of lost along the branch (Abbreviation as in Table 3.1).

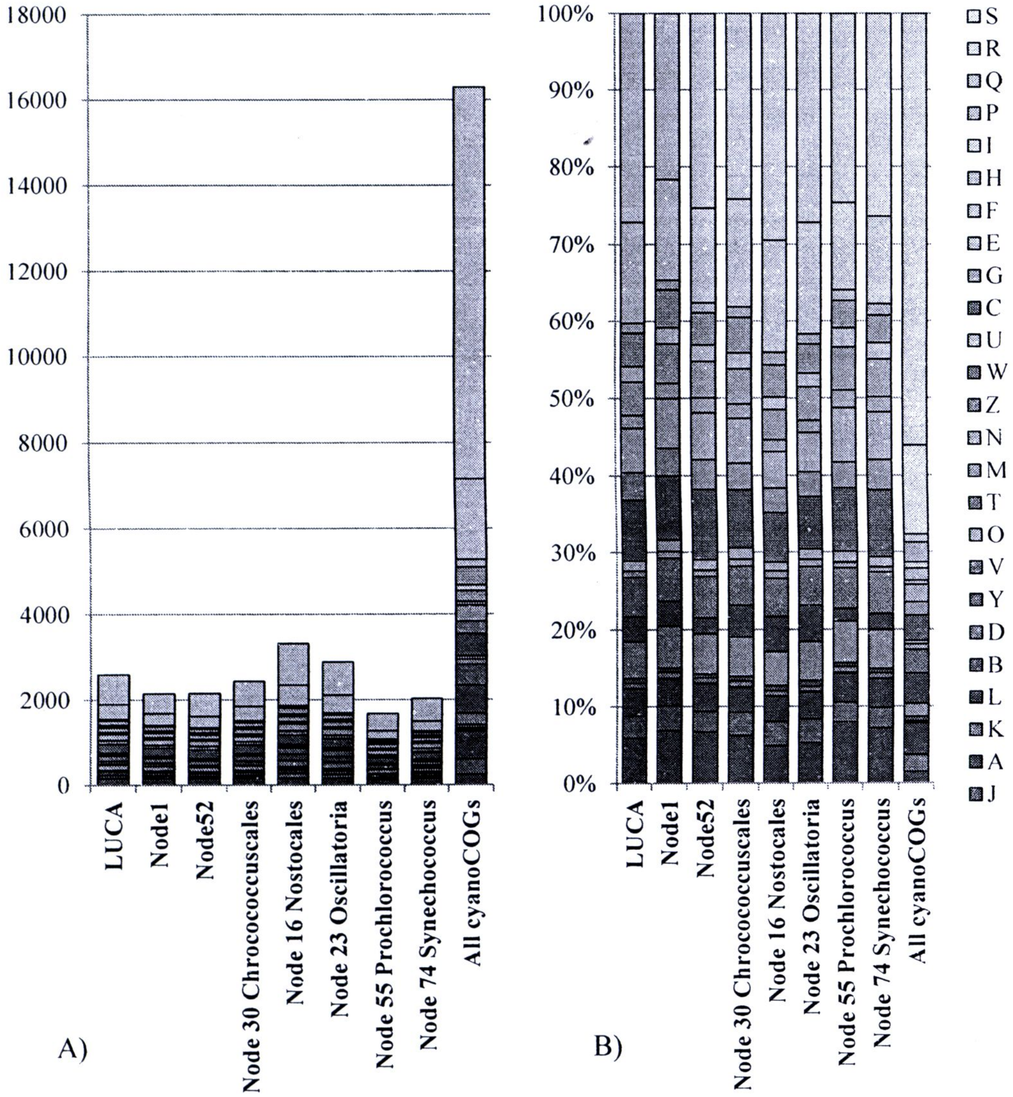

Figure 4.12 The functional beakdown of the entire set of cyanoCOGs and the each level of evolutionary timeline the cyanoCOGs (the ancestral node on the X-axis and the abbreviation of COGs categories on the right Y-axis as shown in the Table 4.1). (A) Represented the number of the cyanoCOGs catecgories. (B) Represented the percentage of the cyanoCOGs categories.

Table 4.3 The description of cyanobacterial group according to figure 4.11 relate to their biological property and environmental niche.

