CHAPTER III

Results

3.1 DNA template

Typically, high molecular weight DNA larger than 23.1 kb in length was usually obtained from the total DNA extraction (Figure 3.1). An absolute amount of DNA extracted was about 1.0-1.5 μg per individual specimen of about 20 mg (comparing its intensity with that of the lamda-*Hin*d III marker). RNA and slightly sheared DNA were observed but these constituents did not interfere the subsequent PCR reaction. Based on the fact that mitochondrial DNA is transmitted matriarchally, it is, therefore, possible to employ a single *A. cerana* worker to represent a genetic pattern of others within the same colony.

3.2 Optimization of PCR conditions

Three regions of mitochondrial genome (sRNA gene, lrRNA gene and inter CO I-COII region) were *in vitro* amplified through PCR using heterospecific primers (e.g. from *A. mellifera*, *D. yakuba*). If these mtDNA gene portions were successfully amplified, the PCR reaction was further optimized for more appropriate primers and $MgCl_2$ concentrations. The other constituents were as described in 2.9.

For amplification of sRNA gene, 0-5.0 mM $MgCl_2$ concentration was optimized at a constant concentration of 0.5 μM primers. High primer

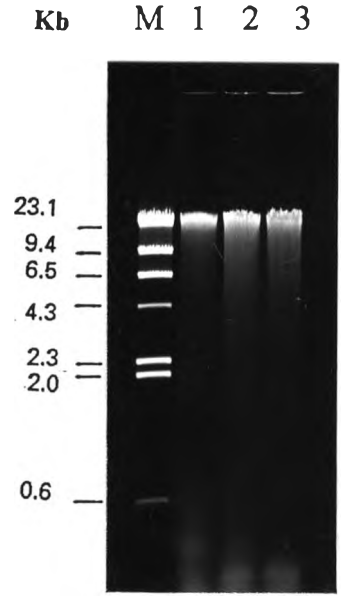


Figure 3.1 High molecular weight DNA extracted from thorax of *A. cerana*.

lane M = λ - *Hin*d III standard DNA marker

lane 1-3 = Total DNA from three *A. cerana* individuals

concentration was used in this optimization, because it was expected that specificity of sRNA primer which was originally developed from distantly related taxon (*D. yakuba*) should be lower than other primers which were developed from closely related species (*A. mellifera*). PCR-amplified DNA was firstly appeared at 2.0 mM $MgCl_2$. Non specific product was observed below PCR-amplified DNA but its amount was not high enough to interfere an interpretation of the restriction enzyme digesting results. Primer dimer was also observed. As can be seen from Figure 3.2, increasing of $MgCl_2$ concentration to be higher than 3.0 mM did not yield significantly higher amount of amplified product, therefore an optimal $MgCl_2$ concentration was chosen at such a concentration. Subsequently, optimal concentration of primers was evaluated in the same manner. Positive relations between primer concentration and amount of primer dimer appeared after PCR-amplified product observed (Figure 3.3). The optimal primer concentration was chosen at 0.12 μM.

Optimization of lrRNA gene amplification was attempted by varying $MgCl_2$ concentration ranging from 0-5.0 mM at a constant concentration of 0.1 μM primers. Faint amplified product was observed at 1.5 mM $MgCl_2$ and consistently increased with higher $MgCl_2$ concentration until 3.0 mM (Figure 3.4). Therefore, the $MgCl_2$ concentration used for amplification of lrRNA gene in present study was 3.0 mM. The primer concentration was then varied in the range of 0-0.2 μM. The amplified product was firstly visualized at 0.04 μM primer. The resulting product was increased in its concentration

Figure 3.2 Agarose gel electrophoresis showing an optimization of $MgCl_2$ concentration used for sRNA gene amplified at a constant primer concentraion (0.5μM).

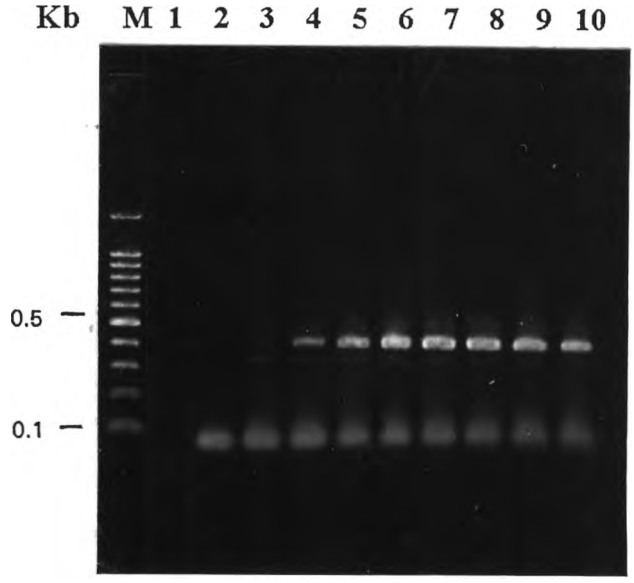
lane M = 100 bp DNA ladder

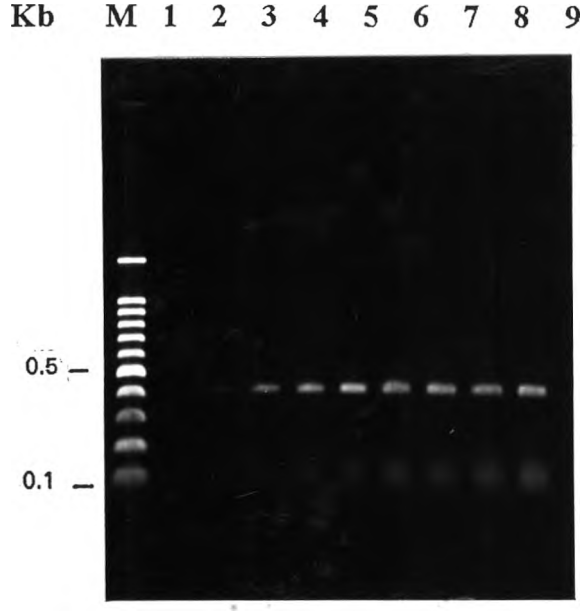
lane 1- 10 = Amplified product in a seies of 0, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5 and 5.0 mM of $MgCl_2$, respectively.

Figure 3.3 Agarose gel electrophoresis showing an optimization of primer concentration used for sRNA gene amplified at a constant $MgCl_2$ concentration (3.0 mM).

lane M = 100 bp DNA ladder

lane 1-9 = Amplified product in a series of 0, 0.04, 0.08, 0.12, 0.16, 0.20, 0.24, 0.28 and 0.32 μM of primers, respectively.


Kb M 1 2 3 4 5 6 7 8 9 10
0.5
0.1


Kb M 1 2 3 4 5 6 7 8 9
0.5
0.1

until 0.10 μM primer (Figure3.5). The optimal primer concentration was then chosen to be 0.10 μM.

Finally, to optimize the amplification condition of inter CO I-CO II region, a series of $MgCl_2$ concentration between 0-5.0 mM at 0.1 μM primer was tested for amplification efficiency. The amplified product of inter CO I-CO II region was initially obtained at 1.5 mM $MgCl_2$ and gradually increased when $MgCl_2$ concentration was elevated. Non-specific products was usually not observed when the $MgCl_2$ concentration was 2.5 mM and less so that, the optimal $MgCl_2$ concentration for amplification of inter CO I-CO II region was 2.5 mM (Figure3.6). Optimal primer concentration was examined as previously decribed. Although no significant different amount of amplified product observed when 0.04-0.18 μM each of primers was included, a concentration of 0.10 μM was selected to ensure their sufficient amount during amplification of the DNA (Figure 3.7).

3.3 PCR-amplified DNA and detection of restriction endonucleases

Total DNA isolated from individuals of 3-5 colonies within each geographic location were used to represent for amplifying and screening of informative restriction endonucleases. As expected, PCR-amplified sRNA gene, lrRNA gene and inter CO I-COII region were about 400, 750 and 1710 bp long, respectively (Figure3.8). Length heteroplasmy was not found in all regions. Four restriction endonucleases consisting of *Eco*R I, *Hin*d III, *Hinf* I and *Dra* I were screened whether they are polymorphic on the amplified DNA fragments.

Figure 3.4 Agarose gel electrophoresis showing an optimization of $MgCl_2$ concentration used for lrRNA gene amplified at a constant primer concentraion (0.1μM).

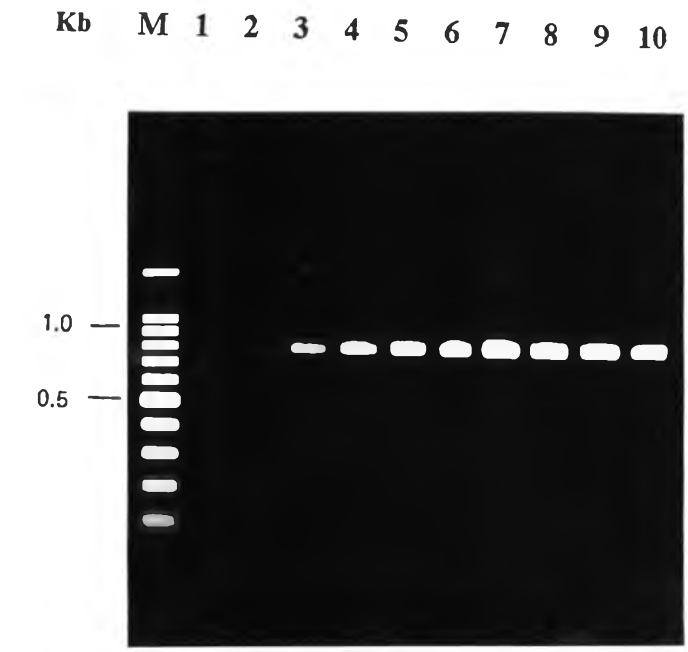
lane M = 100 bp DNA ladder

lane 1- 10 = Amplified product in a seies of 0, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5 and 5.0 mM of $MgCl_2$, respectively.

Figure 3.5 Agarose gel electrophoresis showing an optimization of primer concentration used for lrRNA gene amplified at a constant $MgCl_2$ concentration (3.0 mM).

lane M = 100 bp DNA ladder

lane 1-11 = Amplified product in a series of 0, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18 and 0.20 μM of primers, respectively.


