CHAPTER 6
LITERATURE REVIEW - MOLECULAR PHYLOGENETIC ANALYSIS

Cladistic

Cladistic is a method of classification that groups taxa hierarchically into discrete set and subset. It can be used to organize any comparative data. Cladistic methods were made explicitly by Willi Hennig, the German Entomologist in 1950 in order to explain the biological relationship and how that relationship can be discovered. Cladistic is the relationship based on shared derived characters. The taxa sharing many derived characters are grouped more closely together than those that do not. The principle of cladistic analysis is parsimony – any hypothesis that requires fewer assumptions is a more defensible (Kitching *et al*., 1998).

There are technical terms used in cladistic analysis. Pleisiomorphy is a character shared with common ancestor. Apomorphy is a derived character. Synapomorphy is shared derived character found only in descendence not in ancestor. Autoapomorphy is a derived character found only in one taxon at terminal branch. There are many methods to determined characters as pleisiomorphy or apomorphy. The simple one is to compare with outgroup, a closely related of taxa of interest but not included. If one character occurs in both ingroup and outgroup, then that character is pleisiomorphic. Otherwise, if that character occurs only in ingoup, not in outgroup, it is apomorphic. Cladistic itself is a relative comparison. Apomorphic characters in one taxonomic level can be pleisiomorphic characters in the lower ones. Group of organism based on derived characters. Monophyletic is a group of all descendence of the same common ancestor. Polyphyletic is a group of taxa from two or more common ancestors. Paraphyletic is a group of descendence of the same common ancestor but exclude some taxa (Stussy, 1990).

The distinction between cladogram and evolutionary trees is important. The evolutionary trees are very precise statements of singular history but their precision is gained from criteria other than character distribution. But cladograms are often been explained as evolutionary trees. To do this, some assumptions have to be accepted. For examples, evolution is parsimonious and evolutionary process occurs in

branching. These assumptions are criticized as they are unrealistic for evolution. The Cladogram itself is the starting point for further analysis (Kitching *et al*., 1998).

Today, cladistic becomes the major analysis in systematics. It has been used in many groups in various taxonomic ranks from division to species and below (Soltis *et al.* 2002).

Support of cladogram

Cladistic is subjected to revision using new data and reinterpret hypotheses of homology of particular group. However, while the most parsimonious cladogram represents the best summary of the available data and is the preferred hypothesis of relationship, it should represent the true relationship among the studied taxa. It is needed to determine which changes should be included in the new and improved classification and which are not. Many methods have been proposed to use to assign degree of support or confidence to the result of cladistic analysis. There are three basic types of approaches that use randomization procedure to assess support for individual clades: Monte Caslo (including Bootstrapping), jackknifing and permutation procedures.

Descriptive statistics of cladogram

Length or number of step is the total number of character changes necessary to support the relationship of the taxa in a tree. The shortest tree is the best tree, according to parsimony principle. It requires fewer hypothesis of taxa relationship in a tree.

Consistency index (CI) is the relative amount of homoplasy within a tree. It can be calculated as the number of steps expected given the number of character states in the data, divided by the actual number of steps multiplied by 100. The formula for the CI is

$$CI = \frac{\text{total number of state changes expected given the data set}}{\text{Actual number of step on the tree}}$$

Retention index (RI) is the amount of synapomorphy expected from a data set that is retained as synapomorphy on a cladogram. The formula for the RI is

$$RI = \frac{\text{maximum number of steps on tree} - \text{number of state changes on the tree}}{\text{maximum number of steps on tree} - \text{number of state changes in the data}}$$

Maximum number of steps on tree is the total number of taxa with state 1 or 0 (whichever is smaller), summed over all the characters.

Rescale consistency index (RC), as its name, is the rescaling CI by using RI. Thus, characters with no similarity interpreted as synapomorphy (RI=0) will be disregarded, irrespective of their level of homoplasy. The formula for RC is

$$RC = CI \times RI$$

Internal Transcribed Spacer and 5.3S of Nuclear Ribosomal DNA

The chloroplast genome is more popular in the past 10 years. Sequences from many taxa are accumulated rapidly (Soltis *et al*., 1992). Now, sequences from many areas in nuclear DNA were used in phylogenetic analysis. This genome is widely used in molecular phylogenetic analysis. It is a major focus of comparative sequencing of the taxonomic rank from Genus to Species and below (Soltis *et al*., 1998).

