CHAPTER III

GENETIC DIVERSITY AND POPULATION STRUCTURE OF *Trigona pagdeni* Schwarz IN THAILAND USING TE-AFLP ANALYSIS

INTRODUCTION

Stingless bees of the genus *Trigona* Jure 1807 (Apidae, Apinae, Meliponini) occur in tropical and subtropical regions worldwide, where they are important pollinators (Michener, 1974; Sakagami, 1982; Michener, 1990; 2000). Thirty two species of *Trigona* are known from Thailand (Schwarz, 1939; Sakagami *et al.*, 1985; Michener and Boongird, 2004; Klakasikorn *et al.*, 2005). *Trigona pagdeni* Schwarz is a common stingless bees found in Thailand, peninsular Malaysia and the Indochina region (Sakagami, 1978; Sakagami and Khoo, 1987). Our goal is to survey genetic diversity and population structure of *T. pagdeni* in Thailand.

Earlier studies of *Apis cerana* in Thailand using mtDNA (Smith and Hagen, 1996; Sihanuntavong *et al.*, 1999; Smith and Hagen, 1999; Warrit *et al.*, 2006) and microsatellites analysis (Sittipraneed *et al.*, 2001) showed genetic differentiation among geographic regions within Thailand, particularly between populations north and south of 11°N in the Isthmus of Kra (Warrit *et al.*, 2006). This region, known as the Kra ecotone, corresponds to the transition between seasonal evergreen or seasonal rainforest and mixed moist deciduous or monsoon forest (Hughes *et al.*, 2003), and between the Indochinese and Sundaic biotas (Woodruff, 2003). Shifts from one species or subspecies to another north and south of this ecotone have been documented in many animal taxa as well (e.g., birds, Hughes *et al.*, 2003). However no such information exists for the numerous stingless bee species of Thailand; therefore the goal of this study is to investigate genetic diversity and population structure in *T. pagdeni*, and to determine if similar biogeographic patterns are shown by both *A. cerana* and stingless bees.

This information will help to establish the generality of the Kra ecotone as a biogeographic boundary between northern and southern populations of bees in the Thai-Malay peninsular, and will also be fundamental for establishing rational management of resources for conservation of this native species.

DNA fingerprinting technique, Three-Enzyme Amplified Fragment Length Polymorphisms (TE-AFLP: van der Wurff *et al.*, 2000) was used to survey genetic diversity because it reveals intraspecific variation across the genome, requires no prior knowledge of genome sequences, and is relatively rapid and inexpensive.

MATERIALS AND METHODS

Sampling

Adult stingless bees workers were collected from 117 nests of *T. pagdeni*, 12 nests of *T. laeviceps* Smith, and one nest of *T. minor* at the localities shown in Figure 3.1 and Table 3.1. Specimens were placed in 95% ethanol and kept at 4°C until required. Taxonomic identification was carried out according to Sakagami (1978) and Sakagami and coworkers (1983), and later confirmed by Professor Dr. Charles D. Michener (University of Kansas). Voucher specimens of *T. pagdeni* from some of the colonies used in this study are deposited in the University of Kansas Snow Entomological Collection.

Table 3.1 Collection sites, populations, map abbreviation (Figure 3.1) and sample sizes of *Trigona pagdeni* used in this study.