Group	Description	Inferred biological property and environmental niche
LCCA	Last Cyanobacterial Common Ancestor	Photoautotrophs
A	GVI7421, SJA33AB, SJA23BA Ancestor	Inhabit in the extreme environmental niches
B	TELOBP1, AMA1017, CPC7425 Ancestor	Uncharacterized
C	Nostocales Ancestor	Filamentous nitrogen fixing cyanobacteria
D	Oscillatoriales Ancestor	Non-nitrogen fixing, filamentous cyanobacteria
E	Chroococcales Ancestor	Unicellular, nitrogen fixing cyanobacteria
F	Synechococcus elongates Ancestor	Obligate photoautotrophs
G	Picocyanobacteria Ancestor	Marine cyanobacteria, small genome size

4.5 Evolutionary scenarios of photosynthetic apparatus in cyanobacterial lineage

In the same way to trace the genomic evolution scenarios in the cyanobacterial lineage in the prior part, the parsimonious evolutionary algorithm was applied to the photosynthetic apparatus genes for tracing the evolutionary scenarios of this genes group in the cyanobacterial lineage. Reconstruction of gene gain and loss along the evolutionary trajectory in cyanobacterial lineage was performed using the parsimonious evolutionary scenarios algorithm (Mirkin, *et al.*, 2003). This algorithm was applied with the entire genomic repertoire or the interested gene sets. For this study, the entire set of cyanoCOGs was applied to delineate the biological properties and genomic function of each projected cyanobacterial ancestral nodes as described in table 4.3. In addition, the photosynthesis apparatus gene sets in cyanoCOGs were used to unravel the evolution dynamics of these important gene sets in cyanobacteria as illustrated by the number of gained and lost genes for each cyanobacterial lineages (as shown in the figure 4.11). These photosynthesis apparatus gene sets in cyanobacteria can be subdivided into nine groups (in total of 259 genes) comprised of photosystem I proteins, photosystem II proteins, phycobillisome proteins, chlorophyll–binding proteins, chlorophyll biosynthesis enzyme, cytochrome bf6 complex subunit, water–soluble electron carriers, Calvin cycle enzymes, regulatory and uncharacterized chloroplast proteins. The existing photosynthetic gene groups for various cyanobacteria ancestral states are summarized in table 4.4. There are 51 genes that inherited into all descendent genomes from the LCCA, on the other hand, 36 genes have been lost during the course of evolution. The loss of ancestral genes was counterbalanced by the emergence of 172 genes via the HGT or duplication of the existed genes.

From our observation, with the photosynthesis apparatus evolution in cyanobacterial lineage, the LCCA is hypothesized to be a photoautotroph, since it contains sufficient photosynthesis machineries. Along the cyanobacterial evolution as illustrated by the number of gene gains and losses in figure 4.13, the extensive gene losses were discovered in the picocyanobacteria ancestor and *Prochlorococcus* ancestor. Conversely, the extensive gene gains were discovered particularly in the freshwater *Synenccococcus* ancestor, the cyanobacteria that habit in the extreme environment niches and the Nitrogen–fixing filamentous cyanobacteria (*Nostocales*). The ancestors of extreme environmental cyanobacteria and *Oscillatoriales* have gained their photosynthesis apparatus accessories, such as protoporphyrinogen oxidase, during the course of evolution. Considering phosphoribulokinase enzyme, there are three scenarios that might have occurred during the evolution of cyanobacterial lineage. The first and the second events were the transferring of this enzyme from LCCA to the fresh water cyanobacteria and the picocyanobacteria. The third event was the acquiring of this gene at the common ancestor node of *Nostocales* and *Oscillatoriales*. Another notable acquisition is a group of phycobilosome linker proteins, which are different among cyanobacterial clades. For instance, fresh water cyanobacteria inherited this group of proteins from LCCA, while picocyanobacteria gained phycobilisome proteins from other sources. Furthermore, the phycocyanin, phycoerythrin, and allophycocyanin were inherited from LCCA to entire cyanobacterial ancestral states, whereas Phrochlorococcus ancestor extensively lost these genes group. This result represents that this organismal group renders the different light harvesting complexes (Guan, *et al.*, 2007).

Table 4.4 The photosynthesis-related genes in ancestral cyanobacterial genomes. The description of cyanobacterial group according to Figure 4.11 relate to their biological property and environmental niche.