5.8S rDNA is rarely used alone to construct phylogeny because it is highly conserved and too small (120-160 bps) to be informative at lower taxonomic level. The proportion of informative positions is 0.14, similar to 0.18 of 18S rDNA.

The Internal Transcribed Spacers (ITS-1 and ITS-2) are part of nuclear ribosomal DNA but not incorporated into ribosome. They appear to play a role in the maturation of nuclear rRNAs, bringing the large and small subunits into close proximity within a processing domain. ITS sequences are used for reconstructing phylogeny of angiosperms, fern and algae. Part of ITS is G+C rich and quite conserved among angiosperms. For example, 40% of ITS-2 nucleotides are conserved across all angiosperms sequences. It provides high information for lower taxonomic level. Sequencing of ITS can be difficult because the region is G+C rich and easy to

form secondary structure. The length of ITS plus 5.8s is fairly short (600-700 bps) and relatively uniform in angiosperms. It is easy to sequence these genome but the amplification primers are located in the conserved coding regions that can amplify the contaminant DNA. It is necessary to compare the acquired sequence with those available in Genbank.

Molecular Phylogeny of Zingiberaceae

Searle and Hedderson (2000) studied molecular phylogeny of tribe *Hedychieae* (*Zingiberaceae*) using Internal Transcribed Spacer (ITS) of 18S-26S nuclear ribosomal DNA. This is the first molecular-based phylogenetic analysis of the family Zingiberaceae. They found that *Zingiber* is the sister group of *Cornukaempferia* and nested within tribe *Hedychieae*. *Siphonochilus* is excluded from *Hedychieae* but appeared as basal group of *Alpinieae*. *Curcuma* comes together with *Stahlianthus*. *Camptandra* is sister group of these two genera with high support.

Rangsiruji *et al*. (2000) used ITS to clarify the relationship within the genus *Alpinia*. The Cladogram showed that *Alpinia* is not monophyletic, but genus *Renealmia* is included as sister group of section *Fax*. The infra-generic classification of *Alpinia* is needed to include more species from every subgenus and section.

Wood *et al*. (2000) conducted phylogenetic analysis of *Hedychium* based on ITS of ribosomal nuclear DNA sequence data. They found that *Hedychium* is monophyletic with strong support. Within *Hedychium*, four major clades are moderately supported. The number of flower per bract and distribution are distinguished among these clades.

Kress *et al*. (2002) constructed the cladogram of *Zingiberaceae* based on nuclear ITS and plastid *matK* sequences. He proposed the new system based on the Cladogram resulted from this study. *Zingiberaceae* in new system comprises four subfamilies and six tribes: *Siphonochiloideae* (*Siphonochileae*), *Tamijioideae* (*Tamijieae*), *Alpinioideae* (*Alpinieae*, *Riedelieae*), *Zingiberoideae* (*Zingibereae*, *Globbeae*). The results also suggested that many large genera, including *Alpinia*, *Amomum*, *Boesenbergia*, *Curcuma*, *Etlingera* and *Globba*, are not monophyletic. *Pommereschea* and *Rhyncanthus* form the previous tribe *Alpinieae* are placed in tribe *Zingibereae* in this system. The genus *Curcuma* appeared in three groups. The first

group is sister of the genus *Hitchenia*, the second is the sister group of *Smithatris* and the last one grouped with *Stahlianthus*.

Ngamriabsakul *et al*. (2004) performed a phylogenetic analysis of tribe *Zingibereae* (sensu Kress) using nuclear DNA (ITS, 5.8S) and chloroplast DNA (*trn*L (UAA) 5' exon to *trn*F (GAA). *Zingibereae* is monophyletic with two major clades, *Curcuma* clade and *Hedychium* clade. *Boesenbergia* is apparently not monophyletic. The *Curcuma* clade is composed of seven genera, *Camptandra*, *Curcuma*, *Hitchenia*, *Paracautleya*, *Pyrgophyllum*, *Smithatris* and *Stahlianthus*. *Camptandra* and *Pyrgophyllum* form a basal clade of the rest. *Curcuma* appeared in three groups. The first group comprises *C. alismatifolia*, *C. harmandii* and *C. parviflora*. It grouped with the genus *Stahlianthus*. The second group composed of *C. amada* and *C. rubescens*. They are sister to the group of genus *Hitchenia* and genus *Paracautleya*. The last group, *C. ecomata* is grouped with *Smithatris* and form a basal group of the first two.

