6. DISCUSSION

Morphological and Morphometric Identification

The two hundred and thirty-seven samples from the eight locations in Thailand were identified using morphological features (the rostrum, the distance between the gastro-orbital ridge and the hepatic spine and the third maxilliped in adult males). The morphological identification of some samples is not clear. Although the rostrum feature differentiated *P. merguiensis* from *P. indicus*, it could not clearly separate *P. indicus* from *P. silasi*. Similarly, the distance between the gastro-orbital ridge and the hepatic spine could not clearly distinguish *P. indicus* from *P. silasi*. The best overall feature is the third maxilliped in the adult male, which can separate these species clearly. However, this feature can not be used in female or juvenile prawns. Another significant limitation is that the morphological keys are often effective only for a particular life stage or gender, and many individuals cannot be positively identified.

Morphometric measurements (especially the ratio of L_1: $L_{1,2}$) were thus applied to separate these species. In samples identified as *P. silasi*, the ratio value was still overlapping the values of *P. merguiensis* and *P. indicus*. In addition, in some samples (M/I group), the ratio value was out of the range of *P. merguiensis* and *P. indicus* and did not clearly indicate the species.

Even when the morphological and morphometric data were compared, the species identification was still often unresolved and problematic. One hundred and twenty-seven samples (53.6%) provided consistent species identification from the morphological and morphometric data (group I) while forty-one (17.3%) and sixty-

nine samples (29.1%) were inconsistent (group II), and could not identify the species (group III), respectively. This shows that relationships based solely on morphology, which is cost-effective and rapidly, but it can be misleading because of deep morphological differentiation (slow rate of morphological divergence). Therefore it is difficult to positively identity these species (*P. merguiensis*, *P. silasi* and *P. indicus*).

According to morphological and morphometric identification, *P. penicillatus* was not found in Thai waters. This finding corresponds to Leelapiyanart (Thesis, 1988), who reported that this species was bought from another peninsular and was not native to Thai waters.

Isozyme study

Fifty-nine samples from morphology and morphometric identification (all 3 groups), additional 6 specimens of *P. silasi* were chosen as the outgroup in isozyme study. The isozyme results from sixty-five-specimens showed that the MDH system could not be used to separate *P. silasi* from the other samples. Two patterns of MDH isozyme were found in *P. silasi*. 5 of 6 specimens of *P. silasi* gave a three-banded pattern and one specimen (NKE18) provided the two-banded pattern. Similar to the samples from groups I, II and III, there were two patterns of MDH isozyme. 58 of 59 specimens provided the two-banded pattern, while only one (SKEB5) showed the three-banded pattern. Groups II and III, for which morphological and morphometric data are inconsistent and the prawn unidentifiable, provided the same patterns in the MDH system. The specimens came from the different locations and variable morphology (*P. merguiensis* and *P. indicus*) but they also provided the same pattern

of MDH isozyme (two-banded). Similar to some isozyme studies that often reveal discordant pattern between morphological data and genetic divergence such as the morphological distinct taxa sometimes show little or no genetic divergence (Turner, 1974; Echelle and Dowling, 1992).

MDH is a dimeric enzyme commonly encoded by two different gene loci. The two bands observed here is the result of both gene products migrating to different positions on the gel, and they are homozygous for each of the two loci (*MDH1*, *MDH2*). The heterozygote pattern was not found in our samples, which is similar to Redfield et al. (1980), who found the low heterozygosity in a number of tropical decapod Crustacea. The low heterozygosity presumably makes them more vulnerable to selection than animals with high heterozygosity, but Mulley and Latter (1980) suggest that this is extremely unlikely for marine species. Nevertheless, dramatic environmental changes, such as the Quaternary Glacial Epoch, could have induced bottlenecks in breeding populations, with resulting in breaks in population structures. In contrast, the high fecundity of penaeids would result in non-lethal mutations becoming rapidly distributed through the population. The low heterozygosity does not mean that the rate of mutation is low, indirectly there is considerable diverged has occurred in Penaeidae in recent geological epochs.

Locus differences (medium allele) were observed in *P. silasi* and SKEB5. The differences were considered to be the isozyme rather than allozyme differences due to the large separation distance between the bands. Thus the medium allele remains useful in identifying individuals from *P. silasi*. It is likely to say that *P. silasi* has three-banded pattern although NKE18 and SKEB5 showed strange pattern. For

populations and species that have diverged recently, a major limitation of protein electrophoresis can be that insufficient polymorphism exists for an accurate assessment of population structure or for reconstruction of phylogenetic relationships (Larson et al., 1984).

