CHAPTER V

THE PARTIAL MITOCHONDRIAL GENOME SEQUENCE OF THE STINGLESS BEE (*Trigona pagdeni* Schwarz)

INTRODUCTION

The mitochondrial DNA (mtDNA) is a circular, double-stranded DNA molecule and transmitted maternally. Generally, animal mtDNA exhibits a high evolution rate compared to that of the nuclear DNA and very conserved gene order comprising two rRNA genes, 22 tRNA genes, 13 protein-coding genes, as well as a non-coding control region (D-loop) (Wolstenholme, 1992). MtDNAs have been widely employed to resolve population structure, phylogeography and phylogenetic relationships at various taxonomic levels (Avise, 2000).

In the past years, mtDNA has been extensively employed in *A. mellifera* to investigate natural range origin, and genetic polymorphisms based on PCR-RFLP method (Moritz *et al.*, 1986; Smith and Brown, 1988; 1990; Smith *et al.*, 1989 and Hall and Smith, 1991). In addition, genetic diversity and population subdivision of *Apis cerana* in Thailand were also determined relied on the mtDNA by using PCR-RFLP (Sihanunthavong *et al.*, 1999; Sittipraneed *et al.*, 2001 and Songram *et al.*, 2006).

In stingless bees, *Melipona* mitochondrial genome has been partially sequenced with the molecular size of 14,422 bp (Silvestre *et al.*, 2008). Earlier study, the size of the *Melipona bicolor* mtDNA has been determined as 18,500 bp by RFLP analysis. Additionally, restriction map data was reported for species belonging to the genera *Plebeia* and *Melipona* (Francisco *et al.*, 2001 and Weinlich *et al.*, 2004). Moreover, restriction size patterns for *M. quadrifasciata quadrifasciata* and *M. quadrifasciata anthidioides* were also identified by RFLP analyses (Moretto and Arias, 2005). The information of mtDNA sequences from *Trigona* species is very limited, only few mitochondrial gene sequences such as the 16S rRNA gene, and had been reported in genetic variability study (Rasmussen and Cameron 2007 and Costa *et al.*, 2003).

Therefore, much more knowledge of mtDNA sequences of this species is required in order to effectively promote the genetic variability researches.

Trigona pagdeni Schwarz, an indigenous stingless bee, is one of the most common stingless bees in Thailand (Sakagami, 1978). This native species was frequently found nesting in various artificial structures in close contact with humans (Franck *et. al.*, 2004), and artificially propagated in boxes for plant pollination and commercial products. Likewise, the traditional beekeeping of many stingless bees has been reported by Crane (1992). The objective of this study is to characterize *Trigona pagdeni* partial mtDNA sequence, and the organization of the mitochondrial genes. The partial of mtDNA genome of the stingless bees, *Trigona pagdeni*, was amplified by using a long PCR technique (Cheng, 1994) and the gene order was verified by the PCR amplification using internal primers.

MATERIALS AND METHODS

Sample and DNA extraction

Trigona pagdeni specimens were collected from different localities of Thailand and was immediately preserved in 95% ethanol and stored at 4°C until required. Total genomic DNA was extracted from the entire bee using phenol-chloroform extraction and ethanol precipitation following Smith and Hagen (1996) with a few modification.

PCR and development primers

A portion of the cyt b gene (432 bp) was known from GenBank (accession no. AY575080), whereas the three mtDNA regions for the COI, 16S rRNA and ATP(6, 8)+COIII genes of *T. pagdeni* were amplified using the specific primers (Table A.6; APPENDIX A). The PCR conditions were described in Table A.7; APPENDIX A. The PCR products were then purified (QIAGEN) and cloned into pGEM® -T easy vector (Hoelzel and Green, 1992). The insert sizes were verified by colony PCR. Plasmid DNA was extracted from recombinant clones and sequenced for both directions. The known sequences of the cyt b, 16S rRNA and COI genes were applied to design internal primers (Fast PCR program version 5.2.21; Kalenda, 2007) used for the inverse PCR or genome

walking technique (Topic 5.2 and 5.3; APPENDIX A) to obtain flanking regions of unknown sequences.

PCR and sequencing

On the basis of the four partial sequences (COI, cyt b, 16S rRNA and ATPase(6,8)+COIII genes), three sets of the primer pairs (LR12647-R+COI2494, LR12677+ cytb10729 and COIII9821+cytb5031, respectively; Table A.13; APPENDIX A) were designed on each gene and used to amplify the three long PCR products; 16S/COI, 16S/cytb and cytb/COIII regions, respectively. PCR was done in a Model PTC-200 Peltier thermal cycler (MJ research Inc.), and the reactions were carried out with 30 cycles of a 25-μl reaction volume containing 4.75 μl of sterile distilled H_2O, 2.5 μl of 10×LA PCR buffer II (Takara), 4.0 μl of dNTP (2.5 mM), 5.0 μl of each primer (0.4 μM), 0.25 μl of 1.25-unit Takara LA Taq™ (Takara Bio, Otsu, Shiga, Japan), and 1.0 μl of template containing approximately 5 ng DNA. The PCR reaction was performed with denaturation at 98°C for 10 seconds and annealing and extension combined at the same temperature (60°C) (Table A.14, APPENDIX A). The PCR products were verified on a 1.0% agarose gel and visualized by ethidium bromide staining via ultraviolet transillumination. The PCR products were purified using Qiaqiuck gel extraction kit (QIAGEN) and cloned into pGEM® -T easy vectors (Promega) subsequently used for sequencing with BigDyeTM terminator cycling conditions on an automatic sequencer 3730xl (sequencing service, Macrogen Inc; Korea).

Internal sequencing primers (Table A.15, APPENDIX A) were designed on each the insert fragment for sequencing. DNA sequencing was performed under BigDyeTM terminator cycling conditions on an automatic sequencer 3730xl (sequencing service, Macrogen Inc; Korea). All sequences were analyzed and compared by the homology search to assure the correct fragments using Blast*N* (nucleotide similarity) available at http:// www. ncbi.nlm.nih.gov..

To verify the gene order, the three overlapping fragments (16S/COI, 16S/cytb and cytb/COIII) were amplified by PCR with following ten internal primer pairs (ND4L-F/ COIII-4W, ND4-3W/LR-3W, LR-1W/COI-1W, COII-2W/ LR-2W, ND4-2W/ UN-2W ,

tRNA-Leu-3W/COIII-4W, SR-2W/ND1-2W, ND4-2W/ND5-3W, ND4-3W/ND5-3W and ND4-3W/ND5-4W; Table A.13; APPENDIX A) using Takara LA Taq™ (Takara Bio, Otsu, Shiga, Japan). The PCR conditions were provided in Table A.14; APPENDIX A.

Sequence analysis

Coding regions were identified using searching for open reading frames, including start and stop codons. The comparisons of nucleotide or amino acid sequences were performed by alignment with those of *Melipona bicolor* using the BLAST*X* algorithm (NCBI). Transfer RNA sequences were identified by eye and comparison with homologues of *Apis mellifera*, *Melipona bicolor* and *Bombus ignitus*. The sequences of ribosomal RNA were identified by alignment with those genes of *Apis mellifera*, *Melipona bicolor* and *Bombus ignitus*.

RESULTS

Genome composition

The 16S/COI, 16S/cytb and cytb/COIII regions amplified using three primer pairs (LR12647-R+COI2494, LR12677+ cytb10729 and COIII9821+cytb5031, respectively), were 5855, 5089 and 4879 bp, respectively (Figure 5.1). The three PCR products were purified, cloned into pGEM® -T easy vectors and used for sequencing by primer walking technique.

The known sequences of each fragment were then overlapped and analyzed (Figure 5.2). The gene order on the overlapped fragments was verified by PCR amplification using ten internal primer pairs by long PCR technique (Table A.13 and Figure 5.3). The expected size of PCR products were revealed in Figure 5.4. The total sequence of the overlapped fragment was 12,802 bp containing 12 protein coding genes (11 complete gene sequences and a COI partial sequence, both rRNA genes and 12 tRNA genes (Figure 5.5). Most of these mitochondrial genes was similar in length to their counterpart genes in other sthingless bees (*M. bicolor*) and honey bees (*A. melliferra*) (Table 5.6).

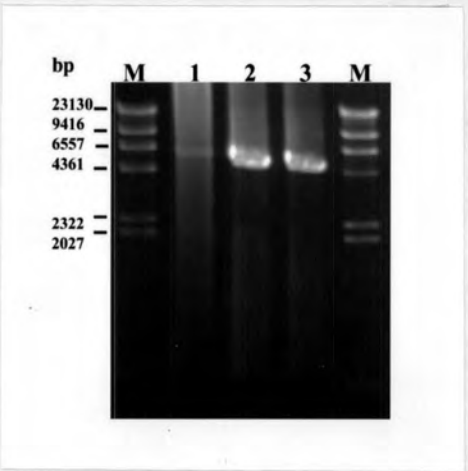


Figure 5.1 The PCR products amplified using primer pairs LR12647-R+COI2494, LR12677+ cytb10729 and COIII9821+cytb5031, respectively. The expected size of PCR products were 5855, 5089 and 4879 bp, respectively. M is λ DNA digested with *Hin*dIII.

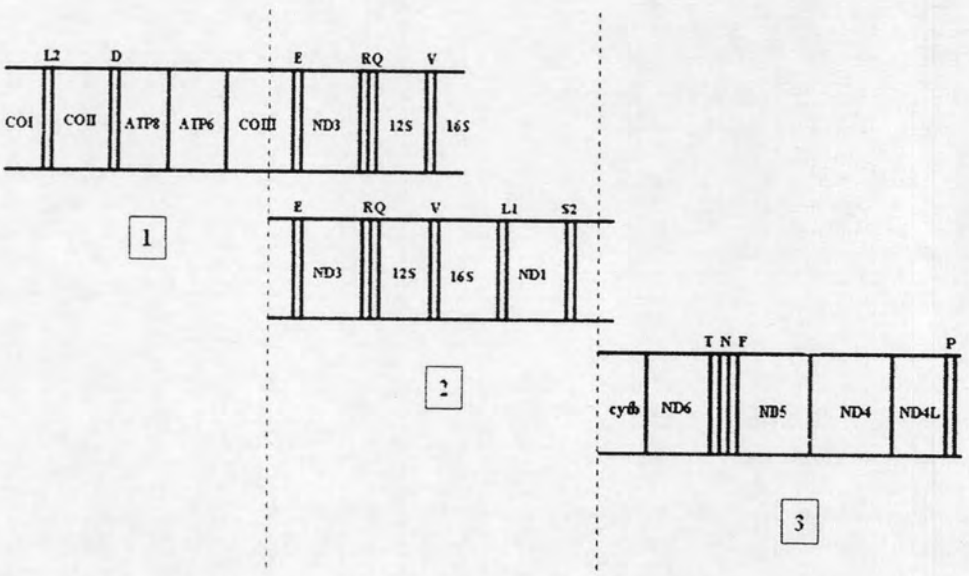


Figure 5.2 Overlapping the three sequenced fragments (5855, 4879 and 5089 bp, respectively) obtained from long PCR amplification with LR12647-R+COI2494, COIII9821+cytb5031 and LR12677+ cytb10729, respectively and sequenced by primer walking.

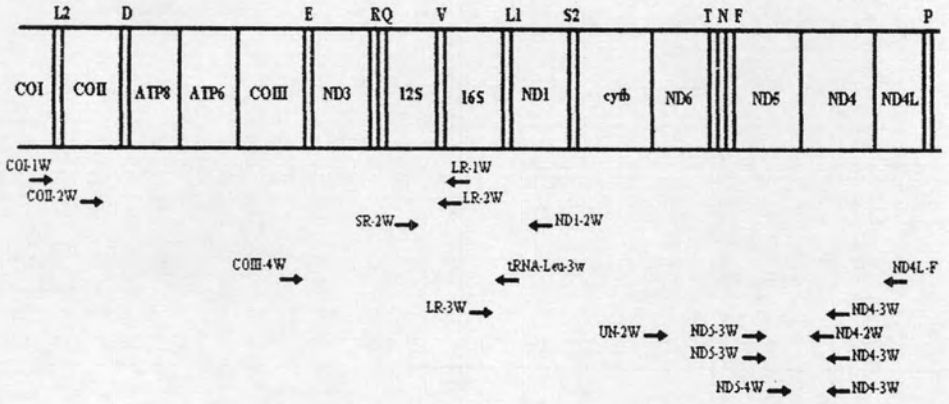


Figure 5.3 The positions of ten internal primer pairs on the overlapped fragement used for verifying the gene order; ND4L-F/COIII-4W, ND4-3W/LR-3W, LR-1W/COI-1W, COII-2W/LR-2W, ND4-2W/UN-2W, tRNA-Leu-3W/COIII-4W, SR-2W/ND1-2W, ND4-2W/ND5-3W, ND4-3W/ND5-3W and ND4-3W/ND5-4W, respectively. The expected sizes of PCR products were 10129, 6563, 4943, 3414, 3380, 3288, 2532, 2450, 1739 and 1100 bp, respectively.