Pedersen (2004) studied the phylogeny of subfamily *Alpinioideae*, particulary *Etlingera*, based on nuclear (ITS) and plastid DNA (*rps*16). The cladogram strongly supports *Etlingera* as a monophyletic genus with four major clades and *Hornstedtia* is a sister group. The inclusion of *Achasma*, *Geanthus* and *Nicolaia* is also strongly supported, even though the genus *Geanthus* forms a monophyletic group and can be separated as a different genus.

Williams *et al.* (2004) studied the genus *Globba* and tribe *Globbeae*. Parsimony and Baysian were conducted on nuclear ITS and plastid *trn*K-*mat*K. The results show that *Gagnepainia*, *Hemiorchis* and *Mantisia* are monophyletic. *Gagnepainia* and *Hemiorchis* are sister group of each other and form a basal group of the rest. *Mantisia* nested within *Globba* and was reduced to section *Mantisia*. Finally, *Globba* was divided into three subgenera, i.e. *Ceratanthera*, *Globba* and *Mantisia*, and eight sections. *Ceratanthera* comprises one section while *Mantisia* composed of three sections, including *Haplanthera*, *Mantisia* and *Substrigosa*. *Globba* is the largest subgenus. It was divided into four sections, *Globba*, *Mediocalcaratae*, *Nudae* and *Sempervirens*. Biogeography of this is notable. Four sections can be found only above Isthmus of Kra while the other two sections distribute only below this barrier. Three of them show the limit of distribution at this barrier too.

Recently, Funakoshi *et al.* (2005) rediscovered the monotypic genus, *Leptosolena*, of the Philippines. He added the sequence of this genus into phylogenetic analysis to confirm that it belongs to tribe *Alpinieae*.

Kress *et al.* (2005) performed parsimony and Baysian analysis of ITS and plastid *mat*K of *Alpinia* and other genera in *Alpinioideae*. The results confirmed that genus *Alpinia* is polyphyletic. Groups of species are more closely related to other genera in the tribe than they are to each others. The genus *Alpinia* can be divided into six major groups. Some groups included small genera within it.

The other molecular techniques use in *Zingiberaceae* systematics.

Apavatjarut *et al.* (1999) used molecular markers to help in identification of problematic species of *Curcuma*, particular early flowering species. Eight enzymes, from 21 evaluated enzymes, show polymorphic patterns. Seven taxa were used in this study; *C. aeruginosa*, *C. elata*, *C. rubescens*, *C. zanthorhiza*, *C. zedoaria* and two unidentified species. *C. rubescens* is a basal taxon of the rest. *C. elata* is grouped with *C. xanthorhiza* while the remaining four taxa formed the core group. Some morphological characters supported this grouping, such as color of rhizome and leaf-sheath and red patch along midrib.

Paisooksantivatana *et al.* (2001) studied genetic diversity of *Curcuma alismatifolia* from eight natural habitats in Thailand and cultivated plants in Japan and Thailand. Isozyme technique was performed with five enzymes, of seven tested enzymes. Eight to eleven plants of each habitat were used. Polymorphic locus of cultivated plants (40-60%) is lower than those of natural populations (80-100%).

Chokthaweepanich and Paisooksantivatana (2003) evaluated isozyme variation and morphological traits of 16 taxa of *Curcuma* and one taxon of *Smithatris*. Eight enzyme systems were used. All of them show polymorphism. The cluster based on 50 morphological characters showed two distinct clusters. The first one comprises species of subgenus *Paracurcuma* and *Smithatris*. The second cluster comprises species of *Eucurcuma* K. Schum. The analysis based on isozyme data resulted in two major clusters but did not agree with the morpho-based cluster. *Smithatris* appeared within one cluster of *Curcuma* in both analyses, not as sister group of all *Curcuma* species.

Vanijajiva *et al.* (2005) employed RAPD technique using five primers, out of ten tested primers, to prove the relationship of *Boesenbergia*, *Kaempferia* and *Scaphoclamys*. Total 53 bands were observed. It was confirmed that *Boesenbergia* and *Scaphoclamys* were more closely related than those with *Kaempferia*.

Objective

– To study the relationship among *Curcuma* species and related genera using molecular data.