Gene products	LCCA	A	B	C	D	E	F	G
Photosystem I Proteins								
• PsaA1, PsaA2, PsaB, PsaD, PsaF, PsaL	Ancestral genes that inherited to all descendent genomes							
• PsaI(b), PsaJ(b), PsaK(b), PsaM(b), PsaX	Gene gained by HGT							
• PsaJ(c), PsaK(c)	Y	N	Y	Y	Y	Y	Y	Y
• PsaM(c)	Y	N	N	Y	Y	Y	N	N
• PsaI(c)	Y	N	N	N	N	N	N	Y
Photosystem II Proteins								
• PsbA(a), PsbD, PsbH, PsbN, PsbO, PabP, PsbQ, PsbU PsbW	Ancestral genes that inherited to all descendent genomes							
• PsaA(b),PsbI, PsbK, PsbL, PsbX(b), PsbY(b), PsbZ(b)	Gene gained by HGT							
• PsbJ	Y	Y	Y	Y	N	N	N	Y
• PsbM	Y	Y	Y	Y	Y	Y	N	Y
• PsbT, PsbY(a), PsbZ(a)	Y	N	N	N	N	N	N	Y
Phycobilisome proteins								
• ApcA§, ApcB(a), §, ApcD(a), §, CpcA(a), §, CpcB(a), §, CpcD(a), §, CpcE§, CpcF§, CpcG§, CpcI, LCM(a), §, LC§	Ancestral genes that inherited to all descendent genomes							
• ApcB(b), ApcD(b), CpcA(b), CpcB(b), CpcD(b), CpcH(b), CpeR, LCM(b), L-RC, NblA	Gene gained by HGT							
• CpcA(c), §,CpcC§, CpeS	Y	N	N	N	N	N	N	Y
• ApcA(c), §, CpcB(c), §	Y	Y	N	N	N	N	Y	Y
• CpcD(c)	Y	Y	Y	Y	Y	Y	Y	N
Chlorophyll-binding proteins								
• CAB/ELIP/HLIP	Ancestral genes that inherited to all descendent genomes							
• PcbB, PcdD, PcdE, PcdF, PcdH	Gene gained by HGT							
• IsiA	Y	N	Y	Y	Y	Y	Y	Y
Chlorophyll biosynthesis enzymes								
• ChlA, ChlB, ChlD, ChlG, ChlI, ChlL, ChlM, ChlN, AscF(a)	Ancestral genes that inherited to all descendent genomes							
• AscF(b)	Gene gained by HGT							
Cytochrome b_6f complex subunit								
• PetA, PetB, PetC, PetD(a), PetF1,	Ancestral genes that inherited to all descendent genomes							
• PetD(b), PetG(b), PetL, PetM	Gene gained by HGT							
• PetG(a)	Y	Y	Y	N	N	N	Y	Y
• PetN	Y	N	N	N	N	N	Y	Y
Water-soluble electron carriers								
• Plastocyanin (PetE), PetJ	Ancestral genes that inherited to all descendent genomes							
Calvin cycle enzymes								
• Phosphoribulokinase(c), RbcX, RbcR	Y	Y	Y	Y	Y	Y	Y	N
• Phosphoribulokinase(c)	Y	N	N	N	N	N	N	Y
Regulatory and uncharacterized chloroplast homolog proteins								
• Ycf48	Ancestral genes that inherited to all descendent genomes							
• GUN4, YcfA	Gene gained by HGT							

The presence or absence of gene coding for proteins of each type in the cyanobacterial ancestral state; Y, gene presence; N, gene absent,

§These gene families are lost at *Prochlorococcus* ancestral state.

The superscripts (a) to (c) indicate the same gene family, but there are clustered in the different cyanoCOGs
based on low amino acid sequences similarity.

(a)The ancestral genes have inherited to all descendent genomes.

(b)The new genes have acquired at some point of the cyanobacterial lineage.

(c)The ancestral genes have inherited to some cyanobacterial lineage.