Kb M 1 2 3 4 5 6 7 8 9 10
1.0
0.5

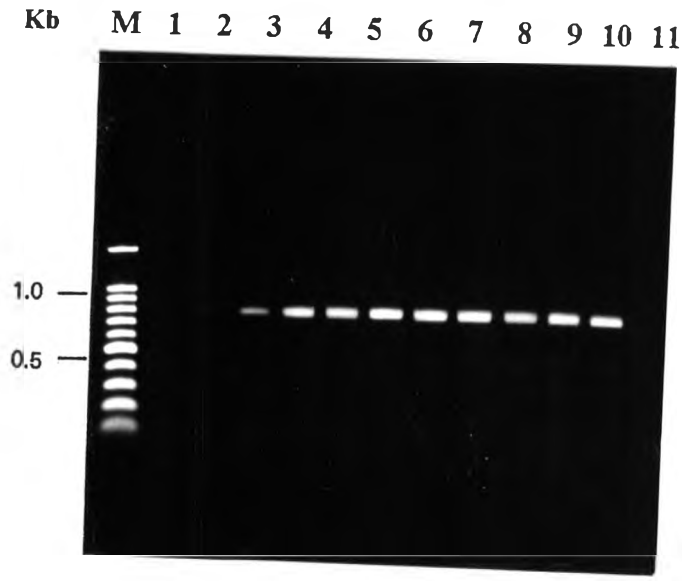

Kb M 1 2 3 4 5 6 7 8 9 10 11
1.0
0.5

Figure 3.6 Agarose gel electrophoresis showing an optimization of $MgCl_2$ concentration used for linter CO I-CO II region amplified at a constant primer concentraion (0.1μM).

lane M1 = λ - *Hin*d III standard DNA marker

lane M2 = 100 bp DNA ladder

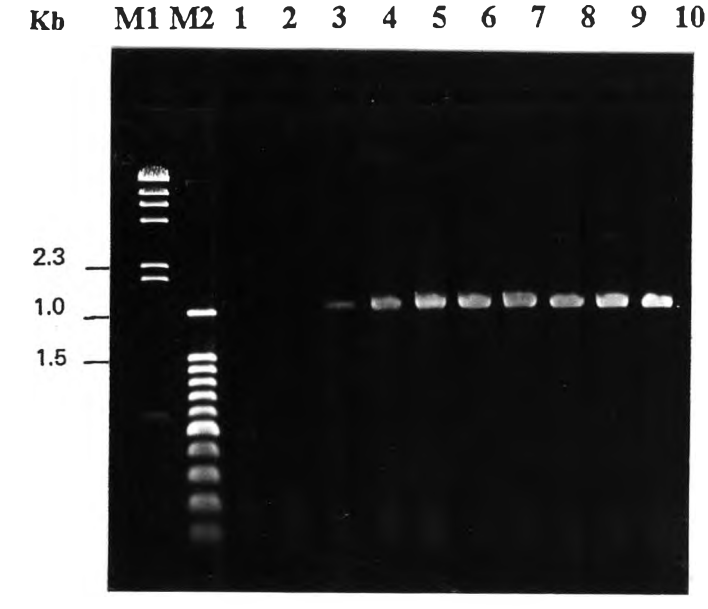
lane 1- 10 = Amplified product in a seies of 0, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5 and 5.0 mM of $MgCl_2$, respectively.

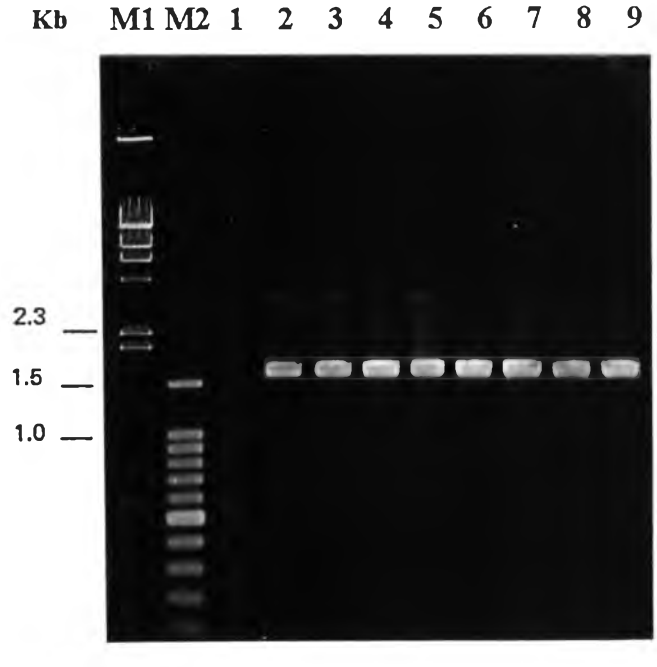
Figure 3.7 Agarose gel electrophoresis showing an optimization of primer concentration used for inter CO I-CO II region amplified at a constant $MgCl_2$ concentration (2.5 mM).

lane M1 = λ - *Hin*d III standard DNA marker

lane M2 = 100 bp DNA ladder

lane 1-9 = Amplified product in a series of 0, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, and 0.16 μM of primers, respectively.


Kb M1 M2 1 2 3 4 5 6 7 8 9 10
2.3
1.0
1.5


Kb M1 M2 1 2 3 4 5 6 7 8 9
2.3
1.5
1.0

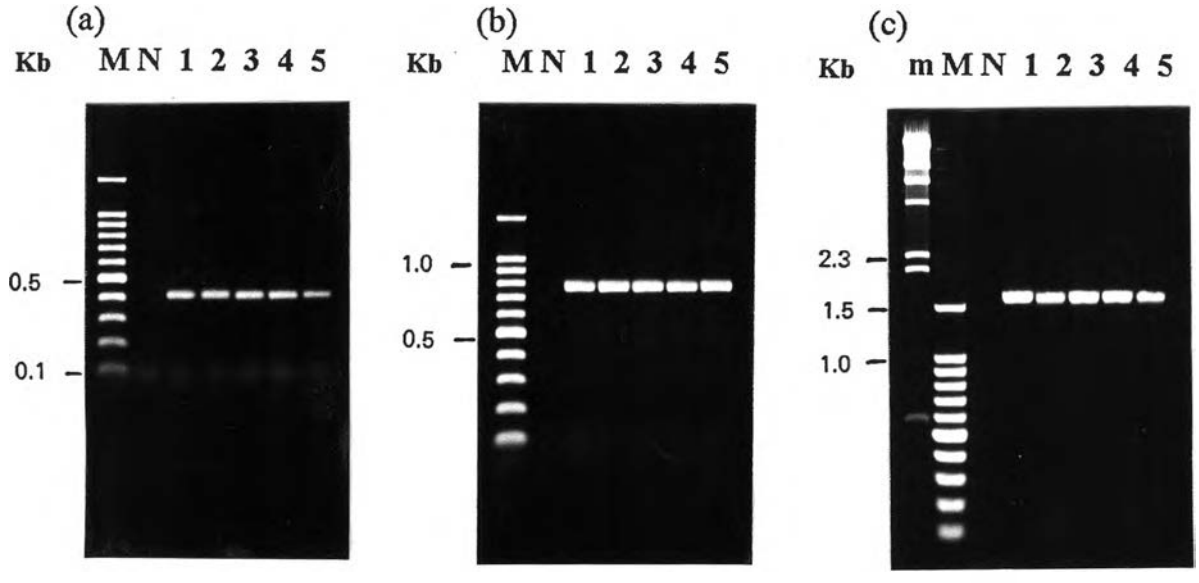


Figure 3.8 Undigested PCR-amplified DNA from each individual was electrophoresed : a. sRNA gene(400bp) b. lrRNA gene (750 bp) and c. inter CO I-CO II region (1710 bp).

lane m = λ - *Hind* III standard DNA marker

lane M = 100 bp DNA ladder

lane N = Negative control

lane 1-5 = An example of PCR-amplifed DNA patterns from 5 *A. cerana* individuals.

Among all restriction endonucleases used to screen sRNA gene, only *Hinf* I and *Dra* I revealed restriction polymorphisms. *Hinf* I exhibited two different haplotypes ; one was uncleaved pattern and the other was a 2 fragment pattern composing of 320 bp and 80 bp. *Dra* I exhibited three different haplotypes. Both *Eco*R I and *Hind* III gave 2 different haplotypes in inter CO I-CO II region consisting of undigested and restricted patterns (1000 bp and 710 bp for *Eco*R I digestion, 930 bp and 780 bp for *Hind* III digestion). Four and eight different haplotypes were resulted from *Hinf* I and *Dra* I digestions, respectively. More importantly, only *Dra* I had restriction sites within an amplified lrRNA gene resulting in 5 different restriction profiles. Accordingly, *Dra* I could be utilised to examine all three regions. The restriction patterns from *Eco*R I, *Hind* III and *Hinf* I digestion of three regions are shown in Figure 3.9.

3.4 Survey of genetic variation in five locations of *A. cerana* in Thailand

Genetic variation of *A. cerana* in Thailand was examined by digestion of three amplified mtDNA regions (sRNA gene, lrRNA gene and inter CO I-CO II region) with *Dra* I. Restriction patterns of all samples are shown in Appendix B. The distributions of each haplotype in each region are shown in Table 3.1. The relationship of all restriction profile digestion of sRNA and lrRNA genes could be explained by either single loss or gain of restriction sites. Interestingly, all haplotypes generated from *Dra* I digestion of CO I-CO II region were still connected if one, two or three point mutation steps were proposed.