Localities	Population	Map Abbreviations	Sample size
Chaing Mai	North	CM	1
Nan	North	NA	2
Uttaradit	North	UT	1
Kamphaeng Phet	North	KP	1
Phitsanulok	North	PL	2
Sukhothai	North	ST	5
Udon Thani	Northeast	UD	1
Khon Kaen	Northeast	KK	3
Ubon Ratchathani	Northeast	UB	3
Si Sa Ket	Northeast	SS	1
Roi Et	Northeast	RE	6
Surin	Northeast	SU	4
Buri Ram	Northeast	BR	2
Maha Sarakham	Northeast	MK	3
Nonthaburi	Central	NB	1
Lop Buri	Central	LB	2
Suphan Buri	Central	SP	1
Sing Buri	Central	SI	2
Nakhon Nayok	Central	NY	1
Bangkok	Central	BK	1
Ayutthaya	Central	AY	3
Phetchaburi	Central	PB	1
Kanchanaburi	Central	KB	10
Prachuap Khiri Khan	Central	PK	15
Trat	Central	TR	2
Chanthaburi	Central	CT	3
Chon Buri	Central	CB	2
Chumphon	Peninsular	CP	17
Ranong	Peninsular	RN	1
Nakhon Si Thammarat	Peninsular	NT	1
Surat Thani	Peninsular	SR	6
Phatthalung	Peninsular	PT	4
Songkhla	Peninsular	SK	9

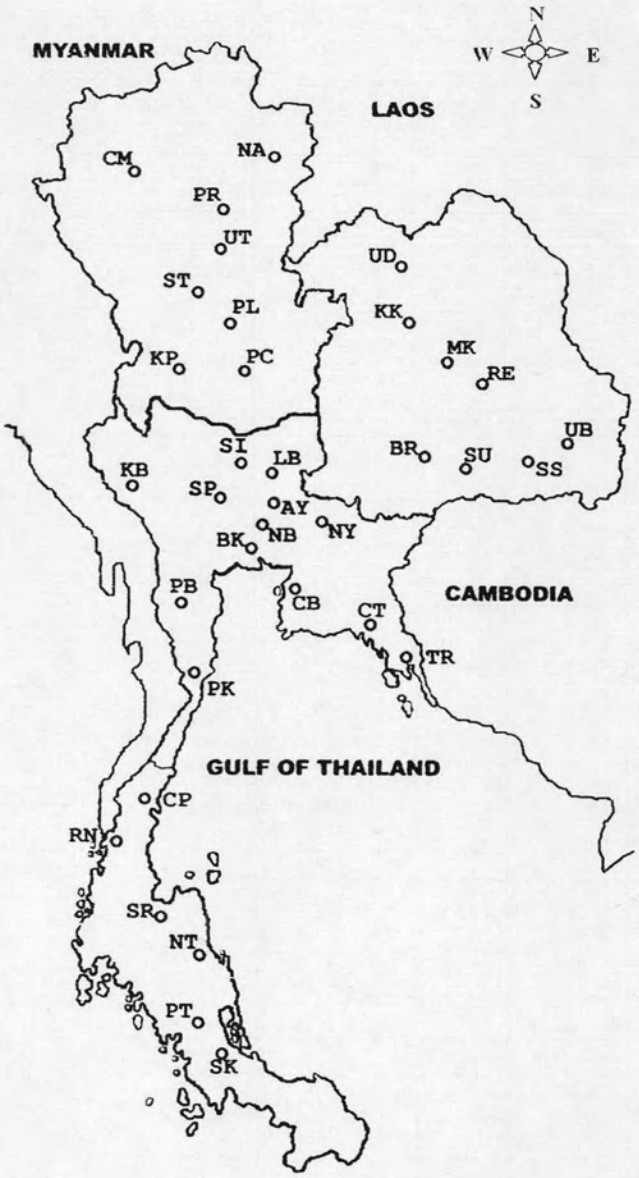

Figure 3.1 Collection sites of *T. pagdeni* Schwarz in Thailand. See Table 3.1 for locality names and sample sizes.

DNA extraction

Total DNA was extracted from one entire bee per nest using standard proteinase K-SDS followed by phenol-chloroform extraction and ethanol precipitation (Smith and Hagen, 1996). DNA concentration was estimated by comparison with a known quantity of digested lambda DNA on a mini-gel (Sambrook *et al.*, 1989).