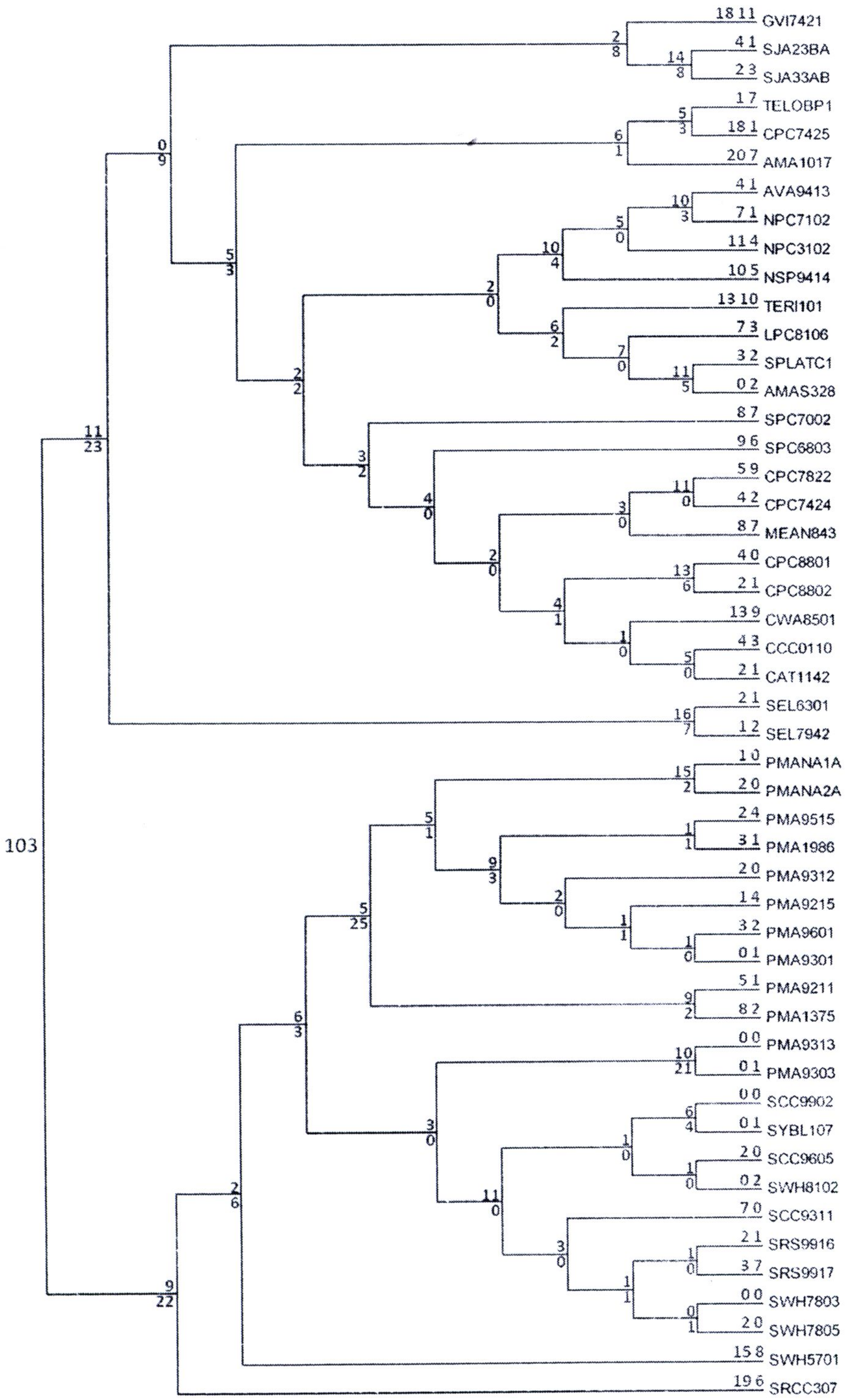

Figure 4.13 The gain and loss events of the photosynthesis apparatus gene along the cyanobacterial lineages. Each branch is labeled by 2 numbers: blue, the number of gained along the branch; red, the number of lost along the branch (genome abbreviation as in Table 3.1).

For modern descendent genomes, we found that the complexity of photosynthesis of these cyanobacteria has increased as evident by the homoplasy of PsbA and PsbD proteins that might be transduced by horizontal gene transfer via cyanophages as proposed to occur in picocyanobacteria (Lindell, *et al.*, 2004; Sandaa, *et al.*, 2008; Millard, *et al.*, 2009; Sharon, *et al.*, 2009). Obviously, the photosynthesis apparatus for each cyanobacterial clade and their evolutionary scenarios of these genes are different. To delineate these situations, the adaptation to the new environmental niches of the descendent genomes was taken into consideration. For example, the marine cyanobacteria or the cyanobacteria that thrive in the extreme environments need to duplicate their ancestral genes or acquire more genes from others for the appropriate functions to meet the habitat or environmental requirements (Luque, *et al.*, 2008).

For the photosynthetic gene sets, the occurrence of the same gene function (name) was clustered into the difference cyanoCOGs. The further analysis of this gene groups was performed, such as the phylogenetic analysis for this gene group. The phylogenetic tree of *PsaI* and *PsaM* was reconstructed by using the same gene from cyanoCOGs and this gene form the photoautotroph organism (as shown in the figure 4.14), this phylogenetic analysis indicated that there are very small conserved domain between these cyanoCOGs and has the distant phylogeny between these cyanoCOGs. Moreover, the analysis of the *PsaI* protein was reconstruction of this gene with the *PsaI* gene from other photoautotroph organism.

For analysis the horizontal gene transfer events along the cyanobacteria lineages, the using of the orthologous protein from cyanobacteria another organimal groups were performed to reconstruct the evolutionary tree. For example, the phylogenetic analysis of the *PsaI* protein in cyanobacteria and plastids of another organismal groups has indicated that the closely related between this gene in cyanobacteria and plastids. In particularly, the *PsaI* proteins from *Nostocales* (Ortho14395) is very closely relate to the *PsaI* of the red algae *Cyanidium caldarium*, and fresh water cyanobacterial *PsaI* proteins are closely related to the *PsaI* from the plastid genomes more than the marine picocyanobacterial genes (as show in the figure 4.15). Then, the horizontal gene transfer of this gene between the red algae and *Nostocales* was taking into consideration. In the environmental perspective, the fresh water cyanobacteria and the fresh water red algae are living in the same habitat and environmental niches, and then it is easily to transfer the genetic material between these two organisms (Rogers, *et al.*, 2007).

