CHAPTER 2

LITERATURE REVIEWS

Cyanobacteria, also known as cyanoprokaryotes, cyanophytes or blue-green algae, are general terms for the microorganism capable of oxygenic photosynthesis that uses H_2O as their electron donor in light, O_2 being a by product of photosynthetic activity. These comprise over 1500 species with various morphologies and species-specific characteristics such as cell movement, cell differentiation, and nitrogen fixation. They can occur in great masses, and are important in many aquatic and terrestrial communities for their substantial biomass and primary production, production of toxic compounds, creation of stromatolites, boring of limestone substrates, and maintain an importance in symbioses (Carr and Whitton, 1982; Whitton and Potts, 2000; Komarek, 2003).

2.1 Thermophilic cyanobacteria

One of the most remarkable features of cyanobacteria is the broad range of thermal habitats in which they can achieve colonization and growth. The upper thermal limit for microbial life has been of great interest not only in terms of the ecology of extreme environments, but also in the search for novel biomolecules of the biotechnological potential, and for providing insights into the processes of cellular evolution (Vincent and Howard-Williams, 2001). Cyanobacteria occur in almost every illuminated hot spring above a pH of 5 and below 73-74°C (Castenholz, 1973). Cyanobacteria accumulate or at least persist in mat-like layers throughout the year. In this sense hot springs offer a unique habitat. Most hot spring water contains a relatively stable quantity of solutes which provides a major source of most essential nutrients. Some nutrients may drain into thermal streams or pools from the surface, and N_2 for nitrogen-fixing forms may enter from the atmosphere. The solute concentration and quality depend on the groundwater, the depth and length of the subterranean aquifer, the wall rocks, the temperature, and the leaching and carrying capacity of the water. In almost all thermal areas, steam-moistened and heated semi-aerobic mats of cyanophytes develop (Edwards, 1990).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Cyanobacteria have mechanisms for promoting protein stability at high temperatures. The protein is stabilized at high temperature (thermostability) by increasing the internal ionic bonding possibly by altering the terminal ionic strength of the cell (Edwards, 1990). There are many advantages for the use of thermophiles in biotechnological processes, for instance the reduced risks of contamination and/or the reduced costs of operating large fermentations by inserting cyanobacterial thermostable gene into bacteria and develop to industrial operation.

2.1.1 Diversity of thermophilic cyanobacteria

The existence of the thermophilic cyanobacteria has been extensively documented during the course of the microbiological characterization of the hot spring, first in Yellowstone National Park, Wyoming, USA (Brock, 1978) and later in other geothermal areas all over the world. Cyanobacteria occur in almost every hot spring below 73-74°C. Different types of unicellular cyanobacteria, classified in the genus *Synecococcus* are the most thermophilic (Oren and Seckbach, 2001). Water temperature was the predominat determinant of community structure in the hot springs. The diversity of cyanobacterial morphotypes fell as temperature increased (Sompong *et al*., 2005).

There are differences in the communities found in hot springs in different parts of the world, for instance, thermophilic *Synechoccocus* species, abundant in Yellowstone, are absent from similar hot springs in Iceland and Alaska (Castenholz, 1996; Ward and Castenholz, 2000). However, it could be found in other hot springs around the world; New Zealand, Northern Italy, Greece, Guatemala, South America, Kenya, Zaire, China (Castenholz, 1996) and also in Thailand's hot springs (Sompong, *et al*., 2005). Other thermophilic cyanobacteria, such as, *Oscillatoria* cf. *amphigranulata* has been found in New Zealand hot springs, *Oscillatoria* cf. *terebriformis* was found in many hot springs in North America (excluding Alaska and western Canada), Japan, Chile, Southwestern Saudi Arabia and Guatemala (Castenholz, 1978). *Mastigocladus* cf. *laminosus* was also found in many states in America, parts of Iceland, Europe, Asia, Africa and New Zealand. *Chlorogloeopsis* sp. has been found in hot springs in New Zealand, Iceland, the Azores and Europe.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Although a few locations worldwide have been studied, distinct phylogeographic groups have only been extensively studied in the continental USA, Iceland, New Zealand and some countries in Asia (Papke *et al*., 2003 and Hongmei *et al*., 2005). While to date studies of the cyanobacterial mats in hot springs in Thailand have been minimally investigated. In 1992, Chansaghavate and Niyomrit reported finding thermophilic cyanobacterial species in Hindad Hot Spring, Thong Pha Poom, Kanchanaburi Province in western Thailand. Cyanobacterial species, *Oscillatoria* spp., *Mastigocladus* sp., *Synechococcus* sp. and *Calothrix* sp. were found at the temperature range of 40-49ºC and the pH range of 6-8. In 1990, Pitugvapi reported that the dominant cyanobacterial species in Pong Hom Hot Spring, Chiang Mai Province in northern Thailand were *Oscillatoria* spp. and *Synechococcus aeruginosus* at a temperature range of 31-77ºC. In 2005, Sompong and colleagues studied the diversity of cyanobacteria along the thermal gradient from 30-80ºC in nine hot springs of northern Thailand. Thirty-six species were identified by morphometric analysis. Dominant species are *Synechococcus* spp., *Cyanothece* sp. and *Phormidium* cf. *boryanum*.

However, a lack of knowledge about the genetic diversity of cyanobacterial populations in Thailand is required a more discriminating analyses using molecular tools to determine some morphotypes.