TE-AFLP Fingerprinting

Approximately 30 ng of each DNA extract was digested with the 6-base restriction enzymes *Xba*I and *Bam*HI and the 4-base enzyme *Rsa*I in a 20 µl reaction. The resulting fragments were ligated to linkers (described in van der Wurff *et al.*, 2000) having sticky ends complementary to the *Xba*I and *Bam*HI sticky ends of the DNA fragments. Next, 0.5 µl of the digestion-ligation mix was used as template in a 12.75 µl PCR reaction containing 2.5 pM each of unlabelled *Xba*I- primer and ^{32}P-labelled *Bam*HI primer, 1X PCR buffer (10 mM Tris-HCl, pH 7.6, 50 mM KCl, 0.1% Triton X-100 and 1.5 mM $MgCl_2$), 50 µM of each dNTP (Promega), and 0.6 units *Taq* DNA polymerase (Promega). The *Xba*I and *Bam*HI primer sequences (van der Wurff *et al.*, 2000) are identical to one strand of the *Xba*I and *Bam*HI linkers, respectively, with arbitrary base extensions. We used a labeled *Bam*HI primer with the arbitrary extension "C", and paired it with two different *Xba*I primers: *Xba*I-CC and *Xba*I-AC.

The polymerase chain reaction was performed on a thermal cycler (TC-412; TECHNE) using 2.5 minutes denaturation at 95°C followed by: 10 cycles of 30 seconds denaturation at 95°C for 30 seconds; annealing at 70°C, 60 seconds; elongation at 72°C, and 40 cycles of 30 seconds; denaturation at 95°C, 30 seconds; annealing at 60°C and 60 seconds; elongation at 72°C, with a final 20 minutes at 72°C (van der Wurff et al., 2000). For analysis, the PCR product was mixed with 3 µl of loading dye (formamide containing 5 mg/ml blue dextran and 2.4 µl of sterile double-distilled water) and denatured for 5 minutes at 95°C. Three µl was loaded on a 8% denaturing polyacrylamide gel with 0.6XTBE electrophoresis buffer and electrophoresed for 5-6 hours at 500 V. The dried gel was exposed to Kodak X-Omat film.

Scoring TE-AFLP variation

Bands generated by each primer combination were scored by eye. The presence of a TE-AFLP band was scored as 1, absence as 0. Thus each bee was characterized by a multi-locus phenotype.

Data analysis

Individuals were grouped into four populations based on collection site: North, Northeast, Central and Peninsular (Figure 3.1, Table 3.1). Genetic diversity, genetic distance among populations, and population structure statistics were calculated using Genetic Analysis in Excel (GenAlEx6; Peakall and Smouse, 2006).

Genetic diversity was estimated as the proportion of polymorphic loci (P), and expected heterozygosity (H_e). A band was considered polymorphic if our samples showed any variation for presence or absence, though "singletons" (bands present or absent in only a single individual) and other rare variants have little effect on assessment of population structure. Estimation of H_e followed the method of Lynch and Milligan (1994), which considers each band position as a different locus with two alleles, band amplified (dominant) and band not amplified (recessive); absence of a band indicates a recessive homozygote. Populations are assumed to be outbreeding or completely inbreeding. At each locus, the frequency of the recessive allele (q) is estimated from the frequency of putative recessive homozygotes (q^2), and the frequency of the dominant allele is estimated as: $p = 1 - q$. Expected heterozygosity at each locus = $h = 1 - \Sigma x_i^2$, where x_i is the frequency of the *i*th allele, or in this case, $1 - (p^2 + q^2)$. Expected heterozygosity averaged over all loci was calculated as:

$$H_e = 1 - 1/m \sum_{y=1}^{m} \Sigma x_i^2$$

where *y* represents loci or bands 1 through *m*.

Analysis of Molecular Variance (AMOVA; Excoffier *et al.*, 1992) implemented in GenAlEx V6 (Peakall and Smouse, 2006) was used to assess the patterns of observed genetic variation. AMOVA can be used with binary or dominant markers such as TE-AFLPs to calculate Φ statistics, which are analogous to Wright's F-statistics for co-dominant data (Wright, 1951; 1965). Φ_{PT}, which can take values

from 0-1, measures the similarity of pairs of individuals drawn at random from the same population, relative to pairs of individuals drawn from the entire collection, and is analogous to Wright's F_{ST}. Significance testing was performed by comparing observed statistics to the distributions generated by 999 random permutations of the data into populations of the same size.