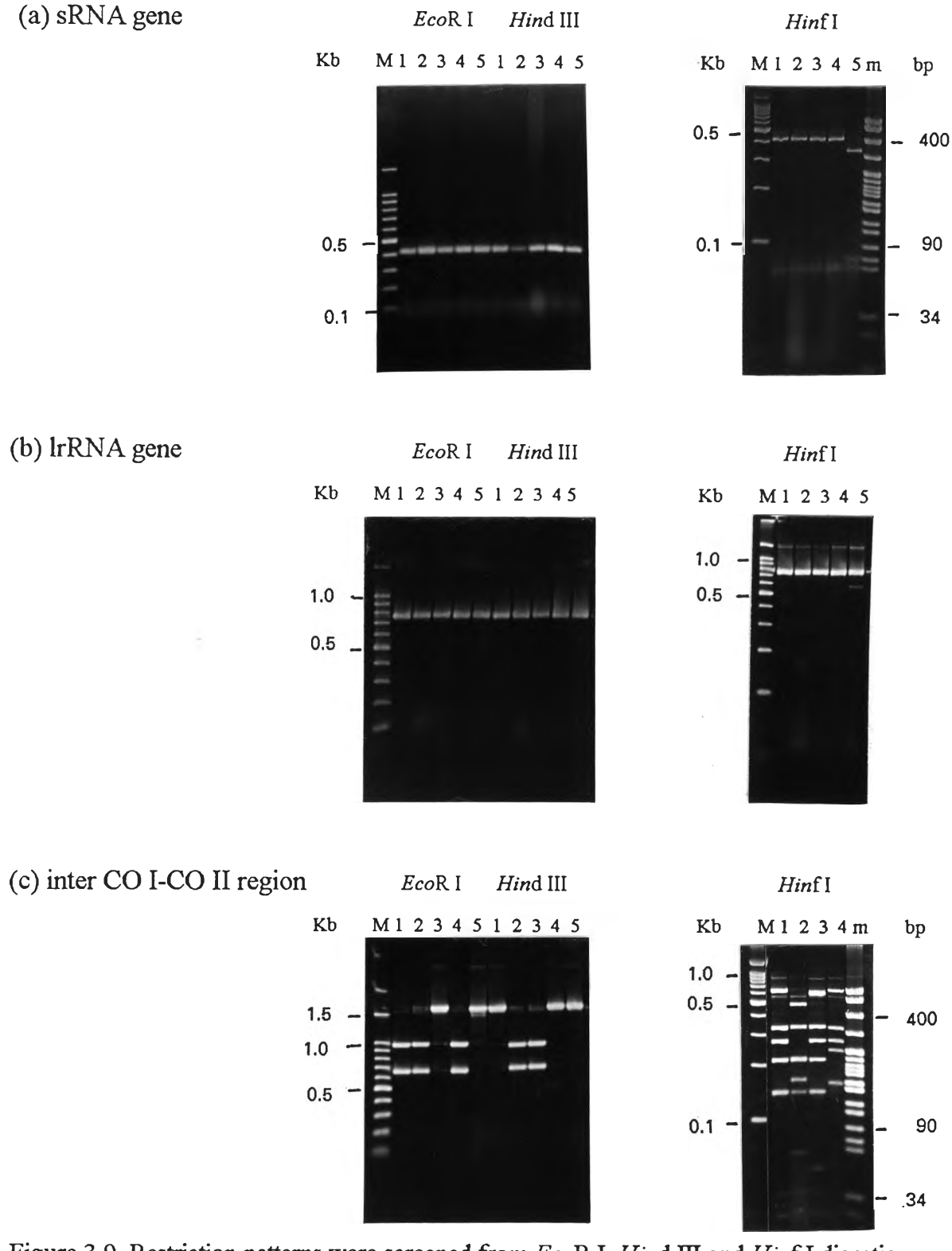


Figure 3.9 Restriction patterns were screened from *Eco*R I, *Hind* III and *Hinf* I digestion of (a) sRNA gene (b) lrRNA gene and (c) inter CO I-CO II region.

lane M = 100 bp DNA ladder

lane m = pBR322-*Msp* I marker

lane 1-5 = An example of restriction patterns from 5 *A. cerana* individuals

Table 3.1 Distributions of single enzyme haplotype observed from digestion of sRNA gene, lrRNA gene and inter CO I-CO II region of five geographic locations *A. cerana* in Thailand.

Geographic location (no. of samples)	Haplotype distribution															
	sRNA gene			lrRNA gene					Inter CO I-CO II region							
	A	B	C	A	B	C	D	E	A	B	C	D	E	F	G	H
North (34)	0.97	0.03	0	0.97	0	0	0.03	0	0.97	0	0	0	0	0	0.03	0
North-East (32)	1.00	0	0	0.94	0	0	0.06	0	0.91	0.03	0	0	0	0	0	0.06
Central (23)	1.00	0	0	0.83	0	0	0.17	0	0.91	0.04	0	0	0	0	0	0.04
South (50)	0	0.96	0.04	0	0.96	0	0	0.04	0	0.84	0.06	0.04	0.04	0.02	0	0
Samui (33)	0	1.00	0	0	0.52	0.48	0	0	0	0.64	0.36	0	0	0	0	0

3.4.1 Analysis of sRNA gene polymorphisms

Three different haplotypes (A, B and C) were obtained from *Dra* I digestion of amplified sRNA gene (Figure 3.10). Haplotype A had discrete bands of 320, 42 and 38 bp, which was commonly distributed in the North, North-East and Central. The interconnection between haplotype A and B can be explained by a loss of the restriction site generating two 160 bp fragments in haplotype B resulted in an appearance of a 320 bp fragment in haplotype A. On the other hand, gaining of restriction site within a 320 bp fragment in haplotype A resulted in an occurance of two 160 bp fragments in haplotype B. Haplotype B was found in the South and Samui Island. Haplotype C was interconnected with haplotype B by a single loss (or gain) of restriction site. A single point mutation step caused a replacement of 42 and 38 bp fragments with a 80 bp band if losing and *vice versa* if gaining. Haplotype C was found in only two specimens in the South. The network showing relationships among three restriction patterns based on minimum number of point mutations is illustrated in Table 3.2 (a) and Figure 3.13 (a).

3.4.2 Analysis of lrRNA gene polymorphisms

Five different haplotypes (A, B, C, D, and E) were observed from *Dra* I digestion of lrRNA gene (Figure 3.11). Discrete bands of 300, 130, 120, 90, 60 and 50 bp fragments were shown in haplotype A. A *Dra* I site gain in a 300 bp fragment of haplotypes A gave 148 and 152 bp fragments in haplotype D. Both of these haplotypes (A and D) were only observed in the North, North-East and Central. On the other hand, haplotype B was commonly found in the South and Samui Island. Only one restriction site

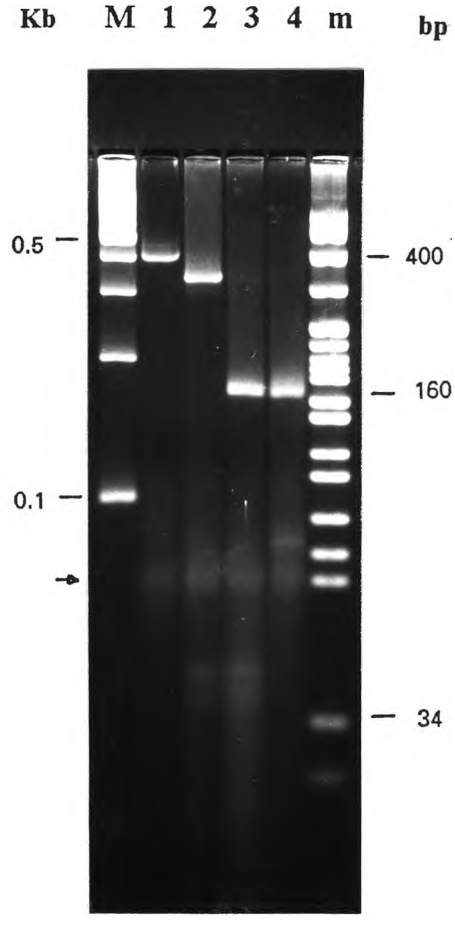


Figure 3.10 Three different haplotypes were observed from *Dra* I digestion of amplified mitochondrial sRNA gene of *A. cerana* in Thailand.

An arrow indicated primer dimer.

lane M = 100 bp DNA ladder

lane m = pBR 322 DNA-*Msp* I marker

lane1-4 = Undigested DNA, Haplotype A, B and C, respectively.

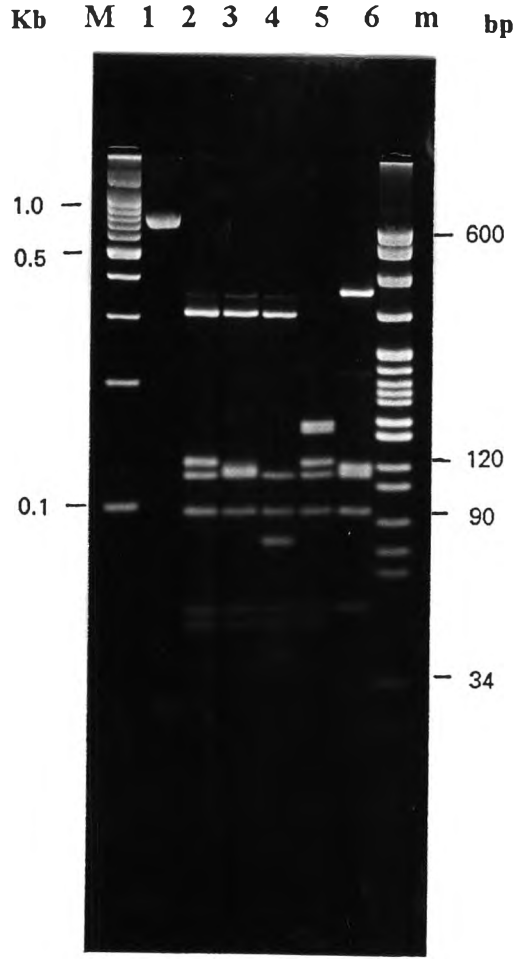


Figure 3.11 Five different haplotypes were observed from *Dra* I digestion of amplified mitochondrial lrRNA gene of *A. cerana* in Thailand.

lane M = 100 bp DNA ladder

lane m = pBR 322 DNA-*Msp* I marker

lane1-6 = Undigested DNA, Haplotype A, B, C, D and E, respectively.

loss that restricted 120 and 10 bp fragments in haplotype B resulting in a 130 bp in haplotype A. Haplotype B and C were interelated by a change of 120 bp fragment in haplotype B to be 80 and 40 bp fragments in haplotype C and *vice versa*. It should be noted that haplotype C was population specific for the Samui Island and found at an equally frequecies of haplotype B. Finally, haplotype E, which was different from haplotype B by only one point mutatuion (losing site generated 300 and 50 bp fragments in haplotype E or gaining site in haplotype B) was found in only two samples in the South. The parsimonious network among haplotypes from digestion of this gene with *Dra* I is summarized in Table 3.2(b) and Figure 3.13(b).