The medium allele possibly arises from the mutation, when gene or parts of genes become duplicated, usually found in an isozyme event. Mutation is a mechanism to increase genetic variation, leading to evolution. The majority of molecular changes in evolution are due to the random fixation of neutral or nearly neutral mutations (Kimura 1968). This hypothesis is known as the neutral theory of molecular evolution. At the molecular level, the majority of evolutionary changes and much of the variability within species are caused neither by positive selection of advantageous alleles nor by balancing selection, but by random genetic drift of mutant alleles that are selectively neutral or nearly so. The absolute value of the selective advantage or disadvantage of an allele must be smaller than $1/(2N_e)$, where N_e is the effective population size. According to the neutral theory, the evolutionary process should be regarded as the result of a continuous process of mutational input and a concomitant random extinction or fixation of alleles. Thus, the neutral theory regards substitution and polymorphism as two facets of the same phenomenon. Substitution is a long gradual process whereby the frequencies of mutant alleles increase or decrease randomly, until the alleles are ultimately fixed or lost by chance. According to the neutral theory, most genetic polymorphism in populations is transient in nature.

Mitochondrial DNA study

12S and 16S rRNA

The A+T content of 12S and 16S data in this study, in the genus *Penaeus* was about 69% and 66%, respectively, which supports the previous finding that the A+T content in the genus *Penaeus* mtDNA molecule is around 70% (Machado et al., 1993).

Mitochondrial ribosomal genes (12S and 16S) were amplified from the samples of strongly identified species*, P. merguiensis* (SREE1), *P. indicus* (STWA6), *P. silasi* (NKE8), *P. indicus* (TA5, from South Africa) and *P. monodon* (D1). Their nucleotide divergence was clearly differentiated and supported by the morphological data, except STWA6, an old sample collected from Satun in 1995 and identified the species as *P. indicus*. The nucleotide sequence of this sample showed a low level of the nucleotide divergence compared with SREE1, which was identified as *P. merguiensis*, both in 12S and 16S data. Nevertheless, it was not close to TA5, which was believed to be the same species (*P. indicus*). Because of the sample (STWA6) was not used the third maxilliped to identify. However, the rostrum, and the distance of gastro-orbital crest between gastro-orbital ridge and the hepatic spine, including the sexual organ, was used for identification. In the case of STWA6, we suspected that it was morphological misleading. Hall (1956) commented "although adults demonstrating the features typical of these species may be identified fairly easily, there are many cases in which features of all three species (*P. merguiensis*, *P. indicus* and *P. penicillatus*) may be exhibited by a single individual". And STWA6 might be a genetic variation of *P. merguiensis*.

The nucleotide divergences of 12S between *P. merguiensis* and *P. silasi* were 2% while the nucleotide divergence of 12S between *P. merguiensis* and *P. indicus* were 5.92%, indicating that *P. merguiensis* (SREE1) is more similar to *P. silasi* (NKE8) than to *P. indicus* (TA5). Palumbi and Benzie (1991) reported the 16S nucleotide divergence between *P. stylirostris* and *P. vannamei* (belonging to the same subgenus) was 8.4% while our 16S nucleotide divergence in the same subgenus was only 0.54% (between *P. merguiensis* and *P. silasi*) and 1.45% (between *P. merguiensis* and *P. indicus*). This indicates that the divergence among our white prawn is earlier than that in *P. stylirostris* and *P. vannamei*.

Our work also indicated that the 12S data provided more phylogenetic usefulness for species identification of the same subgenus than the 16S data, which provided a low level of phylogenetic information at the subgenus level. The 12S sequences data partition contained greater phylogenetic signal than did the 16S data partitions among deeply diverged taxa corresponding to other organisms (Austin et al., 2002; Georges et al., 1999).

The low level of 16S data was obscured phylogenetic signal because of inappropriate choice of gene or insufficient sequence data. In addition, one possibility is that the radiation among *P. merguiensis*, *P. silasi* and *P. indicus* occurred very rapidly such that there was insufficient time for accumulations of large numbers of characters that would be phylogenetically informative (Brown et al., 1994; Georges et al., 1999; Austin et al., 2002). Evidence for rapid divergence is often inferred from low variation in pairwise divergence among lineage (e.g. Kraus and Miyamoto, 1990; Shaffer et al., 1997).

