# CHAPTER IV

## Discussions

Recently, commercial honey bee derived from *A. mellifera* strain is actively maintained in hive by beekeepers to produce honey and other bee products whereas feral honey bee is not. MtDNA variation is potentially used to genetically characterize commercial honey bee queen (Schiff *et al.*, 1994; Schiff and Sheppard, 1995). Based on the fact that the honey bee queen is multiple-inseminated, analysis of genetic polymorphism in *A. cerana* using mtDNA is much simpler than that from nuclear DNA. Accordingly, PCR-RFLP and direct sequencing of specific regions in mitochondrial genome were used to evaluate genetic variability of eastern honey bee, *A. cerana,* from five geographic locations in Thailand.

Thorax rather than a whole body was utilized as a source for total DNA isolation because it is an active organ which is believed to contain high ratio of mitochondrial / nuclear DNA. Since mtDNA specific primers were used, nuclear DNA did not interfere in PCR amplification. RNA was unncessary to carry out digestion because it was present in very small quantity that it can not interfere in PCR amplification (Hoelzel and Green, 1992).The rapid extraction method used in this present study gave tiny amount of extracted DNA and contaminated RNA resulted in the purity and concentration of DNA were not spectrophotometrically estimated.

Three regions (sRNA gene, lrRNA gene and inter CO I-CO II region) in mitochondrial genome (about 17.54 % out of whole genome) were chosen for this study. Previously, it has been reported that sRNA and lrRNA genes are quite conserved therefore these regions are usually suitable to be used interspecifically (Hall and Smith, 1991). On the other hand, an inter CO I-CO II region evolves rapidly reflecting its widely used for studies of genetic polymorphism at the intraspecific level (Cornuet *et al.*, 1991;Garnery *et al.*, 1993; Garnery *et al.*,1995). Theoretically, investigation of genes having different evolutionary rates should yield an average in genetic polymorphism of a particular taxon. Heterologous primers originating from *A. mellifera* and *D. yakuba* were then employed as they have successfully been used in genetic variation analysis of species which is closely related to *A. cerana*. Moreover, these primers had also worked well in *A. cerana* obviating a need to develop homospecific primers for this study. Primer dimer occured in sRNA gene can be misscoring to be a restricted band, including of undigested sRNA gene amplified product in the same gel eleminated its interference.

The expected amplification products for sRNA gene, lrRNA gene and inter CO I-CO II region inferred from *A. mellifera* are 360, 738 and 1720 bp, respectively (Sheppard and McPherin, 1991; Hall and Smith, 1991; Crozier and Crozier, 1993). However, the resulting products in *A. cerana* were 400, 750 and 1710 bp which were slightly longer or shorter than those expected from *A. mellifera* mtDNA sequence. Sequencing of amplified lrRNA gene in *A. cerana* verified its homologue to previously studied sequences of this regions. The similarity of lrRNA gene sequence portions between *A. cerana* and *A. mellifera* was 73 % out

of 432 nucleotides alignment in BLAST search. Moreover, the single intense band observed in amplification of sRNA gene and inter CO I-CO II region implied specificity of PCR reactions. These verified that all investigated genes in the present study are the actual homologous amplified products.

During the restriction enzyme screening process, hexanucleotide rather than tetranucleotide enzymes were of interest. The reason for this was that four base cutters may produce too many digested fragments. Some of which may not be surely identified by gel electrophoresis. This may be resulted from an inability to dissociate fragments which are too closed in size and some extremely small size fragments may be move out of the gel. Therefore, analysis of genetic polymorphism in *A. cerana* may require both agarose and polyacrylamide gel electrophoresis to confirm the results. This is tedious and time consuming. On the other hand, restricted DNA fragment using hexanucleotide restriction enzymes produced reasonably large fragment, consequently, analysis of these fragments could be rapidly examined using agarose gel electrophoresis. From a preliminary screening data, *Dra* I was the most informative restriction enzyme for three regions of *A. cerana*. Moreover, it was previously reported that 103 *Dra* I restriction sites were available in *A. mellifera* mitochondrial genome(Hillis, Moritz and Mable, 1996). Regarding the information described above, *Dra* I was chosen to be employed for analysis of genetic variation in *A. cerana*. This research is believed to be the first population genetic study of *A. cerana* in Thailand.