2.2 Identification of cyanobacteria

Cyanobacteria occur in a wide variety of morphologies and ecological forms. Hence, historically their classification has been based on simple morphological characteristics without the use of cultures (Geitler, 1932; Desikachary, 1959), whereas many bacteriologists have applied physiological and biochemical properties for those species that exist in culture (Castenholz, 2001). During the last four decades, ecological characteristics, ultrastructural features, and molecular evidence have substantially influenced our knowledge of this group. As a result, species concepts and classification are undergoing radical changes (Anagnostidis and Komarek, 1985; Castenholz, 1992; Castenholz, 2001). Many well-known taxa e.g. *Gloeocapsa*, *Chroococcus*, and many others have not yet been studied in culture, which is a first step to experimental revisions of the taxonomic status of species (Rippka *et al*., 1979; Komarek, 1994;

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Castenholz, 2001). However, the traditional approach is still necessary, especially for field studies.

Cell morphology include colony characteristics and trichome morphology, cell ultrastructure, genetic characters, cell physiology and biochemistry, culture conditions and ecology of habitats which are all used to characterize cyanobacterial species (Castenholz and Waterbury, 1989; Castenholz, 2001).

Besides sharing the basic cellular structure of other bacteria, the cyanobacteria posses several unique and diagnostic characteristics [Anagnostidis and Komárek (1985; 1988; 1990); Komárek and Anagnostidis (1989; 1999); Castenholz and Waterbury, 1989; Castenholz, 2001]. Morphological characteristics of cyanobacteria are used to determine species or genus of cyanobacteria such as; cell envelope, cell division, cell exterior and motility, cell interior, heterocysts and akinetes. The classification is based on phenotypic characteristics, the results of pure cultures and phylogenetic interpretations (not only sequences of 16S rDNA or rRNA, but also others genes). Cyanobacteria are subdivided into five major subsections (“Orders”). These include Orders Chrooccales and Pleurocapsales, coccoid forms (Figure 1); Oscillatoriales, Nostocales and Stigonematales; filamentous forms (Figure 2) (Castenholz, 2001).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

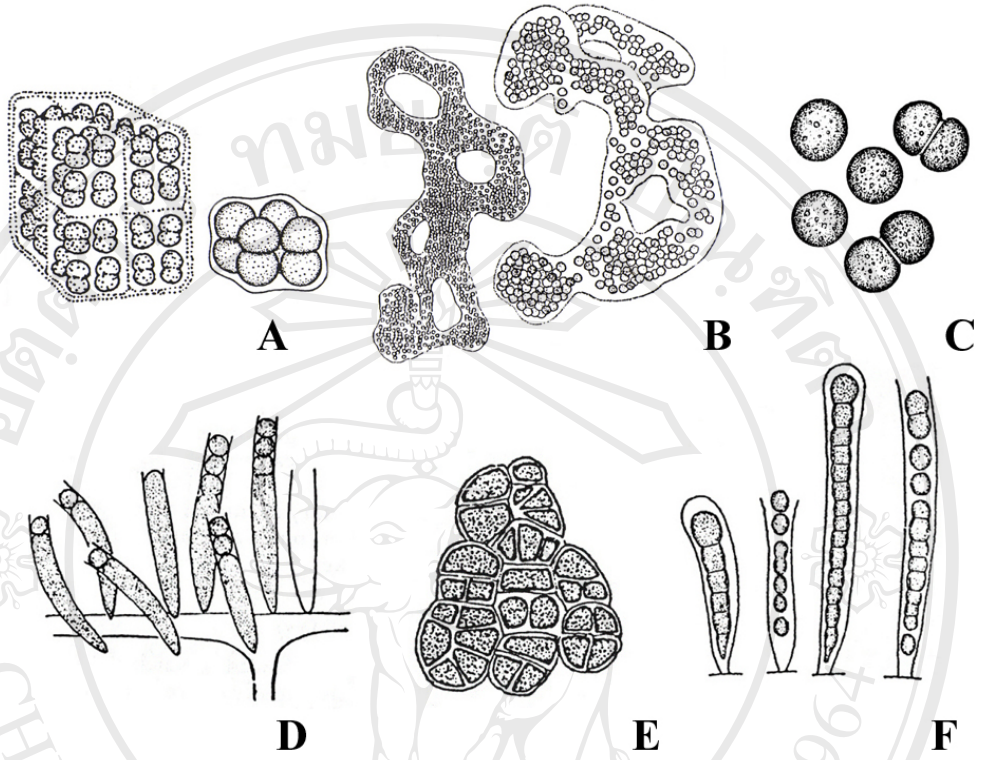

Figure 1. Some coccoid cyanobacteria in Orders Chroococcales and Pleurocapsales.