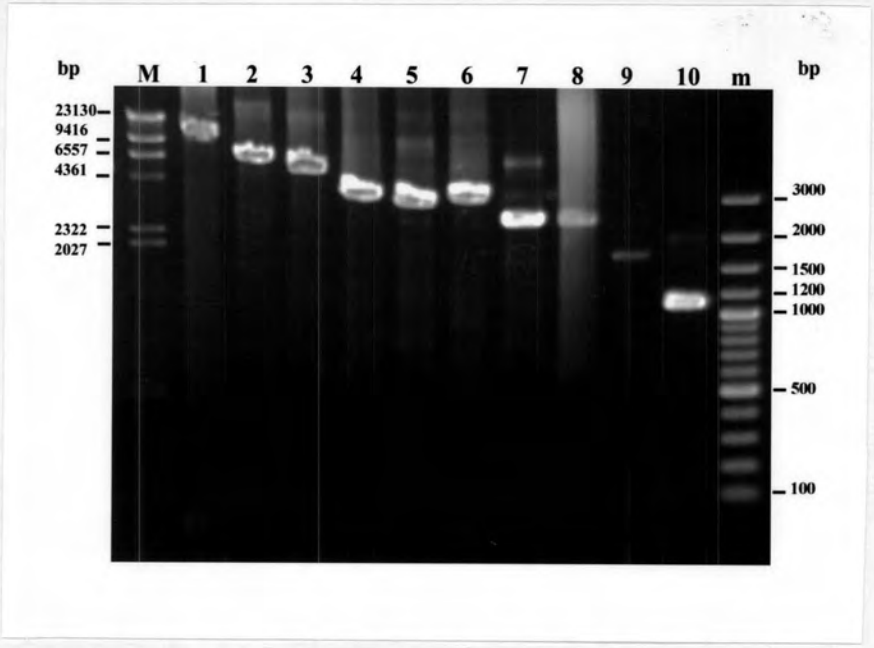
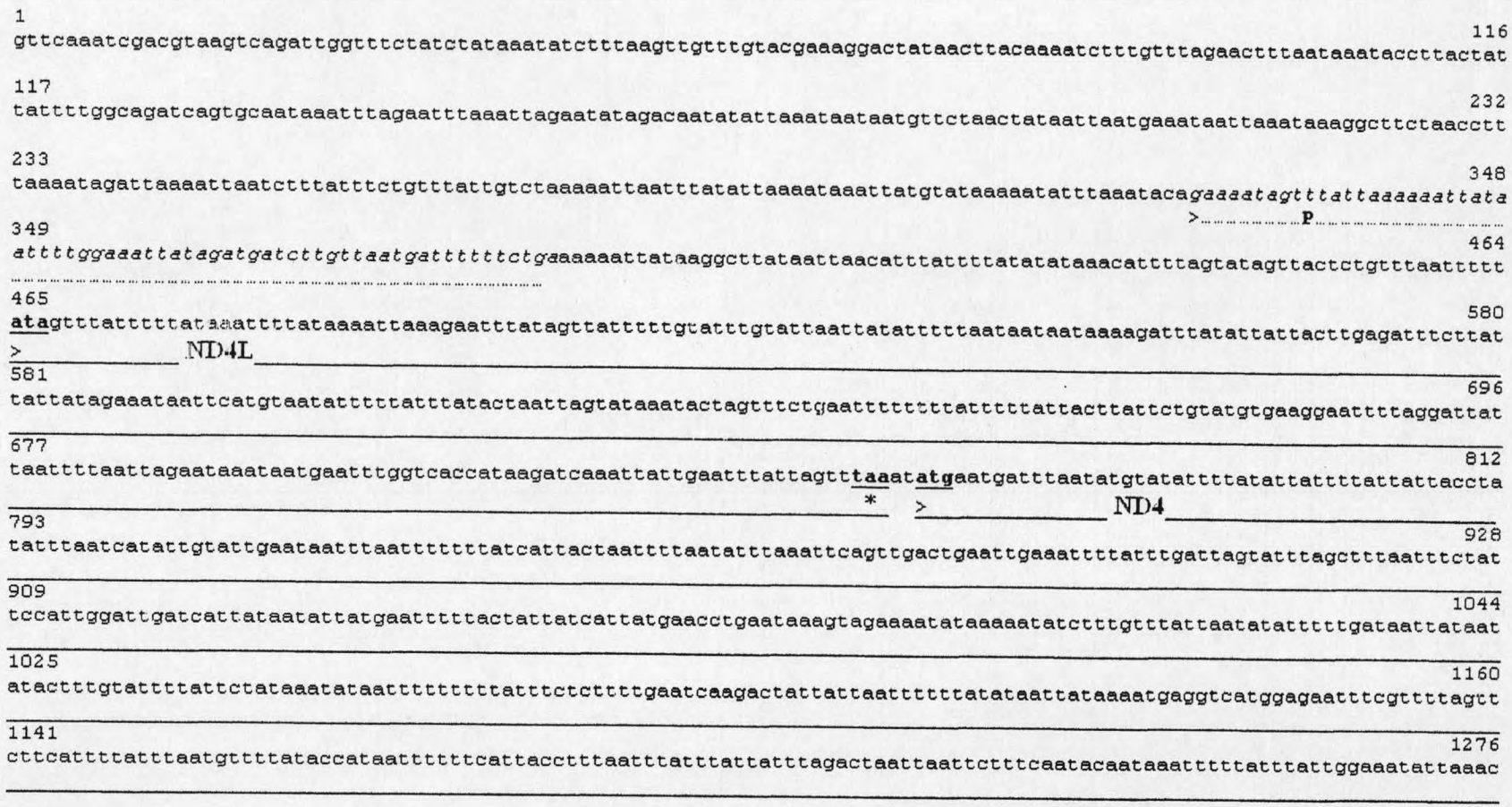


Figure 5.4 The PCR products amplified using ten internal primer pairs. Lane 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 were the PCR products obtained from reactions primed by the following primers, ND4L-F/COIII-4W, ND4-3W/LR-3W, LR-1W/COI-1W, COII-2W/LR-2W, ND4-2W/UN-2W, tRNA-Leu-3W/COIII-4W, SR-2W/ND1-2W, ND4-2W/ND5-3W, ND4-3W/ND5-3W and ND4-3W/ND5-4W, respectively. The expected sizes of PCR products were 10129, 6563, 4943, 3414, 3380, 3288, 2532, 2450, 1739 and 1100 bp, respectively. M is λ DNA digested with *Hin*dIII.


1
116
gttcaaatcgacgtaagtcagattggtttctatctataaatatctttaagttgtttgtacgaaaggactataacttacaaaatctttgtttagaactttaataaataccttactat
117
232
tattttggcagatcagtgcaataaatttagaatttaaattagaatatagacaatatattaaataataatgttctaactataattaatgaaataattaaataaaggcttctaaccttt
233
348
taaaatagattaaaattaatctttatttctgtttattgtctaaaaattaatttatattaaaataaattatgtataaaaatatttaaatacagaaaatagtttattaaaaaattata
>P
349
464
attttggaaattatagatgatcttgttaatgatttttttctgaaaaaattataaggcttataattaacatttatttttatataaacattttagtatagttactctgtttaatttttt
465
580
atagtttatttttataaatttataaaattaaagaatttatagttatttttgtatttgtattaattatatttttaataataataaaagatttatattattacttgagatttcttat
> ND4L
581
696
tattatagaaataattcatgtaatatttttatttatactaattagtataaatactagtttctgaatttttttttatttttattacttattctgtatgtgaaggaattttaggattat
677
812
taattttaattagaataaataatgaatttggtcaccataagatcaaattattgaatttattagtttaaatatgaatgatttaatatgtatattttatattatttttattattaccta
* > ND4
793
928
tatttaatcatattgtattgaataatttaatttttttatcattactaattttaatatttaaattcagttgactgaattgaaatttttatttgattagtatttagctttaatttctat
909
1044
tccattggattgatcattataatattatgaatttttactattatcattatgaacctgaataaagtagaaaatataaaaatatctttgtttattaatatatttttgataattataat
1025
1160
atactttgtattttattctataaatataatttttttttatttctcttttgaatcaagactattattaatttttatataattataaaatgaggtcatggagaatttcgttttagtt
1141
1276
cttcattttatttaatgttttataccataatttttttcattacctttaatttatttattatttagactaattaattctttcaatacaataaatttttatttattattggaaatattaaac

(continued)

1277 1392
attaaagaaatcagtaattttaaatttatttatattatttttttcttttttagtaaaaattcctatatatatagttcatggatgacttcttaaagctcatgtagaagcatccttctt

1393 1508
taattctataattctagcttcagtaatattaaaattaggaggatatgggctaatacgaataatatttttttataaaatatatattcaataaattctatagttatttttattataatta

1509 1624
atttattcggtatattatcactaagaataatatgtttatttcaaatggatattaaactaattattgcaatttcttcaattgtacatataggaattatactcataggaatttttatta

1625 1740
ataaccaaaataagggtatatggaagattctatataataattagtcatggattcatttcatctggattattttatttttgttaatttgatttatagtcaaactaatagacgactagt

1741 1856
atatatcaataaaggaataattagattagttccgtcagttataatattttgatcattaatttgtttttgcaatgtaggttctccttttttcaataaatttgattagagaaattattc

1857 1972
taataataagtttaatctgttgatttaaatatataatattatttattataatatatggtatattctcttttgtttattcagtatatttatttatattaatttttcatggaaaaaag

1973 2088
tttgtaaaatttaaaatatttaattgtaaatcaataaattatttgagattaattaatcatcttgttccagtaaatctgatatttttaaatctactttgtttaaatattttaataaa

2089 2204
aatattgaattgtgatttcaatgatacaatagtatttgttttaaaccatcataaaaatattgattttagaataatattgcttattacaagattaattattttattattttcaata

ND5

2205 2320
atatttttatcattaaatattgaattaataatagaatgaaatgtcttaagaattaattcaataaaaataaacataatttttagtattaaattataaaactttattatacatatttttt

2321 2436
agttatatttatctcatcaataattttcatgtatagaattgaatatatagaattggaaaaatttttagttaaacgcttttattatttaataataatgtttttgatatcaataattt

2437 2552
tactaattatcagacctaacatgcttactattatacttggatgagatatattaggattgacatcatattgcttaattatttactatagaacaattaattcatataactcaggaata

(continued)

2553 2668

actactgttctgcttaatcgtattggagatataaggctattaataattatttcgataatatcaatgtttggaagatgaaatcttttaatatacagaataaataaacctataatagt

2669 2784

tataattattattatagtttttactaagagtgcacagtttccttttttttgtatgactaccaatagcaatgatagctcccactccagtatcatcacttgttcattcatcaacactag

2785 2900

taactgcaggtgtatatttaataatttgatataataaaataattgatttaaaatatataggatttattatatcaatttctagaattacaatacttttttcaggtataatagctaat

2901 3016

tccgaaatagattttaaaaagatcattgccttttctacattaagacaattaggatttataattagaattttatctataggattaaatgaattagctttccttcatttatttattca

3017 3132

tgctttatttaaatcaataatatttatatgtgtaggaagatttattcacaatataaaaggaattcaaaatttccgattttatagtggaatattttatatctatcctattaaaagat

3133 3248

ctataattattttatcattaatgatactttgtggttttccttttcttgtaggattttattctaaagatttaataattgagatatttatatacaataaaattagtatttttaatttt

3249 3364

attgtaatcataattggtacaataataactatttcatattcatttcgtattttattaaaatttttttctaataattacataataaattccataattaaaaaagaatcggatattat

3365 3480

aagattcgtaatagtatttataatgatttttatattattaataagaaaaattgtttataatataaatttaattttatttaattgtaatttaataaatatttacaagtattttgtta

3481 3596

ttaaaatatttatttttaggatatttattaaatattatcattaacaacataatatacaataaaattgtaaatattataaaaaaactacttttatataataaatatattcaaattattg

3597 3712

aaaaaaaactattttattgtattaattaaatatgaaaataattatgaaaaaatgtttaatgaaataattatatcaaaccttataatatttatttccatagttagatataaaaatat

3713 3828

agtaaaagtaaatgtatctatttattctataatatttttttttatatttacttaatcatgattttatttatattataaatatagtataattatttatttaaatagcttatataattag

* F

(continued)

3829 3944

agcataatattgaaaatattaatgagatcatagatttttaaatatagttctaacatataatttaaagttagaataataaattcattaattttattaacaataaaatgcctcaatct

<...N

3945 4060

ttggctttaatttaacaaacaactatatataaagtaatgtaaatatgtaaagatatagaaataaaaagatatgtaacttcatgttaattatcaatttacaaaaccgatgttttta

<...T

4061 4176

ttaaactacataacttatataaatatattttaattttaatgtagagtagctagaaaaaatgaaaacatttgcaattatatattttatttaagaataaattcttttatattgttaa

>_____ND6_____

4177 4292

ttttctttgtaaatatatatatttctcagatttcttcggttccaataaatcaattgattaacattatttttttcatgatttgaagatcttttgttttgtttatggtaagaagtcat

4293 4408

ataattatttatgtttttataatttaattgtaatgatcagaggtataataatttgttttcttattttgtatgtttaagtaatataatgaatataaaaaaaaattaagttcaaata

4409 4524

catttatttattgaattacatattaatttatatttcaataatttttttctgagaatttataattataatatataataaaatatttcgaaatcaagatttaaatcatgaccaattaat

4525 4640

tattaaattattcaattttcctaattattatatattattaattattattgtgtttatatttttgatattaatgttatcttcaaaaatttgctttagtaaaaaaaaatctttacgaa

4641 4756

ataaattgtttt**taa**tgaaaaatttaaaaatttatcttattcaaatactttgtctaaacttgtatcttccttc**att**tcgattccaattccaataaacattaattatttctgaaatt

*

>_____cytb_____

4757 4872

ttggttcaatattaggaatatttttagtaattcaaattatttcaggtttgtttctctcgatacattattgtccaaatattgattatgcatttcaaagagtttcatatattataaaa

4873 4988

gatgtaaattcaggctgattagttcgattgatccatataaatggggcttcattttatttatttaatttatgcacatattatacgaggaatatattattattctttttaaattaac

4989 5104

tagagtatgattaattggaaggataattacatttttatcaatagctacagcatttcttgggtatgtacttccatgaggaccaatatcattttgaggagctatagtaattacaaatt

(continued)

5105 5220

tattatcagcaattccatatgtaggtaatataattgtagaatgattatgaggaggattctcaattaataattccactcttaatcgatttttttcttttcactttattcttcctttc

5221 5336

attatcttatttttgtaatttacatttattaatattacacaaatctggttcttcaaacccttacattcaaaaatcgatgtttataaaattgctttccacccttattttatgat

5337 5452

taaggatttagtgacaattactttaattttatcattatttataattgtaaatcttcaagtaccgtatttttaggtgatccagataactttaaaatagctgatcctatagttactc

5453 5568

cattacatattaaacctgagtgatactttctttttgcctattcaattttacgatctattcctaataagctaggaggagtaattatacttttatatcaatttttatactttatctt

5569 5684

cttcctatgttaaatataaacaacataaaaaatattaaattttatccaatcaaccattttatttattggacatttattaataatgttattgtgttaacatgactaggagggaaagc

5685 5800

tattgaaaaccctttattgaattgaacattgtatttacatttatatacttttttattatttatttcatttgtattaaataatttaattgatattttaatgtacaataaatatt

5801 5916

*aa*attaatgaacttgaataagtatatgttttgaaaacgtaatatagaaataaaattttctattaatt*tta*tatatttgttattataattaattctttaaataaataacaataaatt

S2

5917 6022

aaataataagttattaatactaataattctattcaacaagtatatattagataatcgtaacgaatacgaggtaatactccacgtattatcacgattaaaattaaattaataataaa

6023 6148

aaatatattgaaaaaaaaatttcaataaaatataccataaaataaaattactaaaattactctcataaaaattatatttatatattcggataaaaaaattaatgtaaataatctac

6149 6264

tgtaatattcaatattaaatccagatactaattcagattctccttctactagatcaaaaggaactcgatttaaatctactaataatctaagaataaagataaaatatattggataa

6265 6380

gaaaaaataaataatttaatatttatttgataattaatagaatttgacatagaatatctgttaaagaatattataagaataaatatcattatgtacattaatacttcgaatgaaat

(continued)

6381 6496

tacttgtgaaactatacgtattactctaattatagaataattatttgtagaaattcatctgattattaaaattggatatacattaattctaataattaaaattatataaagtatac

6497 6612

tgtaattaaaatttaaattattgaaaattcatggatataataatcacaaaattgatgatacaataaatattaataatggactataatagaatatataatttatactgaaatcaaac

6613 6728

atttccttagttagaagtttgattatatcgctaaatggttgtagaataccttataatataatttatttggacctttacgacattgcatatatcttaaaatcttacgttcaaataa

6729 6844

agttagaaaagctactctaatcataactattaatactataattaataaattaattagtacaaaaat*tat**tat**ttaatatattgtctatattctaatttaaattctaaatttattgc*

ND1 <

<…… L1 ……

6845 6960

*actgatctgccaaaataat*agtaaggtatttattaaagttctaaacaaagattttgtaagttatagtccttccgtacaaacaacttaaagatatttatagatagaaaccaatctga

……> 16S

6961 7076

cttacgtcgatttggaactcaaatcatgtaagattttaaaagtcgaacagacttaattattaaactcctgcgttcaaaatttatcttaattcaacatcgaggtcgcaatcacttttt

7077 7192

ttctatacgatctataaaaaatattacgctgttatccctaaggtaatttactctagctatcccaaattgagggtcattgaaaaatttttctttaattaaagttatcagaaattaaa

7193 7308

gtttataaaatttaatagtcctcccaaccaaatattgctacatttttttatattaactattaagatgtaactatataaaattctatagggtcttatcgtcccataaaaaaatttttaa

7309 7424

cattttaattaaaatttttaagttcatgaaatatatatgagacagtgtatatttcattaatcctttcattccagatttcatttaaaaaccaattgattatgctacctttgtacagtc

7425 7540

agttatactgcagccattcaagtactcattgggcaggtcgacttaaaattattatcaataagacatgtttttgttaaacaggtgaatatacttgttttgcctagttccttatatat

7541 7656

atatagtatattggactgatattatacattaattttgtatttacactactaatttaatcattataattaatcctatataatttagggttatttttttctaaataatattcattgaaa

(continued)

7657 7772

atttcagaattatgagtctctttcaaaatttttaaattaagtcaataaactatgagattttatcaattaatgtttaatgttaaaaccataaatttacaatcaactagagttaataa

7773 7888

actgaatattctaaactttcagcttgtcccgaaagtttttttgcaaataaaaatttaggtaatttttatttttacctaaattaatttatatgcctaaattcaatttaattagaaaaa

7889 8004

aactagatataaataagttcgaaaaacatctaatttctaaatatatatttatattctttcttccacaggaacaattatatatatttagatcacttattttcgagacttctattagt

8005 8120

taatagaatatattaatcaaccctgatacaaaaggtacaaaaaaaaattttacttatttttttttgaaaaatttcattttcattactttaaactataaataatataattgattttta

8121 8236

atcgttacttgaacttttcccatttatataactttgttaagacatatttctttaatttctaaaatcgacatataagttaattaataatttcaaa*taaaattttctagaataaactct*

<.....V.....