3.4.3 Analysis of inter CO I-CO II region polymorphisms

As can be seen in Figure 3.12, the maximum number of eight single enzyme digestion profiles were derived from *Dra* I digestion of inter CO I-CO II region. Haplotype A was the most common pattern interrelated to its variants; haplotype D, G, H with either two or three point mutation steps whereas the other common haplotype B was directly interconnected with haplotype C, D, E and F. Generally, all but between haplotype C and F were related with more than one point mutation step causing complex relationships among *Dra* I generated restriction patterns. Theoretically, intermediate haplotypes between pairs of patterns related with more than a single mutations should be observed if the sample sizes were large enough. The band patterns and parsimonious network among haplotype are summarized in Table 3.2(c) and Figure 3.13 (c). Distribution of all eight haplotypes was geographically related. The most common, haplotype A was generally distributed in the North,

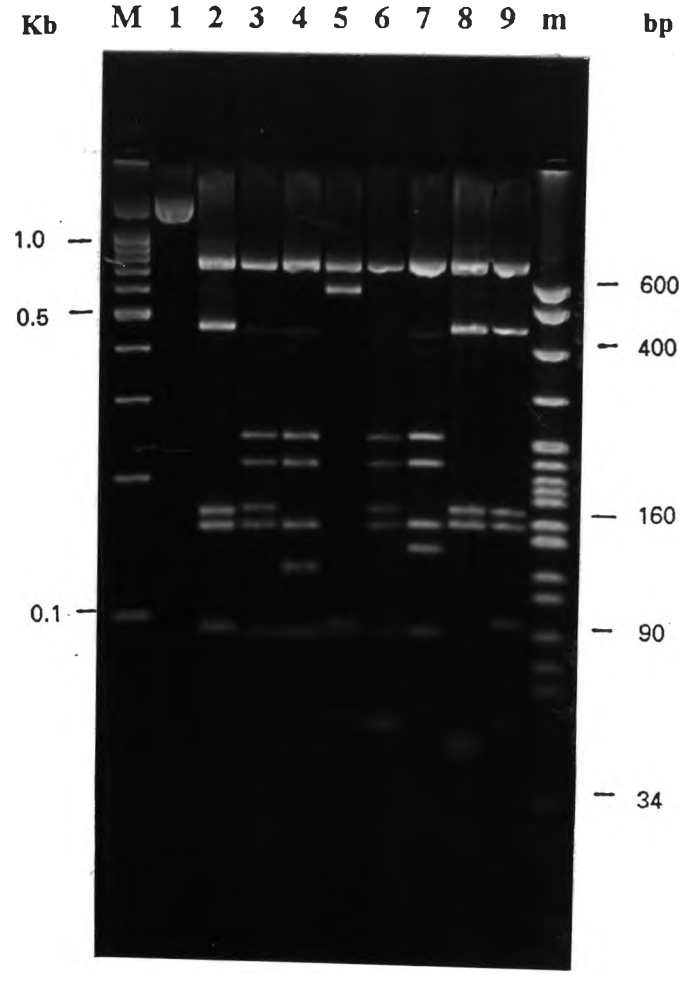


Figure 3.12 Eight different haplotypes were observed from *Dra* I digestion of amplified mitochondrial inter CO I-CO II region of *A. cerana* in Thailand.

lane M = 100 bp DNA ladder

lane m = pBR 322 DNA-*Msp* I marker

lane1-9 = Undigested DNA, Haplotype A, B, C, D, E, F, G and H, respectively.

Table 3.2 Summary of restriction patterns of *Dra* I digested amplified DNA from three different regions in mitochondrial genome of *A. cerana.*

a. sRNA gene

Haplotype	Number of fragments at indicated size in base pairs						Total length
	320	160	160	80	42	38	(bp)
A	1	0	0	0	1	1	400
B	0	1	1	0	1	1	400
C	0	1	1	1	0	0	400

b. lrRNA gene

Haplotype	Number of fragments at indicated size in base pairs													Total length
	350	300	152	148	130	120	120	90	80	60	50	40	7*	(bp)
A	0	1	0	0	1	1	0	1	0	1	1	0	0	750
B	0	1	0	0	0	1	1	1	0	1	1	0	1	747
C	0	1	0	0	0	1	0	1	1	1	1	1	1	747
D	0	0	1	1	1	1	0	1	0	1	1	0	0	750
E	1	0	0	0	0	1	1	1	0	1	0	0	0	740

* this fragment was not observed on the electrophoretic gel but inferred from DNA sequencing data of this gene region.

Table 3.2 (continue)

c. Inter CO I-CO II region

Haplotype	Number of fragments at indicated size in base pairs																							Total length (bp)
	750	620	450	240	210	170	160	140	130	100	90	90	90	85	85	60	50	50	45	45	45	45	40	30	
A	1	0	1	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1710
B	1	0	0	1	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1700
C	1	0	0	1	1	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1700
D	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	1700
E	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1715
F	1	0	0	1	1	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1700
G	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1710
H	1	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1680

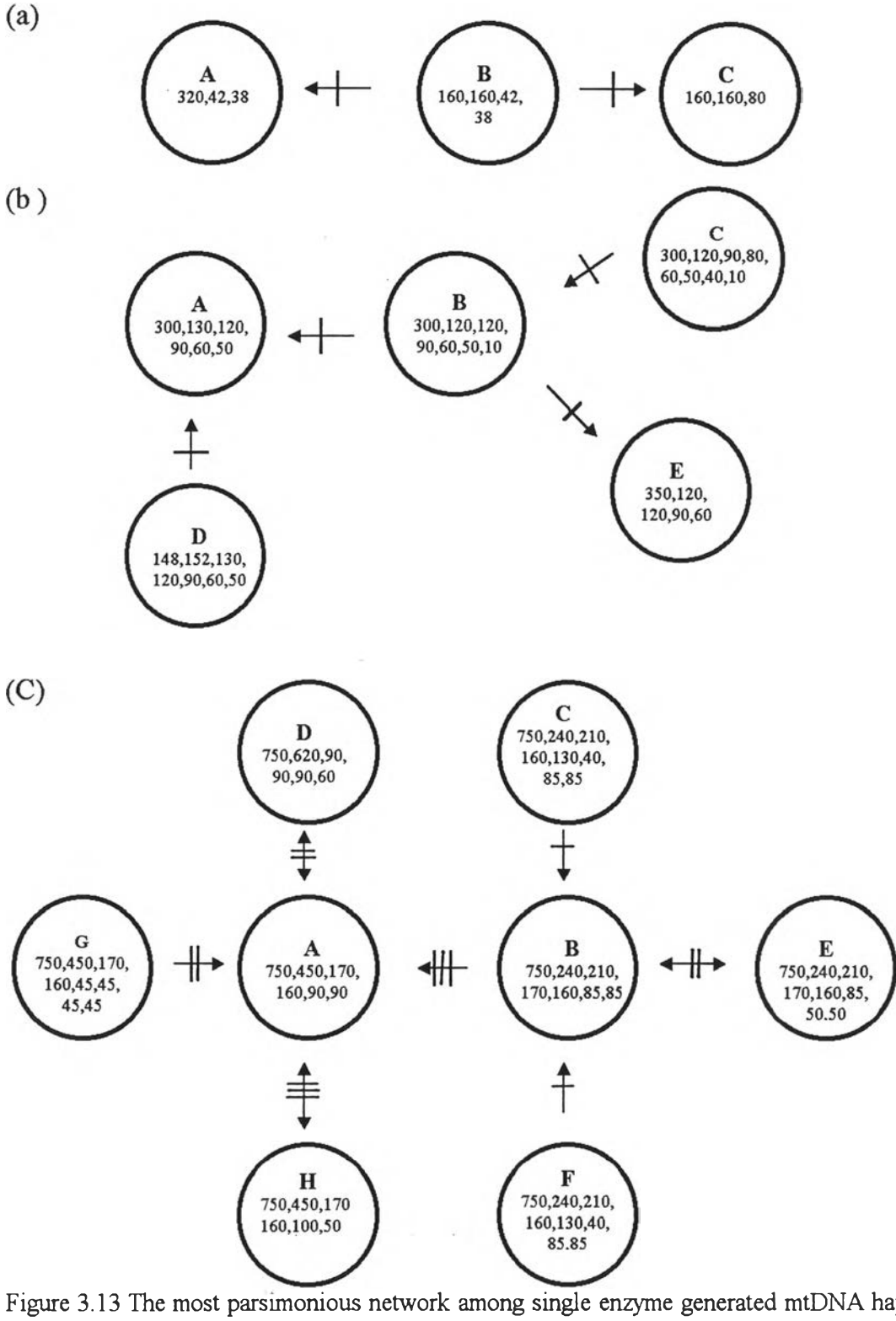


Figure 3.13 The most parsimonious network among single enzyme generated mtDNA haplotypes based on *Dra* I digestion of PCR-amplified DNA of (a)sRNA gene, (b)lrRNA gene and (c)inter CO I-CO II region of *A. cerana* in Thailand. Arrows indicated restriction site losses and not necessarily indicated evolutionary direction. Cross bars indicated the number of point mutations.

North-East and Central while haplotype G was found in single individual from the North. Haplotype H was found in the North and North-East. Haplotype B, C, D, E and F were all distributed in the South in which haplotype B was the most common haplotype of *A. cerana* from the South of Thailand. Only two haplotypes, B and C, were found in the Samui Island individuals.

3.5 Analysis of geographic population structure based on mtDNA-RFLP approach

Thirteen composite haplotypes were generated from *Dra* I digestion of amplified product from *A. cerana* sRNA gene, lrRNA gene and inter CO I-CO II region. Their distribution frequencies among geographically different locals are shown in Table 3.3. The most common composite haplotype, AAA, was found in 77 (45%) out of 172 individuals investigated. This haplotype was only distributed in the North, North-East and Central. The other common composite haplotype, BBB (34 % of investigated individuals) was only observed in the South and Samui Island. Distribution of composite haplotypes within each location is illustrated in Figure 3.14. Composite haplotypes found in the North, North-East and Central were not overlapping from those found in the South and Samui Island. The relationship among all composite haplotypes found in this study are illustrated in Figure 3.15. Five mutation steps were found between the most two common haplotypes (AAA and BBB). Composite haplotype AAB was an intermediate haplotype between AAA and BBB while composite haplotype BBA was a hypothetical ancester betweem BBB and CED. In addition, only composite haplotype BCC and BCB were found in the Samui Island. The genetic distance of composite haplotype is shown in Table 3.4. The values of genetic distance were ranged

Table 3.3 Geographic distribution of 13 composite haplotypes resulted from *Dra* I digestion of sRNA gene, lrRNA gene and inter CO I-CO II region.