(A-C) Chroococcales; A) *Eucapsis alpina* Clements and Shantz, B) *Microcystis aeruginosa* Kuzing, C) *Synechocystis septentrionalis* Skuja, (D-F) Pleurocapsales; D) *Chamaesiphon incrustans* (Grunow) Rabenhorst, E) *Pleurocapsa minor* Hansgirg, F) *Stichosiphon sansibaricus* Bourrelly

Source: Komarek and Anagnostidis (1999); Komarek (2003).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

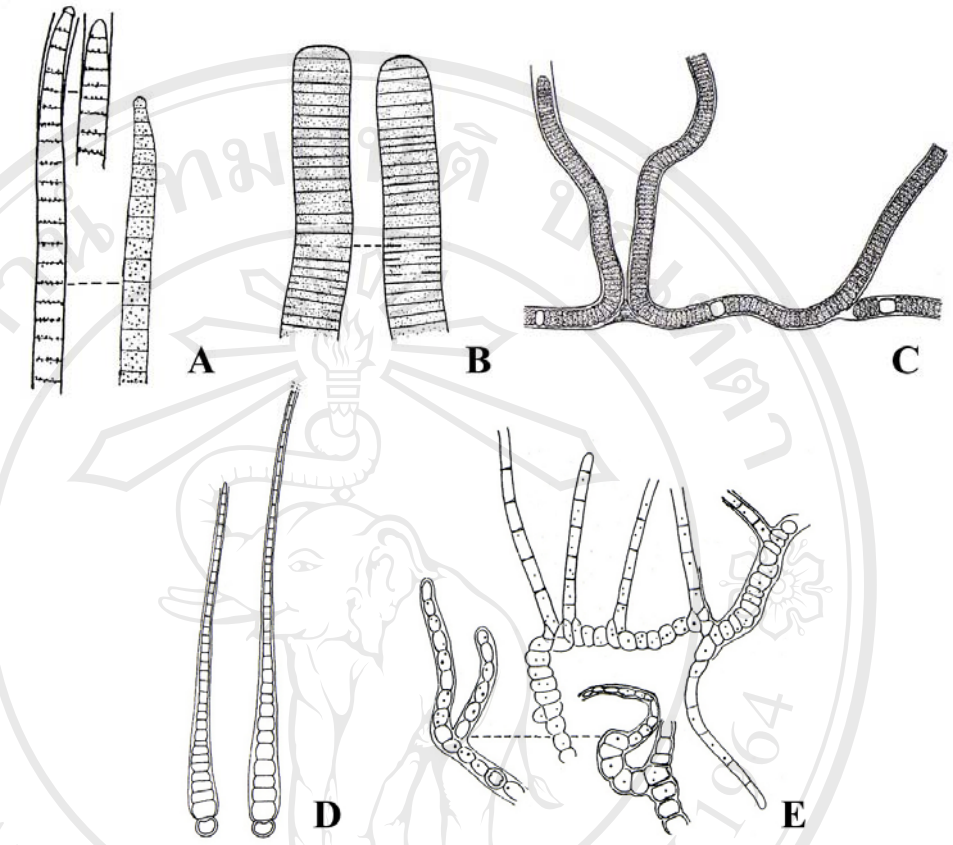

Figure 2. Some filamentous cyanobacteria in Orders Oscillatoriales, Nostocales and Stigonematales.

(A-B) Oscillatoriales; A) *Phormidium autumnale* (Ag.) Gomont, B) *Oscillatoria obtusa* Gardner, (C-D) Nostocales; (C) *Scytonema crispum* (Agardh) Bornet, D) *Calothrix parietina* (Thuret) Bornet et Flahault, (E) Stigonematales; *Mastigocladus laminosus* Cohn

Source: Komarek *et al*. (2003); Whitton (2002)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

The main problems in applying morphological criteria in cyanobacterial classification arise from morphological features that vary with environmental conditions (Pearson and Kingsbury, 1966; Willmotte and Golubic, 1991). Sometimes microscopy and enrichment cultures have limited usefulness since distinct species of cyanobacteria can share similar, simple morphological and cultivation limitations (Ferris *et al*.,

1996b). This is difficult when the phenotypes are morphologically similar or essentially identical, or when the prospective organism is difficult or presently impossible to culture. However, in many groups of cyanobacteria, a large diversity of morphological complexity occurs, which allows some freedom in terms of visual identification. For example, in the case of the largely morphologically-defined genus *Anabaena, Phormidium, Oscillatoria* and *Calothrix*. They are often difficult for non-experts to be confident of their diagnosis. However, it is unsound to assume that morphologically similar (e.g. unicellular) cyanobacteria are the same genotype at the species, or even genus, level (Castenholz, 1996).

One way to better characterize the phylogenetic relationship between these morphologically similar species is to use molecular diversity information, and the development of techniques for the analysis of 16S rRNA sequences in natural samples has already greatly enhanced detection and identification of cyanobacteria in nature (Willmotte and Golubic, 1991; Ferris *et al*., 1997; Nübel *et al*., 1997; Norris *et al*., 2002).

2.3 Molecular diversity

The most recent exciting developments in biological evolution have come about through advances in molecular phylogeny and DNA sequencing. Until recently, it was not possible to scientifically evaluate the evolution of bacteria, cyanobacteria and other microorganisms. Unlike plants and animals, these organisms are structurally too simple to enable their scant fossil record to be useful for identifying species and activities. However, it became apparent that the sequences of the subunits of proteins and nucleic acids retained the evolutionary history of an organism.