8237 8352

*agttaaatttattgtaaataaaaagttttatataaactaaaatttt*tggctaaaaacactttccagtacctttactatgttacgacttattccttttatgaaattgacgggcgattt

.....>

12S

8353 8468

gtacaccaattattactttattcagctaaagaaattggattagcttactattaaatcctctttctagacattattaaaaatgtctattcaaaaataaaatttattataactcattt

8469 8584

gtccttagatatactacattttgatttaaagtactattttattaaatatcactatatatgaaaaatttttgtttatatattaaaataacaatacataaattattaatccaagtatt

8585 8700

tcctatcgtggtttatcaaattaaacaacaagtttccctaaaaagattaagaggccgccatattattttatttttaatgattctacatttaatggtaaaaacttgagagtatcctag

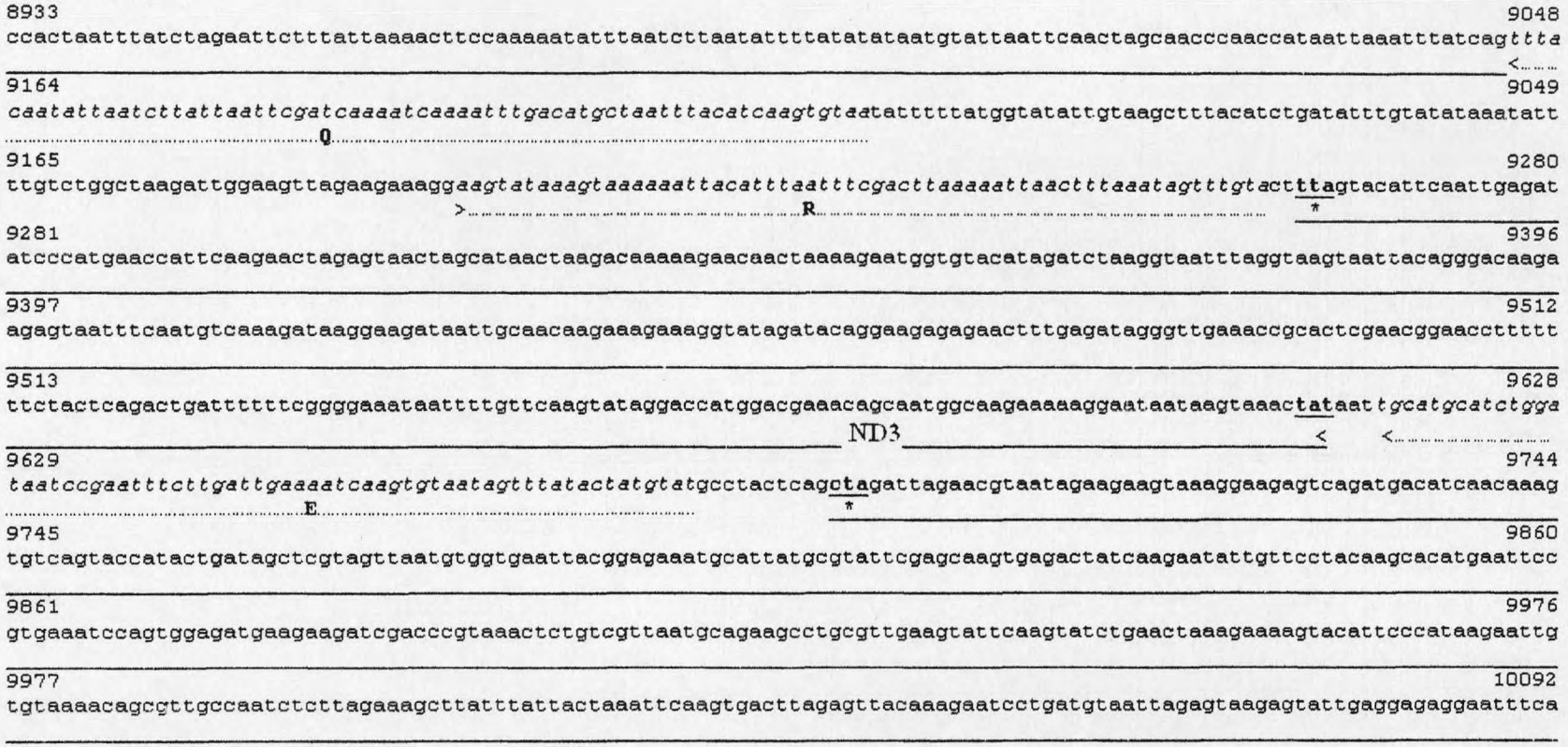
8701 8816

aataatagggtatctaatcctagtctatgttccagttttttacatactttatatttaacctatcctatgttcataaatttaaatttcacctatgatttatgattcaaaaataagc

8817 8932

tatacttcaaggtaaattaggtagagtttcgaagattgatagcttataatcaaagtttaaccgctattgctggcactttatttatcattattaattttaattcccatatttaattt

(continued)


8933
9048
ccactaatttatctagaattctttattaaaacttccaaaaatatttaatcttaatattttatatataatgtattaattcaactagcaacccaaccataattaaatttatcagtttta
9164
9049
caatattaatcttattaattcgatcaaaatcaaaatttgacatgctaatttacatcaagtgtaatattttttatggtatattgtaagctttacatctgatatttgtatataaatatt
Q
9165
9280
ttgtctggctaagattggaagttagaagaaaggaagtataaagtaaaaaattacatttaatttcgacttaaaaattaactttaaatagtttgtacttttagtacattcaattgagat
R
9281
9396
atcccatgaaccattcaagaactagagtaactagcataactaagacaaaaagaacaactaaaagaatggtgtacatagatctaaggtaatttaggtaagtaattacagggacaaga
9397
9512
agagtaatttcaatgtcaaagataaggaagataattgcaacaagaaagaaaggtatagatacaggaagagagaactttgagatagggttgaaaccgcactcgaacggaacctttttt
9513
9628
ttctactcagactgatttttttcggggaaataatttttgttcaagtataggaccatggacgaaacagcaatggcaagaaaaaggaataataagtaaactataattgcatgcatctgga
ND3
9629
9744
taatccgaatttcttgattgaaaatcaagtgtaatagtttatactatgtatgcctactcagctagattagaacgtaatagaagaagtaaaggaagagtcagatgacatcaacaaag
E
9745
9860
tgtcagtaccatactgatagctcgtagttaatgtggtgaattacggagaaatgcattatgcgtattcgagcaagtgagactatcaagaatattgttcctacaagcacatgaattcc
9861
9976
gtgaaatccagtggagatgaagaagatcgacccgtaaactctgtcgttaatgcagaagcctgcgttgaagtattcaagtatctgaactaaagaaaagtacattcccataagaattg
9977
10092
tgtaaaacagcgttgccaatctcttagaaagcttatttattactaaattcaagtgacttagagttacaaagaatcctgatgtaattagagtaagagtattgaggagaggaatttca

(continued)

10093 10208
gtgtaattaaatacctgaactatttttggaggccatactatgttaatttcaatagatggagaaattgatctgtgaaagaaagttcagaaaaatgagataaaaaaaaataattcgga

10209 10324
taaaataaatataaatatgctaaacttgatcattatttttacaactagggtatggcctcctatgaaagtactttctcgaatgatatctcgatttcataaagcaaaaactatggaca

10325 10440
aattaaataatgtgaaagatacaattgagtattctttaagattgattcaaattacaattctaacaagaaagtttattagattaaaagaagcaatgataggccatgggctgatagtc

10441 10556
actaagaggtatggaaaatttttttt**cat**tgg**cta**gatcctttctttcagatacagagctagaagaatagagaatacgtaagcttgaattactgacattgaaacctctagcaacaa

COIII < *

10557 10672
cagaatgttctcaacgactgatatgatagggaaagtatatacgaagtttattacaaagtttcttagaagaataagaatcaggtgtcccgagatgagatttgaggaaagtcgaatag

10673 10788
acaatgtaagaggacggatgagatatctgatcagttcaataattaccatgaaatttatcaaggctttaggagagttgtgagggacaagatgagcaaagaattttatcgggtttgta

10789 10904
ataattgagtataagaaaaatcttgaccataggggtaaagacaatgatagattgaacagaagatgtcttgtagttggaaaaatatatggaaataggcttagaaagttcattaatat

10905 11020
ggtataaaatattaggctgataaaaatcaatgagtttggagtaagcttcttgtatcacattgctagagaaaactcgttgaacagaatagacaagagtaaagtacatattatgacaa

11021 11136
ggcgggaaggaatgactcagtatcctccagctagaaaaaccagggaagatagggagaatacccagttaagctggaaagtatagacgtagtaaactgatggatcaaatctttcaaac

11137 11252
aggtttaacactt**tta**att**cat**tctattttgaaaatttaacctttttagggtctttggagaaaatagaattacatgaatttactatgtaaaatgaattaagtaaaacaattact

ATP6 <

*

(continued)

11253 11368

agaattaaaaaaataatatttataaggaaaactaagattcaattaactggcattatttgaggaatggaaaattaacttcaacttggtttacttcaaaatgacatttcgatgttatt

ATP8 < <..... D .....

11369 11484

tattaactaaatttttcacaagtttcctaaattgaaatttactcaatttacaaagttagagtaatctgttctttctagcataataggtataaatctatggttactccgcagatctca

..... *

11485 11600

gagcactgtccaaaatataaaccaggtcgagtaggtattaggttgagctgattgattcgtccaggaactgcgtctactttaactccaagagattgaatagttcaagagtgaattac

11601 11716

gtcaagagatgaaacgatcagtcgaattgaaatcttgaacggaacgactaatcgattatcagtttcaatcaagcgaaaatgatctatgtccgagtaaatctgtatataagaacttta

11717 11832

tttcatgattgttaaattcggggtattcataagatcaatatcattgatgtccaatagccttaattgaaaagtaggggttagaaatttcatcaataaaatacaagattttgagagaa

11833 11948

gggtaacaaattaggagaagaataattattggaataatagtccagactacttcgacagtgtgatttttaagaacccttaagtttaaaaatgaattaagtacaaaatcaatgataaa

11949 12064

aaacatagtcaatgaagtaattacaataataaatattatgacaaaattatgaaatgaaattagattgtctgagtaaatagaatttgagtcctgaaatgagtatatgtttcatgttg

COII

12065 12180

aaatttttaacgaaaacaaaagttttatacttgaatcttaaattcaaagcactaatctgccacattaaattatagctacttgttcaaaatatttttaactttatttattgaaatta

< ..... L2 ..... < * COI

12181 12296

gaggaatctcattgattgagtgattaagcggaggatatacattcagccattcaagagatgattgactaaatttaaaaacaaccagccgtttagataagagtctttcgtaaataata

12297 12412

taaataaggaacaatagtctattaattgaaattattgatccaactgaggatataaaatttcagcaatagtatgaatcagggtaatcagagtatcgtcgaggcattcctattagacc

12413 12528

taagaaatgttgagggaaaaaggtcaaatttacaccaataaatataaggaaaaattgaaacttcaaccatttctgattcatcattagtccagtaaataatggaaaccagtggataa

(continued)

```
12529                                                                                                12644
accttgcaataattgagaacactgcccctatagataaaacgtaatgaaaatgcccgactacatagtatgtatcatgtaatacaatatcaattgaagaatttgaaagtatgataccc
```

```
12645                                                                                                12760
gttaagcctccaaatgtaagtattagaataaatccgatagatcacacgaaagaaatattaaaatttaacttagatccatgataagtggcaagtcatctgaataccttgattcctgt
```

```
12761                                  12802
aggaactgcaataattattgttgctgatgtaaaatatgctcg
```

Figure 5.5 The partial mtDNA sequence of *T. pagdeni,* numbers above the sequence indicated nucleotide positions. For protein-coding genes, the initiation codons were indicated with the bold and underlined genetic code. Stop codons were also marked with the bold and underlined genetic code and by an asterisk below. A dart (>) marked the first nucleotide of the initiation codon in each gene and presented the direction of transcription. The protein-coding genes were underlined and inferred to be abbreviated. The tRNA genes were marked in italic nucleotides, a row of dots below and the abbreviated amino acid code; P= tRNA-Pro, ND4L= NADH dehydrogenase subunit 4 (light chain), ND4= NADH dehydrogenase subunit 4, ND5= NADH dehydrogenase subunit 5, F= tRNA-Phe; N= tRNA-Asn, T= tRNA-Thr, ND6= NADH dehydrogenase subunit 6, cytb = cytochrome b, S2= tRNASer (UCN), ND1= NADH dehydrogenase subunit 1, L1= tRNA-Leu(CUN), 16S= large rRNA, V= tRNA-Val, 12S= small rRNA, Q= tRNA-Gln, R= tRNA-Arg, ND3= NADH dehydrogenase subunit 3, E= tRNA-Glu, COIII = cytochrome c oxidase subunit III, ATP6= ATP synthase subunit 6, ATP8 = ATP synthase subunit 8, D= tRNA-Asp, COII= cytochrome c oxidase subunit II, L2= tRNALeu(UUR) and COI= cytochrome c oxidase subunit I.

Base composition and codon usage

The A+T contents of each protein-coding gene in *T. pagdeni* were high, ranged from 66 to 87%. The 12S and 16S rRNA genes of *T. pagdeni* had the number of A+T content with 77 and 76%, respectively. The mitochondrial genetic codes used in *T. pagdeni* were similar to that of *Melipona bicolor* and *Apis mellifera*, but different in codon frequencies (Table 5.1). The frequent amino acid used in *T. pagdeni* was Leu (387), Ile (333) and Met (306), respectively. Whereas Gln was the lowest frequent (27). However, the AT bias in codon usage could be revealed by the ratio of "G+C" (Pro, Ala, Arg and Gly) to "A+T" rich codons (Phe, Ile, Met, Tyr, Asn and Lys) (Crozier and Crozier, 1993). That ratio of "G+C" (Pro, Ala, Arg and Gly) to "A+T" rich codons (Phe, Ile, Met, Tyr, Asn and Lys) of *T. pagdeni* was 0.20. When A+T content observed in each protein-coding gene was calculated, *T. pagdeni* showed high A+T nucleotides (Table 5.2). The AT bias on mitochondrial protein-coding genes could be confirmed by the nucleotide usage on first, second and third codon positions (Table 5.3). This implied that *T. pagdeni* had a bias towards AT-rich codons in the protein-coding genes identified.