Composite haplotype	Geographic distribution frequency (no. of individuals)					Total
	North	North-East	Central	South	Samui	
AAA	0.912(31)	0.844(27)	0.739(17)	0	0	0.448(77)
AAB	0	0.031(1)	0.043(1)	0	0	0.011(2)
AAG	0.029(1)	0	0	0	0	0.006(1)
AAH	0	0.062(2)	0.043(1)	0	0	0.006(1)
ADA	0.029(1)	0.062(2)	0.174(4)	0	0	0.041(7)
BAA	0.029(1)	0	0	0	0	0.006(1)
BBB	0	0	0	0.840(42)	0.515(17)	0.343(59)
BBC	0	0	0	0.060(3)	0	0.017(3)
BBE	0	0	0	0.040(2)	0	0.011(2)
BBF	0	0	0	0.020(1)	0	0.006(1)
BCB	0	0	0	0	0.121(4)	0.023(4)
BCC	0	0	0	0	0.364(12)	0.070(12)
CED	0	0	0	0.040(2)	0	0.011(2)
Total	1.000(34)	1.000(32)	1.000(23)	1.000(50)	1.000(33)	1.000(172)

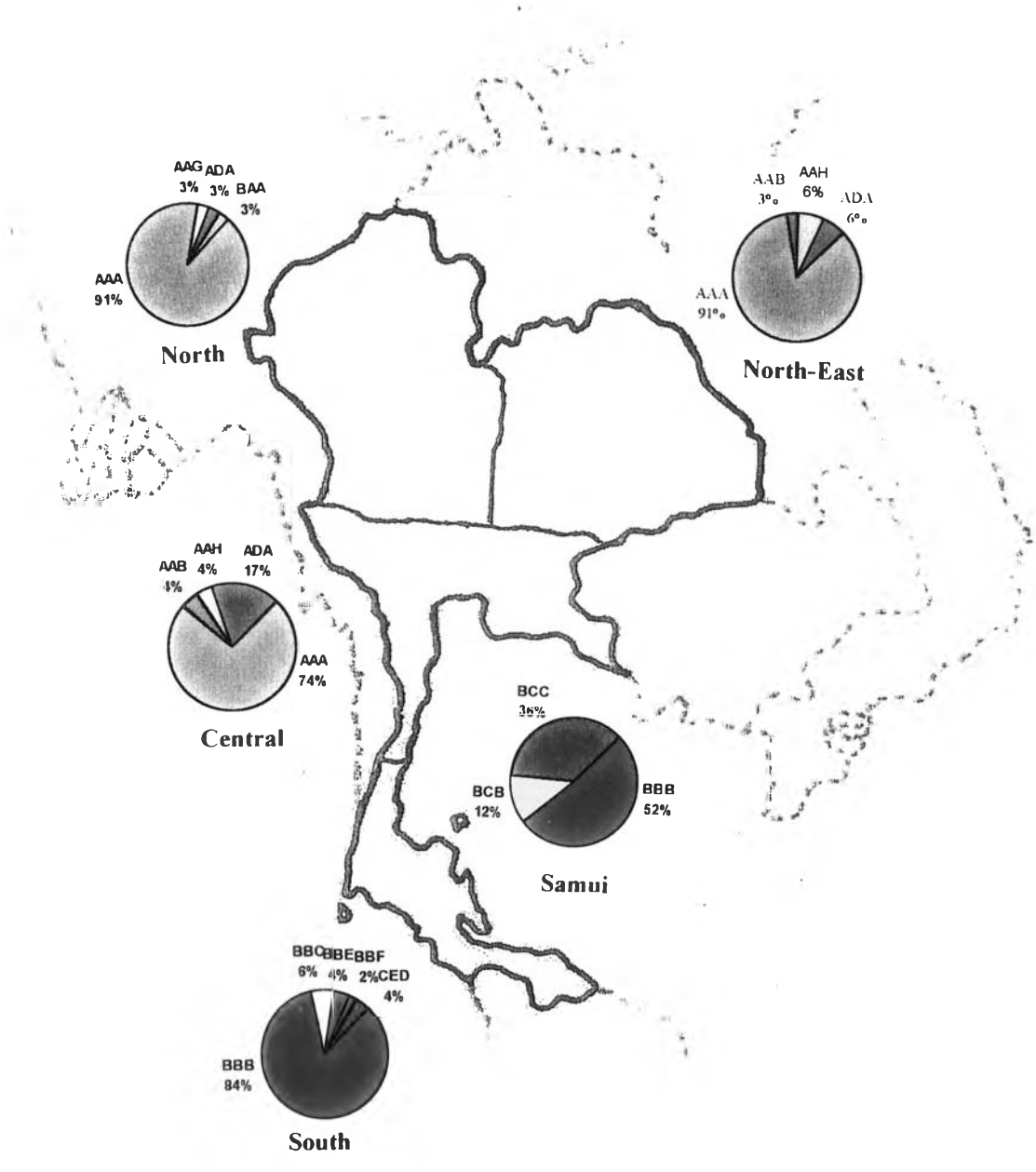


Figure 3.14 Distribution of mtDNA composite haplotypes within each sampling location of *A. cerana* in Thailand.

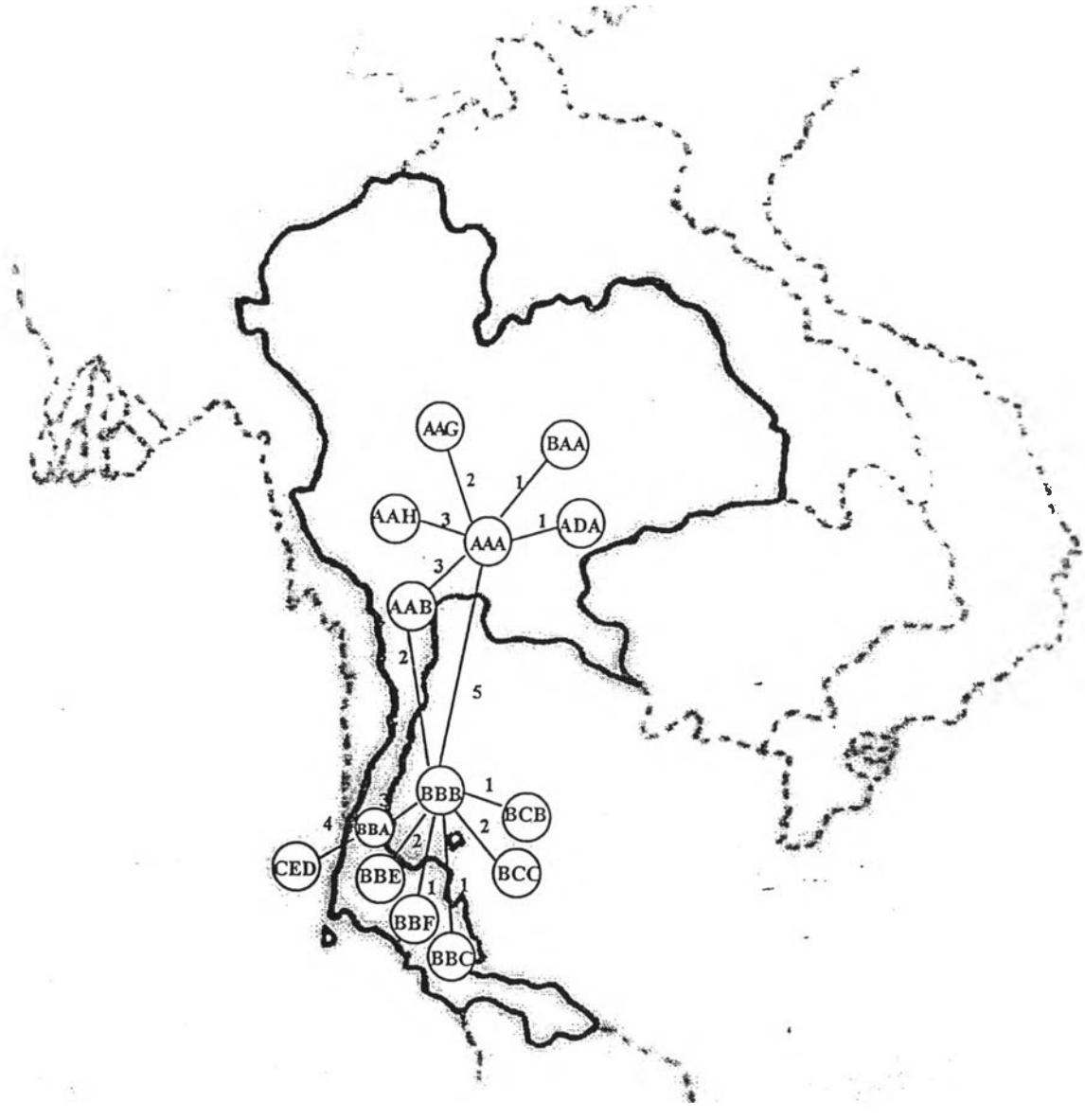


Figure 3.15 Phylogeographic pattern deduced from mtDNA composite haplotypes of *A. cerana* in Thailand. Numbers along connected lines indicated inferred mutation steps. The Northern (North, North-East and Centeral) group was different from the Southern group by five mutation steps.

Table 3.4 Estimated genetic distances among 13 composite haplotypes obtained from *Dra* I digested amplified DNA fragment of sRNA gene, lrRNA gene and inter CO I-CO II region in *A. cerana* mitochondrial genome.

	AAA	BAA	ADA	AAG	AAB	AAH	BBB	CED	BBE	BBC	BBF	BCB	BCC
AAA	-												
BAA	0.0057	-											
ADA	0.0057	0.0117	-										
AAG	0.0117	0.0181	0.0181	-									
AAB	0.0145	0.0214	0.0214	0.0181	-								
AAH	0.0080	0.0145	0.0145	0.0117	0.0145	-							
BBB	0.0289	0.0199	0.0371	0.0324	0.0110	0.0289	-						
CED	0.0523	0.0366	0.0544	0.0825	0.0804	0.0782	0.0488	-					
BBE	0.0307	0.0216	0.0388	0.0342	0.0169	0.0250	0.0047	0.0507	-				
BBC	0.0371	0.0267	0.0461	0.0406	0.0169	0.0371	0.0047	0.0507	0.0097	-			
BBF	0.0371	0.0267	0.0461	0.0406	0.0169	0.0371	0.0047	0.0507	0.0097	0.0062	-		
BCB	0.0307	0.0216	0.0388	0.0341	0.0126	0.0307	0.0047	0.0604	0.0097	0.0097	0.0097	-	
BCC	0.0388	0.0284	0.0478	0.0422	0.0186	0.0388	0.0097	0.0623	0.0149	0.0045	0.0112	0.0045	-

Highest genetic distance : AAG-CED = 0.0825

Lowest genetic distance : BBC-BCC & BCB-BCC = 0.0045

from 0.0045 to 0.0825. A pairwise comparison of AAG and CED yielded the highest genetic distance whereas pairwise comparisons of BCC versus either BBC or BCB gave the lowest genetic distance. Distance values were then subjected to phylogenetic reconstruction using UPGMA method. As can be seen in Figure 3.16, three major groups (A, B and C) were observed. Group A was composed of BBC, BCC, BCB, BBB, BBF, BBE, and AAB. All except AAB specifically found in the South and Samui Island. On the other hand, Group B was composed of AAA, BAA, ADA, AAH, and AAG only found in the North, North-East and Central. Finally, group C was composed of only one composite haplotype, CED found in only two samples in the south. These two individuals were suspected to be the other species, *A. koschenikovi*. Nevertheless, a further proof using already taxonomically classified specimens will provide a strong evidence to support whether individuals carrying haplotype CED are actually *A. cerana*.