During the mid 1980's, studies of microorganisms necessitated their isolation in the laboratory, which is now known to capture less than 1% of naturally occurring species (Embry and Stackebrandt, 1997; Ward *et al*., 1990). The culture-independent method commonly used to detect microbial community changes in the environment is the analysis of 16S ribosomal gene sequences (Amman *et al*., 1995). The use of 16S rRNA or of the gene coding 16S rRNA (16S rDNA) is a powerful approach to microbial taxonomy and microbial community analysis for many reasons (Head *et al*., 1998).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Firstly, ribosomal RNAs are used essentially for protein synthesis, making them ubiquitous in all organisms as well as being structurally and functionally conserved. It is ribosomal RNA (rRNA), which are found in all organisms on Earth. Second, 16S rRNA contains both highly conserved and highly variable regions; a gene that is too variable would not allow its detection across a broad range of organisms, while one entirely conserved would show no differences among taxa. Third, relationships among 16S rRNA genes are indicative of evolutionary relationships, because it is sufficiently long (1,500 or more nucleotides), it contains a large amount of evolutionary studies (Woese, 1987; Willmotte, 1994). Two additional assumptions are basic for the validity of this approach, namely that lateral gene transfer has not occurred between rRNA genes, and that the amount of evolution or dissimilarity between rRNA sequences of a given pair of organisms is representative of the variation shown by the corresponding genomes (Goodfellow *et al*., 1997; Rossello-Mora and Amann, 2001). This allows one to target specific taxonomic groups that are known to be responsible for a particular ecosystem function, such as cyanobacteria.

2.3.1 Revealing microbial diversity

The advances in molecular biology, principally in the development of the polymerase chain reaction (PCR) for amplifying DNA, DNA sequencing and data analysis, have resulted in powerful techniques which can be used for characterization, screening and evaluation of genetic diversity (Vandamme *et al*., 1996).

The polymerase chain reaction (PCR) is an *in vitro* technique which allows the amplification of a specific DNA (Newton and Graham, 1997). This versatile technique was invented in 1985 (Saiki *et al*., 1985). Since the first thermostable DNA polymerase to be introduced into PCR *Taq* DNA polymerase and the development of a variety of thermal cyclers or PCR machines which led to the automation of the PCR, the use of PCR is a key procedure in research and clinical laboratories (Saiki *et al*., 1988). The target DNA sequence is amplified exponentially. Thus, a lot of copies of one or more target DNA sequence are amplified.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

The basic components of a typical PCR, template DNA, buffer, dNTPs and primers are mixed and heated to denature the DNA thermally, DNA polymerase is added and mixed. The reaction components are overlaid with or without mineral oil and the tube is placed in a thermal cycler. At the end of the amplification, the products are analyzed to confirm that the PCR has been successfully amplified.

Figure 3 shows the first few cycles of PCR. The first extension products result from DNA synthesis on the original template and these do not have a distinct length as the DNA polymerase will continue to synthesize new DNA until it stops. The second cycle extension products are also of indeterminate length; however, at the third cycle, fragments of target sequence are synthesized which are of defined length corresponding to the positions of the primers on the original template. From the fourth cycle onwards the target sequence is amplified exponentially.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