Table 5.1 The total number of occurrences of codons on the 11 protein-coding genes in *T. pagdeni* mtDNA (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=cusp).

Amino acid	Codon	ND4L	ND4	ND5	ND6	cytb	ND1	ND3	coIII	ATPase6	ATPase8	COII	Total number
Ala	GCA	0	2	3	0	4	0	1	2	1	0	1	14
	GCC	0	0	1	0	1	0	1	0	1	0	0	4
	GCG	0	0	0	0	0	0	0	0	0	0	0	0
	GCT	0	2	4	0	6	1	1	3	4	0	1	22
Cys	TGC	0	1	1	1	0	0	1	1	0	0	3	8
	TGT	1	7	3	1	1	3	0	0	1	1	1	19
Asp	GAC	0	0	0	1	0	0	1	1	0	1	5	9
	GAT	1	5	7	1	8	5	0	2	1	0	5	35
Glu	GAA	4	7	11	0	3	10	3	6	3	1	6	54
	GAG	0	0	1	1	1	0	1	1	3	0	1	9
Phe	TTC	1	12	5	4	4	6	7	14	10	3	1	67
	TTT	13	48	45	16	34	25	4	19	12	2	13	231
Gly	GGA	2	12	13	0	10	3	1	6	3	0	5	55
	GGC	0	0	0	0	1	0	0	2	0	0	0	3
	GGG	0	1	0	0	3	0	0	1	0	0	0	5
	GGT	1	4	4	1	5	3	1	0	0	0	1	20
His	CAC	1	0	1	0	3	0	0	6	2	0	2	15
	CAT	2	10	4	2	7	0	0	3	2	0	4	34
Ile	ATC	1	5	7	1	4	4	3	10	6	1	5	47
	ATT	13	47	61	25	40	33	8	16	14	7	22	286
Lys	AAA	5	21	28	10	11	5	4	5	5	4	3	101
	AAG	1	1	3	1	2	3	2	3	2	1	2	21
Leu	CTA	1	10	4	0	2	6	4	6	11	1	4	49
	CTC	0	1	0	0	1	0	1	4	3	0	3	13
	CTG	0	3	1	0	0	0	0	1	9	0	1	15
	CTT	1	4	12	0	10	4	11	9	10	2	5	68
	TTA	11	39	45	16	29	33	4	7	6	3	6	199
	TTG	2	9	6	7	3	1	1	4	7	0	3	43
Met	ATA	12	57	84	17	23	32	2	14	8	4	9	262
	ATG	0	4	8	5	3	7	5	5	5	1	1	44
Asn	AAC	0	3	8	1	8	1	2	5	10	0	5	43
	AAT	4	32	34	13	19	20	2	9	3	6	8	150
Pro	CCA	0	1	2	1	9	6	0	4	4	1	1	29
	CCC	0	0	1	0	0	0	0	0	1	0	2	4
	CCG	0	1	0	0	1	0	2	0	1	0	1	6
	CCT	0	4	5	1	8	2	4	3	5	2	4	38
Gln	CAA	0	2	2	3	3	4	0	0	1	1	2	18
	CAG	0	0	1	1	0	0	0	2	1	0	4	9

continued from Table 5.1

Amino acid	Codon	ND4L	ND4	ND5	ND6	Cytb	ND1	ND3	COIII	ATPase6	ATPase8	COII	Total number
Arg	AGA	2	8	16	4	2	9	1	4	5	0	0	51
	AGG	0	1	1	0	1	0	0	0	1	0	1	5
	CGA	0	2	1	2	4	1	0	3	1	0	4	18
	CGC	0	0	1	0	0	0	0	1	1	0	1	4
	CGG	0	0	0	0	0	0	0	0	0	0	0	0
	CGT	0	1	2	0	0	6	0	0	1	0	0	10
Ser	AGC	0	1	0	0	0	1	0	2	2	0	0	6
	AGT	2	7	3	3	0	4	0	2	0	0	1	22
	TCA	0	13	21	2	15	5	2	6	9	2	10	85
	TCC	0	2	3	0	1	1	2	4	4	1	0	18
	TCG	0	0	2	1	2	0	1	1	1	0	1	9
	TCT	1	11	8	8	5	7	3	6	7	1	9	66
Thr	ACA	0	2	7	0	7	3	0	4	3	0	1	27
	ACC	0	2	0	0	0	0	1	1	1	0	1	6
	ACG	0	0	0	0	0	0	0	1	0	0	0	1
	ACT	2	2	8	0	4	4	3	8	5	0	7	43
Val	GTA	4	15	16	4	13	15	3	10	6	2	6	94
	GTC	0	0	1	0	0	0	4	2	5	0	4	16
	GTG	0	0	0	1	2	1	0	2	1	0	0	7
	GTT	3	7	9	3	5	3	9	8	4	4	5	60
Trp	TGA	1	8	5	1	9	5	2	5	2	2	4	44
	TGG	0	0	0	0	1	0	1	3	2	0	1	8
Tyr	TAC	1	1	8	2	3	4	6	11	6	1	7	50
	TAT	6	25	27	10	21	24	1	1	6	0	5	126

Abbreviations: ND4L= NADH dehydrogenase subunit 4 (light chain), ND4 = NADH dehydrogenase subunit 4, ND5 = NADH dehydrogenase subunit 5, ND6 = NADH dehydrogenase subunit 6, cytb = cytochrome b, ND1 = NADH dehydrogenase subunit 1, ND3 = NADH dehydrogenase subunit 3, COIII = cytochrome c oxidase subunit III, ATP6 = ATP synthase subunit 6, ATP8 = ATP synthase subunit 8, COII = cytochrome c oxidase subunit II, Ala = Alanine, Cys = Cysteine, Asp = Asparagine, Glu = Glutamic acid, Phe = Phenylalanine, Gly = Glycine, His = Histidine, Ile = Isoleucine, Lys = Lysine, Leu = Leucine, Met = Methionine, Asn = Asparagine, Pro = Proline, Gln = Glutamine, Arg = Arginine, Ser = Serine, Thr = Threonine, Val = Valine, Trp = Tryptophan and Tyr = Tyrosine

Table 5.2 Presentation of base compositions (%) in each gene of *T. pagdeni* mtDNA

Genes	Length	A+T content (%)	G+C content (%)
ND4L	300	87	13
ND4	1392	83	17
ND5	1665	83	17
ND6	522	86	14
cyt b	1089	77	23
ND1	933	83	17
ND3	351	66	34
COIII	780	68	32
ATPase6	687	66	34
ATPase8	168	80	20
COII	627	69	31
16S rRNA	1351	77	23
12S rRNA	762	76	24
tRNA-Pro	67	84	16
tRNA-Phe	67	85	15
tRNA-Asn	68	82	18
tRNA-Thr	67	79	21
tRNA-Ser(2)	67	85	15
tRNA-Leu(1)	69	83	17
tRNA-Val	68	88	12
tRNA-Gln	68	79	21
tRNA-Arg	61	85	15
tRNA-Glu	65	71	29
tRNA-Asp	81	78	22
tRNA-Leu(2)	65	75	25

Protein coding genes

The mitochondrial protein-coding genes were analyzed and nucleotide composition, codon usage and size were compared with *M. bicolor* and *A. mellifera*. The 11 complete protein-coding genes of *T. pagdeni* (ND4L, ND4, ND5, ND6, cyt b, ND1, ND3, COIII, ATPase6, ATPase8 and COII) were identified in this study excluding one partial protein-coding gene (COI). We detected four overlapping regions between genes. There were genes involving the reading-frame overlaps on the same strand (ND4 and ND5 shared twenty three nucleotides; cytb and tRNA-Ser (2) shared two nucleotides; ND1 and tRNA-Leu (L1) shared six nucleotides; ATPase6 and ATPase8 shared ten nucleotides) (Table 5.4). Furthermore, fourteen non-coding regions were also detected with sizes ranging form 2 to 85 bp (Table 5.5). The 3 protein-coding genes (ND1, ND3, and ND4L) started with ATA, whereas ND4, COIII and ATPase6 genes with ATG, ND5 with ATC and ND6, cyt b, ATPase8 and COII gene with ATT, which have been commonly found in other bees and other animal mtDNAs (Wolstenholme, 1992). Open-reading frames of *T. pagdeni* ended with TAA (ND4L, ND5, ND6, cyt b, ND1, ND3 and ATPase8) or TAG (ND4, COIII, ATPase6 and COII) (Table 5.6). The sequences of 11 protein-coding genes (ND4L, ND4, ND5, ND6, cyt b, ND1, ND3, COIII, ATPase6, ATPase8 and COII) of *T. pagdeni* mtDNA were analyzed and translated into amino acid sequences (Figure 5.6). Stop codons (TAA or TAG) were not detected within the 10 protein-coding gene sequences (ND4L, ND4, ND5, cyt b, ND1, ND3, COIII, ATPase6, ATPase8 and COII), whereas two stop codons (TAA) were observed within the ND6 gene sequence of *T. pagdeni*. However, the ND6 gene sequence was compared to those of *M. bicolor*, it showed 70 % similarity. However, mtDNA genes sequenced of *T. pagdeni* could be applied to resolve relationship among bees or other insects (Figure 5.7).

Table 5.3 A/T compositions (%) observed in the first, second and third codon position on each protein-coding gene.

Proteins	1 st letter AT (%)	2 nd letter AT (%)	3 rd letter AT (%)
ND4L	80	88	93
ND4	80	80	89
ND5	81	80	89
ND6	87	85	84
cyt b	70	72	88
ND1	77	79	91
ND3	59	76	64
COIII	66	72	66
ATPase6	63	71	62
ATPase8	73	82	84
COII	63	71	73

Table 5.4 Overlapping regions between mitochondrial genes of *T. pagdeni*: involved genes, overlap size (bp) and coding strand.

Genes	Size (bp)	Strand
ND4/ND5	23	+/+
cytb/tRNA-Ser(S2)	2	+/+
ND1/tRNA-Leu(L1)	6	-/-
ATPase8/ATPase6	10	-/-

Table 5.5 Non-coding regions between mitochondrial genes of *T. pagdeni*: flanking genes and size (bp).

Genes	Size (bp)
ND4L/ tRNA-Pro	74
ND4/ND4L	2
tRNA-Phe/ND5	5
tRNA-Phe/tRNA-Asn	18
tRNA-Asn/tRNA-Thr	51
tRNA-Thr/ND6	57
ND6/cytb	59
tRNA-Gln/tRNA-Arg	85
tRNA-Arg /ND3	2
ND3/tRNA-Glu	10
tRNA-Glu /COIII	10
COIII/ATP6	3
ATP8/ tRNA-Asp	2
tRNA-Asp/COII	41

Table 5.6 Size, start codon and stop codon comparisons of protein-coding genes between *T. pagdeni* (Tp), *M. bicolor* (Mb) and *A. mellifera* (Am).

	Size (bp)			Start codon			Stop codon		
Gene	Tp	Mb	Am	Tp	Mb	Am	Tp	Mb	Am
ND4L	300	279	264	ATA	ATA	ATT	TAA	TAA	TAA
ND4	1392	1323	1344	ATG	ATT	ATA	TAG	TAA	TAA
ND5	1665	1647	1665	ATC	ATT	ATT	TAA	TAA	TAA
ND6	522	540	504	ATT	ATT	ATT	TAA	TAA	TAA
cyt b	1089	1050	1152	ATT	ATT	ATG	TAA	TAA	TAA
ND1	933	930	918	ATA	ATA	ATT	TAA	TAA	TAA
ND3	351	354	354	ATA	ATA	ATA	TAA	TAA	TAA
COIII	780	780	777	ATG	ATG	ATG	TAG	TAA	TAA
ATPase6	687	684	681	ATG	ATG	ATG	TAG	TAA	TAA
ATPase8	168	168	159	ATT	ATT	ATT	TAA	TAA	TAA
COII	627	678	676	ATT	ATT	ATT	TAG	TAA	T

> NADH dehydrogenase subunit 4 (light chain) (ND4L)

```
1     ata gtt tat ttt tat aaa ttt tat aaa att aaa gaa ttt ata gtt   45
1      I   V   Y   F   Y   K   F   Y   K   I   K   E   F   I   V   15

46    att ttt gta ttt gta tta att ata ttt tta ata ata ata aaa gat   90
16     I   F   V   F   V   L   I   I   F   L   I   I   I   K   D   30

91    tta tat tat tac ttg aga ttt ctt att att ata gaa ata att cat   135
31     L   Y   Y   Y   L   R   F   L   I   I   I   E   I   I   H   45

136   gta ata ttt tta ttt ata cta att agt ata aat act agt ttc tga   180
46     V   I   F   L   F   I   L   I   S   I   N   T   S   F   w   60

181   att ttt ttt att ttt att act tat tct gta tgt gaa gga att tta   225
61     I   F   F   I   F   I   T   Y   S   V   C   E   G   I   L   75

226   gga tta tta att tta att aga ata aat aat gaa ttt ggt cac cat   270
76     G   L   L   I   L   I   R   I   N   N   E   F   G   H   H   90

271   aag atc aaa tta ttg aat tta tta gtt taa   300
91     K   I   K   L   L   N   L   L   V   *
```

> NADH dehydrogenase subunit 4 (ND4)

```
1     atg aat gat tta ata tgt ata ttt tat att att tta tta tta cct   45
1      M   N   D   L   I   C   I   F   Y   I   I   L   L   L   P   15

46    ata ttt aat cat att gta ttg aat aat tta att ttt tta tca tta   90
16     I   F   N   H   I   V   L   N   N   L   I   F   L   S   L   30

91    cta att tta ata ttt aaa ttc agt tga ctg aat tga aat ttt att   135
31     L   I   L   I   F   K   F   S   W   L   N   W   N   F   I   45

136   tga tta gta ttt agc ttt aat ttc tat tcc att gga ttg atc att   180
46     W   L   V   F   S   F   N   F   Y   S   I   G   L   I   I   60

181   ata ata tta tga att ttt act att atc att atg aac ctg aat aaa   225
61     I   I   L   W   I   F   T   I   I   I   M   N   L   N   K   75

226   gta gaa aat ata aaa ata tct ttg ttt att aat ata ttt ttg ata   270
76     V   E   N   I   K   I   S   L   F   I   N   I   F   L   I   90

271   att ata ata tac ttt gta ttt tat tct ata aat ata att ttt ttt   315
91     I   I   I   Y   F   V   F   Y   S   I   N   I   I   F   F   105

316   tat ttc tct ttt gaa tca aga cta tta tta att ttt tat ata att   360
106    Y   F   S   F   E   S   R   L   L   L   I   F   Y   I   I   120

361   ata aaa tga ggt cat gga gaa ttt cgt ttt agt tct tca ttt tat   405
121    I   K   W   G   H   G   E   F   R   F   S   S   S   F   Y   135

406   tta atg ttt tat acc ata att ttt tca tta cct tta att tat tta   450
136    L   M   F   Y   T   I   I   F   S   L   P   L   I   Y   L   150

451   tta ttt aga cta att aat tct ttc aat aca ata aat ttt tat tta   495
151    L   F   R   L   I   N   S   F   N   T   I   N   F   Y   L   165
```