Haplotype and nucleotide diversity within population for five geographic locations is illustrated by Table 3.5. The samples from Samui Island showed the highest haplotype diversity (0.5967), whereas those of Central, South, North-East and North were 0.4290, 0.2901, 0.2837 and 0.1686, respectively. The average haplotype diversity within the Thai *A. cerana* was 0.3536. On the other hand, nucleotide diversity showed that the highest nucleotide diversity (0.49%) was observed in the South following by 0.47%, 0.36%, 0.25% and 0.13% observed in the Samui Island, Central, North-East and North, respectively. The average nucleotide diversity was 0.34%.

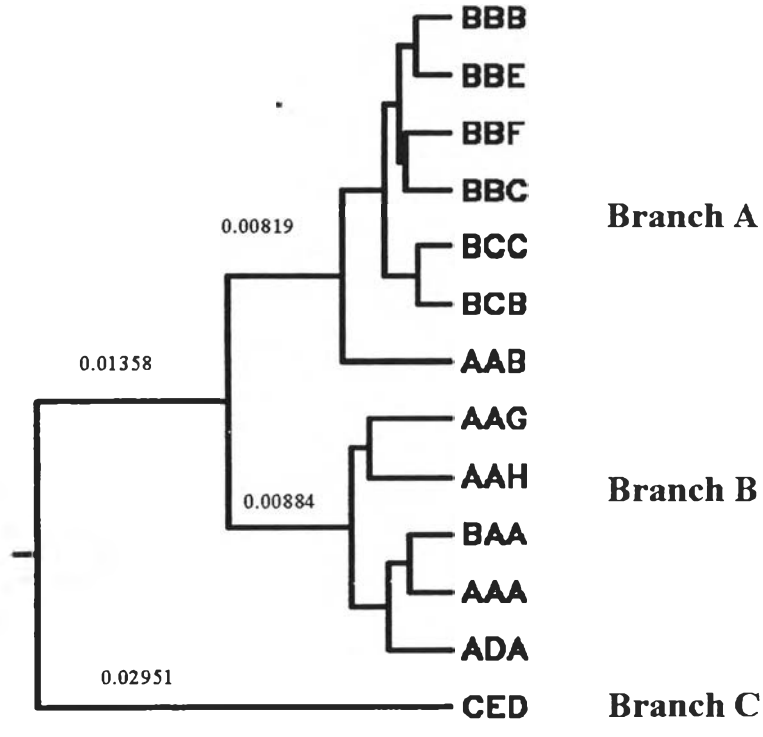


Figure 3.16 A UPGMA dendrogram showing the relationship among thirteen composite haplotypes based on *Dra* I digestion of PCR-amplified DNA of sRNA gene, lrRNA gene and inter CO I-CO II region of *A. cerana* in Thailand.

Table 3.5 Haplotype and nucleotide diversity within population for five geographic locations of *A. cerana* in Thailand.

Population	Haplotype diversity (h +/-SE)	Nucleotide diversity
North	0.1686 +/-0.06075	0.00136
North-East	0.2837 +/-0.07163	0.00255
Central	0.4290 +/-0.07996	0.00367
South	0.2901 +/-0.05836	0.00490
Samui Island	0.5967 +/-0.03320	0.00476
Average	0.3536 +/-0.00540	0.00345

As can be seen in Table 3.6, nucleotide diversity between populations was much higher than that within population indicating population differentiation within this species. Pairwise comparisons of both the Samui and the South versus the remaining population were relatively high (3.062% - 3.340%). Relatively low nucleotide diversity were observed from all pairwise comparisons among North, North-East and Central population. An average nucleotide diversity between populations was 2.056%.

Nucleotide divergence of *A. cerana* was quite high (1.711 % in average). The highest nucleotide divergence (2.976%) was found from a comparison between the North and Samui Island. Phylogenetic reconstruction based on nucleotide divergence is shown by Figure 3.17. Major genetic break between the Northern * (North, North-East and Central) and the Southern* (South and Samui Island) parts were clearly observed. The estimated percent sequence divergence between these two groups was 1.410 %. Moreover, A consensus tree was reconstructed on a distant approach using bootstrapping data (2000 times). As can be seen in Figure 3.18, all bootstrapping values were higher than 1000 times reflecting a statistical support for the tree topology.

* Thereafter, "the Northern or Northern population" is used to refer to the North, North-East and Central *A. cerana* whereas "the Southern or Southern population" is used to refer to the South and Samui Island *A. cerana*.

Table 3.6 Nucleotide diversity (above diagonal) and divergence (below diagonal) between populations for five geographic locations of *A. cerana* in Thailand.

	North	North-East	Central	South	Samui island
North	-	0.001957	0.002584	0.030671	0.032815
North-East	0.000000	-	0.003065	0.030623	0.032685
Central	0.000065	0.000000	-	0.031300	0.033397
South	0.027541	0.026901	0.027016	-	0.006473
Samui Island	0.029755	0.029033	0.029183	0.001647	-

Average Nucleotide diversity between populations = 0.020557 +/-0.0000217

Average Nucleotide divergence = 0.017110 +/-0.0000207

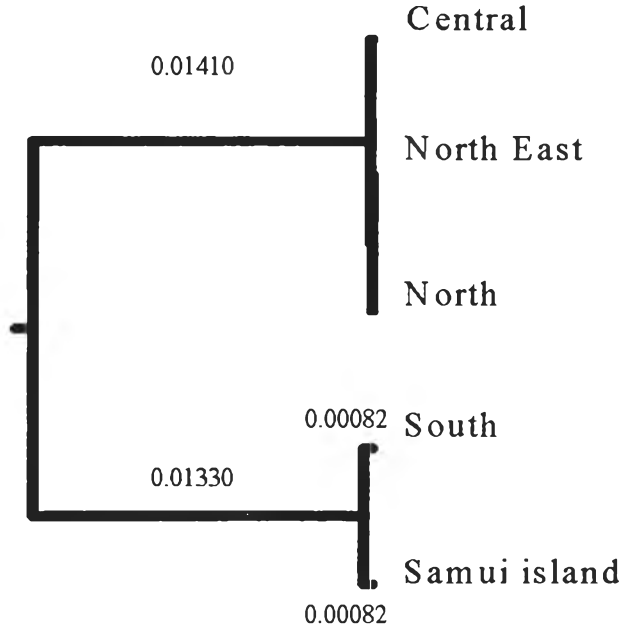


Figure 3.17 A UPGMA dendrogram showing the relationship among 5 geographic locations of *A. cerana* in Thailand.

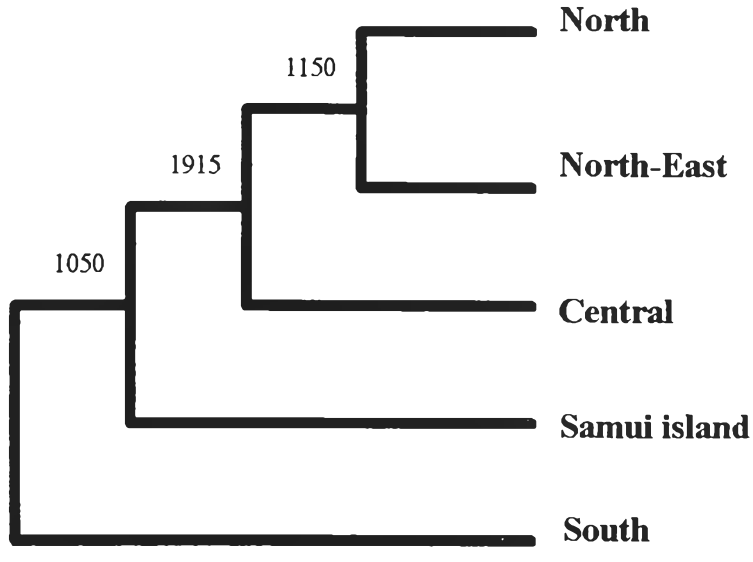


Figure 3.18 A consensus tree of five *A. cerana* locations in Thailand. Values at the nodes represent the number of times that clustering occured out of 2000 bootstrapping data.

Finally, a pseudo chi-square (χ^2) test was performed using Monte carlo simulation for geographic heterogeneity analysis. As can be seen in Table 3.7, significant differences in composite haplotype distribution frequencies were observed between either the South or Samui island populations with the rest ($p = 0.0000$). Accordingly, North, North-East and Central were pooled to one group and reanalysed. The result showed highly significant different heterogeneity between the pooled samples versus both South and the Samui Island.

3.8 DNA sequencing

According to restriction pattern analysis, lrRNA gene was chosen for further study on DNA sequencing. The reason for this was that this DNA region provided enough information on genetic population structure as the same as that from *Dra* I digested CO I-CO II region but the number of composite haplotypes resulted from the former were less than that of the latter. Two *A. cerana* samples showing the same RFLP genotype (haplotype A, B, C, D and E) were chosen as representatives of a population. After amplification, DNA template was purified from agarose gel. The recovery yield was obtained about 80 % (Figure 3.19). Five primers were used for sequencing of lrRNA gene. Two of those were external primers originally used to amplify lrRNA gene, whereas others were internal primer desinged by Oligo programme (Figure 3.20). Two samples of each haplotype showed all identical nucleotides. Approximately, one hundred and fifty nucleotides were obtained from single sequencing of each pair of primer (Figure 3.21). A total of six hundred and fifty - four nucleotides were read from haplotype A and D while six hundred and fifty - three nucleotides were read from

Table 3.7 (a) Geographic heterogeneity analysis in distribution frequency of composite haplotype among 5 *A. cerana* locations based on *Dra* I digestion of sRNA gene, lrRNA gene and inter CO I-CO II region.

	North	North-East	Central	South	Samui Island
North	-				
North-East	$p=0.7510^{ns}$	-			
Central	$p=0.0572^{ns}$	$p=0.3657^{ns}$	-		
South	$p=0.0000^{***}$	$p=0.0000^{***}$	$p=0.0000^{***}$	-	
Samui Island	$p=0.0000^{***}$	$p=0.0000^{***}$	$p=0.0000^{***}$	$p=0.0000^{***}$	-

ns = not significant * = $p< 0.05$, ** = $p< 0.01$, *** = $p< 0.001$

(b) Geographic heterogeneity analysis in distribution frequency of composite haplotype between pooled samples (North, North-East and Central) versus South and Samui Island based on *Dra* I digestion of sRNA gene, lrRNA gene and inter CO I-CO II region.