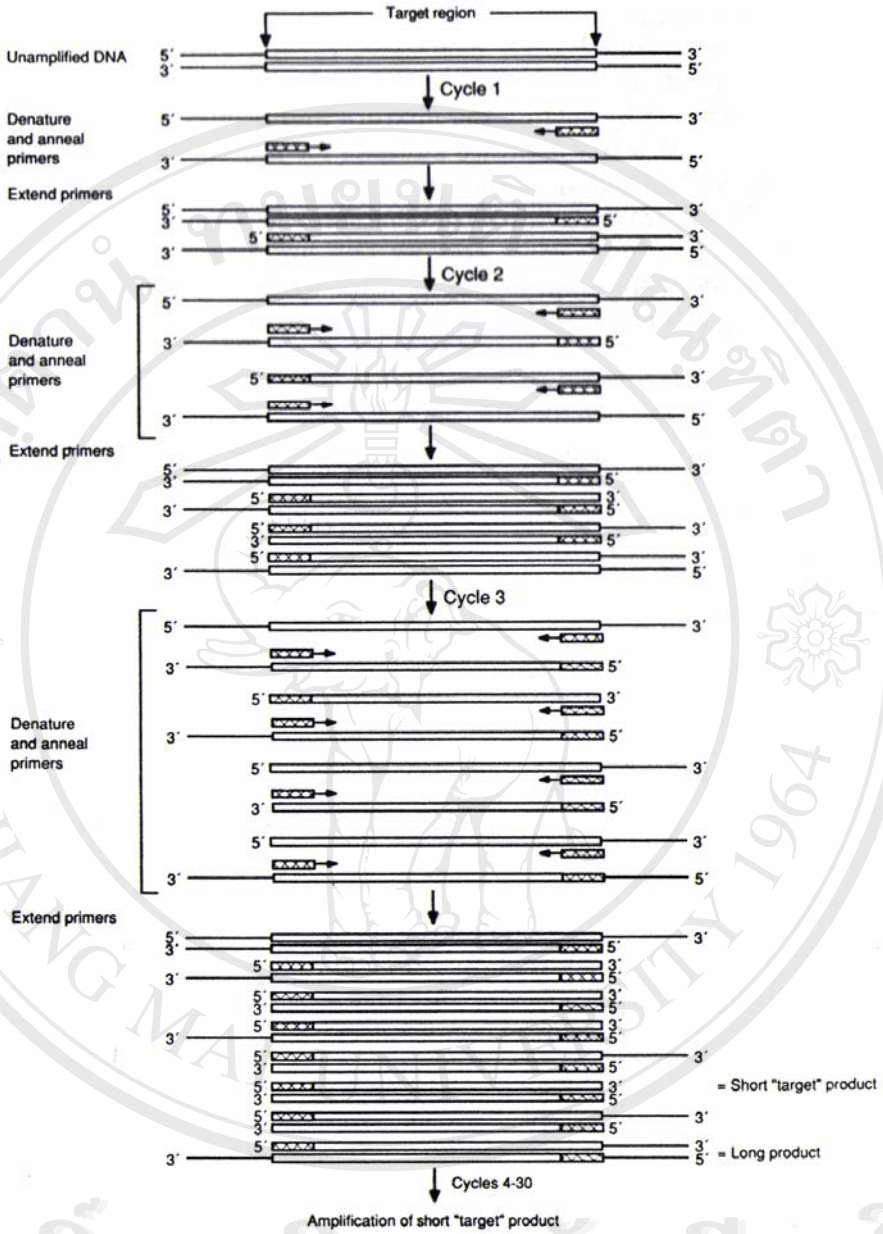

Figure 3 The polymerase chain reaction. PCR is a cycling process and with each cycle the number of DNA targets doubles. This cycle is normally repeated for 20-40 cycles (Newton and Graham, 1997).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Initial denaturation of template DNA at 95-100ºC is sufficient to completely denature complex genomic DNA so that the primers can anneal after cooling. There are three steps of PCR amplification. In the first step of PCR cycle, denaturating step, the DNA template is made of a single strand by using temperatures of

92-95 ºC. The second step, annealing step, oligonucleotide primers are annealed to the target sequence on the template DNA by lowering temperature to 37-65ºC (depending on the T_m of primer sequence). For the third step, extension step, the primers are extended by using the DNA polymerase, usually performed at 72ºC. Repeating these three cycle steps 25-35 times, the target sequence is amplified exponentially.

The PCR-target approaches could be applied to study microbial communities. Now, the development of techniques to extract genomic DNA and ribosomal RNA, amplified DNA or RNA sequence from environmental samples has allowed new ways to study microbial communities independently from cultivation. Most of the results in these studies have been obtained by cloning and sequencing of 16S rDNA fragments obtained after PCR amplification of cyanobacterial genomic DNA extracted from environmental samples.

For this reason the cloning approach is not a main approach to study the mixed microbial communities, because it is time consuming, cumbersome and impractical for multiple sample analysis. For this purpose a quick and easy approach, which gives direct visualisation of predominant constituents of mixed microbial communities, and which allows multiple sample analysis is needed. DNA fingerprinting techniques, such as denaturing gradient gel electrophoresis (DGGE), developed to detect sequence variation in DNA fragments, have been introduced into microbial ecology to determine the genetic diversity of complex microbial communities and to monitor population shifts after environmental perturbations.

2.3.2 The denaturing gradient gel electrophoresis (DGGE)

The DGGE method is based upon the analytical separation of DNA fragments of identical or near-identical lengths but with varying sequence compositions. This method was first developed to detect single based changes in genes for the diagnosis of human genetic diseases and in genetic linkage studies (Liu and Stahl, 2002). DGGE has been extended to resolve environmental populations of microorganisms by separating PCR amplification products generated using primers targeting conserved genes. PCR primers designed for the amplification and cloning of SSU rRNA genes were first used to demonstrate the technique as applied to

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

environmental microbiology. Another study used the same approach to resolve natural populations of sulfate-reducing bacteria that share a highly conserved hydrogenase. Others have used the methods to examine the diversity of genes encoding phenol hydroxylase and nitrogenase.

This method constitutes direct extraction of the community DNA and amplification of typically 200-600 bp long 16S rDNA. Separation is based on the changing electrophoretic mobilities of DNA fragments migrating in a gel containing a linearly increasing gradient of DNA denaturants (urea and formamide). Changes in fragment mobility are associated with partial melting of dsDNA in discrete regions, the so-called melting domains. For example, in the DGGE analysis under a constant temperature, the Tm for each domain varies according to the concentration of denaturant and therefore by position in the gel. When the DNA enters a region of the gel containing sufficient denaturant, a transition of helical to partially melted molecule occurs, and migration is severely retarded. Sequence variation within such domains alters the melting behavior, and sequence variants of the different amplification products stop migrating at different positions in the denaturing gradient (Figure 4).