(continued)

```
496   ttg gaa ata tta aac att aaa gaa atc agt aat ttt aaa ttt att   540
166    L   E   I   L   N   I   K   E   I   S   N   F   K   F   I   180

541   tat att att ttt tct ttt tta gta aaa att cct ata tat ata gtt   585
181    Y   I   I   F   S   F   L   V   K   I   P   I   Y   I   V   195

586   cat gga tga ctt ctt aaa gct cat gta gaa gca tcc ttc ttt aat   630
196    H   G   W   L   L   K   A   H   V   E   A   S   F   F   N   210

631   tct ata att cta gct tca gta ata tta aaa tta gga gga tat ggg   675
211    S   I   I   L   A   S   V   I   L   K   L   G   G   Y   G   225

676   cta ata cga ata ata ttt ttt ata aaa tat ata ttc aat aaa ttc   720
226    L   I   R   I   I   F   F   I   K   Y   I   F   N   K   F   240

721   tat agt tat ttt att ata att aat tta ttc ggt ata tta tca cta   765
241    Y   S   Y   F   I   I   I   N   L   F   G   I   L   S   L   255

766   aga ata ata tgt tta ttt caa atg gat att aaa cta att att gca   810
256    R   I   I   C   L   F   Q   M   D   I   K   L   I   I   A   270

811   att tct tca att gta cat ata gga att ata ctc ata gga att tta   855
271    I   S   S   I   V   H   I   G   I   I   L   I   G   I   L   285

856   tta ata acc aaa ata agg gta tat gga aga ttc tat ata ata att   900
286    L   I   T   K   I   R   V   Y   G   R   F   Y   I   I   I   300

901   agt cat gga ttc att tca tct gga tta ttt tat ttt gtt att tga   945
301    S   H   G   F   I   S   S   G   L   F   Y   F   V   I   W   315

946   ttt ata gtc aaa cta ata gac gac tag   972
316    F   I   V   K   L   I   D   D   *
```

> NADH dehydrogenase subunit 5 (ND5)

```
1     atc ata aaa ata ttg att ttt aga ata ata ttg ctt att aca aga   45
1      I   I   K   I   L   I   F   R   I   I   L   L   I   T   R   15

46    tta att att tta tta ttt tca ata ata ttt tta tca tta aat att   90
16     L   I   I   L   L   F   S   I   I   F   L   S   L   N   I   30

91    gaa tta ata ata gaa tga aat gtc tta aga att aat tca ata aaa   135
31     E   L   I   I   E   W   N   V   L   R   I   N   S   I   K   45

136   ata aac ata att tta gta tta aat tat aaa act tta tta tac ata   180
46     I   N   I   I   L   V   L   N   Y   K   T   L   L   Y   I   60

181   ttt tta gtt ata ttt atc tca tca ata att ttc atg tat aga att   225
61     F   L   V   I   F   I   S   S   I   I   F   M   Y   R   I   75

226   gaa tat ata gaa ttg gaa aaa ttt tta gtt aaa cgc ttt tat tat   270
76     E   Y   I   E   L   E   K   F   L   V   K   R   F   Y   Y   90

271   tta ata ata atg ttt ttg ata tca ata att tta cta att atc aga   315
91     L   I   I   M   F   L   I   S   I   I   L   L   I   I   R   105
```

(continued)

```
316   cct aac atg ctt act att ata ctt gga tga gat ata tta gga ttg   360
106    P   N   M   L   T   I   I   L   G   W   D   I   L   G   L   120

361   aca tca tat tgc tta att att tac tat aga aca att aat tca tat   405
121    T   S   Y   C   L   I   I   Y   Y   R   T   I   N   S   Y   135

406   aac tca gga ata act act gtt ctg ctt aat cgt att gga gat ata   450
136    N   S   G   I   T   T   V   L   L   N   R   I   G   D   I   150

451   agg cta tta ata att att tcg ata ata tca atg ttt gga aga tga   495
151    R   L   L   I   I   I   S   I   I   S   M   F   G   R   W   165

496   aat ctt tta ata tac aga ata aat aaa cct ata ata gtt ata att   540
166    N   L   L   I   Y   R   I   N   K   P   I   I   V   I   I   180

541   att att ata gtt ttt act aag agt gca cag ttt cct ttt ttt gta   585
181    I   I   I   V   F   T   K   S   A   Q   F   P   F   F   V   195

586   tga cta cca ata gca atg ata gct ccc act cca gta tca tca ctt   630
196    W   L   P   I   A   M   I   A   P   T   P   V   S   S   L   210

631   gtt cat tca tca aca cta gta act gca ggt gta tat tta ata att   675
211    V   H   S   S   T   L   V   T   A   G   V   Y   L   I   I   225

676   tga tat aat aaa ata att gat tta aaa tat ata gga ttt att ata   720
226    W   Y   N   K   I   I   D   L   K   Y   I   G   F   I   I   240

721   tca att tct aga att aca ata ctt ttt tca ggt ata ata gct aat   765
241    S   I   S   R   I   T   I   L   F   S   G   I   I   A   N   255

766   tcc gaa ata gat ttt aaa aag atc att gcc ttt tct aca tta aga   810
256    S   E   I   D   F   K   K   I   I   A   F   S   T   L   R   270

811   caa tta gga ttt ata att aga att tta tct ata gga tta aat gaa   855
271    Q   L   G   F   I   I   R   I   L   S   I   G   L   N   E   285

856   tta gct ttc ctt cat tta ttt att cat gct tta ttt aaa tca ata   900
286    L   A   F   L   H   L   F   I   H   A   L   F   K   S   I   300

901   ata ttt ata tgt gta gga aga ttt att cac aat ata aaa gga att   945
301    I   F   I   C   V   G   R   F   I   H   N   I   K   G   I   315

946   caa aat ttc cga ttt tat agt gga ata ttt tat atc tat cct att   990
316    Q   N   F   R   F   Y   S   G   I   F   Y   I   Y   P   I   330

991   aaa aga tct ata att att tta tca tta atg ata ctt tgt ggt ttt  1035
331    K   R   S   I   I   I   L   S   L   M   I   L   C   G   F   345

1036  cct ttt ctt gta gga ttt tat tct aaa gat tta ata att gag ata  1080
346    P   F   L   V   G   F   Y   S   K   D   L   I   I   E   I   360

1081  ttt ata tac aat aaa att agt att ttt aat ttt att gta atc ata  1125
361    F   I   Y   N   K   I   S   I   F   N   F   I   V   I   I   375

1126  att ggt aca ata ata act att tca tat tca ttt cgt att tta tta  1170
376    I   G   T   I   I   T   I   S   Y   S   F   R   I   L   L   390
```

(continued)

```
1171  aaa ttt ttt tct aat aat tac ata ata aat tcc ata att aaa aaa  1215
391    K   F   F   S   N   N   Y   I   I   N   S   I   I   K   K   405

1216  gaa tcg gat att ata aga ttc gta ata gta ttt ata atg att ttt  1260
406    E   S   D   I   I   R   F   V   I   V   F   I   M   I   F   420

1261  ata tta tta ata aga aaa att gtt tat aat ata aat tta att tta  1305
421    I   L   L   I   R   K   I   V   Y   N   I   N   L   I   L   435

1306  ttt aat tgt aat tta ata aat att tac aag tat ttt gtt att aaa  1350
436    F   N   C   N   L   I   N   I   Y   K   Y   F   V   I   K   450

1351  ata ttt att tta gga tat tta tta aat att atc att aac aac ata  1395
451    I   F   I   L   G   Y   L   L   N   I   I   I   N   N   I   465

1396  ata tac aat aaa att gta aat att ata aaa aac tac ttt tat ata  1440
466    I   Y   N   K   I   V   N   I   I   K   N   Y   F   Y   I   480

1441  ata aat ata ttc aaa tta ttg aaa aaa aac tat ttt att gta tta  1485
481    I   N   I   F   K   L   L   K   K   N   Y   F   I   V   L   495

1486  att aaa tat gaa aat aat tat gaa aaa atg ttt aat gaa ata att  1530
496    I   K   Y   E   N   N   Y   E   K   M   F   N   E   I   I   510

1531  ata tca aac ctt ata ata ttt att tcc ata gtt aga tat aaa aat  1575
511    I   S   N   L   I   I   F   I   S   I   V   R   Y   K   N   525

1576  ata gta aaa gta aat gta tct att tat tct ata ata ttt ttt tta  1620
526    I   V   K   V   N   V   S   I   Y   S   I   I   F   F   L   540

1621  tat tta ctt aat cat gat ttt att tat att ata aat ata gta taa  1665
541    Y   L   L   N   H   D   F   I   Y   I   I   N   I   V   *
```

> NADH dehydrogenase subunit 6 (ND6)

```
1     att ata tat ttt att tta aga ata aat tct ttt ata ttg tta att   45
1      I   I   Y   F   I   L   R   I   N   S   F   I   L   L   I   15

46    ttc ttt gta aat ata tat att tct cag att tct tcg gtt cca ata   90
16     F   F   V   N   I   Y   I   S   Q   I   S   S   V   P   I   30

91    aat caa ttg att aac att att ttt ttc atg att tga aga tct ttt  135
31     N   Q   L   I   N   I   I   F   F   M   I   W   R   S   F   45

136   gtt ttg ttt atg gta aga agt cat ata att att tat gtt ttt ata  180
46     V   L   F   M   V   R   S   H   I   I   I   Y   V   F   I   60

181   att tta att gta atg atc aga ggt ata ata att ttg ttt tct tat  225
61     I   L   I   V   M   I   R   G   I   I   I   L   F   S   Y   75

226   ttt gta tgt tta agt aat ata atg aat ata aaa aaa att aag ttc  270
76     F   V   C   L   S   N   I   M   N   I   K   K   I   K   F   90

271   aaa tac att tat tta ttg aat tac ata tta att tat att tca ata  315
91     K   Y   I   Y   L   L   N   Y   I   L   I   Y   I   S   I  105
```

(continued)

```
316   att ttt tct gag aat tta taa tta taa tat tat aat aaa ata ttt   360
106    I   F   S   E   N   L   *   L   *   Y   Y   N   K   I   F   120

361   cga aat caa gat tta aat cat gac caa tta att att aaa tta ttc   405
121    R   N   Q   D   L   N   H   D   Q   L   I   I   K   L   F   135

406   aat ttt cct aat tat tat ata tta tta att att att gtg ttt ata   450
136    N   F   P   N   Y   Y   I   L   L   I   I   I   V   F   I   150

451   ttt ttg ata tta atg tta tct tca aaa att tgc ttt agt aaa aaa   495
151    F   L   I   L   M   L   S   S   K   I   C   F   S   K   K   165

496   aaa tct tta cga aat aaa ttg ttt taa   522
166    K   S   L   R   N   K   L   F   *
```

> Cytochrome b (cyt b)

```
1     att tcg att cca att cca ata aac att aat tat ttc tga aat ttt   45
1      I   S   I   P   I   P   I   N   I   N   Y   F   W   N   F   15

46    ggt tca ata tta gga ata ttt tta gta att caa att att tca ggt   90
16     G   S   I   L   G   I   F   L   V   I   Q   I   I   S   G   30

91    ttg ttt ctc tcg ata cat tat tgt cca aat att gat tat gca ttt   135
31     L   F   L   S   I   H   Y   C   P   N   I   D   Y   A   F   45

136   caa aga gtt tca tat att ata aaa gat gta aat tca ggc tga tta   180
46     Q   R   V   S   Y   I   I   K   D   V   N   S   G   W   L   60

181   gtt cga ttg atc cat ata aat ggg gct tca ttt tat ttt att tta   225
61     V   R   L   I   H   I   N   G   A   S   F   Y   F   I   L   75

226   att tat gca cat att ata cga gga ata tat tat tat tct ttt aaa   270
76     I   Y   A   H   I   I   R   G   I   Y   Y   Y   S   F   K   90

271   tta act aga gta tga tta att gga agg ata att aca ttt tta tca   315
91     L   T   R   V   W   L   I   G   R   I   I   T   F   L   S   105

316   ata gct aca gca ttt ctt ggg tat gta ctt cca tga gga cca ata   360
106    I   A   T   A   F   L   G   Y   V   L   P   W   G   P   I   120

361   tca ttt tga gga gct ata gta att aca aat tta tta tca gca att   405
121    S   F   W   G   A   I   V   I   T   N   L   L   S   A   I   135

406   cca tat gta ggt aat ata att gta gaa tga tta tga gga gga ttc   450
136    P   Y   V   G   N   I   I   V   E   W   L   W   G   G   F   150

451   tca att aat aat tcc act ctt aat cga ttt ttt tct ttt cac ttt   495
151    S   I   N   N   S   T   L   N   R   F   F   S   F   H   F   165

496   att ctt ccc ttc att atc tta ttt ttt gta att tta cat tta tta   540
166    I   L   P   F   I   I   L   F   F   V   I   L   H   L   L   180

541   ata tta cac aaa tct ggt tct tca aac cct tta cat tca aaa atc   585
181    I   L   H   K   S   G   S   S   N   P   L   H   S   K   I   195
```

(continued)

```
586   gat gtt tat aaa att gct ttc cac cct tat ttt atg att aag gat   630
196    D   V   Y   K   I   A   F   H   P   Y   F   M   I   K   D   210

631   tta gtg aca att act tta att tta tca tta ttt ata att gta aat   675
211    L   V   T   I   T   L   I   L   S   L   F   I   I   V   N   225

676   ctt caa gta ccg tat ttt tta ggt gat cca gat aac ttt aaa ata   720
226    L   Q   V   P   Y   F   L   G   D   P   D   N   F   K   I   240

721   gct gat cct ata gtt act cca tta cat att aaa cct gag tga tac   765
241    A   D   P   I   V   T   P   L   H   I   K   P   E   W   Y   255

766   ttt ctt ttt gcc tat tca att tta cga tct att cct aat aag cta   810
256    F   L   F   A   Y   S   I   L   R   S   I   P   N   K   L   270

811   gga gga gta att ata ctt ttt ata tca att ttt ata ctt tat ctt   855
271    G   G   V   I   I   L   F   I   S   I   F   I   L   Y   L   285

856   ctt cct atg tta aat ata aac aac ata aaa aat att aaa ttt tat   900
286    L   P   M   L   N   I   N   N   I   K   N   I   K   F   Y   300

901   cca atc aac cat ttt att tat tgg aca ttt att aat aat gtt att   945
301    P   I   N   H   F   I   Y   W   T   F   I   N   N   V   I   315

946   gtg tta aca tga cta gga ggg aaa gct att gaa aac cct ttt att   990
316    V   L   T   W   L   G   G   K   A   I   E   N   P   F   I   330

991   gaa ttg aac att gta ttt aca ttt ata tac ttt ttt tat tat tta  1035
331    E   L   N   I   V   F   T   F   I   Y   F   F   Y   Y   L   345

1036  ttt tca ttt gta tta aat aat tta att gat att tta atg tac aat  1080
346    F   S   F   V   L   N   N   L   I   D   I   L   M   Y   N   360

1081  aaa tat taa   1089
361    K   Y   *
```

> NADH dehydrogenase subunit 1 (ND1)

```
1     ata ata att ttt gta cta att aat tta tta att ata gta tta ata   45
1      I   I   I   F   V   L   I   N   L   L   I   I   V   L   I   15

46    gtt atg att aga gta gct ttt cta act tta ttt gaa cgt aag att   90
16     V   M   I   R   V   A   F   L   T   L   F   E   R   K   I   30

91    tta aga tat atg caa tgt cgt aaa ggt cca aat aaa tta tat tat  135
31     L   R   Y   M   Q   C   R   K   G   P   N   K   L   Y   Y   45

136   aag ggt att cta caa cca ttt agc gat ata atc aaa ctt cta act  180
46     K   G   I   L   Q   P   F   S   D   I   I   K   L   L   T   60

181   aag gaa atg ttt gat ttc agt ata aat tat ata ttc tat tat agt  225
61     K   E   M   F   D   F   S   I   N   Y   I   F   Y   Y   S   75

226   cca tta tta ata ttt att gta tca tca att ttg tga tta tta tat  270
76     P   L   L   I   F   I   V   S   S   I   L   W   L   L   Y   90
```