	North/North-East/Central	South	Samui Island
North/North-East/Central	-		
South	$p=0.0000^{***}$	-	
Samui Island	$p=0.0000^{***}$	$p=0.0000^{***}$	-

ns = not significant * = $p< 0.05$, ** = $p< 0.01$, *** = $p< 0.001$

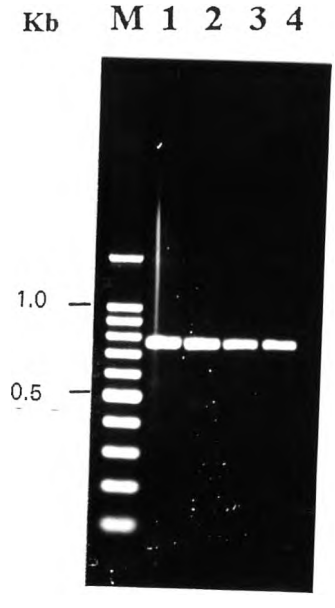


Figure 3.19 Amplified lrRNA gene was electrophoresed through a 1.2 % agarose gel and recovered for sequencing using Geneclean II Kit (Bio101).

lane M = 100 bp DNA ladder

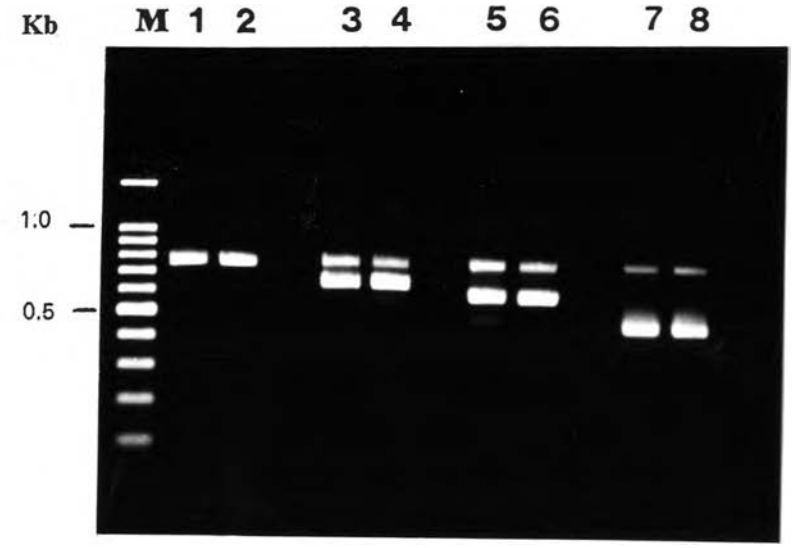
lane 1-2 = PCR-amplified DNA

lane 3-4 = Eluted DNA

Figure 3.20 (a) Diagram of five primers used for DNA sequencing.

5′ TTTTGTACCTTTTGTATCAGGGTTG.....5′ CATGGAGCAGATCGACCTAA..... 5′ TTTTTATTACTAATTAATCATTATCAC.....AGCTTACTGTAAATTAGAGA5′.....CCCTGCTATTCTGGGATATC 5′

primer L1 primer L3 primer L4 primer L5 primer L2



(b) Agarose gel electrophoresis illustrating of PCR-amplified lrRNA gene portion using designed primers (primer L3, L4 and L5).

lane M = 100 bp DNA ladder

lane 1-2 = PCR-amplified lrRNA gene using primers L1-L2 lane 5-6 = PCR-amplified DNA using primers L2-L3

lane 3-4 = PCR-amplified DNA using primers L1-L5 lane 7-8 = PCR-amplified DNA using primers L2-L4

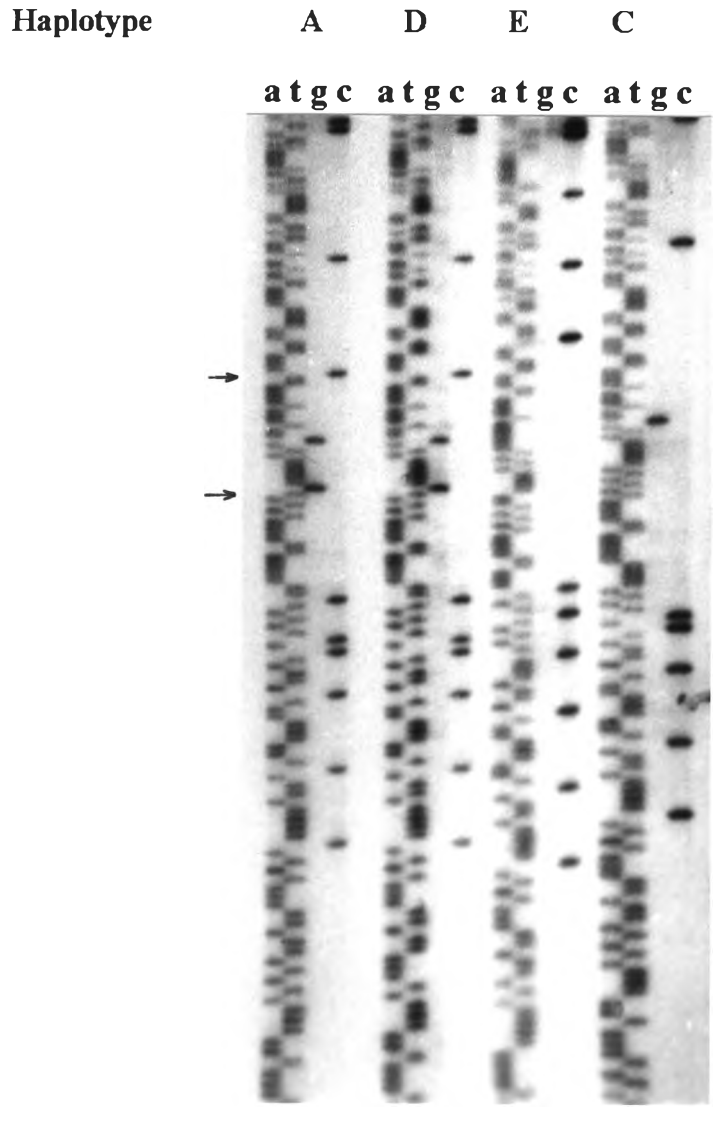


Figure 3.21 An autoradiography of partial lrRNA gene sequence derived from primer L3. The arrow indicated mutation point such as between haplotype A and D versus C.

haplotype B and C. The smallest number of nucleotides (651 readable bases) were found in haplotype E. The average base composition for the lrRNA gene in *A. cerana* was 43.03% A, 41.44% T, 5.61% G and 9.92% C. All sequences were then aligned in Figure 3.22. The homology between pairs of sequences is shown in Table 3.8. A comparison of lrRNA gene sequences of individuals carrying haplotype A and D was 99.85 % homology whereas the lowest percent homology (the most diverse sequence) were observed between individuals having haplotype E verus A and E versus D. Seventy-one mutation steps were found among all alinged sequences. These were composed of fifty-seven point mutations and fourteen gap of either deletions or insertions. Nineteen of point mutations were transitions while the remaining point mutations were transversions. Therefore, transversional mutations in lrRNA gene of *A. cerana* were twice higher than transitional mutations.

To verify that the PCR product obtained were the real lrRNA gene in mitochondrial genome. DNA sequences from an *A. cerana* carrying the most common haplotype A were compared to those of previously studied taxa using BLAST search in web site (http://www.genome. ad.jp/SIT/BLAST.html ; see Appendix C). The result indicated the highest homology of lrRNA gene sequences obtained in the present study was those from *A. mellifera*. The probability to be indentical by change alone between the two sequences was 5.5×10^{-143}.

```
Haplotype A     ATAGAGACAGTTGTTATTTCATCAATTCATTCATTCAATTCTTCAATTAA   50
Haplotype B     ATAGAGACAGTTGTTATTTCATCAATTCATTCATTCAATTCTTCAATTAA
Haplotype C     ATAGAGACAGTTGTTATTTCATCAATTCATTCATTCAATTCTTCAATTAA
Haplotype D     ATAGAGACAGTTGTTATTTCATCAATTCATTCATTCAATTCTTCAATTAA
Haplotype E     ATAGAGACAGTTACTATTTCATCAATTCGTTCATTCAATTCTTCAATTAA
                ************  ************** ********************

Haplotype A     AAGACAATTTATTATGCTACCTTTGTACAGTCAATATACTGCAGCTATTT  100
Haplotype B     AAGACAATTTATTATGCTACCTTTGTACAGTCAATATACTGCAGCTATTT
Haplotype C     AAGACAATTTATTATGCTACCTTTGTACAGTCAATATACTGCAGCTATTT
Haplotype D     AAGACAATTTATTATGCTACCTTTGTACAGTCAATATACTGCAGCTATTT
Haplotype E     AAGACAATTTATTATGCTACCTTTGTACAGTCAATATACTGCAGCTATTT
                **************************************************

Haplotype A     AAATTTATTTCATGGAGCAGATCGACCTAAAATTATACTCAATAGGCCAT  150
Haplotype B     AAATTTATTTCATGGAGCAGATCGACCTAAAATTATATTCAATAGGCCAT
Haplotype C     AAATTTATTTCATGGAGCAGATCGACCTAAAATTATATTCAATAGGCCAT
Haplotype D     AAATTTATTTCATGGAGCAGATCGACCTAAAATTATACTCAATAGGCCAT
Haplotype E     AAATTTATTTCATGGAGCAGATCGACCTAAAATTAAA-TCAATAGGCCAT
                ********************************** * ************

Haplotype A     GTTTTTGTTAAACAGGTGAATAATCAATTTTGCCGAGTTCCTTTAAATTA  200
Haplotype B     GTTTTTGTTAAACAGGTGAATAATCAATTTTGCCGAGTTCCTTTAAATTA
Haplotype C     GTTTTTGTTAAACAGGTGAATAATCAATTTTGCCGAGTTCCTTTAAATTA
Haplotype D     GTTTTTGTTAAACAGGTGAATAATCAATTTTGCCGAGTTCCTTTAAATTA
Haplotype E     GTTTTTGATAAACAGGTGAATAATTTATTTTGCCGAGTTCCTTTAAATTA
                ******* *************** ************************

Haplotype A     TATATATATAAATAATTTATATATTATTAAATATACTTTTATTACTAATT  250
Haplotype B     TATATATATAAATAATTTATATATTATTAAATATACTTTTATTACTAATT
Haplotype C     TATATATATAAATAATTTATATATTATTAAATATACTTTTATTACTAATT
Haplotype D     TATATATATAAATAATTTATATATTATTAAATATACTTTTATTACTAATT
Haplotype E     TATATAAATAAA--ATTT-TATATTATTAAATATACTTTTATTACTAATT
                ****** *****  **** ********************************

Haplotype A     TAATCATTATCACTATATCTTAAAAATTAAAATATATGTTTTTATAGAAT  300
Haplotype B     TAATCATTATCACTATATTTCAAAAATTAAAATATATATTTTTATAGAAT
Haplotype C     TAATCATTATCACTATATTTTAAAAATTAAAATATATATTTTTATAGAAT
Haplotype D     TAATCATTATCACTATATCTTAAAAATTAAAATATATGTTTTTATAGAAT
Haplotype E     TAATCATTATTTCTATATCTATCAAATTAAAATATATAATTTTATATAAA
                ********** ****** *  ************** ******* **

Haplotype A     AAATAAAATTCAAAATTTAAATTTTTAAAAATTAATAA--CTAAATTATT  350
Haplotype B     AAATAAAATTTAAAATTTAAATTTTTTAAA-TTAATAA--CTAAATTATT
Haplotype C     AAATAAAATTTAAAATTTAAATTTTTTAAA-TTAATAA--CTAAATTATT
Haplotype D     AAATAAAATTCAAAATTTAAATTTTTAAAAATTAATAA--CTAAATTATT
Haplotype E     AAATAAAATT--AAATTTAAATCTTTAAAATTTATTAATACTAAATTATT
                ********** ********** *** *** *** ***  **********

Haplotype A     AAATTTTTTA-TATTAATAAAAAATATTAACTTCATAATATTATAAATAA  400
Haplotype B     AAATTTTTTA-TATTAATAAAAAATATTAACTTCATAATATTATAAATAA
Haplotype C     AAATTTTTTA-TATTAATAAAAAATATTAACTTCATAATATTATAAATAA
Haplotype D     AAATTTTTTA-TATTAATAAAAAATATTAACTTCATAATATTATAAATAA
Haplotype E     AAATTTTTAAGCATTAAAAAAAAATATTAATTTCATAATATTATAATCTA
                ******** *  ***** ************ ***************   *
```