Recent study indicated that several photosynthesis apparatus genes could be found in the marine virus genomes and it cause the horizontal gene transfer across genome (Sharon, *et al.*, 2009). The phylogenetic analysis of *PsaA* and *PsaD* were performed by using the cyanobacteria proteins and the cyanophages proteins in order to determine the genomic distance between those organisms. Phylogenetic tree of *PsaA* and *PsaD* proteins, which done by using the neighbor joining methods with 1,000 bootstrapping are shown in the figure 4.16. The result indicated that *PsaA* and *PsaD* proteins of cyanophages are closely related to the genes that found in *Prochlorococcus*. The horizontal gene transfer between the cyanophages and *Prochlorococcus* genome was taken into consideration.

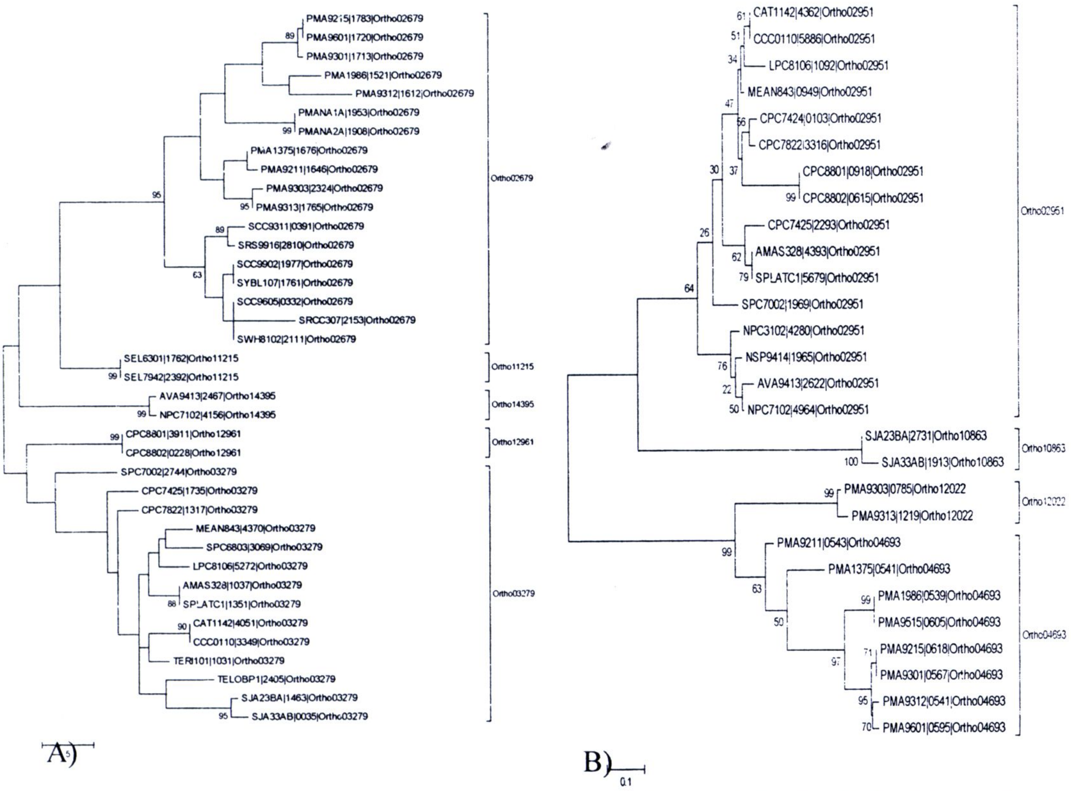

Figure 4.14 The photosynthetic tree reconstructed from (A) PsaI, (B) PsaM proteins depicts that the protein from difference cyanoCOGs has clustered into the difference clades.