(continued)

```
271   cca tga att ttc aat aat tta aat ttt aat tac agt ata ctt tat   315
91     P   W   I   F   N   N   L   N   F   N   Y   S   I   L   Y   105

316   ata att tta att att aga att aat gta tat cca att tta ata atc   360
106    I   I   L   I   I   R   I   N   V   Y   P   I   L   I   I   120

361   aga tga att tct aca aat aat tat tct ata att aga gta ata cgt   405
121    R   W   I   S   T   N   N   Y   S   I   I   R   V   I   R   135

406   ata gtt tca caa gta att tca ttc gaa gta tta atg tac ata atg   450
136    I   V   S   Q   V   I   S   F   E   V   L   M   Y   I   M   150

451   ata ttt att ctt ata ata ttc ttt aac aga tat tct atg tca aat   495
151    I   F   I   L   I   I   F   F   N   R   Y   S   M   S   N   165

496   tct att aat tat caa ata aat att aaa tta ttt att ttt tct tat   540
166    S   I   N   Y   Q   I   N   I   K   L   F   I   F   S   Y   180

541   cca ata tat ttt atc ttt att ctt aga tta tta gta gat tta aat   585
181    P   I   Y   F   I   F   I   L   R   L   L   V   D   L   N   195

586   cga gtt cct ttt gat cta gta gaa gga gaa tct gaa tta gta tct   630
196    R   V   P   F   D   L   V   E   G   E   S   E   L   V   S   210

631   gga ttt aat att gaa tat tac agt aga tta ttt aca tta att ttt   675
211    G   F   N   I   E   Y   Y   S   R   L   F   T   L   I   F   225

676   tta tcc gaa tat ata aat ata att ttt atg aga gta att tta gta   720
226    L   S   E   Y   I   N   I   I   F   M   R   V   I   L   V   240

721   att tta ttt tat ggt ata ttt tat tga aat ttt ttt ttc aat ata   765
241    I   L   F   Y   G   I   F   Y   W   N   F   F   F   N   I   255

766   ttt ttt att att aat tta att tta atc gtg ata ata cgt gga gta   810
256    F   F   I   I   N   L   I   L   I   V   I   I   R   G   V   270

811   tta cct cgt att cgt tac gat tat cta ata tat act tgt tga ata   855
271    L   P   R   I   R   Y   D   Y   L   I   Y   T   C   W   I   285

856   gaa tta tta gta tta ata act tat tat tta att tat tgt tat tta   900
286    E   L   L   V   L   I   T   Y   Y   L   I   Y   C   Y   L   300

901   ttt aaa gaa tta att ata ata aca aat ata taa   933
301    F   K   E   L   I   I   I   T   N   I   *
```

> NADH dehydrogenase subunit 3 (ND3)

```
1     ata gtt tac tta tta ttc ctt ttt ctt gcc att gct gtt tcg tcc   45
1      I   V   Y   L   L   F   L   F   L   A   I   A   V   S   S   15

46    atg gtc cta tac ttg aac aaa att att tcc ccg aaa aaa tca gtc   90
16     M   V   L   Y   L   N   K   I   I   S   P   K   K   S   V   30

91    tga gta gaa aaa aag gtt ccg ttc gag tgc ggt ttc aac cct atc   135
31     W   V   E   K   K   V   P   F   E   C   G   F   N   P   I   45
```

(continued)

```
136   tca aag ttc tct ctt cct gta tct ata cct ttc ttt ctt gtt gca   180
46     S   K   F   S   L   P   V   S   I   P   F   F   L   V   A    60

181   att atc ttc ctt atc ttt gac att gaa att act ctt ctt gtc cct   225
61     I   I   F   L   I   F   D   I   E   I   T   L   L   V   P    75

226   gta att act tac cta aat tac ctt aga tct atg tac acc att ctt   270
76     V   I   T   Y   L   N   Y   L   R   S   M   Y   T   I   L    90

271   tta gtt gtt ctt ttt gtc tta gtt atg cta gtt act cta gtt ctt   315
91     L   V   V   L   F   V   L   V   M   L   V   T   L   V   L   105

316   gaa tgg ttc atg gga tat ctc aat tga atg tac taa   351
106    E   W   F   M   G   Y   L   N   W   M   Y   *
```

> Cytochrome c oxidase subunit III (COIII)

```
1     atg aaa aaa aat ttt cca tac ctc tta gtg act atc agc cca tgg    45
1      M   K   K   N   F   P   Y   L   L   V   T   I   S   P   W    15

46    cct atc att gct tct ttt aat cta ata aac ttt ctt gtt aga att    90
16     P   I   I   A   S   F   N   L   I   N   F   L   V   R   I    30

91    gta att tga atc aat ctt aaa gaa tac tca att gta tct ttc aca   135
31     V   I   W   I   N   L   K   E   Y   S   I   V   S   F   T    45

136   tta ttt aat ttg tcc ata gtt ttt gct tta tga aat cga gat atc   180
46     L   F   N   L   S   I   V   F   A   L   W   N   R   D   I    60

181   att cga gaa agt act ttc ata gga ggc cat acc cta gtt gta aaa   225
61     I   R   E   S   T   F   I   G   G   H   T   L   V   V   K    75

226   ata atg atc aag ttt agc ata ttt ata ttt att tta tcc gaa tta   270
76     I   M   I   K   F   S   I   F   I   F   I   L   S   E   L    90

271   ttt ttt ttt atc tca ttt ttc tga act ttc ttt cac aga tca att   315
91     F   F   F   I   S   F   F   W   T   F   F   H   R   S   I   105

316   tct cca tct att gaa att aac ata gta tgg cct cca aaa ata gtt   360
106    S   P   S   I   E   I   N   I   V   W   P   P   K   I   V   120

361   cag gta ttt aat tac act gaa att cct ctc ctc aat act ctt act   405
121    Q   V   F   N   Y   T   E   I   P   L   L   N   T   L   T   135

406   cta att aca tca gga ttc ttt gta act cta agt cac ttg aat tta   450
136    L   I   T   S   G   F   F   V   T   L   S   H   L   N   L   150

451   gta ata aat aag ctt tct aag aga ttg gca acg ctg ttt tac aca   495
151    V   I   N   K   L   S   K   R   L   A   T   L   F   Y   T   165

496   att ctt atg gga atg tac ttt tct tta gtt cag ata ctt gaa tac   540
166    I   L   M   G   M   Y   F   S   L   V   Q   I   L   E   Y   180

541   ttc aac gca ggc ttc tgc att aac gac aga gtt tac ggg tcg atc   585
181    F   N   A   G   F   C   I   N   D   R   V   Y   G   S   I   195
```

(continued)

```
586   ttc ttc atc tcc act gga ttt cac gga att cat gtg ctt gta gga   630
196    F   F   I   S   T   G   F   H   G   I   H   V   L   V   G   210

631   aca ata ttc ttg ata gtc tca ctt gct cga ata cgc ata atg cat   675
211    T   I   F   L   I   V   S   L   A   R   I   R   I   M   H   225

676   ttc tcc gta att cac cac att aac tac gag cta tca gta tgg tac   720
226    F   S   V   I   H   H   I   N   Y   E   L   S   V   W   Y   240

721   tga cac ttt gtt gat gtc atc tga ctc ttc ctt tac ttc ttc tat   765
241    W   H   F   V   D   V   I   W   L   F   L   Y   F   F   Y   255

766   tac gtt cta atc tag   780
256    Y   V   L   I   *
```

> ATP synthase subunit 6 (ATPase 6)

```
1     atg aaa tta aaa gtg tta aac ctg ttt gaa aga ttt gat cca tca   45
1      M   K   L   K   V   L   N   L   F   E   R   F   D   P   S   15

46    gtt tac tac gtc tat act ttc cag ctt aac tgg gta ttc tcc cta   90
16     V   Y   Y   V   Y   T   F   Q   L   N   W   V   F   S   L   30

91    tct tcc ctg gtt ttt cta gct gga gga tac tga gtc att cct tcc   135
31     S   S   L   V   F   L   A   G   G   Y   W   V   I   P   S   45

136   cgc ctt gtc ata ata tgt act tta ctc ttg tct att ctg ttc aac   180
46     R   L   V   I   I   C   T   L   L   L   S   I   L   F   N   60

181   gag ttt tct cta gca atg tga tac aag aag ctt act cca aac tca   225
61     E   F   S   L   A   M   W   Y   K   K   L   T   P   N   S   75

226   ttg att ttt atc agc cta ata ttt tat acc ata tta atg aac ttt   270
76     L   I   F   I   S   L   I   F   Y   T   I   L   M   N   F   90

271   cta agc cta ttt cca tat att ttt cca act aca aga cat ctt ctg   315
91     L   S   L   F   P   Y   I   F   P   T   T   R   H   L   L   105

316   ttc aat cta tca ttg tct tta ccc cta tgg tca aga ttt ttc tta   360
106    F   N   L   S   L   S   L   P   L   W   S   R   F   F   L   120

361   tac tca att att aca aac ccg ata aaa ttc ttt gct cat ctt gtc   405
121    Y   S   I   I   T   N   P   I   K   F   F   A   H   L   V   135

406   cct cac aac tct cct aaa gcc ttg ata aat ttc atg gta att att   450
136    P   H   N   S   P   K   A   L   I   N   F   M   V   I   I   150

451   gaa ctg atc aga tat ctc atc cgt cct ctt aca ttg tct att cga   495
151    E   L   I   R   Y   L   I   R   P   L   T   L   S   I   R   165

496   ctt tcc tca aat ctc atc tcg gga cac ctg att ctt att ctt cta   540
166    L   S   S   N   L   I   S   G   H   L   I   L   I   L   L   180

541   aga aac ttt gta ata aac ttc gta tat act ttc cct atc ata tca   585
181    R   N   F   V   I   N   F   V   Y   T   F   P   I   I   S   195
```

(continued)

```
586   gtc gtt gag aac att ctg ttg ttg cta gag gtt tca atg tca gta   630
196    V   V   E   N   I   L   L   L   L   E   V   S   M   S   V   210

631   att caa gct tac gta ttc tct att ctt cta gct ctg tat ctg aaa   675
211    I   Q   A   Y   V   F   S   I   L   L   A   L   Y   L   K   225

676   gaa agg atc tag   687
226    E   R   I   *
```

> ATP synthase subunit 8 (ATPase 8)

```
1     att cct caa ata atg cca gtt aat tga atc tta gtt ttc ctt ata   45
1      I   P   Q   I   M   P   V   N   W   I   L   V   F   L   I   15

46    aat att att ttt tta att cta gta att gtt tta ctt aat tca ttt   90
16     N   I   I   F   L   I   L   V   I   V   L   L   N   S   F   30

91    tac ata gta aat tca tgt aat tct att ttc tcc aaa gac cct aaa   135
31     Y   I   V   N   S   C   N   S   I   F   S   K   D   P   K   45

136   aag gtt aaa att ttc aaa ata gaa tga aat taa   168
46     K   V   K   I   F   K   I   E   W   N   *
```

>Cytochrome c oxidase subunit II (COII)

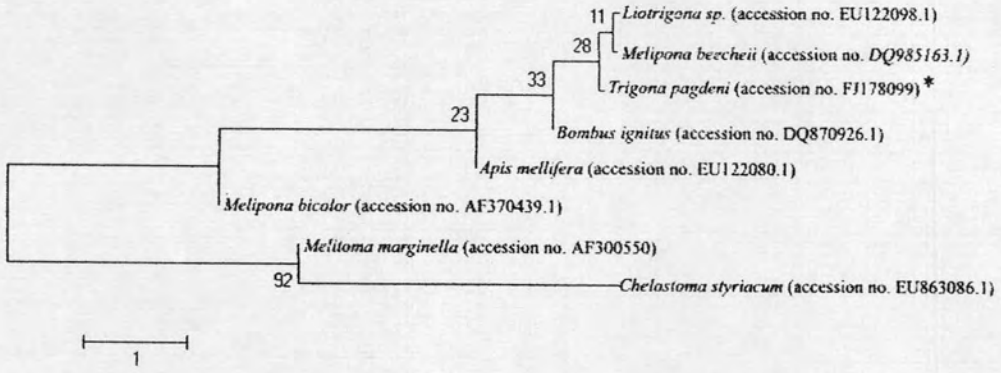
```
1     att tca aca tga aac ata tac tca ttt cag gac tca aat tct att   45
1      I   S   T   w   N   I   Y   S   F   Q   D   S   N   S   I   15

46    tac tca gac aat cta att tca ttt cat aat ttt gtc ata ata ttt   90
16     Y   S   D   N   L   I   S   F   H   N   F   V   I   I   F   30

91    att att gta att act tca ttg act atg ttt ttt atc att gat ttt   135
31     I   I   V   I   T   S   L   T   M   F   F   I   I   D   F   45

136   gta ctt aat tca ttt tta aac tta agg gtt ctt aaa aat cac act   180
46     V   L   N   S   F   L   N   L   R   V   L   K   N   H   T   60

181   gtc gaa gta gtc tgg act att att cca ata att att ctt ctc cta   225
61     V   E   V   V   W   T   I   I   P   I   I   I   L   L   L   75

226   att tgt tac cct tct ctc aaa atc ttg tat ttt att gat gaa att   270
76     I   C   Y   P   S   L   K   I   L   Y   F   I   D   E   I   90

271   tct aac ccc tac ttt tca att aag gct att gga cat caa tga tat   315
91     S   N   P   Y   F   S   I   K   A   I   G   H   Q   w   Y   105

316   tga tct tat gaa tac ccc gaa ttt aac aat cat gaa ata agt tct   360
106    w   S   Y   E   Y   P   E   F   N   N   H   E   I   S   S   120

361   tat ata cag att tac tcg gac ata gat cat ttt cgc ttg att gaa   405
121    Y   I   Q   I   Y   S   D   I   D   H   F   R   L   I   E   135

406   act gat aat cga tta gtc gtt ccg ttc aag att tca att cga ctg   450
136    T   D   N   R   L   V   V   P   F   K   I   S   I   R   L   150
```

(continued)

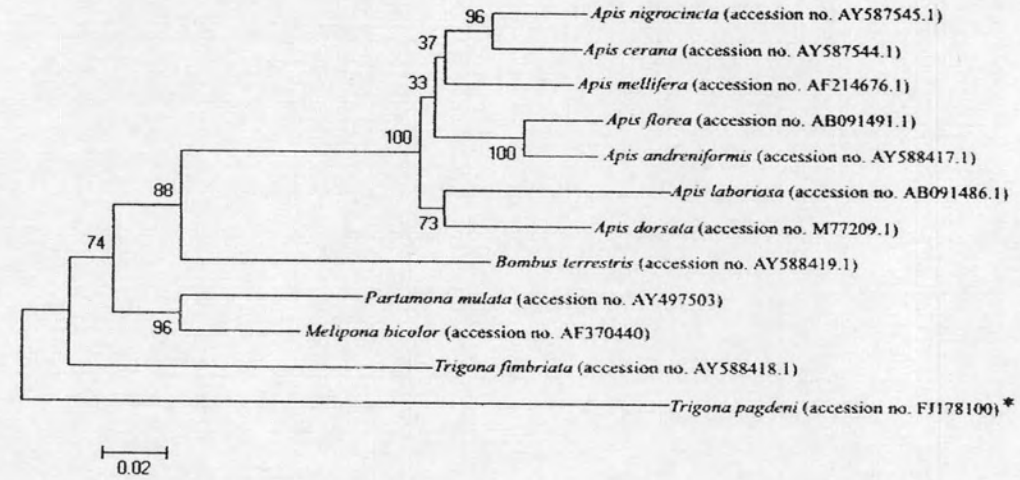
```
451   atc gtt tca tct ctt gac gta att cac tct tga act att caa tct   495
151    I   V   S   S   L   D   V   I   H   S   w   T   I   Q   S    165

496   ctt gga gtt aaa gta gac gca gtt cct gga cga atc aat cag ctc   540
166    L   G   V   K   V   D   A   V   P   G   R   I   N   Q   L    180

541   aac cta ata cct act cga cct ggt tta tat ttt gga cag tgc tct   585
181    N   L   I   P   T   R   P   G   L   Y   F   G   Q   C   S    195

586   gag atc tgc gga gta acc ata gat tta tac cta tta tgc tag       627
196    E   I   C   G   V   T   I   D   L   Y   L   L   C   *
```

Figure 5.6 A presentation for the protein-coding genes of the mtDNA sequence of *T. pagdeni,* numbers on both besides of the sequences indicate nucleotide positions. The corresponding amino acid is indicated with the genetic codes and amino acid codes. The initial codons were the first three nucleotides of each gene. The stop codons are marked by an asterisk.