The number of specimens and populations needed per group to resolve relationships among groups depends critically on the amount of polymorphism relative to the extent of divergence. If the sequence indicates that virtually all of the variation occurs among groups, then it is appropriate to use small samples per groups. However, here the number of specimens examined per group was quite small (one specimen) and with the low divergence of the 12S and 16S rRNA genes, it was necessary to examine a larger sample size of closely related species. In closely related species, particularly when non-recombining sequences such as mtDNA are being used, it may be more efficient to conduct the sampling and analysis in two steps, first to identify clusters of closely related taxa and second to add geographically remote populations for each of the members of such clusters (Hillis et al., 1996).

The protein-coding gene in the animal mitochondrial genome (COI) is a better target because higher variation of nucleotide sequences in the conserve amino acid sequence.

COI

Phylogentic relationships

To resolve the incongruent tree of 12S and 16S data and to confirm species of STWA6, COI is one of the protein-coding genes in mtDNA, which was chosen to identify these species. The COI nucleotide sequence provided a high level of divergence between species and was also easily and clearly differentiated (Baldwin et al., 1998). COI could be used either to study in the interspecific or intraspecific level (Zhao et al., 2002). In our case, the branching topologies agree with the species tree. It

should be clear that nucleic acid sequencing could be used to study virtually any systematic problem, from studies of evolutionary processes to the phylogeny of life.

Specimens representing each species (*P. merguiensis*, *P. silasi* and *P. indicus*) were chosen. Two samples, SREE1 and SREE2, were strongly identified as *P. merguiensis,* five specimens (NKE2, NKE5, NKE8, NKE17 and NKE19 from Nakhon si thummarat) of certain *P. silasi* species and three samples, STWA5, STWA6 and STWA7 strongly identified as *P. indicus* were selected based on morphological characteristics. We needed *P. indicus* as a reference to compare with the sequence data of STWA6. So we contacted Jacopo Querci (Ph.D. student) in South Africa (where the major species of white prawn is *P. indicus*) and obtained the DNA of *P. indicus* from there to use as a reference. If STWA6 was *P. indicus,* the results would be the same as the reference. Four specimens of *P. indicus* (PIZA, PIOMAN, PIMZ and PITA5) were chosen. In addition, construction of a tree may be aided by getting outgroup sequences (*P. monodon* and *P. vannamei*) as well. *P. indicus* and *P. merguiensis* sequences in database is required for cluster same species together and separate from different species. After that the larger sample sizes (Group I, II and III)will be needed to estimate phylogeny.

The COI data provide four clusters, A, B, C and D. The phylogenetic tree shows that the differentiation of *P. merguiensis*, *P. silasi* and *P. indicus*. *P. silasi* and *P. indicus* are monophyletic whereas *P. merguiensis* is paraphyletic. Two clusters (clusters A and B) are identified within *P. merguiensis*. Clusters C and D are clearly differentiated for *P. silasi* and *P. indicus*, respectively. *P. indicus* (samples from South Africa, PITA5, PIZA, PIOMAN, PIMZ) are clustered in the same cluster of *P.*

indicus sequence from the GenBank database (PIGB), indicating that these samples are the true *P. indicus*. Cluster A is *P. merguiensis* because the *P. merguiensis* sequences from the GenBank database are clustered in this cluster. In addition, STWA6 was clustered in cluster A (*P. merguiensis*) as we suspected.

Cluster B contains the samples from group I, II and III (morphological and morphometric comparison) and samples (SREE1 and SREE2, which were strongly identified as *P. merguiensis*). In addition, the MDH pattern of these samples is *MDH1/MDH2*. Most samples of cluster B come from the West Coast of Thailand while most samples of cluster A come from the East Coast of Thailand.

The samples from Nakhon Si Thummarat (NKE) were both *P. silasi* (NKE2, NKE5, NKE8, NKE17, NKE18 and NKE19) and *P. merguiensis* (NKE1 and NKE9). It is likely to be the sympatric speciation, the formation of two or more descendant species from a single ancestral species all occupying the same geographic location. Sympatric speciation is strongly suggests that ecological factors (e.g. food) in a sympatric population can cause speciation. In the depths of Siberia's Lake Baikal, sympatric speciation driven by ecological factors may also account for the extraordinary diversity of crustaceans (Rundle et al., 2000). Evidence is accumulating that ecology is important for speciation (Orr and Smith, 1998). The ecological factor provide for understanding the puzzle evidence for many sympatric species, including cichlids (Meyer et al., 1990; Schliewen et al., 1994), sticklebacks (Schluter and McPhail 1993; Schluter, 1994; Taylor and McPhail, 1999), snail (Johannesson et al., 1995). Individuals vary in a quantitative character determining resource use, in the case of *P. merguiensis* and *P. silasi*, their third maxilliped character, which are used

extensively for grooming and holding large pieces of food, are different. This different feature show that sympatric speciation is a likely outcome of competition for resource. The other characters may be the effect of ecological factor therefore the more details of morphological information should be further study. In addition, the evolution of assortative mating (where individuals mate preferentially with like individuals) is considered. It depends either on an ecological character affecting resource use or on a selectively neutral marker trait. In both cases, evolution of assortative mating often leads to reproductive isolation between ecologically diverging subpopulations.