The inter CO I-CO II region of *A. cerana* showed the highest polymorphic level. A total of eight haplotypes were found from *Dra* I digestion of this region.

This indicated that mtDNA sequences at this region are much less conserved than those of other two genes where the lowest polymorphic level was found in sRNA gene portion. Similar to *A. mellifera*, the inter CO I-CO II region showed 23 different haplotypes(Garnery *et al.*,1993;Garnery *et al.*,1995). Four distinguishable variability lengths of non-restricted amplified DNA have been observed in amplified inter CO I- CO II region of *A. mellifera*, that was not found in *A. cerana* samples. For other regions, sRNA and lrRNA genes of *A. cerana* were found to have three and five haplotypes from *Dra* I digestion which were lower polymorphic than inter CO I-CO II region. Moreover, lrRNA gene of *A. cerana* did not contain restriction enzyme cleavage site such as *Eco*R I which was found in *A. mellifera* mitochondrial lrRNA gene and clearly discriminated East European from West European and African subspecies (Hall and Smith, 1991).

For scoring fragment size, some characters of polymorphic bands in three regions may be lost. Too small fragments (below 30 bp) can not be detected but a few missing band did not significantly disturb the genetic variation level. Restricted fragment having the same size may be resulted from different origin (not homologous). Theoretically, restriction fragments which are different in size due to insertions or deletions can be misscored to be different. An example of this was observed in *Dra* I digestion of amplified lrRNA gene. A 120 bp fragment in haplotype B was scored to be the same as a 120 bp fragment in haplotype E. Data from DNA sequencing indicated that the latter are the homologous fragment but deletions of five nucleotides were occured (Figure 3.22). This demonstrated that homologous fragments could have been wrongly scored if insertions or deletions

are involved. When needed, restriction site analysis or alternatively, DNA sequencing can be utilised to confirm the results.

On the basis of restriction analysis in three mtDNA gene regions, five geographic locations of *A. cerana* in Thailand can be roughly allocated into two distincitve groups (Table 3.1). One was the Northern *A. cerana* and the other was the Southern *A. cerana* . Haplotype A from sRNA gene, A and D from lrRNA gene, and A,D,G and H from inter CO I-CO II region were found in the Northern group whereas the remaining haplotypes from each region (B and C from sRNA gene, B and E from lrRNA gene, and B,C,E and F from inter CO I-CO II region) were specific to the Southern group. More specifically, haplotype C from lrRNA gene was population-specifically observed only in the Samui Island samples. This evidence clearly illustrated the genetic break between the Northern and Southern populations.

The South and Samui Island *A. cerana* were closely related. Genetic distance between these two populations was estimated to be 0.08% sequence divergence (Figure 3.17). Shared single and composite haplotypes (haplotype B from lrRNA gene, haplotype C from an inter CO I-CO II region and composite haplotype BBB) indicated the closed relationship between the South and the Samui Island *A. cerana.* While higher frequency of haplotype B was found in the South, distribution frequency of haplotype C was on the opposite direction.

Based on the genetic difference between composite haplotypes, three distinctive mtDNA groups could be indentified. The most common composite

haplotype AAA and BBB were allocated into group A and B, respectively, in which AAB was expected to be an intermediate haplotype (Figure 3.16). It should be emphasized that CED (group C) was allocated into different clone. This composite haplotype was carried by 2 individuals in the South. The genetic distance between group C and A or B was 2.95 % sequence divergence. This reflected a large genetic difference within taxon therefore a further study of individuals possessing CED haplotypes with other *Apis* species is required to answer a clearer circumstance whether such individual should be systematically allocated to be *A. cerana*. The interconnection between BBB and CED could not be directly related unless a hypothetical ancestor (BBA) was proposed. In the present study, BBA was not observed in any Thai *A. cerana* populations (Figure 3.15). As a result, an increase in sample size investigated may be required.

According to haplotype diversity within populations (Table 3.5), the highest variation was observed in the Samui Island corresponding to distribution frequencies of its common composite haplotypes. The Samui Island had two major composite haplotypes (BBB and BCC), while the others had only one major composite haplotype. Nucleotide diversity between populations indicated that large genetic differences were illustrated when each of the Northern group is compared with the Southern group but not within each group (Table 3.6). Theoretically, when nucleotide diversity among populations was higher than that within populations, it implied that degree of population differentiation in an investigated species exists. Phylogenetic reconstruction showed high level of estimated percent sequence divergence (1.41%) between the Northern and the Southern groups (Figure 3.16).