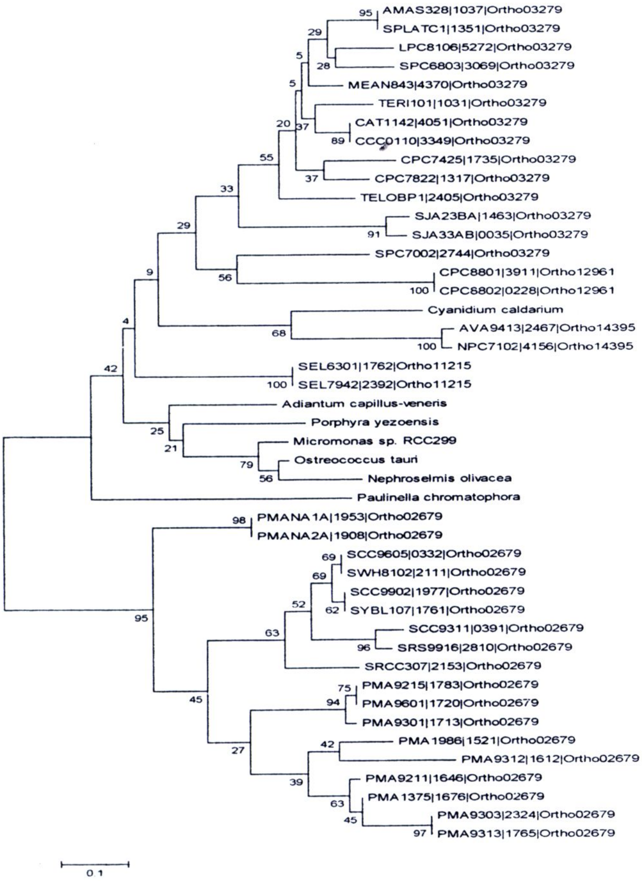

Figure 4.15 The evolutionary tree of *PsaI* proteins from cyanobacteria with the *PsaI* protein from plastids of other organismal groups (*Cyanidium caldarium*, *Adiantum capillus-veneris*, *Porphyra yezoensis*, *Micromonas* sp., *Ostreococcus tauri*, *Nephroselmis olivacea*, and *Paulinella chromatophora*).