The two clusters, A and B are, on average, 5.0% divergence. Although the morphology of cluster A and clde B was not different significantly (both identified as *P. merguiensis*), the COI data showed the significant differences. It is possible to form two hypotheses 1) *P. merguiensis* is a complex of two cryptic species, and 2) *P. merguiensis* is a single species with a strong phylogeographic subdivision. The first hypothesis could be explained by the apparent differences in molecular data and morphological evolution (Palumbi and Benzie, 1991): (1) the rate of mtDNA evolution might be accelerated in prawn, or (2) the rate of morphological divergence might be slow in prawn. The observations indicate that the Indian-Pacific region contains the greatest species diversity with morphologically similar species, and the occurrence of recent speciation has also been observed in marine animals (Palumbi, 1992; Palumbi, 1997). A large number of sibling species have been detected in what were previously thought to be a single taxa (Knowlton, 1993; Knowlton and Jackson, 1994). It is possible that allopatric speciation (the formation of two or more species

often requirs geographic isolation of subpopulation of the species) may be occurred between cluster A and cluster B because of the land barrier. In addition, the congeneric species of animals possess substantial sequence divergence in their COI genes. In fact, more than 98% of species pairs show greater than 2% sequence divergence (Hebert et al., 2003). A second hypothesis (Phylogeographic hypothesis) derived from our data that most samples of cluster A came from the East (Pacific Ocean) and those of cluster B came from the West (Indian Ocean). It is possible that the genetic distance between cluster A and B (about 5%) is because of the geographic distribution. Significant genetic differentiation of populations in the Indian and Pacific Oceans has been reported for coconut crabs (Lavery et al. 1996), the starfish *Linckia laevigata* (Williams and Benzie, 1997), butterfly fish (McMillan and Palumbi, 1995), and the starfish *Acanthaster planci* (Benzie, 1999). These two hypotheses will be completely answered by combining data between mtDNA and nuclear DNA (Wanna et al., 2004). If the nuclear DNA shows the significant difference between cluster A and B, the first hypothesis will correct. If the nuclear DNA shows no significant between cluster A and B, the second hypothesis will correct.

However, five samples (STWA5, STWA6, STWA7, SREE1 and SREE2) provided inconsistent data between COI and geographic distribution. Samples STWA5, STWA6 and STWA7, collected from the West Coast, were placed into cluster A while samples SREE1 and SREE2, collected from the East Coast, were placed into cluster B. The low level admixture appearance between cluster A and B could be due to a limited amount of genetic exchanges or the transportation of larvae from one coast to the other by human, or even to collecting artifacts.

Recently separated species often display paraphyletic gene tree patterns (DeSalle et al., 1987; Satta and Takahata, 1990; Brown et al., 1996). Most geographic and demographic scenarios for speciation initially result in paraphyletic taxa when reproductive isolation forms the basis for species definition. Our topology show the paraphyletic group among *P. merguiensis*, *P. silasi* and *P. indicus*, therefore the COI phylogenetic tree indicates that they are recently separated species, which corresponds to the isozyme data.

In this study, samples of *P. silasi* in clade C came from Nakhon si thummarat, the Gulf of Thailand only. Several attempts to isolate *P. silasi* from the other locations were not successful. At this time, it is not clear whether there is restricted distribution of this species. It may be extant in other time intervals or other locations that we did not collect specimens from. More time and study is needed. This is an interesting point for conservation biology of this species and should be studied further.

The COI nucleotide divergence in the subgenus *Fenneropenaeus* is higher than the 12S and 16S data. The divergence between *P. merguiensis* and *P. silasi* is about 8%, between *P. merguiensis* and *P. indicus* about 12% and between *P. silasi* and *P. indicus* about 16%. This data support that *P. merguiensis* is more similar to *P. silasi* than to *P. indicus*.