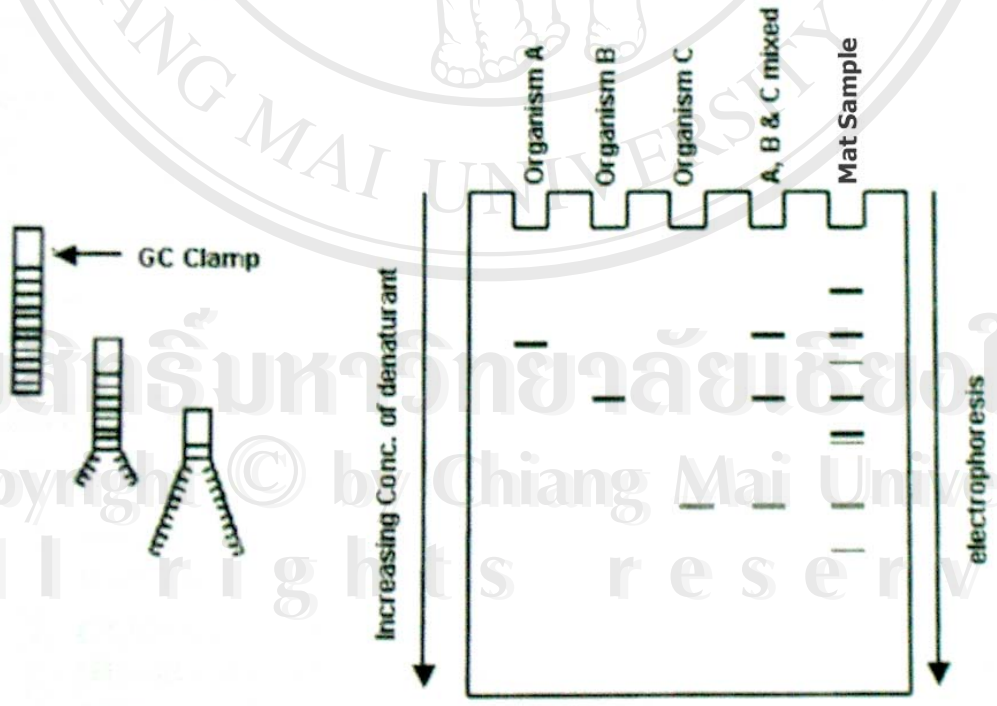

Figure 4. Separation bands in DGGE

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Using the method as first developed, approximately 50% of the sequence variants studied could be detected in DNA fragments up to 1,000 bp in length. However, virtually all variants can be resolved by the attachment of a GC-rich sequence to the DNA fragment. This terminal appendage, the GC clamp, acts as a high-temperature melting domain. The 40-nucleotide GC-rich sequence (GC clamp) is added to the 5′end of the one of PCR primers (Muyzer *et al.*, 1993). Now, it can detect over 99% of single-nucleotide sequence variation (Sheffield *et al*., 1989).

The use of DGGE to characterize a population of DNA amplified from a particular environment generally requires a preliminary study to establish appropriate running conditions. Amplification products are first characterized by perpendicular gels with an increasing gradient of denaturants from left to right (perpendicular to the direction of electrophoresis) (Muyzer *et al*, 1993). The sample is applied across the entire width of the gel, electrophoresed, and visualized by ethidium bromide staining, silver staining, or SYBR green (Norris *et al*., 2002). DNA molecules at the side of the gel containing a low range of denaturants or temperatures migrate as dsDNA, whereas those on the other side of the gel containing a low range of denaturants or temperatures melt upon entry and stop. At intermediate ranges of denaturants or temperatures, the molecules have different degrees of melting, as reflected in different mobilities. This preliminary analysis provides some information about population complexity and defines a narrower range of denaturants or temperatures for use in subsequent higher-resolution studies. The later studies are generally done using a parallel gradient gel with an increasing gradient of denaturants or temperatures from top to bottom. Analyses of multiple samples are generally preceded by a time travel experiment to establish an appropriate electrophoresis run time(s). This simply involves loading the same sample repeatedly in different lanes of a parallel gradient gel with a defined time interval between sample loadings.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

DGGE analysis of PCR-amplified SSU rRNA fragments provides a rapid method for the characterization and monitoring of community population structure and dynamics (Muyzer and Smalla, 1998; Ferris *et al*., 1996b). The initial study by Muyzer *et al.* (1993) demonstrated the presence of several distinguishable bands (between 5 and 10) in the gel separation pattern which were most likely derived from the predominant

species within the communities characterized. Sequences of individual bands (fragments) are determined following their extraction from the gel, a second round of PCR amplification, and sequencing (direct or after cloning). Group-specific PCR primers have been used to restrict population analysis to specific microbial groups. Also, as discussed in relationship to T-RFLP analysis, DGGE can be used to evaluate the relative activities of individual populations based on changing rRNA/rDNA ratios. The method has been used to evaluate this ratio among different natural populations by comparing the patterns and intensities of bands derived using either rDNA or rRNA (using reverse transcriptase to generate cDNA) as templates (Norris *et al*., 2002).

2.3.3 Advantages of DGGE

DGGE is fast, reproducible, reliable and advantageous when compared to the RFLP technique which cannot be used to detect single base changes as these do not alter any particular cleavage site, so many changes may go undetected (Muyzer, 1999). When compared to hybridisation techniques with specific probes that can be used to detect changes at defined sites but are not useful for screening randomly distributed polymorphisms (Muyzer, 1999) and compared with Dot Blot Hybridisation, these studies only focus on particular groups of organisms for which probes are developed and this technique is not useful for studying population shifts.

2.3.4 Disadvantages of DGGE

The inaccuracy in DGGE analysis is due to artificial or natural microheterogeneity in the DNA sequence (Speksnijder *et al*., 2001) resulting in that a single band may be composed of several species (Van Hannen *et al*, 1998; Sekiguchi *et al*., 2001) or that several bands are degenerated from a single species. They may display only the dominant populations excluding populations present of 1% or less.

2.3.5 Biodiversity applications

Since the first publication by Muyzer *et al*. (1993) an increasing number of studies in microbial ecology have used DGGE. There are many investigations into microbial biodiversity. They can describe aspects of composition, structure and function

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

of the ecosystem, determine the effect of a perturbation on community composition structure or function (Pennanen *et al.*, 2004; Cheung and Kinkle, 2001), determine the spatial or temporal dynamics (Vanhoutte *et al*., 2004), determine taxonomic or phylogenetic relatedness, provide an analysis of composition in order to determine the source of a specific function (Ji *et al.*, 2004), search for markers for identification of the presence of specific microorganism, search for new species/genera and they can be used for evaluating the utility of methods (Boon *et al.*, 2002; Dunbar *et al*., 2002). Muyzer *et al*. (1995); Moyer *et al*. (1995) used DGGE analysis of PCR-amplified rDNA fragments to provide information on the genetic diversity of microbial communities found around hydrothermal vents.