In order to search the data set for geographic patterns, we examined (1) differentiation among the four geographic populations and (2) differentiation between Peninsular samples and those from north of the Isthmus of Kra. Because mean expected heterozygosity and pairwise genetic distances among the four populations indicated the Northeast population differed from the others, we also examined (3) differentiation between Northeast samples and the others. In each case, significance testing was performed by comparing observed statistics to the distributions generated by 999 random permutations of the data into populations of the same size.

RESULTS

The presence or absence of 51 TE-AFLP bands was scored for each individual. Eight bands (16%) were fixed (present in all 117 *T. pagdeni*). The remaining 43 bands, each of which were present in at least one individual and absent in at least one individual, were considered polymorphic. Thirty five of these bands (69%) had inferred allele frequencies in the range of 5% to 95%. The mean expected heterozygosity, H_e, was 0.195 (st. error 0.022). We observed 87 unique banding patterns or phenotypes for the primer pair *Bam*HI-C and *Xba*I-CC, and 77 for the primer pair *Bam*HI-C and *Xba*I-AC (APPENDIX G). Sample banding patterns are shown in Figure 3.2. For population structure, genetic diversity in the four populations of *T. pagdeni* was compared in Table 3.2. Within populations, 14-19 bands were fixed. Two bands were found only in the Northeast population; no other population showed "private" bands (data not shown). The Northeast population shows the highest percentage of polymorphic bands and highest mean expected heterozygosity.

Pairwise genetic distances among populations are shown in Table 3.3; the highest values are those comparing the Northeast population to others. AMOVA results are shown in Tables 3.4 - 3.5. Differentiation was observed among all four populations (Φ_{PT} = 0.18, p = 0.001), though 82% of observed variation occurred among individuals within populations. If bees collected from north of the Kra ecotone are combined into a single population, small but significant differentiation between bees north and south of the Kra ecotone can be detected (Φ_{PT} = 0.13, p = 0.001). However, the greatest differentiation was observed in the contrast of bees from the Northeast with the other combined samples (Φ_{PT} = 0.21, p = 0.001).

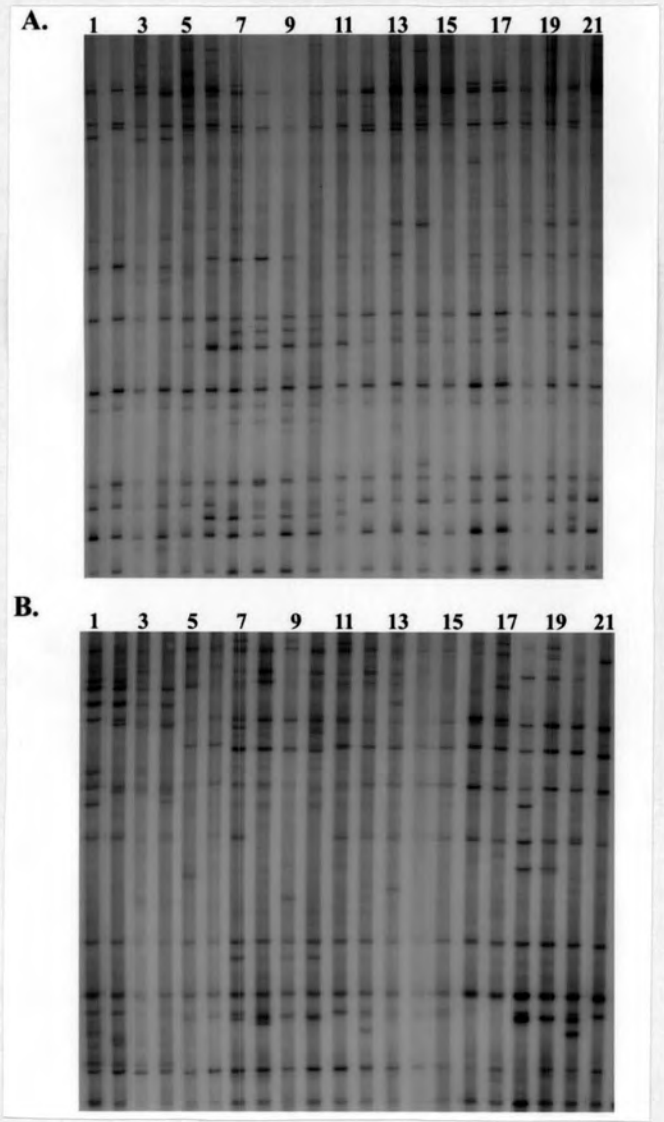

Figure 3.2 Autoradiogram of some TE-AFLP patterns obtained after amplification of *Trigona* DNA with primers (A) *Bam*HI-C and *Xba*I-CC and (B) *Bam*HI-C and *Xba*I-AC. Lanes 1-4 are a related species, *T. laeviceps*, lanes 5-21 are *T. pagdeni*; (lanes 5-6, North population; lanes 7-11, Northeast population; lanes 12-17, Central population; and lanes 18-21, Peninsular population).

Table 3.2 Comparison of TE-AFLP bands generated by primer pairs *Bam*HI-C and *Xba*I-CC, and *Bam*HI-C and *Xba*I-AC in four populations of small *T. pagdeni* in Thailand. Fixed bands were present in every individual sampled, polymorphic bands include any band that is present in at least one individual and absent in at least one individual. Allele frequencies are inferred from the frequency of individuals in which a band is absent (considered recessive homozygotes) assuming alleles are in Hardy-Weinberg equilibrium (Lynch and Milligan, 1994).

	Locality (N)			
Bands:	**NORTH (12)**	**CENTRAL (44)**	**NORTHEAST (23)**	**PENINSULAR (38)**
Number scored	41	45	48	43
Number fixed	19	19	14	17
Number absent	10	6	3	8
# polymorphic	22	26	34	26
% polymorphic	54% of 41 (43% of 51)	58% of 45 (51% of 51)	71% of 48 (67% of 51)	60% of 43 (51% of 51)
Loci:				
# loci with inferred allele frequencies 5-95%	16	23	28	19
% with inferred allele frequencies 5-95%	39% of 41 (31% of 51)	51% of 45 (45% of 51)	58% of 48 (55% of 51)	44% of 43 (37% of 51)
mean *He* (std. error)	0.130 (0.026)	0.137 (0.025)	0.207 (0.027)	0.139 (0.026)

Table 3.3 Pair-wise genetic distances among populations of *T. pagdeni* in Thailand, as estimated in GenAlEx V6 (Peakall and Smouse, 2006).

	North	**Northeast**	**Central**	**Peninsular**
North (N = 12)	*			
Northeast (N = 23)	0.061	*		
Central (N = 44)	0.018	0.057	*	
Peninsular (N = 38)	0.029	0.079	0.033	*

Table 3.4 Analysis of Molecular Variance (AMOVA) of TE-AFLP banding patterns generated by two primer pairs in 117 *T. pagdeni* from 4 population groupings (North, Northeast, Central and Peninsular). df = degrees of freedom, SS = sums of squares, MS = mean squares, Est. Var. = estimated variance within and among populations, % = percentage of observed variance within or among populations, Φ_{PT} = correlation of individuals within populations relative to the collection as a whole. Significance testing was carried out by comparison of observed values to those generated by 999 permutations in which individuals were randomly assigned to four populations of the same size as the four study populations. Prob. = the probability of obtaining a value of Φ_{PT} as large or larger when individuals are randomly assigned to groups.

	df	**SS**	**MS**	**Est. Var.**	**%**	**Φ_{PT} (p)**
Among Populations	3	80.47	6.82	0.84	18%	0.18 (p = 0.001)
Within Populations	113	428.30	3.79	3.79	82%	

Table 3.5 Regional differentiation: AMOVA of TE-AFLP banding patterns generated by two primer pairs in 117 *T. pagdeni.* Upper chart (A): samples grouped in two populations: Northeast and Others; Lower chart (B) samples grouped into two populations: South (same as Peninsular population) and North of the Kra ecotone Significance testing by comparison of observed values to those generated by 999 random permutations of individuals into populations of the same size. Prob. = the probability of obtaining a value of Φ_{PT} as large or larger when individuals are randomly assigned to groups. Other abbreviations are as in Table 3.4.

		df	SS	MS	Est. Var.	%	Φ_{PT} (p)
A. Northeast vs. others	**Among populations**	1	44.11	44.11	1.08	21%	0.21 (p = 0.001)
	Within populations	115	464.66	4.04	4.04	79%	
B. North vs. South of Kra Ecotone	**Among Populations**	1	35.84	35.84	0.62	13%	0.13 (p = 0.001)
	Within Populations	115	472.93	4.11	4.11	87%	

DISCUSSION

The stingless bees *Trigona pagdeni* Schwarz (Apidae, Apinae, Meliponini) is one of the most common stingless bees in Thailand. We surveyed genetic diversity and population structure of *T. pagdeni* in Thailand using a DNA finger-printing technique, Three-Enzyme Amplified Fragment Length Polymorphisms (TE-AFLP, van der Wurff *et al.*, 2000). Since it reveals intraspecific variation across the genome, requires no prior knowledge of the genome, and is relatively rapid and inexpensive. In this study we showed that the TE-AFLP procedure can be used to investigate genetic diversity and population structure of *Trigona* bees.

DNA fingerprinting (TE-AFLPs) revealed substantial genetic variation among *T. pagdeni* collected throughout Thailand. Analysis of population structure showed differentiation among the Peninsular, Central, North and Northeast population, and between *T. pagdeni* north and south of the Kra ecotone. However, the greatest differentiation was between *T. pagdeni* from the Northeast and the other samples (Φ_{PT} = 0.21, p = 0.001, Table 3.5); we also found higher pairwise genetic distances between the Northeast population and each of the other three populations, than between other pairs of populations (Table 3.3). The samples from the Northeast region also show a higher level of polymorphism and expected heterozygosity compared to other samples (Table 3.2).

This pattern seems to be different from that found for *Apis cerana*, the bee species which has been most thoroughly investigated in Thailand. DNA sequence data (e.g., Warrit *et al.*, 2006, Smith and Hagen, 1996) indicate that the Kra ecotone marks a major shift in mitochondrial lineages of *A. cerana*. Microsatellite data (Sittipraneed *et al.*, 2001) also showed differences between populations of *A. cerana* north and south of this ecotone. The most direct comparison of *T. pagdeni* and *A. cerana* would use the same data set for both species, but published data do not yet exist for either TE-AFLP data for Thai *A. cerana*, or mitochondrial sequence data for *T. pagdeni*. Furthermore, *A. cerana* from the Northeastern region have not been well-sampled yet, so we cannot say whether they are conspicuously diverse or strongly differentiated from other Thai *A. cerana*

populations. However, it is known that *A. cerana* from neighboring Cambodia and southern Vietnam have more diverse mitochondrial haplotypes than populations found in North, Central and Peninsular Thailand (Smith *et al.*, 2005).

In this study we showed that the TE-AFLP procedure can be used to investigate genetic diversity and population structure of *Trigona* bees. This may provide useful information for management and conservation of this important pollinator of tropical plants, whose original habitat is decreasing because of human development (see, for example, Araujo *et al.*, 2004). For example, it has recently become popular to cultivate *T. pagdeni* in boxes for plant pollination, and human transportation of such hives could lead to mixing of genetically differentiated populations.