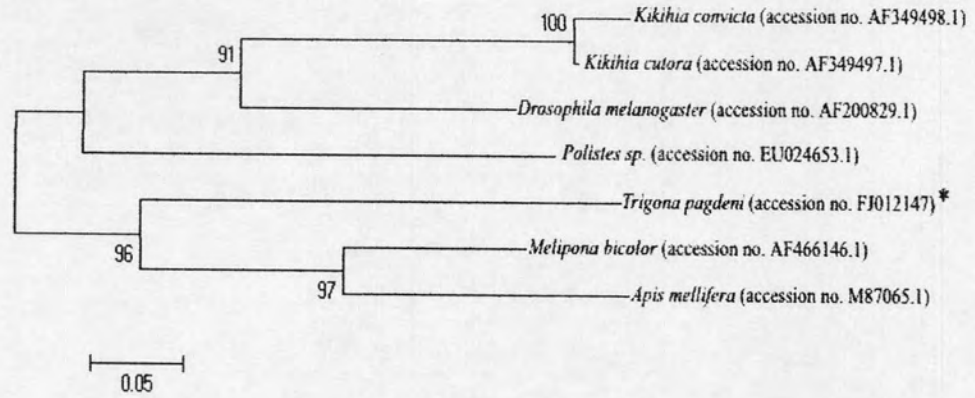
A.


11
Liotrigona sp. (accession no. EU122098.1)
28
Melipona beecheii (accession no. DQ985163.1)
33
Trigona pagdeni (accession no. FJ178099) *
23
Bombus ignitus (accession no. DQ870926.1)
Apis mellifera (accession no. EU122080.1)
Melipona bicolor (accession no. AF370439.1)
Melitoma marginella (accession no. AF300550)
92
Chelostoma styriacum (accession no. EU863086.1)
1

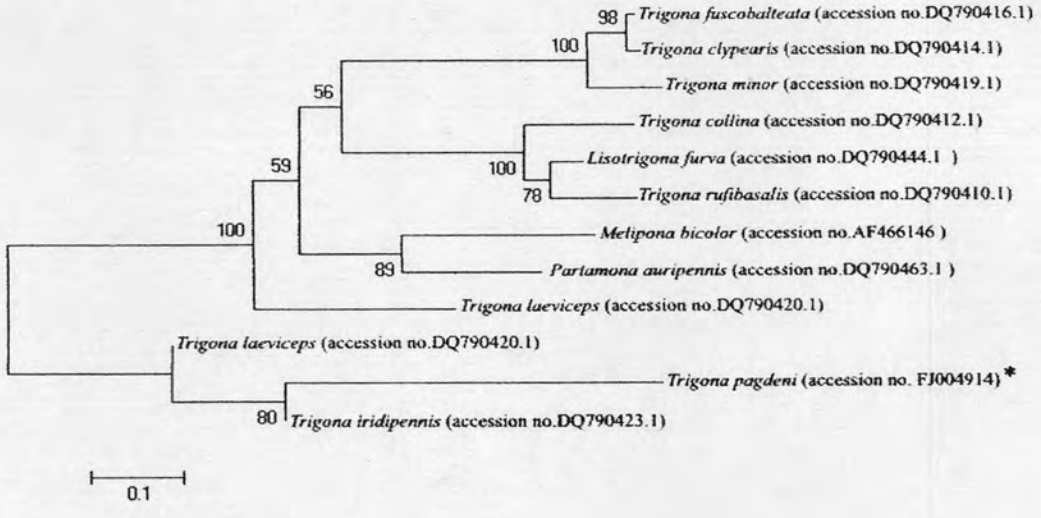
B.


96
Apis nigrocincta (accession no. AY587545.1)
37
Apis cerana (accession no. AY587544.1)
33
Apis mellifera (accession no. AF214676.1)
Apis florea (accession no. AB091491.1)
100
100
Apis andreniformis (accession no. AY588417.1)
88
Apis laboriosa (accession no. AB091486.1)
73
Apis dorsata (accession no. M77209.1)
74
Bombus terrestris (accession no. AY588419.1)
Partamona mulata (accession no. AY497503)
96
Melipona bicolor (accession no. AF370440)
Trigona fimbriata (accession no. AY588418.1)
Trigona pagdeni (accession no. FJ178100) *
0.02

C.


100
Kikihia convicta (accession no. AF349498.1)
91
Kikihia cutora (accession no. AF349497.1)
Drosophila melanogaster (accession no. AF200829.1)
Polistes sp. (accession no. EU024653.1)
Trigona pagdeni (accession no. FJ012147) *
96
Melipona bicolor (accession no. AF466146.1)
97
Apis mellifera (accession no. M87065.1)
0.05

D



E.

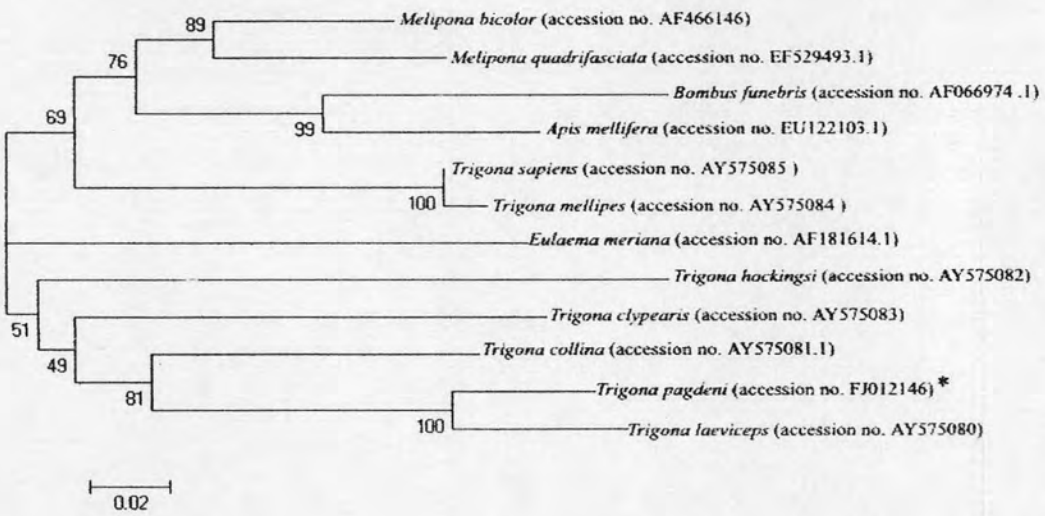


Figure 5.7 A neighbor-joining tree summarizing genetic relationships between insect species analyzed from each gene sequence (deposited in GenBank); COI gene (A), COII gene (B), ATPase6 gene (C), 16S rRNA gene (D),cytb gene (E). *Trigona pagdeni* was used as our samples in this study and indicated by asterisk *. Numbers above the branches are measures (0-100) of the robustness of the branch determined from the frequency the branch was obtained in 1000 bootstrap replications.

Ribosomal RNA genes

The 12S and 16S rRNA genes of *T. pagdeni* mtDNA could be estimated with length of 762 and 1351 bp, respectively. Our sequence showed 83% and 78% similarity from the partial 12S nucleotide rRNA sequences (437 bp) and complete 16S rRNA nucleotide sequences (1354 bp) genes of *Melipona bicolor*, respectively. The srRNA and lrRNA genes of *T. pagdeni* were located between the tRNA-Leu (1) and tRNA-Val gene and between the tRNA-Val and tRNA-Gln genes, respectively (Figure 5.9).

Transfer RNA genes

Twelve tRNA genes were identified in the mitochondrial sequence of *T.pagdeni* by eye using comparison with homologues of *Apis mellifera*, *Melipona bicolor* and *Bombus ignitus* (Figure 5.8). All tRNA sequences had 61- 81 nucleotides with 12 to 29% G+C and 70.77 to 88% A+T (Table 5.2). The anticodon nucleotides were identical to those commonly found for the corresponding tRNA genes in *Apis mellifera*, *Melipona bicolor* and *Bombus ignitus* mtDNA (Crozier and Crozier, 1993; Silvestre *et al.*, 2008; Young Cha *et al.*, 2007). However, the tRNA-Gln found in *T. pagdeni* mtDNA resembled those in *Apis mellifera* and *Bombus ignitus* mtDNA, but not seen in *Melipona bicolor* mtDNA (Crozier and Crozier, 1993; Silvestre *et al.*, 2008; Young Cha *et al.*, 2007). The tRNA- Gln gene was located between 12S and tRNA- Arg (Figure 5.9).

tRNA-Pro

```
>tRNA-Pro Bi     CAAAAAATAGTTTAATTAAAA-TAATAATTTTGGGAATTATTGATATTTTTA--AGAAAATTTTTTTGA
>tRNA-Pro Mp     -..............AA....A.............G...G.A....A...-.--.A...T.........
>tRNA-Pro Tp     --G...........T.A....A.T..........A..........GA.C..GTT.ATG.T.....C...
```

tRNA-Phe

```
>tRNA-Phe Ap     ATTTAAATAGCTTATAT--TTAGAGCGTAATATTGAAAATATTAATGAAAATTTTTTAAATTTTTTAAATA
>tRNA-Phe Bi     ..............TA--G......A..................G....---..A.TTTA..........
>tRNA-Phe Mp     ................--.-.....A.........................---...A...A..........
>tRNA-Phe Tp     ..............AA.......A.....................G.----.CAT.GA..........
```

tRNA-Asn

```
>tRNA-Asn Ap     TTTAGTTAGAATTTTTTAAATTCATATGATTTATTAACAATAAATTGCTAAT-TATTTAGCTTTAACTAA
>tRNA-Asn Bi     A.A.A.......A...-.....ATGAT.................A...CTC--.T...G.......T.T.
>tRNA-Asn Mb     A.A.A.......----.TT.....ATAAT.............A...CTC--.T...G.......T.T.
>tRNA-Asn Tp     .AA.........AA.--.........TAAT.............A...CTCAA.C...G.......T.T.
```

tRNA-Thr

```
>tRNA-Thr Ap     TAGCTAAAATATTATAATGA-ATTTTAATATAATTTT-AATTTACAAAATTAATGTTTTTAAATTTTAAACTATTTAGCT-
>tRNA-Thr Bi     --------T.G.....T.A.TG..A....TA......T.....................----.........AA.A.A-
>tRNA-Thr Mb     --------.AG.....-.A.TTAT.A..AT.....A.T...................AT----.........A..A..T
>tRNA-Thr Tp     --------.AGA...G-.A.CT--.C.TGT.....A.C...........CCG........T.---.........CA..A..T
```

tRNA-Ser(2)

```
>tRNA-Ser(2)Am   -AGTTAATGAACTTGAATAAGTATATATTTTGAAAATATAATATAGAAATAAAATTTTCTATTAACTT
>tRNA-Ser(2)Bi   A.......AG.AAAT.T..TT.C....ATA...TTTC.A.CC...A.CT..TTCAAG.TC.......-
>tRNA-Ser(2)Mb   A.A.....AG.AAATTT..TT.C....T.....TTTC.A.TC...T.CT..T-CAAG.TC.....T.-
>tRNA-Ser(2)Tp   A.A.....................G.........G.........................T.-
```

(continued)

tRNA-Leu(1)

```
>tRNA-Leu(1)Bi   ----ATATATTATAAATAATTATAATTTA-TTTAAATTCTAAATTTAATGCACTAAATATGCTAATATA--
>tRNA-Leu(1)Mb   TATT..T........T.T..-........A.................AA......-....T...A...AT
>tRNA-Leu(1)Tp   TATT..T..A....-T.GTC....T.C..A.................T......G.-.C..CC..A...AT
```

tRNA-Val

```
>tRNA-Val Am     TAAAATTTTAAAAACTAATTTTTTAATCTTTTCACTGTAAAAGAAATATTTTACTTT-----AAACTAAAATTTT
>tRNA-Val Bi     ...........TC.TC..A..---..A.......T......T..........TT...TATTT.........C..
>tRNA-Val Mb     .........C...TA.C....A--.T...A..T.A......T-.....A......A.-----.............
>tRNA-Val Tp     ........CT.G..TA..C.CTA-GT.AAA..T.T......TA...AG.....TA.A------.............
```

tRNA-Gln

```
>tRNA-Gln Am     ----ATTTATTAATT-TAGTTAATACTATAAAATTCAAAATTTTATGTGC-TTTAAACACTAAAATAAA
>tRNA-Gln Bi     TTTATA.........A.TA.....T.AT.T...A.........A..A...-AA..T...T....TAT..
>tRNA-Gln Tp     TTTACAA.......C-.TA.....T.G..C...A.........G.CA...TAA.TT...TC..GTGT..
```

tRNA-Arg

```
>tRNA-Arg Am     ATATAAGAAGTAATTATTTACAATTTAATTTCGACTTAAATATTGATTT-ATAAAAAATCCTTATATT
>tRNA-Arg Bi     .............AGT.A.TACT...................A.....T.................A
>tRNA-Arg Mb     -A...T........AT--.TAC..................C..AA..CA--.TTTTG..T.....T.-
>tRNA-Arg Tp     .AG..TA......AA.-A.TAC..................A...A.C..--.....T.G.TTG..----
```

tRNA-Glu

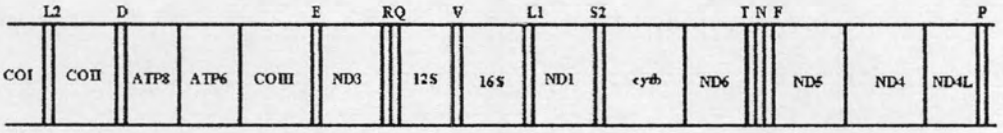
```
>tRNA-Glu Am     -ATTTATATAGTTTAAAAA--AAACATTATATTTTCAATATAAAA------ATAATTAAATT-TA-ATTTATAAATA
>tRNA-Glu Bi     A................TT-..........C.............TTATTA....A.....AA.T-.A.........
>tRNA-Glu Mb     -.AAG..........TT.TTA..........T.............---------..A.TT..AATT-.A.....TT..
>tRNA-Glu Tp     ---A..C.....A...-.CTATT...C.TG.........CA.G..---------.T.CGG...A.CC.GA.GC.TGC.
```

(continued)

```
tRNA-Asp

>tRNA-Asp Am     TAAAAAATAATTTATGAATAAATTTATTTTAGTTTGACAAACTAATGTTATAT---TATTAACTAATTTTTT----------------
>tRNA-Asp Bi     AG.........AA..T..ATT...A.......AA......T..........T.A---.............C----------------
>tRNA-Asp Mb     --.......T.A....-T..............AAA.....TTT...........ATA...............ATTATTAATAGATT--
>tRNA-Asp Tp     ---.....T.AC.TCA.C.TGG....C..C.AAA.....TTTCG.......T.----.............CACAAGTTTCCTAAATT

tRNALeu(2)

>tRNA-Leu(2)Bi   TATTAATAAAAGAAAAAAAAATCTTTATTATTGAATTTTAAATTCAAAGCACT-AATCTGCCATATTAAT
>tRNA-Leu(2)Mb   .T.........ATT.T.T.---T.....A......A................T..............A
>tRNA-Leu(2)Tp   .T....CC....C....G----T.....AC......C...............-..............A
```

Figure 5.8 Alignment of the mitochondrial tRNA genes from *T. pagdeni*, mitochondrial tRNA genes were identified and aligned according to the criteria described in the text, using the reported complete mtDNA sequences for honey bees *(A. mellifera*; Crozier and Crozier, 1993 and *Bombus ignites*;Young Cha *et al.*, 2007) and stingless bees (*Melipona bicolor*; Silvestre *et al.*, 2008). Dots indicate identity to the *A. mellifera* mtDNA sequences and the bases are revealed only where different from *A. mellifera*. The optimal alignments are shown with internal gaps in these regions. Anticodon positions in tRNA genes were marked. in bold.