DGGE of PCR-amplified 16S rRNA gene fragments has been applied to profile the distribution of microbial populations inhabiting regions with different temperatures in a hot spring cyanobacterial community (Ferris *et al*., 1996b). Different profiles were found for samples from sites with different temperatures indicating different populations. Heuer *et al*. (1997) used a group-specific PCR and DGGE to analyze actinomycete communities in different soils, and to monitor shifts in their abundances in the potato rhizosphere. Nübel *et al*. (1997) designed a primer pair for the specific amplification of 16S rDNA fragments from cyanobacteria. Combined with DGGE analysis of these fragments it was possible to determine the cyanobacterial diversity in nonaxenic cultures, lichens and complex microbial assemblages, such as those present in microbial mats samples. Curtis and Craine (1998) used DGGE to compare the diversity of microbial communities present in different activated sludge plants.

By using DGGE, many samples taken at different time intervals during a study can be simultaneously analyzed for monitoring community behaviour after environmental changes. Ferris *et al.* (1997) used PCR-DGGE to study the re-establishment of a microbial mat after removal of the entire cyanobacterial layer. The results showed that previously undetected cyanobacterial colonized the remaining part of the mat, and that other cyanobacteria which were present before the disturbance remained undetected for up to 40 days.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

In a subsequent study, DGGE was used to evaluate seasonal distributions of bacteria populations along thermal gradients in a hot spring microbial mat (Ferris and Ward, 1997). Similar DGGE patterns were found for samples collected at the same site and for sites with the same temperature, regardless of the season. However, different profiles were seen for samples from sites with different temperatures.

A polyphasic approach, including phenotypical and phylogenetical analyses, was used to investigate the diversity of geographically distant field populations and cultured strains of the cyanobacterium *Microcolues chthonoplastes* (Garcia-Pichel *et al*., 1996). Identical DGGE patterns of 16S rDNA fragments from the field populations and cultures as well as similar morphology, and the presence or absence of certain biochemical markers, demonstrated that *M. chthonoplastes* represent a single, well-defined taxon with a ubiquitous distribution.

2.4. Phylogenetic analysis

Phylogenetic tree construction has become increasingly popular in many fields of biology. Their inclusion reflects the growing recognition of trees as a tool for understanding biological processes. Phylogenetic trees allow you to organize your thinking about a protein of interest in terms of its relationship to other DNA or proteins, and may allow you to draw conclusions about its biological functions that would not otherwise be apparent (Hall, 2001). Phylogenetic trees are used for purposes such as knowing the evolution of genes and proteins, knowing the phylogenetic relationship of organisms, clarifing the taxonomic position of an organism, determining the origin of hybrid species, estimating the functions of molecules, finding out the origin of organelles (chloroplasts and mitochondria), detecting a lateral gene transfer and identifying organisms in a particular environment. The phylogenetic reconstruction is also widely accepted to apply the rRNA technology as an integrated part of a polyphasic approach for new descriptions of cyanobacterial species or higher taxa (Ludwig *et al*., 1998).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

An important feature of the 16S rRNA molecule in its use as a universal standard parameter for phylogenetic inferences is the relative ease of sequence alignment (Embry and Stackebrandt, 1997). Alignment is the first critical step to sequence-based phylogenetic analyses. Given that positions with a common ancestry

have to be compared for reliable phylogenetic conclusions, homologous positions have to be arranged in common columns in correct alignment (Ludwig *et al*., 1998).

2.4.1 Phylogenetic trees

A tree is composed of lines called branches that intersect and terminate at nodes. The nodes at the tips of the branches represent the taxa or the sequences that exist today. The internal nodes represent ancestral taxa, whose properties can only be inferred from the existing taxa.

2.4.2 Methods for constructing phylogenies

There are two primary approaches to tree construction: algorithmic and tree searching. The algorithmic approach uses an algorithm to construct a tree from data. The tree-searching method constructs many trees, then uses some criterion to decide which is the best of the trees.

The algorithmic approach has two advantages: It is fast, and it yields only a single tree from any given data set. The two algorithmic methods in current use are Neighbor Joining, and UPGMA. All the other methods are tree-searching methods. These methods such as Parsimony, Maximum likelihood and Bayseian analysis search for the tree that best meets some optimality criterion by evaluating individual trees. The best we can hope for is a tree that well reflects what happened in the past, so we can never be entirely sure how accurate the inspection of the tree is. Tree-searching methods may yield one tree or several, but all methods implicitly acknowledge that the trees produced are only a subset of the possible trees that are consistent with the data.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

2.4.3 Distance methods

Distance methods convert the aligned sequence into a distance matrix of pairwise differences (distances) between the sequences. Distance methods use that matrix as the data from which branching order and branch lengths are computed. While there are statistical corrections used to estimate corrected distances from the number of observed differences, distances always underestimate the actual amount of change in the lineages.