A chi-square analysis was used to statistically indicate genetic differentiation from composite haplotype distribution frequencies (Table 3.7). Five geographic locations of *A. cerana* could be statistically allocated into three genetically different groups. The first group was composed of North, North-East and Central whereas the second group consisted of only the South. Samui Island *A. cerana* was also separated to be another group.

The genetic variation and population structure of *A. mellifera* in Morocco and Spain have also been analysed using mtDNA PCR-RFLP of inter CO I- CO II region with *Dra* I digestion. The Moroccan *A. mellifera* was a large single population representing the African lineage whereas the Spainish *A. mellifera* could be further devided into two populations including the African and the West European lineages (Gernery *et al.*, 1995). RFLP analysis in mitochondrial 16S RNA gene of *Crassostrea virginica* have been used to investigate genetic diversity within the species using ten restriction enzymes. Low genetic variation level was found among different geographic samples (Small and Chapman, 1997). On the orther hand, genetic diversity of mtDNA lrRNA gene of *A. cerana* in this study is sensitive enough for analysis of its genetic variation and population structure.

An amplified mitochondrial portion of lrRNA gene can be directly sequenced for comparisons of distantly related taxa. Purification of amplified DNA is necessary to remove dNTPs, primers and non specific product. Seven-deaza dGTP, modified version of dGTP is used to prevent forming of intrastrand hairpin loop affecting gel electrophoresis. In order to visualize sequencing bands,

$^{32}P$ was chosen to be end labelled at 5′ end of primer instead of $^{35}S$. Degradation of the sequencing products by radiolysis is not a problem when using end-labelled sequencing primer (Brown, 1994).

For a long DNA, internal primers were designed to carry on further sequencing. In this study, these were directly designed from the former sequences. Theoretically, no less than 15 nucleotides should be used. Moreover, the GC content of synthesised primers should also be between 40-60 %. Based on the fact that the lrRNA gene in *A. cerana* is AT rich, internal primers was designed to be long enough to provide sufficiently high annealing temperature ( > 45 °C). Δ G was also considered for spontenous reaction (Hoelzel and Green, 1992).

Base composition of *A. cerana* lrRNA gene was extremely AT-biased (84.47%). Transversional mutations between investigated sequences were about twice as abundant as transitional mutations. Nevertheless, this ratio was still slightly lower than that in *A. mellifera* where transversions / transitions ratio of lrRNA gene was about 3.66 (AT content of lrRNA gene in *A. mellifera* was 85.3 %). Most of substitutions found in lrRNA gene of *A. cerana* were A ↔ T (73.68% of all transversion sites). In general, high frequencies of A and T in lrRNA gene seem to be a prominent character for hymenopterous insects (Crozier and Crozier, 1993). Insertions or deletions of amplified lrRNA gene were not found by gel electrophoresis due to a limitation of agarose gel electrophoresis. However, those could be unambiguously identified by DNA sequencing.

Clustering of sequences of different lrRNA gene haplotype was generally in accord with that of PCR-RFLP datum. Three different groups were also observed from the sequencing data (Figure 3.23). The estimated percent sequence divergence in each case was not enormously different. Therefore, a reasonably accurate and precise genetic polymorphism level can be estimated through PCR-RFLP approach. Moreover, ninty-seven nucleotides out of seventeen samples were obtained from inter CO I-CO II region of *A. cerana* in Thailand and used to determine for phylogenetic analysis. The result of these sequences could geographically divid *A. cerana* into only two distinctive groups : the Northern and the Southern bee (Bugharung, 1996).

Although PCR-RFLP analysis may theoretically offer less information than nucleotide sequencing, it is a simple, powerful, cost-effective and less time consuming alternative where a large number of individuals, loci or large segments of genome can be examined in a short period of time. Therefore, PCR-RFLP is a promising approach to be utilised for population studies in various taxa so an advantage of PCR-RFLP is using for rapid discriminate *A. cerana* in Thailand. Furthermore, this technique can be helpfully used to investigate *A. cerana* strain which has an important biological characters such as disease resistance, hive maintainance and more honey production. Selected breeder queen can be then applied for beekeeping improvement of *A. cerana* in Thailand.