A.



B.

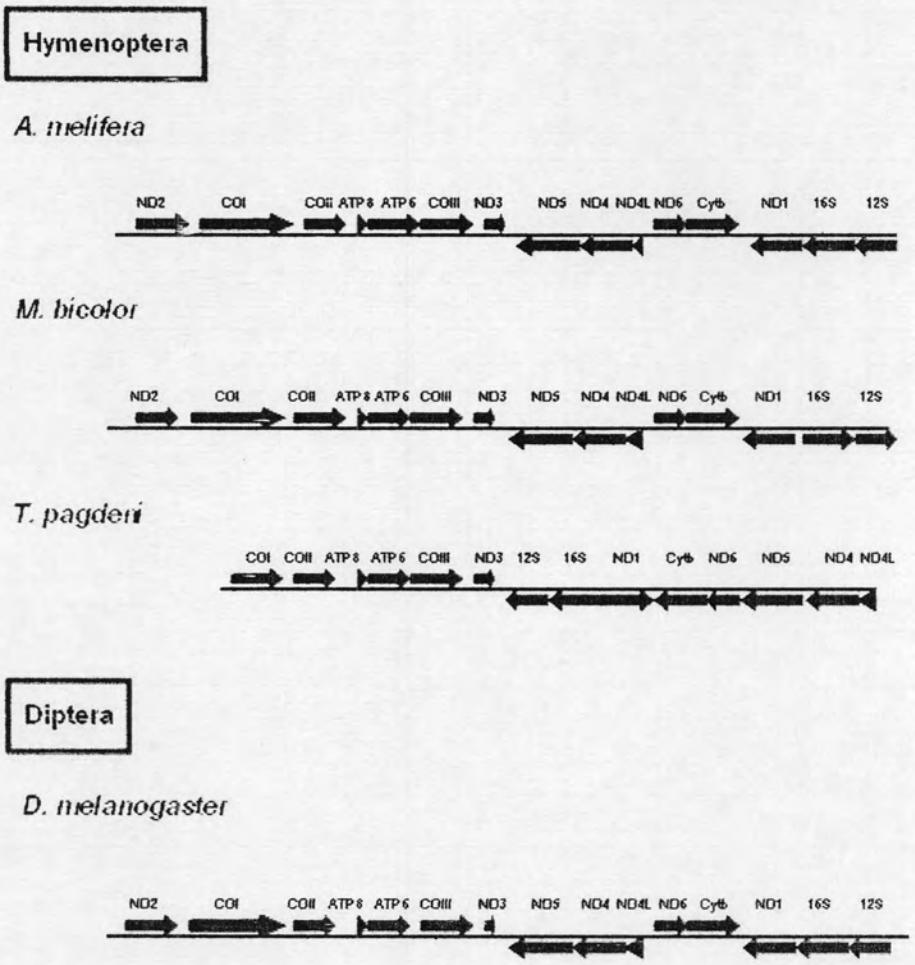


Figure 5.9 Gene organization of the *T. pagdeni* mtDNA from (A) and comparison of mitochondrial coding and non-coding gene among 4 insect species including one insect order (B). All tRNA genes were indicated by the amino acid they encode.

DISCUSSION

The entire mtDNA of *T. pagdeni* was not obtained in this study. However, the total size of stingless bees mtDNA, *M. bicolor* has been estimated to be 18,500 bp by RFLP analysis (Weinlich *et al.*, 2004) and 14,422 bp has been sequenced (Silvestre *et al.*, 2008). From this study, overlapped fragment of *T. pagdeni* mtDNA of 12,802 bp was sequenced. This overlapped fragment contained the 12 protein-coding genes (complete sequence was obtained for the ND4L, ND4, ND5, ND6, cyt b, ND1, ND3, COIII, ATP6, ATP8 and COII genes, and a partial sequence for COI) from 13 protein- coding genes, 12 of 22 tRNA genes and the two rRNA (the large subunit-16S and the small subuinit-12S). We detected four overlapping regions (ND4/ND5, cytb/tRNA-Ser (S2), ND1/tRNA-Leu (L1) and ATP8/ATP6) between genes. Some of the overlapped genes have been reported for *M. bicolor* mtDNA by Silvestre and coworkers (2008) such as ND1 and tRNA-Leu (L1) shared six nucleotides; ATP6 and ATP8 shared ten nucleotides. Moreover, fourteen non-coding regions were observed with total intergenic region of 419 bp. In honey bees, the number of non-coding nucleotides of *A. mellifera* mtDNA, excluding the COI-COII intergenic region and the control region, is greater as 618 bp (Crozier and Crozier, 1993). For stingless bees, *M. bicolor* presents a more compact arrangement (the total intergenic region is 486 bp; Silvestre *et al.*, 2008).

When we analyzed the non-coding region found in *T. pagdeni* mtDNA sequenced, they showed no significant similarities with any regions of the mtDNA of *M. bicolor* or other organisms. The intergenic region between COI and COII genes of *A. mellifera* mtDNA is known as hypervariable region. The COI-COII intergenic region has been widely studied in *A. mellifera*, and size polymorphism has been reported (from 200 to 650 bp) among subspecies (Garnery *et al.*, 1992, 1995 and Franck *et al.*, 1998). It has also been referred as a possible second origin of mtDNA replication and transcription (Cornuet *et al.*, 1991). However, this region was absent in *T. pagdeni*. Likewise, the COI-COII intergenic region investigated in *M. bicolor* mtDNA, is also absent. Furthermore, the COI-COII intergenic region is also disappeared in at least 16 other Meliponini species (Arias *et al.*, 2006).

The A+T content were very high in *T. pagdeni* mtDNA, as same as that of *M .bicolor* (87% Silvestre *et al.*, 2008) and *A. mellifera* (85% Crozier and Crozier, 1993). *A. mellifera* has been known as an insect containing the most AT biased mitochondrial genome (Simon *et al.*, 1994). The advantage of AT bias in mitochondrial genome could be explained by one hypothesis that the DNA polymerase could use those bases in a more efficient way during mtDNA replication (Clary and Wolstenholme, 1985), because the energetic cost to break the A-T links would be lower than the G-C links. The AT bias would be generated on organisms to preserve a high metabolic rate during mtDNA replication and transcription (Xia, 1996).

The sequences of 11 protein-coding genes were analyzed and nucleotide composition, codon usage and size were compared with those of *M. bicolor* and *A. mellifera*. The initiation codons in *T. pagdeni* were 4 ATT, 1 ATC (both for isoleucine), 3ATA, and 3 ATG (both for methionine). The incomplete stop codons (T or TA) were not detected in the 11 protein-coding genes genes of *T. pagdeni*, but they have been found in two genes of *A. mellifera* and four of *D. yakuba* (Crozier and Crozier, 1993). Frame translation of *T. pagdeni* was stopped with TAA (ND4L, ND5, ND6, cyt b, ND1, ND3 and ATP8) or TAG (ND4, COIII, ATP6 and COII). The results of optimum frame translations in each protein-coding gene were summarized in Figure 5.6. The standard insect mitochondrial genetic codes were used to analyze 10 protein-coding genes (ND4L, ND4, ND5, cyt b, ND1, ND3, COIII, ATP6, ATP8 and COII) of *T. pagdeni* mtDNA successfully, since it yielded no stop codons within the gene sequences. Whereas two stop codons (TAA) were detected within the ND6 gene sequence of *T. pagdeni*. Base substitution or even deletion may influence to the absence and presence of the stop codon within the gene sequence than expected. However, the ND6 gene and amino acid sequences were very similar to those of *M. bicolor* (70%, 45%, respectively) and the ND6 gene sequences (522 bp) did not differ substantially from those reported for *M. bicolor* (540 bp) and *A. melliferra* (540 bp). The codon usage of 11 protein-coding genes of *T. pagdeni* showed a preferred codon for each amino acid, generally ending with A or T, and there are two non-used codons ending with G (GCG and CGG). In *A. mellifera*,

there are seven non-used codons (Crozier and Crozier, 1993) and in *M. bicolor* there are 12 codons that are not used, all ending with C or G (Silvestre *et al.*, 2008). Generally, the AT bias in codon usage can be revealed by the ratio of "G+C" (Pro, Ala, Arg and Gly) to "A+T" rich codons (Phe, Ile, Met, Tyr, Asn and Lys) (Crozier and Crozier, 1993). That ratio was 0.18 for *A. mellifera,* 0.14 for *M. bicolor* and 0.20 for *T. pagdeni.* The AT bias on mitochondrial protein-coding genes could be confirmed by the nucleotide usage on first, second and third codon positions. Base composition was biased towards A/T nucleotides in the first, second and third codon positions in *T. pagdeni* mtDNA. The bias A and T nucleotides in the first, second codon positions have been observed in *M. bicolor* (Silvestre *et al.*, 2008). The bias A and T nucleotides in the third codon position has been detected in mtDNA protein-coding genes of other invertebrates, specially in *D. yakuba* (94%), *Apis mellifera* (95%), and *Caenorhabditis elegans* (Nematoda; 86%) (Wolstenholme, 1992 and Crease, 1999).

For ribosomal RNA genes, it is generally difficult to analyze the size of ribosomal RNA transcript precisely with inferring the DNA sequence by itself, so it is assumed that their ends are located on the boundaries of the flanking genes (Boore, 2001). The 16S rRNA gene of *T. pagdeni* was completely sequenced with size 1,351 bp, 3 bp smaller than the 16S rRNA gene of *M. bicolor* (Sivestre *et al.*, 2008) and 20 bp smaller than the 16S rRNA gene of *A. mellifera* (Crozier and Crozier, 1993), and their nucleotide similarity was 78%. The G+C content of 16S rRNA in *T. pagdeni* was 23%, while *M. bicolor* was 13% and *A. mellifera,* 15%. Earier study, the 12S rRNA gene of *M. bicolor* was partially sequenced (437 bp) by Silvestre *et al.* (2008). However, the 12S rRNA sequence of *T. pagdeni* was obtained completely. It had 762 bp, with 24 bp smaller than the 12S rRNA gene of *A. mellifera.* The 12S rRNA sequence of *T. pagdeni* similarity between *A. mellifera* and *M. bicolor* was 71% and 83 %, respectively. The 12S rRNA gene presented a high A+T content (76%) similar to 12S rRNA gene of *A. mellifera* (81%). The difference in size observed for the 16S and 12S rRNA gene was quite small comparing to *A. mellifera* (1371 and 786 bp), respectively. However, the size differences of rRNA genes could be accepted more than those of protein-coding genes. Due to it have

not to maintain a frame to read, and only the secondary structure needs to their function (Wolstenholme, 1992).

The A+T-rich region is major non-coding region for the initiation of replication in vertebrate and invertebrate and generally located between the 12S rRNA gene and tRNA-Met (Brown, 1985). In honey bees, *A. melifera*, A+T-rich region is located between the 12S rRNA gene and tRNA-Glu (Crozier and Crozier, 1993). The A+T-rich region contains several short repeating sequences (6–13 bp) with varying copy number (two to four copies) scattered through the whole region. It also contains a polythymidine stretch. This polythymidine stretch has been reported to be a transcription control or the initiation of replication (Zhang and Hewitt, 1997). The A+T-rich region of stingless bee, *M. bicolor*, could not be sequenced because of difficulties in amplification (Sivestre *at al.*, 2008). However, the A+T-rich region of *M. bicolor* may also be described by its size using RFLP analysis (Weinlich *et al.*, 2004). It was estimated in size approximately 3,300 bp, about 2.5 kb longer than in *A. mellifera* (Crozier and Crozier, 1993). In this study, because we were not able to amplify the entire mtDNA of *T. pagdeni*, the A+T-rich region, ND2 and 10 tRNA genes was not found on the mitochondrial DNA fragment sequenced. We assumed that this A + T-rich region of *T. pagdeni* had the same length as *M. bicolor*. The large size of this region can influence to partial duplications inside this region, which is a common characteristic of insect mtDNA (Simon *et al.*, 1994) and may lead to amplification failure. These may be explained by generally difficult to amplify in insects, mainly because of tandem repeats, heteroplasmy and great length variation at intra and inter-specific levels (Zhang and Hewitt, 1997).

From the 22-23 tRNA genes regularly found in animal mitochondrial genomes,12 tRNA genes could be identified and positioned on *T. pagdeni* mtDNA, though the mtDNA sequence was not entirely obtained. The 12 tRNA genes were sequenced. Of these, 4 tRNA genes (tRNA-Glu, tRNA-Gln, tRNA-Thr and tRNA-Pro) of *T. pagdeni* were on different positions when compared with *M. bicolor*. The tRNA gene arrangement of *T. pagdeni* mtDNA varied from that of *Melipona bicolor*. Particularly, the presence of tRNA-Glu gene between the ND3 and COIII gene instead of tRNA-Gly

found in *Melipona bicolor* mtDNA, and the tRNA-Gln gene was found between tRNA-Arg and 12S rRNA genes in *T. pagdeni* mtDNA. The tRNA-His gene in *T. pagdeni* was not found between the ND4 and ND5 genes. While the tRNA-His gene in other bee mtDNAs is located between the ND4 and ND5 genes. Generally, the arrangement of tRNA gene position was diverged within insects between the orders Diptera (Clary and Wol-Stenhome, 1985) and Hymenoptera (Crozier *et al.*, 1989) and within the order Diptera, with differences observed between Aedes (Hsuchen *et al.*, 1984 and Hsuchen and Dubin, 1984) and *Drosophila yakuba* (Clary and Wol-Stenhome, 1985). Likewise, the rearrangement of 12S rRNA, 16S rRNA, ND1, cyt b and ND6 gene of *T. pagdeni* mtDNA (Figure 5.8) was different from those of *M. bicolor* (Silvestre *et al.*, 2008). The mitochondrial gene rearrangement could be explained by several mechanisms. One of the most extensively-accepted mechanisms is tandem duplication of gene regions as a result of slipped-strand mispairing, followed by deletions of genes (Levinson and Gutman, 1987; Moritz and Brown, 1987 and Macey *et al.*, 1998).