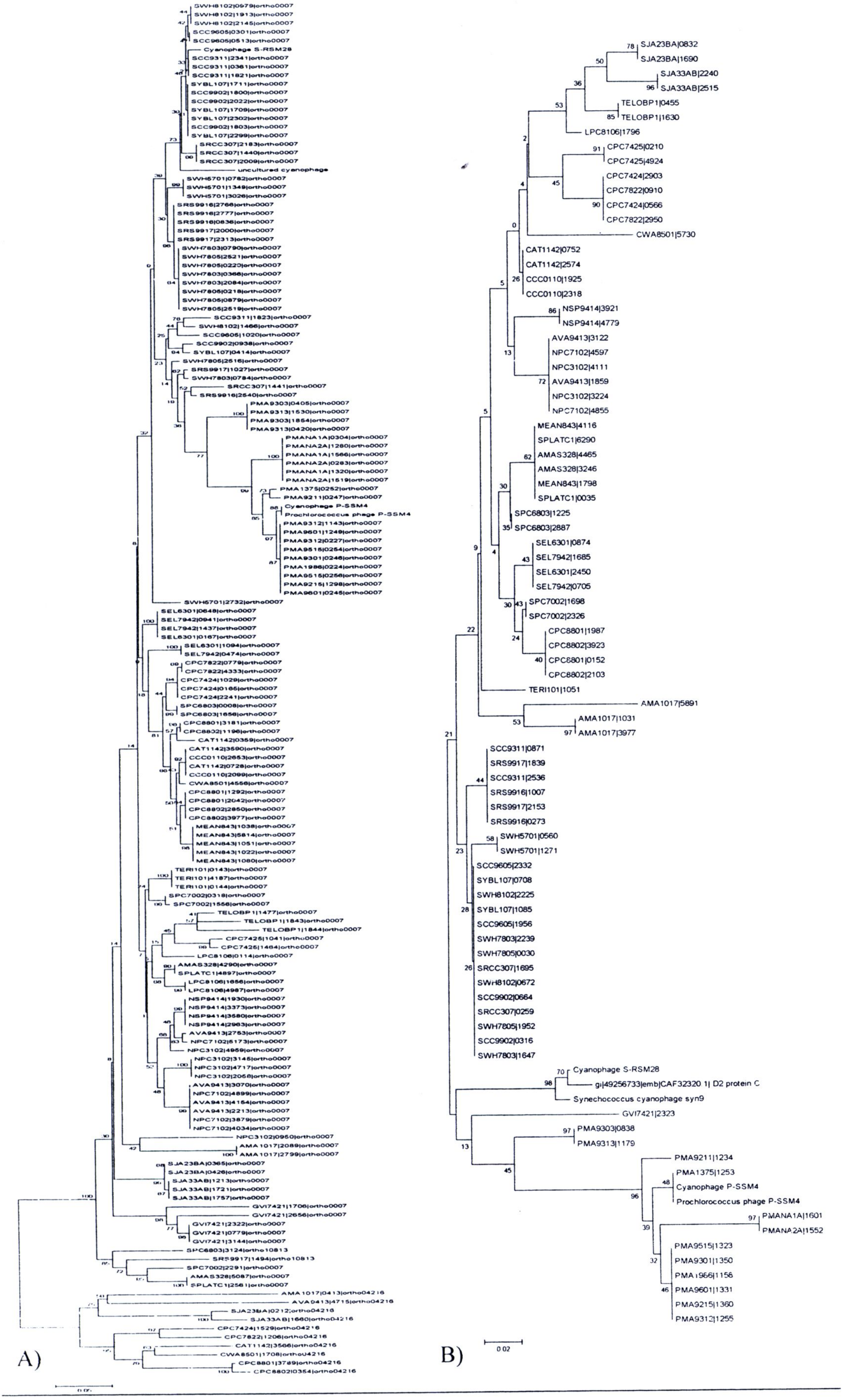
A)
SWH8102|0979|ortho0007
SWH8102|1913|ortho0007
SWH8102|2145|ortho0007
SCC9605|0301|ortho0007
SCC9605|0513|ortho0007
Cyanophage S-RSM28
SCC9311|2341|ortho0007
SCC9311|0361|ortho0007
SCC9311|1821|ortho0007
SYBL107|1711|ortho0007
SCC9902|1800|ortho0007
SCC9902|2022|ortho0007
SYBL107|1709|ortho0007
SYBL107|2302|ortho0007
SCC9902|1803|ortho0007
SYBL107|2299|ortho0007
SRCC307|2183|ortho0007
SRCC307|1440|ortho0007
SRCC307|2009|ortho0007
uncultured cyanophage
SWH5701|0782|ortho0007
SWH5701|1349|ortho0007
SWH5701|3026|ortho0007
SRS9916|2766|ortho0007
SRS9916|2777|ortho0007
SRS9916|0836|ortho0007
SRS9917|2000|ortho0007
SRS9917|2313|ortho0007
SWH7803|0790|ortho0007
SWH7805|2521|ortho0007
SWH7805|0220|ortho0007
SWH7803|0366|ortho0007
SWH7803|2084|ortho0007
SWH7805|0218|ortho0007
SWH7805|0879|ortho0007
SWH7805|2519|ortho0007
SCC9311|1823|ortho0007
SWH8102|1466|ortho0007
SCC9605|1020|ortho0007
SCC9902|0938|ortho0007
SYBL107|0414|ortho0007
SWH7805|2516|ortho0007
SRS9917|1027|ortho0007
SWH7803|0784|ortho0007
SRCC307|1441|ortho0007
SRS9916|2540|ortho0007
PMA9303|0405|ortho0007
PMA9313|1530|ortho0007
PMA9303|1854|ortho0007
PMA9313|0420|ortho0007
PMANA1A|0304|ortho0007
PMANA2A|1280|ortho0007
PMANA1A|1566|ortho0007
PMANA2A|0283|ortho0007
PMANA1A|1320|ortho0007
PMANA2A|1519|ortho0007
PMA1375|0252|ortho0007
PMA9211|0247|ortho0007
Cyanophage P-SSM4
Prochlorococcus phage P-SSM4
PMA9312|1143|ortho0007
PMA9601|1249|ortho0007
PMA9312|0227|ortho0007
PMA9515|0254|ortho0007
PMA9301|0246|ortho0007
PMA1986|0224|ortho0007
PMA9515|0256|ortho0007
PMA9215|1298|ortho0007
PMA9601|0245|ortho0007
SWH5701|2732|ortho0007
SEL6301|0648|ortho0007
SEL7942|0941|ortho0007
SEL7942|1437|ortho0007
SEL6301|0167|ortho0007
SEL6301|1094|ortho0007
SEL7942|0474|ortho0007
CPC7822|0779|ortho0007
CPC7822|4333|ortho0007
CPC7424|1029|ortho0007
CPC7424|0165|ortho0007
CPC7424|2241|ortho0007
SPC6803|0008|ortho0007
SPC6803|1656|ortho0007
CPC8801|3181|ortho0007
CPC8802|1196|ortho0007
CAT1142|0359|ortho0007
CAT1142|3590|ortho0007
CCC0110|2653|ortho0007
CAT1142|0728|ortho0007
CCC0110|2099|ortho0007
CWA8501|4556|ortho0007
CPC8801|1292|ortho0007
CPC8801|2042|ortho0007
CPC8802|2850|ortho0007
CPC8802|3977|ortho0007
MEAN843|1038|ortho0007
MEAN843|5814|ortho0007
MEAN843|1051|ortho0007
MEAN843|1022|ortho0007
MEAN843|1080|ortho0007
TERI101|0143|ortho0007
TERI101|4187|ortho0007
TERI101|0144|ortho0007
SPC7002|0318|ortho0007
SPC7002|1556|ortho0007
TELOBP1|1477|ortho0007
TELOBP1|1843|ortho0007
TELOBP1|1844|ortho0007
CPC7425|1041|ortho0007
CPC7425|1464|ortho0007