The data on divergence within species of cluster A (0.0-2.25%), B (0.18-1.49%), C (0.74-1.1%) and D (0.18-1.5 %) are consistent with Bucklin et al. (1998), who reported that the COI sequence variation within species ranges from 1-2%. It showed the high genetic divergence of the COI gene, especially in cluster A. These results are similar to reports for genera of tropical marine vertebrates, which showed

high genetic variation of widespread marine organism in the Indo-Pacific (i.e. Briggs, 1974; Bowen et al., 1998). In addition, our morphological and COI data supported the hypothesis that the genus arose in the Indo-Pacific (Dall et al., 1990). This hypothesis was originally formulated based on the criteria that biogeographic centers of origin have the highest species diversity (high genetic divergence) and the deepest morphological differentiation (Briggs, 1995; Avise, 2000).

PCR-RFLP

Many studies of gene evolution require DNA sequencing because no other technique provides the necessary information to infer relationships among individual alleles. Although it has become easier to obtain sequences from many individuals for certain loci (particularly the mitochondrial genome) by amplifying the DNA, it is still expensive and time-consuming. The nucleotide sequence provides a detailed restriction site map that allows for precise interpretation of fragment changes (e.g., Cann et al., 1984; Hugall et al., 1994). The PCR-RFLP technique was used rapidly and cost effectively for population analysis (Dowling et al., 1990; Wilding et al., 1999) and species differentiation (Schroeder et al., 2003; Elliott et al., 2002). The high sequence variation of COI gene both between and within clusters also can be examined indirectly by electrophoretically comparing DNA fragments to look for the number or distribution of restriction sites.

The 3'COI fragment obtained from primers mtD-8 and mtD-12, about 900 bp covering the 558 bp region, was amplified and then cleaved by the restriction endonuclease *Rsa*I and the closely related species identified. This fragment was not

only used to identify other taxa in subgenus *Fenneropenaeus* (*P. merguiensis, P. indicus*, *P. silasi*), but also to diagnose in genus *Penaeus* (*P. monodon* and *P. semisulcatus*). In addition, the *Rsa*I restriction enzyme in the COI sequences was consistent with the phylogenetic tree that can separate this same subgenus clearly.

A 5'-COI fragment was obtained from primers mtD-4 and mtD-9, about 800 bp, which could be applied as a marker for studying the geographic distribution of *P. merguiensis*. The PCR-RFLP patterns of the west and the east coasts of Thailand are different. Such a marker is likely to be useful in searching for whether an individual came from the east (cluster A) or the west (cluster B) of peninsular Thailand. Clusters A and B have a 5% sequence divergence. Hillis et al. (1996) reported that when sequence divergence in animal mtDNA is relatively high (>2-4%), it is usually possible to obtain enough characters from 6 bp restriction enzymes. Therefore we can search the restriction sites from the sequences and successfully differentiate these two clusters by *Bgl*II and *Mbo*I sites. This result also corresponds with the phylogenetic tree that can separate clusters A and B.

The restriction pattern from the agarose gel electrophoresis was incongruent with the restriction site search from nucleotide sequence because of the low efficiency of agarose gel electrophoresis technique. The limit of this method is the low different in fragment sizes (*P. merguiensis*, 142 and 152 bp, and *P. indicus*, 282, 248 and 334 bp) are difficult to separate. However, this is preliminarily and roughly tests and takes a short time to confirm the expected differentiation from these species. To confirm thoroughly, polyacrylamide gel electrophoresis should be applied.

In PCR-RFLP, comparison of mapped restriction sites allow for interpretation of fragment pattern differences as individual mutations that affect the presence and position of restriction sites. The 558 bp could be use to search for the restriction site map and we can predict the fragment size of these closely related species. In general, the fragment comparison should be restricted to very closely related sequences but the length variation fragment of PCR product (Table 23) was found in our closely related species. This evidence will affect the fragment patterns produced by RE. Then the expected fragment size and the fragment pattern on the gel are different.

The small sample size was chosen in PCR-RFLP so that the base substitution would be constant within species. If there is high variation within the species such as *P. monodon* (Klinbunga et al., 1999), it may interfered with species identification. So the sample size in each species should be larger in order to be confident of the species identification.

Using combined data from morphology, isozyme and DNA, we can separate *P. merguiensis* from other close species (*P. silasi* and *P. indicus*). COI data is suitable for species–specific identification, including female samples. The basis of a diagnostic PCR-RFLP assay can be used to study both inter- and intra-specific levels of white prawn. The COI marker is suitable for population genetics in further studies, and it is also possible to apply to larvae identification in aquaculture.