Figure 3.22 Alingment of lrRNA gene nucleotide sequences of *A. cerana* in Thailand. Asterisk indicated the same nucleotides among five haplotypes.

```
Haplotype A     AATCAAAAATTTTTATAAATAAATTTATAGTTTATCCCATAAATTTTTAA  450
Haplotype B     AATCAAAAATTTTTATAAATAAATTTATAGTTTATCCCATAAATTTTTAA
Haplotype C     AATCAAAAATTTTTATAAATAAATTTATAGTTTATCCCATAAATTTTTAA
Haplotype D     AATCAAAAATTTTTATAAATAAATTTATAGTTTATCCCATAAATTTTTAA
Haplotype E     TAAATTAAATTTTTATAAATAAATTTATAGGTTATCCCATAAATTTTTAA
                 *    ********************** ******************

Haplotype A     TATAAAAATTAATACTATAAAT--AAATTTTAAGGTATTAAAAATTTTAT  500
Haplotype B     TATAAAAATTAATACTATAAAT--AAATTTTAATGTATTAAAAATTTTAT
Haplotype C     TATAAAAATTAATACTATAAAT--AAATTTTAATGTATTAAAAATTTTAT
Haplotype D     TATAAAAATTAATACTATAAAT--AAATTTTAAAGTATTAAAAATTTTAT
Haplotype E     TATAAAAATTAATAA-ATAAATTAAAATTTTAATTTATTAAAATTTTTAT
                ************** ******  *********  ******** ******

Haplotype A     ATCTAAATTAAATTTATTTCTAAAAAAACTAGATATCAATAACTTCGAAT  550
Haplotype B     ATCTAAATTAAATTTATTTCTAAAAAAACTAGATATCAATAACTTCGAAT
Haplotype C     ATCTAAATTAAATTTATTTCTAAAAAAACTAGATATCAATAACTTCGAAT
Haplotype D     ATCTAAATTAAATTTATTTCTAAAAAAACTAGATATCAATAACTTCGAAT
Haplotype E     ATCTAAATTAAATTTATTTCTAAAAAAACTAGATATCAATACCTTCGAAT
                ****************************************** ********

Haplotype A     GACATTTAATCTCTAAATTTATATTTATAATTTTATTGCAACAAAAAAAA  600
Haplotype B     GACATTTAATCTCTAAATTTATATTTATAATTTTTATGCAACAAAAAAAA
Haplotype C     GACATTTAATCTCTAAATTTATATTTATAATTTTTATGCAACAAAAAAAA
Haplotype D     GACATTTAATCTCTAAATTTATATTTATAATTTTATTGCAACAAAAAAAA
Haplotype E     GACATTTAATCTCTAAATTTATATTTATAATTTTAATGCCACAAAAAAAA
                ************************************  *** **********

Haplotype A     TATTACAAATTTAGCTCACTTATTTTCGAGATATTTAAATTTATTAAATA  650
Haplotype B     TATTATAAACTTAGCTCACTTATTTTCGAGATATTTAAATTTATTAAATA
Haplotype C     TATTATAAACTTAGCTCACTTATTTTCGAGATATTTAAATTTATTAAATA
Haplotype D     TATTACAAATTTAGCTCACTTATTTTCGAGATATTTAAATTTATTAAATA
Haplotype E     TTTTATAAACTTAGCTCCCCTATTTTCGAGATATTTTAATAT-TAAAACA
                * *** *** ******* * ***************** *** * * *** *

Haplotype A     AATTTTAAT  659
Haplotype B     AATTTTAAT
Haplotype C     AATTTTAAT
Haplotype D     AATTTTAAT
Haplotype E     TATTTTAAT
                 ********
```

Figure 3.21 continue.

Table 3.8 Percentage of homology among of lrRNA gene sequences

	Haplotype A	Haplotype B	Haplotype C	Haplotype D	Haplotype E
Haplotype A	-				
Haplotype B	97.3966	-			
Haplotype C	97.7029	99.8469	-		
Haplotype D	99.8471	97.3966	97.7029	-	
Haplotype E	84.1782	84.3318	84.3318	84.1782	-

3.9 Data analysis of DNA sequencing

Genetic distance between sequences was calculated using Kimura's two parameter approach. The range of genetic distance was between 0.0015 - 0.0817 (Table 3.9). These values were used to construct phylogenetic tree using UPGMA (Figure 3.23). Three distinctive groups (A, B and C) were clearly seperated. Group A was composed of haplotype A and D only found in the North, North-East and Central. Group B was composed of haplotype B and C specifically found in the South and Samui Island. Group C was composed of only one haplotype (haplotype E). The estimated percent sequence divergence between group A and B was 0.74 %, and between group C with either A and B was 3.96 %. These isolated groups were in accord with the result from PCR-RFLP. Original data was bootstapped for 1000 times and subjected to phylogenetic reconstruction using the most parsimoniuos approach. The tree topology obtained was generally identical to that from UPGMA method (Figure 3.24).

Table 3.9 Estimated genetic distances among haplotypes derived from lrRNA gene sequences of *A. cerana* in Thailand.

	Haplotype A	Haplotype B	Haplotype C	Haplotype D	Haplotype E
Haplotype A	-				
Haplotype B	0.0170	-			
Haplotype C	0.0155	0.0015	-		
Haplotype D	0.0015	0.0170	0.0155	-	
Haplotype E	0.0817	0.0767	0.0767	0.0817	-

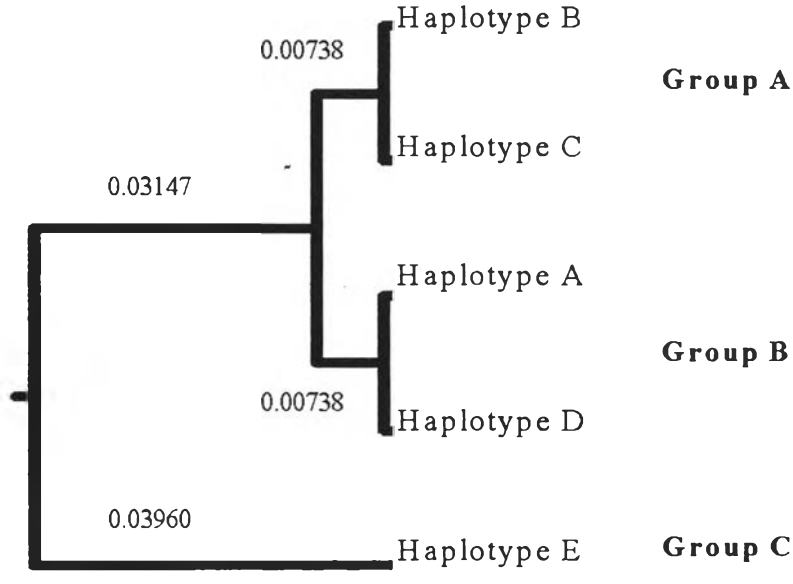


Figure 3.23 A UPGMA dendrogram showing relationship among 5 haplotypes derived from lrRNA gene sequence of *A. cerana* in Thailand.

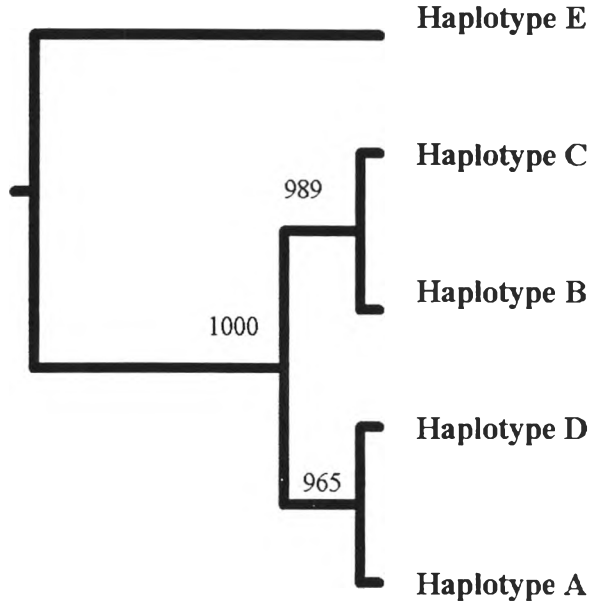


Figure 3.24 The most parsimonious tree showing relationship among 5 haplotypes derived from lrRNA gene sequences of *A. cerana* in Thailand. The upper number indicated number of time the node was occured out of 1000 trees.