LPC8106|0114|ortho0007
AMAS328|4290|ortho0007
SPLATC1|4897|ortho0007
LPC8106|1656|ortho0007
LPC8106|4987|ortho0007
NSP9414|1930|ortho0007
NSP9414|3373|ortho0007
NSP9414|3580|ortho0007
NSP9414|2963|ortho0007
AVA9413|2753|ortho0007
NPC7102|5173|ortho0007
NPC3102|4959|ortho0007
NPC3102|3145|ortho0007
NPC3102|4717|ortho0007
NPC3102|2058|ortho0007
AVA9413|3070|ortho0007
NPC7102|4899|ortho0007
AVA9413|4154|ortho0007
AVA9413|2213|ortho0007
NPC7102|3879|ortho0007
NPC7102|4034|ortho0007
NPC3102|0950|ortho0007
AMA1017|2089|ortho0007
AMA1017|2799|ortho0007
SJA23BA|0365|ortho0007
SJA23BA|0426|ortho0007
SJA33AB|1213|ortho0007
SJA33AB|1721|ortho0007
SJA33AB|1757|ortho0007
GVI7421|1706|ortho0007
GVI7421|2656|ortho0007
GVI7421|2322|ortho0007
GVI7421|0779|ortho0007
GVI7421|3144|ortho0007
SPC6803|3124|ortho10813
SRS9917|1494|ortho10813
SPC7002|2291|ortho0007
AMAS328|5087|ortho0007
SPLATC1|2561|ortho0007
AMA1017|0413|ortho04216
AVA9413|4715|ortho04216
SJA23BA|0212|ortho04216
SJA33AB|1660|ortho04216
CPC7424|1529|ortho04216
CPC7822|1206|ortho04216
CAT1142|3566|ortho04216
CWA8501|1708|ortho04216
CPC8801|3789|ortho04216
CPC8802|0354|ortho04216
0.05
B)
SJA23BA|0832
SJA23BA|1690
SJA33AB|2240
SJA33AB|2515
TELOBP1|0455
TELOBP1|1630
LPC8106|1796
CPC7425|0210
CPC7425|4924
CPC7424|2903
CPC7822|0910
CPC7424|0566
CPC7822|2950
CWA8501|5730
CAT1142|0752
CAT1142|2574
CCC0110|1925
CCC0110|2318
NSP9414|3921
NSP9414|4779
AVA9413|3122
NPC7102|4597
NPC3102|4111
AVA9413|1859
NPC3102|3224
NPC7102|4855
MEAN843|4116
SPLATC1|6290
AMAS328|4465
AMAS328|3246
MEAN843|1798
SPLATC1|0035
SPC6803|1225
SPC6803|2887
SEL6301|0874
SEL7942|1685
SEL6301|2450
SEL7942|0705
SPC7002|1698
SPC7002|2326
CPC8801|1987
CPC8802|3923
CPC8801|0152
CPC8802|2103
TERI101|1051
AMA1017|5891
AMA1017|1031
AMA1017|3977
SCC9311|0871
SRS9917|1839
SCC9311|2536
SRS9916|1007
SRS9917|2153
SRS9916|0273
SWH5701|0560
SWH5701|1271
SCC9605|2332
SYBL107|0708
SWH8102|2225
SYBL107|1085
SCC9605|1956
SWH7803|2239
SWH7805|0030
SRCC307|1695
SWH8102|0672
SCC9902|0664
SRCC307|0259
SWH7805|1952
SCC9902|0316
SWH7803|1647
Cyanophage S-RSM28
gi|49256733|emb|CAF32320 1| D2 protein C
Synechococcus cyanophage syn9
GVI7421|2323
PMA9303|0838
PMA9313|1179
PMA9211|1234
PMA1375|1253
Cyanophage P-SSM4
Prochlorococcus phage P-SSM4
PMANA1A|1601
PMANA2A|1552
PMA9515|1323
PMA9301|1350
PMA1966|1156
PMA9601|1331
PMA9215|1360
PMA9312|1255
0.02

Figure 4.16 The phylogenetic tree of *PsaA* (A) and *PsaD* (B) proteins of cyanobacteria with the same proteins from cyanophages. The number in each node represents the Bootstrapping values.