1) UPGMA

UPGMA (Unweighted Pair-Group Method with Arithmetic Mean) is an example of a clustering method. The program first finds the pair of taxa with the smallest distance between them and defines the branching between them as half of that distance. It combines the two taxa into a cluster and rewrites the metrix with the distance from the cluster to each of the remaining taxa. The number of entries in the metrix is reduced until the metrix consists of a single entry. Then set of metrics is used to reconstruct the tree. The distance between any two nodes equals the sum of the lengths of all branches between them and all taxa are equally distant from the root.

2) Neighbor Joining

NJ is similar to UPGMA in that it manipulates a distance metrix, reducing it in size at each step, then the tree is reconstructed from that series of metrics. It differs from UPGMA in that it does not construct clusters but directly calculates distances to internal nodes. From the original matrix, NJ first calculates for each taxon its net divergence from all other taxa as the sum of the individual distances from the taxon. It then used that net divergence to calculate a corrected distance metrix. Then, NJ finds the pair of taxa with the lowest corrected distance and calculates the distance from each of those taxa to the node that joins them. NJ does not assume that all taxa are equidistant from the root (Hillis *et al*, 1996; Swofford *et al*, 1996).

2.4.4 Character-based methods

Character-based methods, Maximum Parsimony, Maximum likelihood and Bayesian methods, all use the multiple alignment directly by comparing characters within each column (each site) in the alignment.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

1) Maximum Parsimony

Maximum Parsimony (MP) is based on the assumption that the most likely tree is the one that requires the fewest number of changes (evolutionary steps, or the number of changes in this case being mutations) to explain the data in the alignment. Parsimony, or minimum change, is the criterion for choosing the best tree. That number

is the score for the tree and the tree or trees with the lowest scores are the most parsimonious trees (Swofford, 1996; Hall, 2001). For nucleotide or protein sequences, the data are the aligned sequences. Each site in the alignment is a character, and each character can have different states in different taxa. Not all characters are useful in constructing a parsimony tree (Li, 1997).

2) Maximum Likelihood

Maximum likelihood (ML) tries to infer an evolutionary tree by finding that tree that maximizes the probability or likelihood of observing the data. For sequences, the data is the alignment of nucleotides or amino acids. ML almost always produces a single tree (Swofford *et al.*, 1996). The advantages of the ML method are that it allows users to specify the evolutionary model they want to use, and that the likelihood of the resulting tree is known. A disadvantage is that ML is considerably slower than either MP or NJ, and it is not difficult to exceed the capacity of even the most up-to-date desktop computer (Hall, 2001).

3) Bayesian Analysis

Bayesian inference is based on the notion of posterior probabilities: probabilities that are estimated based on a model, after learning something about the data. Bayesian analysis is a variant of Maximum Likelihood. Instead of seeking the tree that maximizes the likelihood of observing the data, it seeks the tree that maximizes the probability of the tree given the data and the model for evolution and searches for the best set of trees. This method is similar to ML in that the user postulates a model of evolution and the program searches for the best trees that are consistent with the alignment (Hall, 2001; Hongmei *et al.*, 2005).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

A perfect match of the tree topologies cannot be expected (Rossello-Mora and Amann, 2001). No method is ideal for all performance criteria. Some of the criteria that have been considered are efficiency, robustness, computational speed, and discriminating ability. Efficiency is a measure of how quickly the method converges on the correct tree as the amount of data (length of the sequences) increases; robustness is a measure of how well the method can tolerate deviations from its assumptions and still

recover the correct tree; computational speed is obvious; and discriminating ability is how well the method guarantees recovering the correct tree (Hillis *et al.*, 1996).

From The Bergey's Manual of Systematic Becteriology 2001 edition, the oxygenic photosynthetic procaryotes (cyanobacteria) comprise a single phylogenetic branch within the domain Bacteria (Eubacteria). On the basis of 16S rDNA sequence analyses, this branch or cluster is most closely related to the Gram-positive bacteria.

At the time of the 1989 edition (Castenholz and Waterbury, 1989), it was thought that the order Prochlorales included in the Phylum "Prochlorophyta" (the oxygenic photosynthetic bacteria that lack phycobilisomes and possess chlorophyll b in addition to chlorophyll a) constituted an evolutionary branch distinct from the cyanobacteria. Now, this hypothesis has been shown to be incorrect (Palenik and Swift, 1996; Castenholz, 2001). The *Prochloron*, *Prochlorothrix* and *Prochlorococcus* genera fall within three distinct polyphyletic branches of the cyanobacterial tree according to 16S rDNA sequence analyses (Litvaitis, 2002). He used Parsimony and Neighbor-joining to analyze 16S rDNA nucleotide sequences of cyanobacteria and prochlorophytes supporting a Nostocales, a Stigonematales and three independent lineages of prochlorophytes within cyanobacteria. It is more parsimonious to assume a common cyanobacteria/prochlorophyte ancestor and it is recommended that Prochlorales be reclassified as cyanobacteria. Therefore, they are more closely related to a number of typical cyanobacteria than to each other (Willmotte, 1994; Turner, 1997).

Miller and Castenholz (2000) constructed the phylogenetic trees from 20 clones of the genus *Synechococcus* isolates from collections made along an Oregon hot spring thermal gradient. They combined a comparative physiological approach with phylogenetic analyses to study the evolution of thermotolerance *Synechococcus*. This study illustrates the utility of using phylogenetic comparative methods to investigate how evolutionary processes have shaped historical patterns of ecological diversification in microorganisms.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved