THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Genetics)

DEGREE

Genetics	Genetics
FIELD	**DEPARTMENT**

TITLE: Population Genetic Study of *Anopheles harrisoni* (Diptera: Culicidae) and Sequence-based Multilocus Phylogenetic Reconstruction of Malaria Vectors in Thailand

NAME: Mr. Prin Phunngam

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR

(Miss Uraiwan Arunyawat, Dr.rer.nat.)

THESIS CO-ADVISOR

(Mrs. Anchanee Kubera, Ph.D.)

DEPARTMENT HEAD

(Associate Professor Somsak Apisitwanich, Dr.Agr.Sci.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

POPULATION GENETIC STUDY OF *Anopheles harrisoni* (DIPTERA: CULICIDAE) AND SEQUENCE-BASED MULTILOCUS PHYLOGENETIC RECONSTRUCTION OF MALARIA VECTORS IN THAILAND

PRIN PHUNNGAM

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Genetics)
Graduate School, Kasetsart University
2012

Prin Phunngam 2012: Population Genetic Study of *Anopheles harrisoni* (Diptera: Culicidae) and Sequence-based Multilocus Phylogenetic Reconstruction of Malaria Vectors in Thailand. Master of Science (Genetics), Major Field: Genetics, Department of Genetics. Thesis Advisor: Miss Uraiwan Arunyawat, Dr.rer.nat. 149 pages.

Investigation of pattern and level of genetic diversity of malaria vector species are important for understanding the evolutionary history in order to achieve vector control. In this study, two *An. harrisoni* populations from Kanchanaburi province and one from north-western of Vietnam were collected and identified by molecular identification. Five putative neutral fragments on the X chromosome were chosen to amplify and sequence. The total length of the studied loci range from 245 bp to 609 bp. and average level of nucleotide variations (π) across all loci varies from 0.151 – 0.268 % based on single nucleotide polymorphism data. This result exhibited low level of nucleotide diversity for all three *An. harrisoni* populations. Moreover, the neutrality tests, e.g. Tajima's *D* and Fu and Li's *D* statistics, did not show significant deviation from standard neutral expectation suggesting that all populations are likely under mutation-random genetic drift equilibrium. Furthermore, no genetic differentiation was observed between Thai populations, while the genetic differentiation between Vietnam and Thai populations was high due to the geographical isolation between them.

An understanding the evolutionary and taxonomic status of closely related malaria vector species is the initial step in a malaria vector control program. In this study, four different approaches were performed to reconstruct phylogenetic trees for the main malaria vectors presented in Thailand based on the six DNA fragments from both the nuclear and mitochondrial regions. The results revealed clear evidence that *Anopheles* species separate into three distinct clades: Dirus group, Minimus group and Maculatus group. Interestingly, phylogenetic trees based on different reconstructed algorithms and different gene regions provided congruent phylogenetic status of the mosquito species studied. The phylogenetic relationships of these malaria vector species follow a pattern based on morphological identification. Moreover, estimation of the divergence time among the studied species inferred that *Anopheles* species probably existed around the Eocene and Miocene.

_____ _____ ___ / ___ / ___

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

I would sincerely like to grateful thanks and indebted to Dr. Uraiwan Arunyawat my thesis advisor for her suggestion, encouragement and opportunity during study and research work, and also to Professor Dr. Theeraphap Chareonviriyaphap from Department of Entomology, Faculty of Agriculture, Kasetsart University for his suggestion, samples collection and kindly providing the mosquito samples for using in this research.

I would sincerely like to thanks Dr. Sylvie Manguin from Faculté de Pharmacie, University Montpellier I, Montpellier, France for kindly providing the mosquito samples and special experiences in France, and also to Dr. Suchitra Changtragoon from Forest Genetics and Biotechnology Division, Department of National Park, Wildlife and Plant Conservation who open me a new perspective for Molecular Biology's world. I would like to thanks Department of Genetics, Faculty of Science, Kasetsart University for good and comfortable laboratory. I am heartfelt thanks to the Lab 4507 members and the members of Department of Genetics for their helpfulness and friendship. I would like to greatly thank members of the Mosquito Lab. under Professor Dr. Theeraphap Chareonviriyaphap from Department of Entomology, Faculty of Agriculture, Kasetsart University for their helpfulness and friendship.

This research was supported by the Thailand Research Fund and Faculty of Science, Kasetsart University (TRF-MAG-WII), TRF-Senior Research Scholar Program (RTA5280007), and the Graduate School of Kasetsart University.

Finally, I am deeply appreciated to my parents and brother for their support and encouragement.

Prin Phunngam

May, 2012

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

AFLP	=	Amplified fragment length polymorphism
bp	=	basepairs
BLAST	=	Basic local alignment search tool
COI	=	Cytochrome c oxidase I
COII	=	Cytochrome c oxidase II
ºC	=	degree Celsius
DNA	=	deoxyribonucleic acid
dNTP	=	deoxyribonucleotide triphosphate
U	=	enzyme unit
indel	=	insertion/deletion
µl	=	microliter
mya	=	million years ago
NCBI	=	National Center for Biotechnology Information
PCR-RFLP	=	Polymerase chain reaction- restriction fragment length polymorphism
RAPD	=	Random amplification of polymorphic DNA
RFLP	=	Restriction fragment length polymorphism
RNA	=	r ibonucleic acid
rDNA	=	ribosomal DNA
rRNA	=	ribosomal RNA
s.l.	=	sensu lato
SSR	=	Simple sequence repeat
SNP	=	Single nucleotide polymorphism
SSCP	=	Single-strand conformation polymorphism
kya	=	thousand years ago
VNTR	=	Variable number tandem repeat
WHO	=	World Health Organization

POPULATION GENETIC STUDY OF *Anopheles harrisoni* (DIPTERA: CULICIDAE) AND SEQUENCE-BASED MULTILOCUS PHYLOGENETIC RECONSTRUCTION OF MALARIA VECTORS IN THAILAND

INTRODUCTION

Malaria is an important cause of illness and death in several countries in tropical and sub-tropical region. *Anopheles* mosquitoes are the most important vectors for transmitting the malaria parasites, *Plasmodium* spp. About 70 to 80 of 484 recognized *Anopheles* species are considered as malaria vectors. There are six *Anopheles* mosquitoes considered as malaria vectors in Thailand. Three species of these have been examined as primary vectors, i.e. *An. dirus, An. minimus*, and *An. maculatus*. The rests of them have been considered as secondary vectors, i.e. *An. aconitus, An. sundaicus*, and *An. pseudowillmori*. *Anopheles* mosquito species mostly exist in temperate, subtropical and tropical areas of the world. Moreover, they can be found in different geographies from hilly to coastal areas. Most *Anopheles* mosquitoes are species complex which is difficult to distinguish from morphological characteristics.

The Dirus Complex occurs in tropical evergreen rainforest, cultivated forests fringes along Southeast Asia, and five Dirus mosquito species are found in Thailand including *An. baimaii*, *An. cracens*, *An. dirus*, *An. nemophilous* and *An. scanloni*. Within the Minimus Complex, *An. minimus* and *An. harrisoni* can be found in large sympatric areas in the Oriental Region. Moreover, *An. aconitus, An. varuna*, and *An. pampanai* are considered as the closely related species of the Minimus Complex due to ambiguity of overlapping morphological characters. The Sundaicus Complex consists of at least four distinct species, including *An. sundaicus* s.s., *An. epiroticus*, *An. sundaicus* D, and *An. sundaicus* E. In Thailand, *An. epiroticus* is predominant species in the Sundaicus Complex and could transmit malaria in coastal areas. Five species of the Maculatus Group are distributed in Thailand, i.e. *An. pseudowillmori*, *An. willmori*, *An. maculatus*, *An. sawadwongporni* and *An. notanandai*.

An. harrisoni (former species C) is one of the major malaria vectors in Southeast Asia. In Thailand, malaria remains the considerable health problem, especially along the international borders where are largely forest areas and difficult management from formal health system. *An. harrisoni* is the member of the Minimus Complex which is often misidentification on the morphology. The others sibling species of the Minimus Complex are *An. minimus* (former species A). This species can be found in large sympatric areas in mainland of Southeast Asia with *An. harrisoni*, and *An. yaeyamaensis* (former species E), endemic in Ishigaki Island in the Ryukyu Archipelago of Japan. This species complex belongs to the Minimus Subgroup, the Funestus Group, the Myzomyia Series, the Laticorn Section, and within Subgenus *Cellia*. In Thailand, *An. harrisoni* is widespread in Kanchanaburi province, western Thailand, especially in Sai Yok district.

Malaria control and management are continuously improved for decreasing the burden of malaria. In Thailand, tendency of reported cases and deaths are becoming decrease. Insecticides usability is one of malaria control programs against the burden of malaria. However, insecticide resistance is still an important problem of malaria control because the malaria vector has rapid capacity of adaptation to insecticide. Genetic machinery changes of the malaria vector are likely causes of vector adaptation to insecticides. Therefore, better understanding in vector evolutionary genetics is necessary for malarial control improvement.

Single nucleotide polymorphisms (SNPs) are the most common DNA variation throughout the genome including coding and non-coding regions in several organisms. Most SNPs locate in the non-coding regions of genome which they are no effect on encoding the proteins and phenotype. Now a day, SNPs are becoming the popular molecular markers for several studies, especially the study of molecular population genetic. One choice of increase interestedness is the technology of sequencing for SNP discovery. The simplest method for SNP discovery is direct sequencing.

An understanding the pattern of genetic diversity and the evolutionary factors which they have effects on the pattern of genetic diversity of organisms is an important requirement for malaria management. At present, the information of genetic diversity of main malaria vectors in Thailand is limited, especially *An. harrisoni*. The available complete genome sequence of *An. gambiae*, African mosquito, could provide the choices of studies and insights into the pattern of genetic diversity and evolutionary factors in *Anopheles* mosquitoes. The aim of this work is to estimate the level of genetic diversity of *An. harrisoni* across different populations in Thailand using single nucleotide polymorphisms (SNPs) as a genetic marker. The genetic variation data will be used to test population genetic model with statistical analysis to reveal the answers to evolutionary questions of this species. This study may provide an understanding of evolutionary forces acting on pattern of genetic variation of *An. harrisoni* mosquitoes in Thailand, and the finding may help in further studies on origin and spread of insecticide resistance in malaria vector species.

An understanding the evolutionary and taxonomic status of closely related malaria vector species is the initial stage in malaria vector control program. This information may help for further studies and better understanding in the capacity to transmit malaria, genetically determined ecological and evolutionary of genes in insecticide resistance for management of malaria vector control. Mosquito-borne diseases are still a major health problem in many countries. Thailand is one of the countries in tropical region where malaria is prevalent. Malaria spreads through mosquitoes, belonging to *Anopheles* genus. The ability to transmit malaria varies within members of this genus. Moreover, the vectorial capacity of *Anopheles* mosquitoes involved in malaria transmission remains poorly understood in Thailand.

Most of taxonomic relationships in the previous studies have been identified based on the morphological characters. However, these taxonomies showed inconsistent and ambiguous relationships among the species. Nowadays, molecular phylogenetic relationships based on DNA and protein sequences comparative can be generated more simple and reliable than only morphological data. Information on the phylogenetic relationships and divergence time of mosquitoes in Thailand, however, is still limited and not covered on the major malaria vector species. Moreover, most

of phylogenetic studies on *Anopheles* species rely on single locus for the phylogenetic construction. In order to obtain precise phylogenetic tree of *Anopheles* species, applying a multilocus approach with fragments of different functions and chromosomal locations has more advantage.

This research consists of two parts. Firstly, molecular population genetic study was presented**,** using multi-fragments located on X chromosome, to investigate the level of genetic diversity and genetic population structure as well as to test population genetic models for understanding evolution processes of *An. harrisoni* spreading in Thailand. Basal genetic diversity information is first step for further studies and managements of interested species. Secondly, molecular phylogenetic approach was presented, using several fragment loci (nuclear genome, mitochondrial genome, coding region, and non-coding region), to investigate the relationships among malaria vectors that are prevalent in Thailand. Moreover, the divergence time among malaria vector and related species was estimated based on the COII mitochondrial gene.

OBJECTIVES

Part A: Population genetic study of the malaria vector *Anopheles harrisoni* in Thailand

1. To evaluate the level of genetic diversity and genetic population structure of *An. harrisoni* using single nucleotide polymorphism (SNPs) data

2. To test population genetic models of malaria vector from genetic diversity data for an understanding evolution processes of *An. harrisoni*

Part B: Phylogenetic reconstruction of malaria vectors in Thailand using multilocus DNA sequences

1. To reconstruct phylogenetic trees among Thai malaria vectors by utilizing multilocus DNA sequences based on both nuclear and mitochondrial regions

2. To estimate divergence time among malaria vectors and related species distributed in Thailand

LITERATURE REVIEW

Malaria

Malaria is the considerable mosquito-borne disease in tropical and sub-tropical region (Mehlotra and Zimmerman, 2006; Dixit *et al.*, 2010). The parasites of genus *Plasmodium* (phylum Apicomplexa) are transmitted by *Anopheles* mosquitoes (Cann, 1996; Rich and Ayala, 2006). These human malaria parasites consist of 5 species; these are *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, Plasmodium malariae* and *Plasmodium knowlesi* (Lowe *et al.*, 1998; Rich and Ayala, 2006; Galinski and Barnwell, 2009). The *Plasmodium falciparum* and *Plasmodium vivax* are most common infection in Thailand (Chareonviriyaphap *et al.*, 2000). Malaria can be found in several areas, e.g. some hilly-forested areas and a few coastal foci of some tropical regions and temperate regions (Trung *et al.*, 2004; World Health Organization [WHO], 2007a; Narain, 2008). In Thailand, forest areas and the international borders (Thai-Myanmar border and Thai-Cambodia border) are highly malaria transmission areas; such as Trat province, Tak province and Kanchanaburi province because of high number of migrants in these areas, which may be difficult to manage from formal health system (WHO, 2010).

Malaria has become world health problem. Three billion people in 109 malarious countries are risk of malaria infection and 250 million cases and 1 million deaths were annually reported (WHO, 2008). Most of malaria burden (population risk, cases and deaths) occurs in African region (Figure 1, Figure 2 and Figure 3), which *An. gambiae* is a major malaria vector and *Plasmodium falciparum* is main human malaria parasite (WHO, 2008; WHO, 2009). In South-East Asia region, Thailand is malaria-endemic areas (WHO, 2009) that there are 33,178 confirmed malaria cases and 97 deaths reporting in 2007 (WHO, 2010).

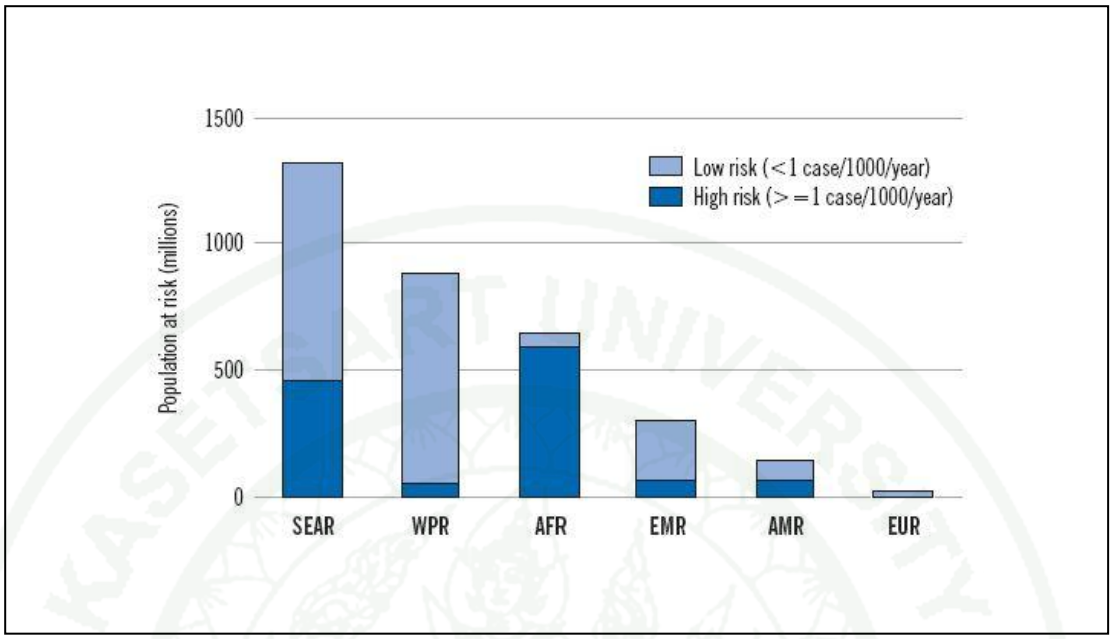


Figure 1 Number of populations in area with low and high risk of malaria (SEAR: WHO South-East Asia Region; WPR: WHO Western Pacific Region; AFR: WHO African Region; EMR: WHO Eastern Mediterranean Region; AMR: WHO Region of the Americas; EUR: WHO European Region).

Source: WHO (2008)

For malaria vector prevention and control, there are two effective procedures which these are the use of insecticide-treated nets (ITNs) and indoor residual spraying (IRS) (WHO, 2008). In Thailand, the tendency of the numbers of malaria case and death from malaria control activities has been uninterruptedly decreased (Figure 4 and Figure 5) (WHO, 2008; WHO, 2009), but malaria burden is still prevalent disease in hilly-forested areas along international borders where formal health system is restricted (WHO, 2010). In addition to insecticide resistance in mosquito and drug resistance in parasite are also becoming an important problem of global malaria control (Mehlotra and Zimmerman, 2006; WHO, 2009)

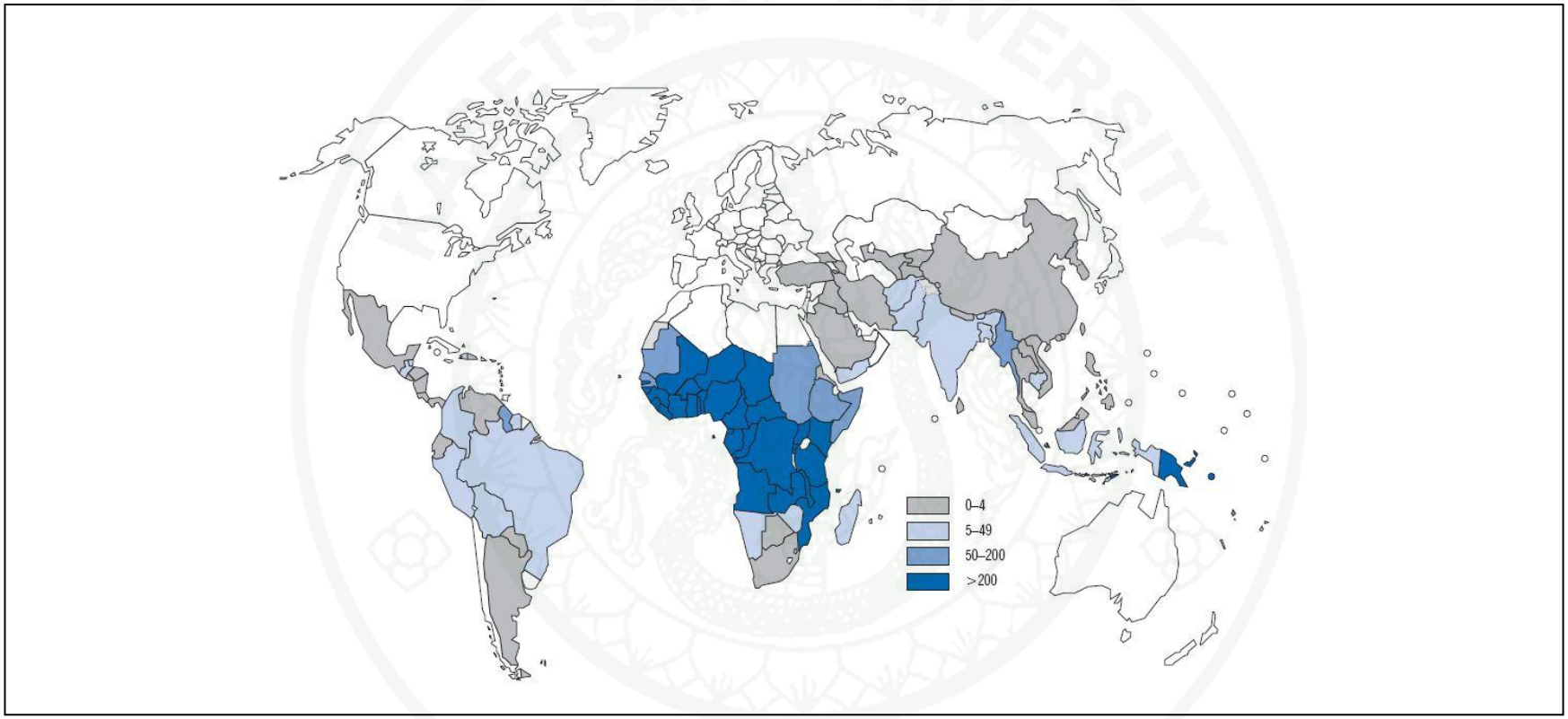


Figure 2 Estimated incidence of malaria per 1000 population, 2006.

Source: WHO (2008)

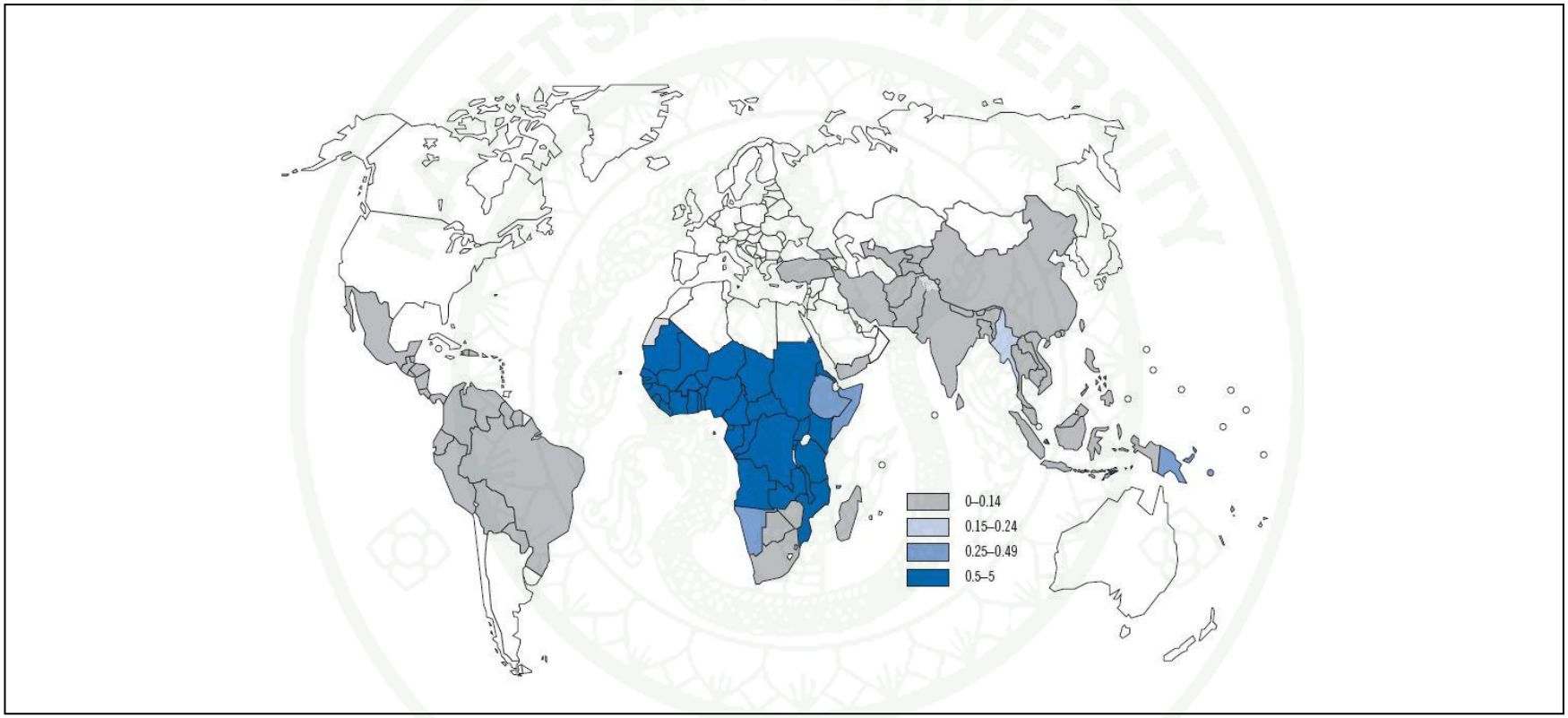


Figure 3 Estimated deaths from malaria per 1000 population, 2006.

Source: WHO (2008)

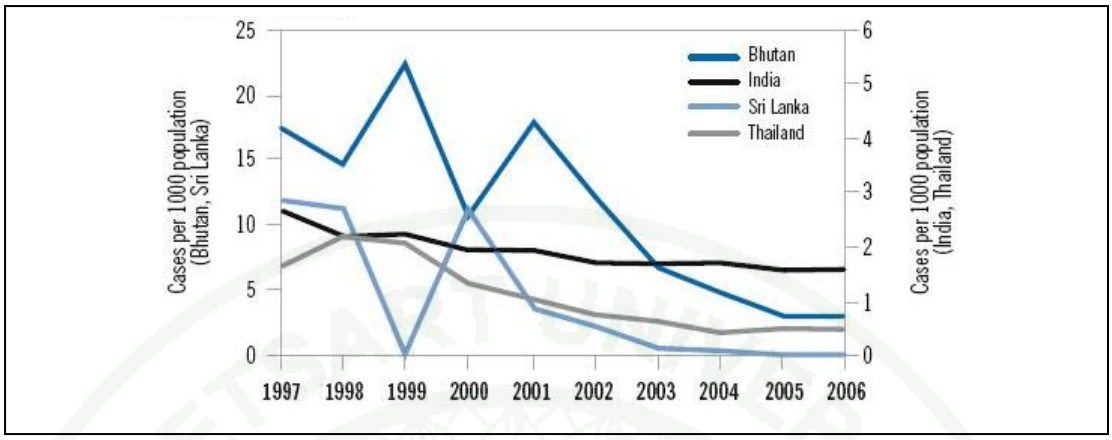


Figure 4 Tendency of number of reported malaria cases in selected countries of WHO region, for the years 1997-2006.

Source: WHO (2008)

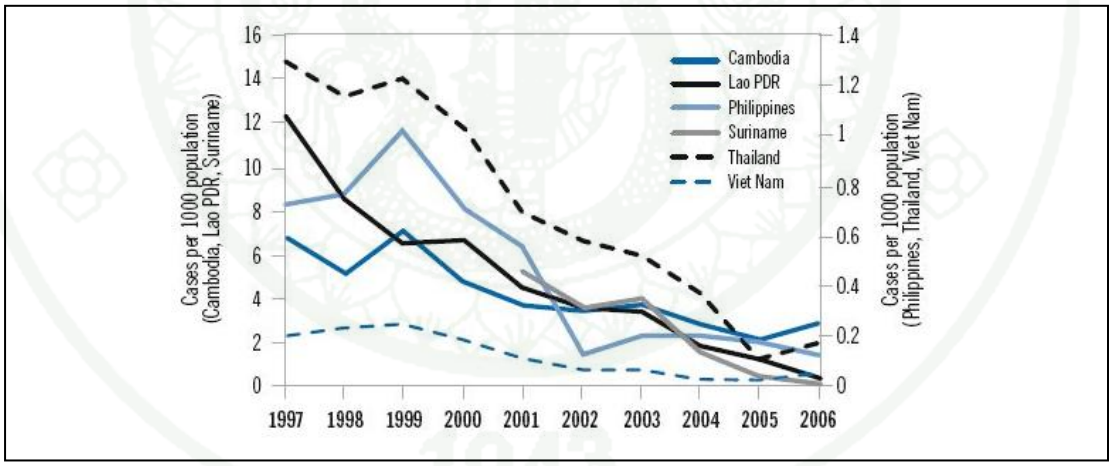


Figure 5 Tendency of number of reported malaria deaths in selected countries of WHO region, for the years 1997-2006.

Source: WHO (2008)

Association between host, parasite and vector

Malaria life cycle has several complex stages (Figure 6). Initially, the infected *Anopheles* mosquito inoculates the *Plasmodium* sporozoites into the human skin when it feeds on blood and then sporozoites migrate through bloodstream to liver (about 30 minutes after injection). When sporozoites invade the liver cells, they differentiate and multiply into merozoites form (infection periods are 6-15 days with no symptoms). Thousands of merozoites are released from liver into bloodstream and infect red blood cells (or erythrocytes). Merozoites multiply using asexual replication within red blood cells following several cycle steps in new erythrocytes. Few of infected the red blood cells leave from asexual cycles. They are parasite gametocytes which they are sexual parasite forms consisting of male and female gametocytes and then they take up in the bloodstream. This human host is the infected carrier. When mosquitoes as transmission vectors feed on blood from infected human, they ingest parasites from blood. Gametocytes develop to gametes and combine to zygotes, then zygotes grow, differentiate and divide into active haploid forms, this is the *Plasmodium* sporozoites form in mosquitoes (Pouniotis *et al.*, 2004; Silvie *et al.*, 2008; National Institute of Allergy and Infectious Diseases [NIAID], National Institutes of Health [NIH], Department of Health and Human Services [HHS], 2009; Anonymous, 2009).

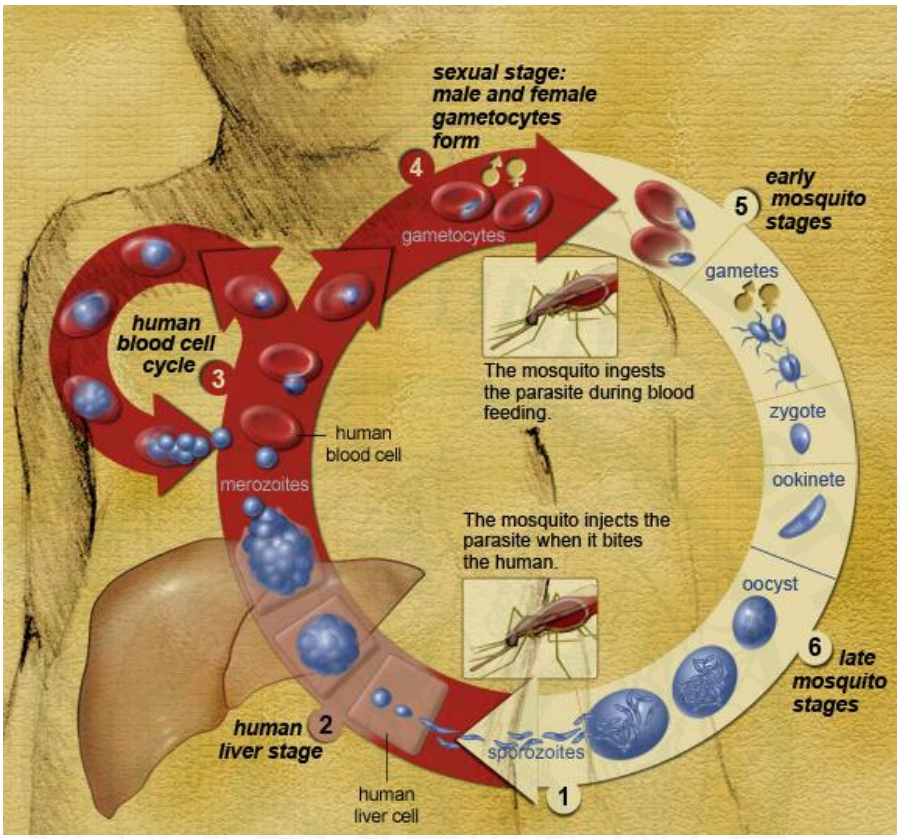


Figure 6 Life cycle of the malaria parasite.

Source: http://www.niaid.nih.gov/topics/malaria/pages/lifecycle.aspx

Malaria Vector

The mosquitoes of genus *Anopheles* are the transmitted vector of malaria diseases (Cann, 1996; Rich and Ayala, 2006). The genus *Anopheles* consists of 484 species of Anopheline mosquitoes including formal and informal named species and are separated into 6 subgenus; these are *Anopheles* (189 species), *Cellia* Theobald (239 species), *Kerteszia* Theobald (12 species), *Lophopodomyia* Antunes (6 species), *Nyssorhynchus* Blanchard (33 species) and *Stethomyia* Theobald (5 species) (Harbach, 2004). Only 70-80 species of the whole *Anopheles* mosquitoes are vectors of malaria transmission. The most of importance malaria vectors belong to subgenus *Anopheles* and *Cellia*, while genus *Lophopodomyia* and *Stethomyia* contain no malaria vectors (Manguin *et al.*, 2008a).

The mosquitoes of subgenus *Cellia* are the major malaria vectors in Southeast Asia (Manguin *et al.*, 2008b). This subgenus comprises seven taxonomic groups; these are the Minimus, Fluviatilis, Culicifacies, Dirus, Leucosphyrus, and Sundaicus Complexes, and the Maculatus Group (Manguin *et al.*, 2008b). In Thailand, six *Anopheles* mosquitoes are considered as malaria vectors. Three species of these have been examined as primary vectors, i.e. *An. dirus*, *An. minimus,* and *An. maculatus.* The rests of them have been considered as secondary vectors, i.e. *An. aconitus*, *An. sundaicus*, and *An. pseudowillmori* (The office of Diseases Prevention and Control (DPC), Department of Disease Control (DDC), Ministry of Public Health (MOPH), 2008).

The Minimus Complex consists of three formal species, i.e. *An. minimus*, *An. harrisoni*, and *An. yaeyamaensis* (Manguin *et al.*, 2008b; Harbach, 2011). Normally, *An. minimus* and *An. harrisoni* can be found in large sympatric areas in the Oriental Region (Manguin *et al.*, 2008b), while in Kanchanaburi province of Thailand, they have been found in restricted sympatric area (Poolprasert *et al.*, 2008). Moreover, *An. aconitus*, *An. varuna*, and *An. pampanai* are considered as the closely related species of this complex due to ambiguity of overlapping morphological characters (Manguin *et al.*, 2008b).

The Sundaicus Complex consists of at least four distinct species, including *An. sundaicus* s.s., *An. epiroticus*, *An. sundaicus* D, and *An. sundaicus* E (Dusfour *et al.*, 2007; Manguin *et al.*, 2008b). In Thailand, *An. epiroticus* is predominant species (Dusfour *et al.*, 2007; Manguin *et al.*, 2008b; Sumruayphol *et al.*, 2010) and could transmit malaria in coastal areas (Sumruayphol *et al.*, 2010).

The Maculatus Group are found in hilly or mountainous areas (Manguin *et al.*, 2008b) and consists of nine species; these are *An. dispar*, *An. greeni*, *An. pseudowillmori*, *An. willmori*, Maculatus Subgroup (*An. dravidicus* and *An. maculatus*), and Sawadwongporni Subgroup (*An. notanandai*, *An. rampae* and *An. sawadwongporni*) (Harbach, 2011). Five species of the Maculatus Group are distributed in Thailand, i.e. *An. pseudowillmori*, *An. willmori*, *An. maculatus*, *An. sawadwongporni* and *An. notanandai* (Manguin *et al.*, 2008b).

***Anopheles harrisoni*: the Minimus Complex**

An. harrisoni is one species of the Minimus Complex. This complex belongs to the Minimus Subgroup, the Funestus Group, the Myzomyia Series, the Laticorn Section, and within Subgenus *Cellia* (Harbach, 2004). Formally, the Minimus Complex consists of three species; these are *An. minimus*, *An. harrisoni*, and *An. yaeyamaensis* (Harbach, 2004). *An. minimus* and *An. harrisoni* are recognized as main malaria vectors in the mainland of the Oriental regions. *An. yaeyamaensis* can be found only in Ishigaki Island in the Ryukyu Archipelago of Japan, which it is a non-malarial area. Previous scientific name of *An. harrisoni* is *An. minimus* species C. In 2007, Harbach and colleagues had changed the name of *An. minimus* species C to *An. harrisoni* Harbach & Manguin. *An. harrisoni* has been named by honoring Dr. Bruce A. Harrison who is professional in mosquitoes (Harbach *et al.*, 2007).

An. harrisoni can be found in Vietnam, China, Myanmar, Laos and Thailand. *An. minimus* can be found in India, Bangladesh, Myanmar, Thailand, Laos, China, and Vietnam. Moreover, *An. minimus* and *An. harrisoni*, sibling species, also can be found in sympatric areas (Figure 7). In Thailand, *An. harrisoni* is widespread in 3 provinces; including Kanchnaburi province (Sharpe *et al.*, 1999; Sharpe *et al.*, 2000; Rwegoshora *et al.*, 2002; Kengluecha *et al.*, 2005; Sungvornyothin *et al.*, 2006a; Poolprasert *et al.*, 2008), Tak province (Green *et al.*, 1990; Rattanarithikul *et al.*, 2006), and Chiang Mai province (Sharpe *et al.*, 1999; Sharpe *et al.*, 2000; Rattanarithikul *et al.*, 2006), while *An. minimus* can be found throughout Thailand (Green *et al.*, 1990; Rwegoshora *et al.*, 2002; Rattanarithikul *et al.*, 2006). For morphological identification, the humeral pale (HP) spots on the costa veins of wings (Figure 8) were used to distinguish between *An. harrisoni* and *An. minimus* species (Rattanarithikul, 2006). The costa veins of *An. harrisoni* usually present the HP spots on the wings, while *An. minimus* usually absent (Figure 9 and Figure 10). However, morphological misidentification is the significant problem of these sibling species in sympatric areas (Kengluecha *et al.*, 2005). Thus, several molecular methods were developed for species identification, e.g. polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) (Van Bortel *et al.*, 2000), multiplex

allele-specific polymerase chain reaction (AS-PCR) (Phuc *et al.*, 2003; Garros *et al.*, 2004).

Habitats or breeding sites of *An. harrisoni* are stream margins with clear, slow-running current, stream pools, ground pools and hilly open areas in deforested agro-ecosystems (Kengluecha *et al.*, 2005; Garros *et al.*, 2006; Manguin *et al.*, 2008b), whereas *An. minimus* can be found in various habitats, i.e. stream margins with clear, slow-running current, stream pools, swamps, ditches, rock pools, rice-field terraces and even in water tanks (Kengluecha *et al.*, 2005; Garros *et al.*, 2006; Manguin *et al.*, 2008b). There are several types of host-feeding preferences of members of the Minimus Complex. In Bangladesh, India and Nepal, the anthropophilic behavior of mosquitoes of the Minimus Complex were observed, whereas zoophilic behavior was reported in north-eastern India (Garros *et al.*, 2006). In western Thailand, *An. harrisoni* and *An. minimus* would prefer zoophilic to anthropophilic behavior (Rwegoshora *et al.*, 2002; WHO, 2007b) and when feeding on human they would prefer outdoors to indoors (WHO, 2007b). Seasonal densities of *An. harrisoni* in western Thailand were reported on 2 periods; the early cool season (October and November), and the end of the cool season (February), while *An. minimus* was abundant at the end of rainy season (October and November) (Rwegoshora *et al.*, 2002). For biting behavior of the Minimus Complex was reported that they bit throughout the nights. In Thailand, large peak of biting activity is from 21.00 hours to 22.00 hours (Garros *et al.*, 2006 *Cited* Harbach *et al.*, 1987). Biting activity of *An. minimus* occurred throughout the night, allocating to two biting peaks; from 18.00 hours to 21.00 hours, and 3.00 hours to 6.00 hours (Chareonviriyaphap *et al.*, 2003).

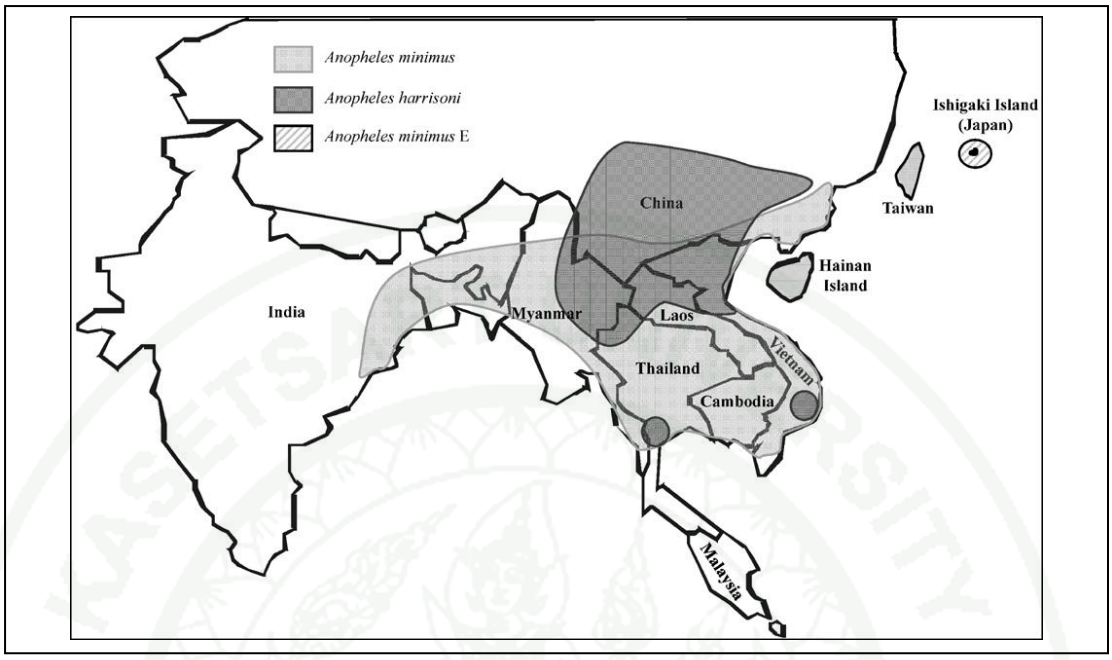


Figure 7 Distribution of the Minimus Complex.

Source: Manguin *et al.* (2008)

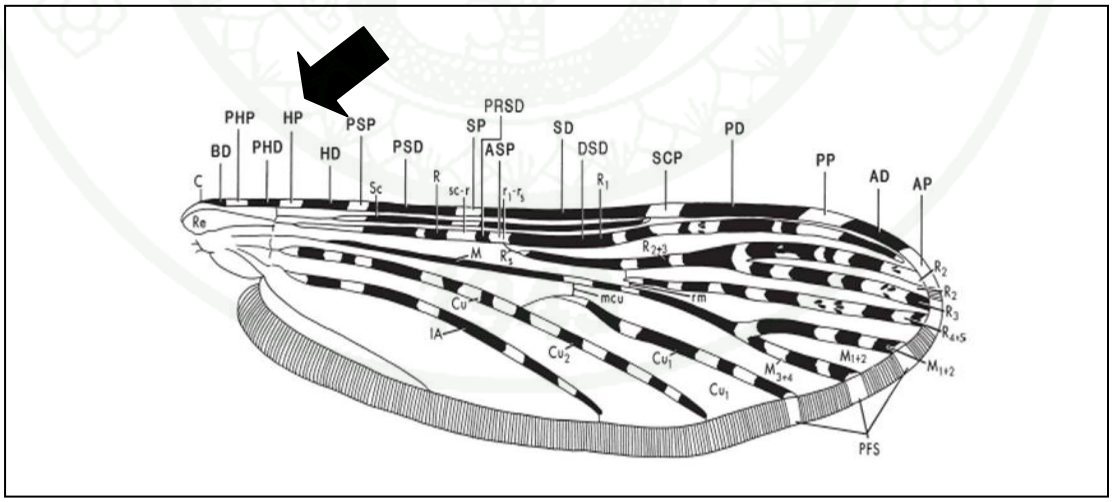


Figure 8 Humeral pale (HP) spots on the costa veins of mosquito wing (black arrow).

Source: http://www.scielo.br/img/revistas/rbent/v49s1/a1fig03.gif

Furthermore, the average flight range of mature *Anopheles* mosquitoes is a few hundred meters to two kilometers. In some mosquito species, the flight range is up to 10 kilometers from their breeding places (Sanofi-aventis, 2008). For the Minimus Complex, the flight ranges of *An. minimus* s.l. in Thailand and *An. yaeyamaensis* are around 2 kilometers (Garros *et al.*, 2006 *Cited* Meek, 1995) and 1 kilometer (Tsuda *et al.*, 1999), respectively.

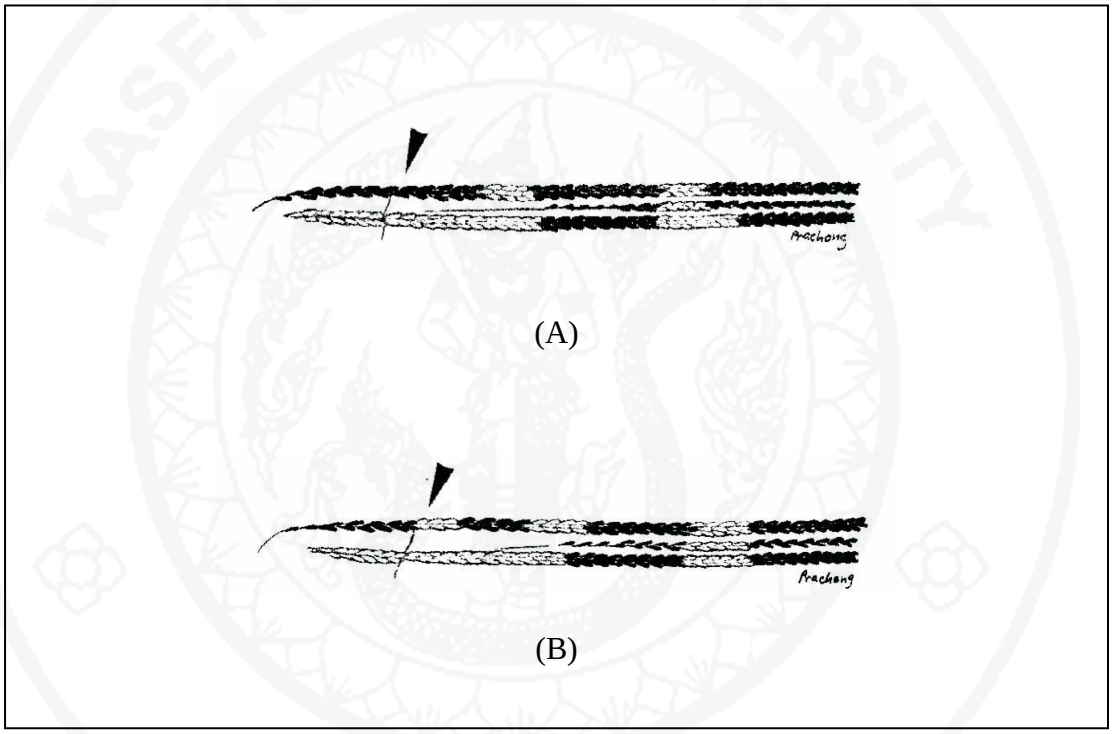


Figure 9 (A) Wing of *An. minimus* lacking the HP spots; (B) Wing of *An. harrisoni* showing the HP spots

Source: Rattanarithikul (2006)

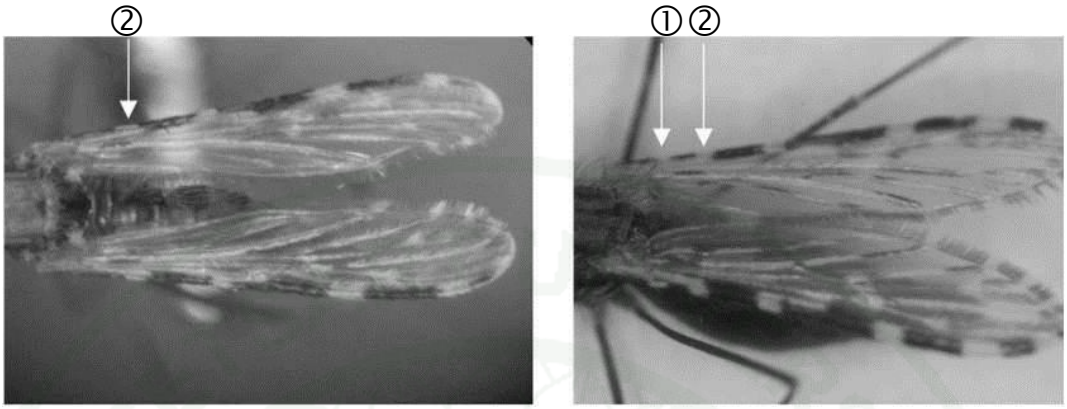


Figure 10 Difference of the costa veins of *Anopheles* mosquitoe wings; (A) wings of *An. minimus* absence the HP spots, while (B) wings of *An. harrisoni* presences the HP spots(①: the humeral pale spot (HP); ②: the presector pale spot).

Source: Sungvornyothin *et al.* (2006b)

Single Nucleotide Polymorphisms (SNPs)

A difference among nucleotides in DNA sequence, alleles in gene, or amino acids in protein sequence between individuals, groups, or populations is called genetic polymorphism (Nei and Kumar, 2000; Smith, 2002). Mutation plays important role for genetic polymorphism, such as nucleotide substitution, insertion/deletion, gene conversion, and interalleric recombination (Nei and Kumar, 2000).

Single Nucleotide Polymorphism (SNP; pronounced snip) is base substitution or changes in single nucleotide (point mutation) in DNA sequence (Figure 11) (Vignal *et al.*, 2002; Halliburton, 2004), occurring by chance (Smith, 2002). SNP is the most common variation at the DNA level (Halliburton, 2004). Two patterns of nucleotide substitutions are transitions and transversions. Transitions are base

substitutions between purine bases (Adenine, Guanine) or between pyrimidine bases (Cytosine, Thymine). Transversions are base substitutions between purine and pyrimidine bases (Nei and Kumar, 2000). SNPs are widespread in many regions including non-coding and coding regions (Smith, 2002; Morin *et al.*, 2004). Most SNPs are located in non-coding regions including regulatory regions (Kohn *et al.*, 2006; Jarukamjorn and Jatupornprasert, 2007). In coding regions, SNPs or point mutations can be divided into 2 effects on the amino acid sequence of protein; replacement and silent mutations. Replacement mutations (non-synonymous) are substitution of the amino acid with another. This mutation has 2 types; missense and nonsense mutations. Missense mutations change the amino acid, and nonsense mutations result in stop codon. Silent mutations result no effect on the amino acid sequence because of the genetic code degeneracy (Smith, 2002; Halliburton, 2004). Most of SNPs are simple bi-allelic, because the mutation rates of base substitutions show the low frequency (about 10^{-8}-10^{-9}). The probability of point mutation taking place at former position is very low (Vignal *et al.*, 2002; Brumfield *et al.*, 2003). In addition to the tri-allelic SNPs can be found in low frequency (less than 0.1%) (Lai, 2001; Petkovski *et al.*, 2003). Sometimes, one base pair indels (insertions or deletions) are also assumed as the SNPs (Vignal *et al.*, 2002). For average SNP frequencies of diverse taxa, SNPs identifying from a non-coding regions (ncSNPs) can be found every 200-500 base pairs and SNPs identifying from coding regions (cSNPs) can be found every 500-1000 base pairs (Brumfield *et al.*, 2003). Moreover, SNPs can be found in diverse organisms, such as vertebrates, invertebrates and plants (Rafalski, 2002; Andersen *et al.*, 2006; Cohuet *et al.*, 2008).

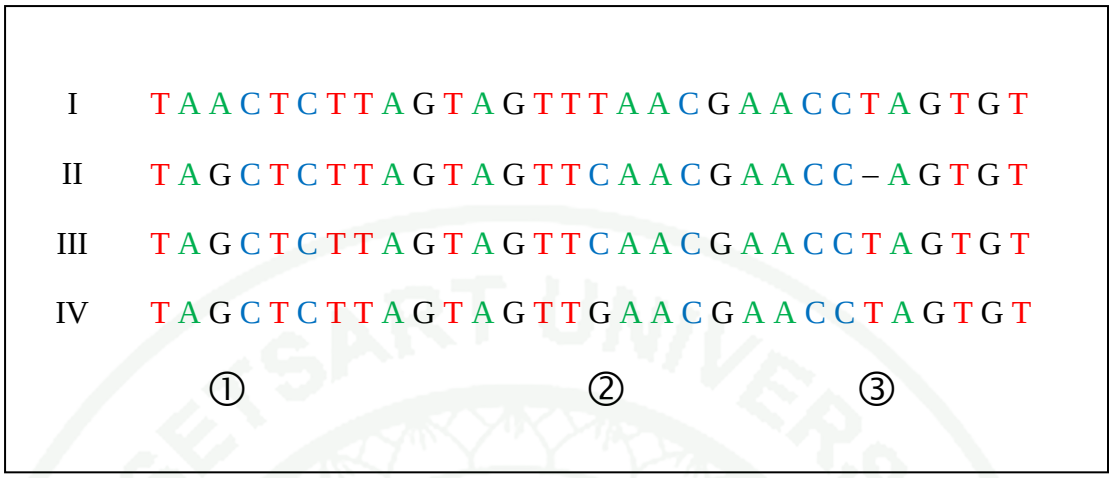


Figure 11 Sequence alignment showing base substitution and insertion/deletion; ①: the bi-allelic SNP; ②: the tri-allelic (multiple hit); ③: indel.

The advance of DNA sequencing technology (fast and cheap) generating abundant sequence information especially SNPs databases is one of many choice for using SNPs (DNA sequencing) (Halliburton, 2004; Morin *et al.*, 2004; Black IV and Vontas, 2007; Lister *et al.*, 2008) which direct sequencing is the simplest process for SNP discovery (Rafalski, 2002). SNPs are considered as an ideal genetic marker (Vignal *et al.*, 2002; Morin *et al.*, 2004), which is extensively used in many research fields, such as human (Guo and Jamison, 2005), wild populations of the endangered Italian wolf (*Canis lupus*) (Andersen *et al.*, 2006; Seddon *et al.*, 2005), immune related genes and randomly chosen genes in *An. gambiae* (Cohuet *et al.*, 2008), crops (corn and soybean) (Rafalski, 2002; Zhu *et al.*, 2003), fungus (Xu *et al.*, 2007), cultivated species (citrus) (Novelli *et al.*, 2004), vertebrates (bird, whale and cattle) (Heaton *et al.*, 2002; Cramer *et al.*, 2008) and insects (Black IV and Vontas, 2007). Table 1 and 2 showed the comparison of characteristics of molecular markers (Vignal *et al.*, 2002).

Table 1 Main categories of molecular markers.

Marker name	Variation type			Information content		
	SNP	**Indel**	**VNTR**	**dominant (2 alleles)**	**co-dominant (2 alleles)**	**co-dominant (multi alleles)**
RFLP	+	+[1]	+	–	+	+[2]
PCR-RFLP	+	+[1]	+	–	+	–
RAPD	+	+[1]	+	+	–	–
AFLP	+	+[1]	+	+	+[3]	–
SSCP	+	+[1]	+	–	+	+[2]
SSR	–	+[4]	+	–	–	+
SNP	+	+[5]	–	–	+	–

[1] Although the RAPD, AFLP, RFLP, PCR-RFLP and SSCP techniques will detect base substitutions in the vast majority of cases, the two other types of DNA variation can also be analysed.

[2] In some instances, more than two alleles can be analysed.

[3] With an automatic sequencer, some markers can be scored as co-dominant.

[4] Variations in PCR product length can be due to a deletion in the sequence flanking the microsatellite.

[5] Many SNP detection techniques can also be used for scoring small insertions or deletions (indels).

Source: Vignal *et al.* (2002)

Table 2 Technical requirements and characteristics.

Marker name	Technical requirement				Technical characteristics			
	Restriction enzyme	PCR	Specific primer	Gel	Development effort	Genotyping effort	Reproducibiliy[1]	Accuracy[2]
RFLP	+	–	–[3]	+	High	High	High	Very high
PCR-RFLP	+	+	+	+	High	Medium	High	Very high
RAPD		+	–	+	Very low	Very low	Low	Very high
AFLP	+	+	–	+	Low	Very low	High	Medium
SSCP	–	+	+	+	Medium	Medium	Medium	Medium
SSR	–	+	+	+	High	Low	High	High
SNP	–	+	+	+/–[4]	High	Variable[4]	High	Very high

[1] Refers to the genotyping error rate of the method: results may vary from one experiment to another.
[2] Refers to the precision at which true allele recognition can be performed.
[3] However, the RFLP technique relies on the use of a specific probe for the Southern-blot technique. Nowadays, RFLPs are usually genotyped by PCR-RFLPs, requiring specific primers.
[4] According to the genotyping technique used.

Source: Vignal *et al.* (2002)

Population genetic studies of the Minimus Complex

Seven enzyme electromorphic loci were used to study in *An. minimus sensu lato* from eleven natural populations in Thailand. The result showed that the absence of heterozygotes at many electromorphic loci indicating two isomorphic species within the taxon *An. minimus* Theobald inThailand (Green *et al.*, 1990).

Isozyme electrophoresis was used to clarify the *An. minimus* sensu lato species composition in northern Vietnam and identify behavioural divergences of individual species. The Octanol dehydrogenase (Odh) enzyme locus could be used to distinguish *An. minimus* s.l. from closely related species, *An. aconitus* and *An. jeyporiensis*. Significant positive F_{IS} values showed clear evidence of nonrandom mating within the *An. minimus* s.l. population. No evidence for restricted gene flow was found between monthly samples, villages, or collection methods in either of the two *An. minimus* species. These sympatric species showed different proportions depending on the collection site, and had dissimilar resting and biting behaviors (Van Bortel *et al.*, 1999).

Population structure of *An. minimus* in Vietnam was studied by using allozyme data. The studied localities are 2 rural and 1 urban areas. The existence of *An. minimus* in both areas suggests that the ability of this species to adapt to anthropogenic environmental changes. The results showed small significant genetic differentiation between rural and urban of *An. minimus* populations. Limited differentiation was also observed between two rural populations at distances of more than 1000 km. This result suggests that geographical distance does not seem to be the primary factor differentiating mosquito populations. Population structure found between the two rural populations indicates that genes may spread over large areas. This could be influenced the spread of possible insecticide resistance (Van Bortel *et al.*, 2003).

Five microsatellite markers were developed for studying the genetic diversity of *An. minimus*. Preliminarily, 60 samples of *An. minimus* from Mae Sod district, Tak province, Thailand were used to test these primer groups. The results showed that all

microsatellite loci showed a high level of polymorphism in Mae Sod population (Boonsuepsakul *et al.*, 2005).

Low genetic differentiation (F_{ST} = 0.061) among 7 collections from Kanchanaburi province, Thailand (6 populations of *An. minimus* and 1 population of *An. harrisoni*) was observed from studying in 8 enzyme systems, 9 loci. The data among all populations showed no correlation between genetic and geographical distance ($P > 0.05$) and the highest variation was detected in Bong Ti Noi Village, as the lowest variation was showed in Tha Kradan Village (Poolprasert *et al.*, 2008).

The eight DNA fragments were developed from available complete genome sequence of *An. gambiae*. These multilocus nuclear DNA markers were used as the choice of genetic markers for studying and inferring the population genetic structure and demography affecting on the pattern of nucleotide polymorphism based on single nucleotide polymorphisms (SNPs) in Indian populations of *An. minimus.* This study showed that these DNA fragments evolved under a neutral model of molecular evolution, linkage disequilibrium and demographic equilibrium model. These DNA fragment were considered as putative neutral markers for inferring population structure and demographic history of *An. minimus* (Dixit *et al.*, 2011).

Phylogeny of *Anopheles* mosquitoes according to subgenus *Cellia*

Most species of genus *Anopheles* (about 80%) belong to subgenus *Cellia* and *Anopheles* (Anthony *et al.*, 1999). These species mostly exist in temperate, subtropical and tropical areas of the world, but cannot be found in the Pacific Islands and isolated islands in the Atlantic. Moreover, they can be found in different geographies from hilly to coastal areas (Krzywinski and Besansky, 2003; Harbach, 2011).

Because of ambiguous relationships among important vectors of genus *Anopheles*, the interrelationships and origin of Australasian mosquitoes were compared to Oriental species based on the mitochondrial COII fragments. Character analysis supplementing with geographic distribution were also considered. The results

showed that a characteristic of anopheline paleobiogeographic in Australasia species arose from two-way exchange rather than only immigration with Orientals, and the Punctulatus Group of Australasian mosquitoes showed monophyly and separated from Orientals. Moreover, these sample populations of this group were supported by dispersal rather than vicariance explanation due to collision in the recent past (Approximately 0.35 to 2.44 million years ago (mya)) of distal Australia and Vanuatu, and species belonging to the Neomyzomyia Series are the most primitive position based on morphological characters (Foley *et al.*, 1998).

The Oriental and the Afrotropical species of the Pyretophorus Series of subgenus *Cellia* were investigated based on morphological characters and the cibarial armature carrying phylogenetic signal was utilized in the first time. A cladistic analysis generated the Oriental species as a monophyletic group, while the Afrotropical are basal and paraphyletic relative to the Oriental. The outgroup taxa comprising the Neocellia, Neomyzomyia and Myzomyia Series were clearly separated from the Pyretophorus. Moreover, the pattern of relationships showed that the capacity of malaria transmission is independent from ancestral (Anthony *et al.*, 1999).

64 species mosquitoes in subfamily Anophelinae covering 3 genera (genus *Anopheles*, *Bironella* and *Chagasia*) were studied phylogenetic relationships based on 163 morphological characters. The results showed monophyletic clade between genus *Anopheles* and *Bironella*. This clade comprised three major lineages. The basal lineage (lineage 1) consists of *An. implexus* (subgenus *Anopheles*) and members of subgenus *Nyssorhynchus* and *Kerteszia*. The second lineage (lineage 2) contains species of the subgenus *Cellia*, and the third lineage (lineage 3) consists of members of subgenus *Anopheles*, *Stethomyia*, *Lophopodomyia* and the genus *Bironella* (Sallum *et al.*, 2000).

Members of the Minimus Group of subgenus *Cellia* (*An. aconitus*, *An. varuna*, *An. minimus* A and *An. minimus* C) were investigated intra- and interspecific variation based on the mitochondrial COII and D3 of rDNA regions. The result confirmed that *An. minimus* in Thailand consists of at least two cryptic species besides species A and C. The divergence time between *An. minimus* A and C based on the mitochondrial COII and ITS2 regions was estimated around 0.57 to 1.5 mya. The result of estimated effective population size suggested that distribution of *An. minimus* C probably has larger than current recognition (Sharpe *et al.,* 2000).

Anopheles species belonging to subfamily Anophelinae were studied phylogenetic relationships based on fragments of the partial COI, COII, 18S rRNA and D2 region of 28S rRNA. These species include 3 genera (*Anopheles, Bironella* and *Chagasia*) and 6 subgenera of *Anopheles* (*Anopheles*, *Baimaia*, *Cellia*, *Kerteszia*, *Lophopodomyia*, *Nyssorhynchus* and *Stethomyia*). The results of phylogeny showed that the combined rDNA data hold strong phylogenetic signal, while the mtDNA hold little phylogenetic signal, except groups of most recently derived species (Sallum *et al.*, 2002).

Fifteen Oriental and two Afrotropical taxa of the Myzomyia Series in subgenus *Cellia* were utilized to investigate the phylogenetic relationships with the Neocellia (*An. maculatus*) and Neomyzomyia (*An. dirus* A) Series considering as outgroup species. Maximum likelihood and maximum parsimony analyses based on separated COII and D3 of 28S rDNA and combination have identified five monophylies; *An. fluviatilis* U and T; *An. fluviatilis* U and T + *An. minimus* A, C, E and #157 + *An. leesoni*; *An. filipinae* + *An. mangyanus*; *An. filipinae* + *An. mangyanus*+ *An. aconitus*; and *An. culicifacies* A and B. The results confirm the specific ststus of *An. flavirostris*, the close relationship of *An. leesoni* with the Minimus Complex, and the exclusion of *An. jeyporiensis*, *An. culicifacies* s.l. and *An. funestus* from the Minimus Group. Morphological species of the Minimus Group formed a single clade comprising 2 subgroups, these are the Minimus Subgroup (*An. minimus* s.l., *An. fluviatilis* s.l. *An. leesoni* and *An. flavirostris*) and the Aconitus Subgroup (*An. filipinae*, *An. mangyanus*, *An. aconitus*, *An. pampanai* and *An. varuna*) supporting with weakly present dataset (Chen *et al*., 2003).

The morphological classification in both of the Afrotropical Funestus and Afro-Oriental Minimus groups has been reflected their ambiguous relationships. Morphological and molecular studies were together utilized for examine. These species groups showed shared morphological characters based on cross-identification. Molecular analysis separated species into five clades (the Aconitus, Culicifacies, Funestus, Minimus, and Rivulorum subgroups), but they were not congruent with geographical distribution. Moreover, the results of morphological and molecular studies (based on ribosomal D3 and mitochondrial COII fragments) were accordance (Garros *et al.*, 2005a) (Now, the Funestus and Minimus Groups have been united into a composite Funestus Group consisting of five subgroups: the Aconitus, Culicifacies, Funestus, Minimus and Rivulorum Subgroups [Harbach, 2004]).

Two groups of *Anopheles* mosquitoes with geographical difference (Afrotropical Funestus and Oriental-African Minimus Groups) were considered the phylogenetic relationships by using the comparison of ribosomal (ITS2 and D3) and mitochondrial (COI) nucleotide sequences. The results showed the clearly three clades of the studied speies, i.e. the Funestus, Minimus and Rivulorum clades. An African species, *An. leesoni*, was confirmed to belong to member of the Minimus Groups with the Oriental species. *An. rivulorum* defining as the Afrotropical species was placed in basal position of phylogenetic relationships relating to the African and Oriental species. Estimated divergence time showed that these species correspond to periods of major tectonic movment as well as periods of great aridity or humidity (Garros *et al.*, 2005b).

The phylogenetic relationships of the Anopheline mosquitoes of subgenus *Cellia* based on the mitochondrial COI and COII fragments, the D3 of the ribosomal 28S RNA and ITS2 region were studied for understanding the interrelationship among various mosquitoes and resolving the unclear morphological characters from proper species position in phylogeny. The results showed that arrangement of the various Anopheline taxa based on the D3 and ITS2 region conforms to morphological character-based classification, while the COI and COII are not. They suggested that selection of proper degree of the evolution rate probably indicates the correct phylogenetic analysis of these species (Mohanty *et al.*, 2009).

Phylogenetic inferences of Indian malaria vector species were investigated based on multilocus DNA sequences, i.e. the COII, CYP, NADPH, NOS and ITS2 genes. The results showed that the tree topologies of COII and ITS2 were congruent, and the estimated divergence time with COII gene support the hypothesis of the species radiation of the *Anopheles* genus during the late Cretaceous period (Dixit *et al.*, 2010).

Accurate species identification is particularly important in vector-borne disease studies. Thus, ambiguous morphological identitification between the Funestus group in the Myzomyia Series (*An. fluviatilis, An. culicifacies, An. varuna* and *An. aconitus*) and the Annularis group in the Neocellia Series (*An. annularis, An. pallidus* and *An. philippinensis*) were addressed by developing a multiplex-PCR assay based on sequence of the D3 region of 28S rDNA. Phylogenetic relationships inferring from maximum parsimony and neighbor joining showed that these species were separated into two distinct monophyletic clades, species of Myzomyia and other of species of the Neocellia Series. This molecular phylogeny was consistent with the classical morphological taxonomy reasonably well (Swain *et al.*, 2010).

MATERIALS AND METHODS

Part A: Genetic diversity and tests of population genetic models

Mosquito samples

Adult female mosquitoes of *An. harrisoni* were observed and collected from 2 sampling sites as shown below in Table 3 and Figure 12.

Likewise, DNA samples of *An. harrisoni* collecting from Vietnam were kindly provided by Dr. Sylvie Manguin from Faculté de Pharmacie, University Montpellier I, Montpellier, France. The sampling sites are also shown in Table 3 and Figure 13.

Table 3 Sampling sites of *An. harrisoni*.

Locality	**No. of samples**	**Latitude (N)**	**Longitude (E)**
(1) Pu Teuy Village, Sai Yok district, Kanchanaburi province, Thailand (KPT)	10	14º 17'	99º 11'
(2) Bong Ti Noi Village, Sai Yok district, Kanchanaburi province, Thailand (KBT)*	2	14 º 15' 53.2"	98 º 55' 52.8"
(3) Khoi Village, Hoa Binh province, Vietnam (VHB)	2	20º 38' 11.1"	105º 09' 58.4"

* Co-ordinates were calculated from Universal Transverse Mercator (UTM) system to Degree-Minute-Second (DMS) system by using GPS Sway program (GPSFileDepot, 2007).

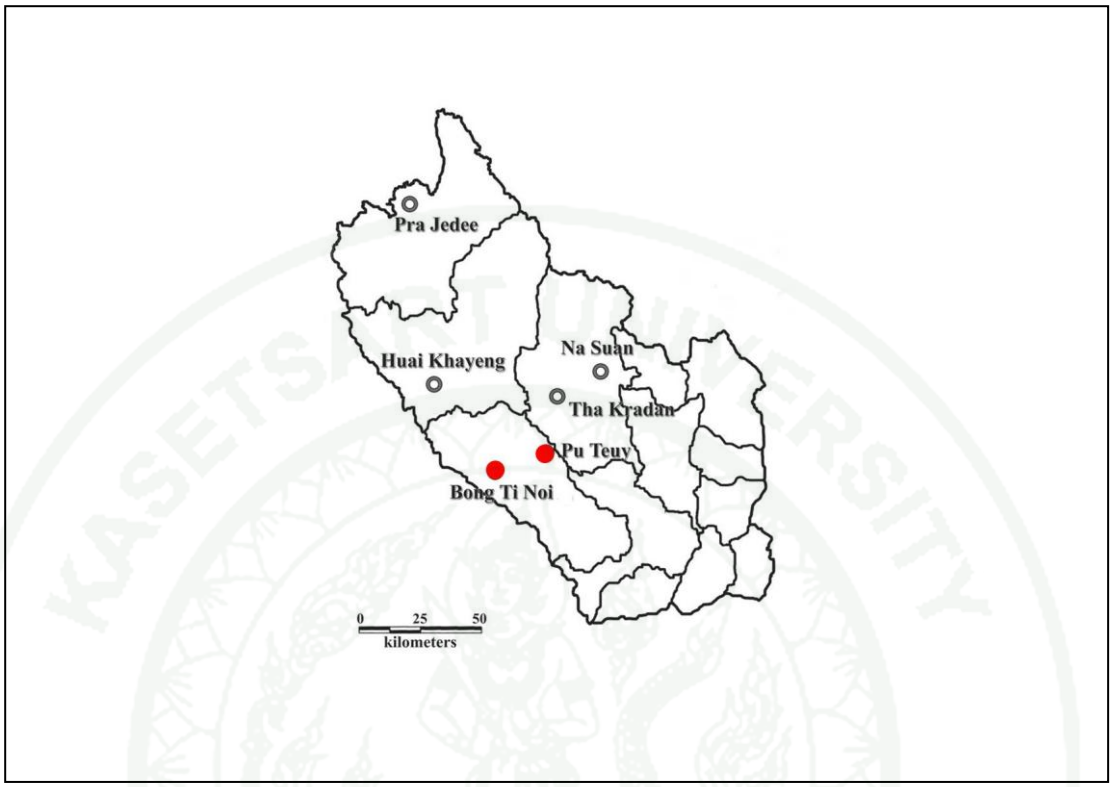


Figure 12 Two sampling sites in Kanchanaburi province; Bong Ti noi Village and Pu Teuy Village (●).

Source: Poolprasert *et al.* (2008)

Figure 13 Sampling sites in Khoi Village, Hoa Binh province, Vietnam (●).

Source: Modified from http://www.savethechildren.net/vietnam/map.jpg

Development of neutral DNA sequence markers

The published complete genome sequence information of *An. gambiae* was used to develop putative neutral DNA sequence markers. Introns (length about 300 to 700 bp) of orthologous genes locating on X-chromosome of *An. gambiae* were chosen for developing the primes. The information of X-chromosome of *An. gambiae* is showed as following (www.ensembl.org):

Length (bps):	24,393,108
Known Protein-coding Genes:	106
Novel Protein-coding Genes:	982
miRNA Genes:	1
tRNA Genes:	5
SNPs:	51,682

The information of the chosen genes is provided in Table 4.

Table 4 Information of the interested genes for developing primers.

Gene ID	Gene Description	Gene Size (Kb)	Location at X chromosome	No. of Exon	Between exons: Intron Length (Primer name)
AGAP000369	Putative GPCR class a orphan receptor 10	2.12	6,791,093-6,793,207	3	2-3: 486 bp (O3)
AGAP000500	NADPH cytochrome P450 reductase	4.75	8,794,244-8,798,989	12	5-6: 415 bp (O4)
AGAP000824	Gbb-60A	2.88	15,336,624-15,339,506	8	2-3: 680 bp (O8)
AGAP001043	Armadillo segment polarity protein	14.12	20,143,880-20,158,004	5	2-3: 435 bp (O9)
AGAP001076	Cytochrome P450 CYP4G16	9.19	22,937,938-22,947,129	6	3-4: 579 bp (O10)

The primers were developed from intron flanking-region of interested gene using EPIC (Exon-Primed Intron-Crossing) method (Bierne *et al.*, 2000) (Figure 14). Two primer pairs developing in Indian *An. minimus* (Primer O8, Gene ID AGAP000824; Primer O10, Gene ID AGAP001076) (Dixit *et al.*, 2011), three primer pairs developing in *An. gambiae* (Primer P9, Gene ID AGAP001076; Primer P12, Gene ID AGAP000124; Primer P22, Gene ID AGAP001039) (Stump *et al.*, 2005), and one primer pairs utilizing in development of PCR identification in the Minimus Complex and the closely related species (Primer ITS2) (Garros *et al.*, 2004) were selected for PCR amplification in *An. harrisoni* from Thailand and Vietnam. Information of the genes, the primer sequences and annealing temperature are listed in Table 5.

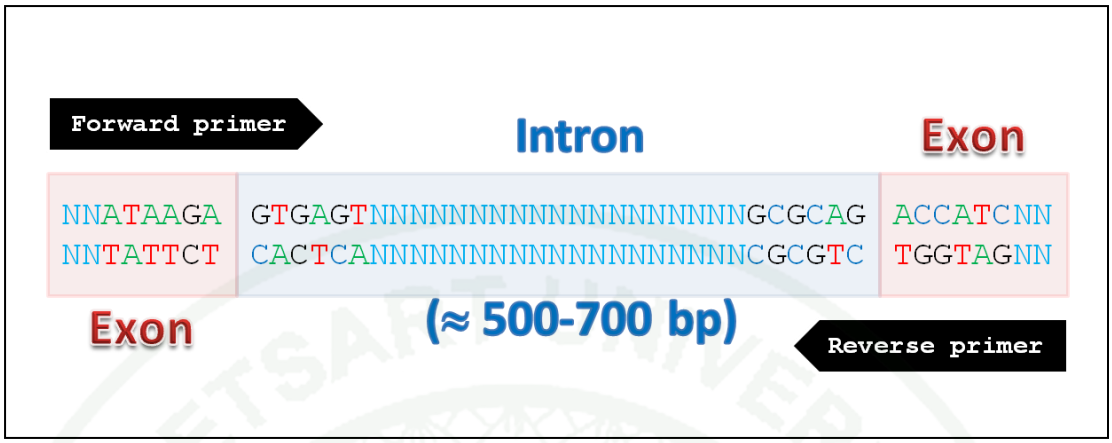


Figure 14 Using EPIC method for primer design, forward and reverse primers were designed from intron flanking-region of interested gene.

Table 5 Information of primers, the sequence of primers and their annealing temperature.

Gene (Primer name)	Gene details		Details of analyzed fragments			Reference
	Chromosomal location	Putative function	Region	Primer sequences	Annealing temperature	
CYP (P9)	X chromosome	Metabolism of xenobiotics	Non-coding	F: TGG AAA GTA GGG GCA ATC AG R: GCT GGG CAA CGG ACT ACT TA	51 ºC	Stump *et al.* (2005)
INPP4 (P12)	X chromosome	Inositol phosphate metabolism	Non-coding	F: TTT ACG TCG TGC AAG AGT GC R: GCC CAC GTT CTT CAT CGT AT	51 ºC	Stump *et al.* (2005)
CYP (P22)	X chromosome	Metabolism of xenobiotics	Non-coding	F: CTC TAC GGA ACG GAG CTG TC R: CAG GGT CGA CCA GTC ACA TA	51 ºC	Stump *et al.* (2005)
Gbb-60A (O8)	X chromosome	Signaling proteins, controls proliferation, differentiation	Non-coding	F: CGG CCA TTG GGC GGA CAT TT R: ACG CAT CTC CGA TCC GTG ATG ATC	60 ºC	Dixit *et al.* (unpublished data)
CYP (O10)	X chromosome	Metabolism of xenobiotics	Non-coding	F: CTG GCT GGG CAA CGG ACT A R: GGA AAG TAG GGG CAA TCA GTT TG	58 ºC	Dixit *et al.* (2011)
ITS2 (ITS2)	Nuclear genome	No function	Non-coding	F: TGT GAACTG CAG GAC ACA T R: TAT GCT TAA ATT CAG GGG GT	50 ºC	Garros *et al.* (2004)

Note: *CYP*, Cytochrome P450; *INPP4*, Inositol polyphosphate 4-phosphatase; *Gbb-60A*, Glass bottom boat 60A (Transforming growth factor beta at 60A); ITS2, Ribosomal internal transcribed spacer 2.

Mosquito collections

Adult female mosquitoes of *Anopheles* spp. were collected using human-landing and cattle-bait collection methods. Aspirator was used to capture the adult mosquitoes. Alive mosquito samples are required for morphological identification. Period for mosquito collections was from 6 p.m. until midnight. The collected samples were kept in a small cup and maintained by controlling the humidity and feeding the 10% sugar solution (weight/ volume; ten grams of sucrose dissolved in 100 ml of solution).

Morphological identification

Adult mosquitoes were identified as member of the Minimus Complex using morphological keys (Rattanarithikul, 2006). The presence or absence of humeral pale (HP) spot on the costa veins of the mosquito wings is primarily used to distinguish species of the Minimus Complex. The wings of *An. minimus* usually absent HP spot, while *An. harrisoni* usually present HP spot on the wings (Figure 9).

Extraction of genomic DNA

Genomic DNA was extracted from individual adult female mosquitoes using the modification of Genomic DNA Mini Kit (Tissue) (Geneaid, Taiwan) describing in Appendix A. Good quality of DNA is required. Three microliters of each DNA sample were separated on 1.0% agarose gel (Vivantis, California) in 0.5X TAE buffer (Vivantis, California) for quality and concentration examination and 1 Kb ladder (Vivantis, California) was use as standard marker. Gel was strained with ethidium bromide and the DNA bands were visualized under ultraviolet light.

Molecular identification

The multiplex allele-specific polymerase chain reaction (AS-PCR) assay was used to distinguish and confirm the species of the Minimus Complex (*An. harrisoni* and *An. minimus*) and the closely related species, i.e. *An. aconitus*, *An. pampanai*, and *An. varuna* as following Garros *et al.* (2004). The primers of AS-PCR are shown below in Table 6. These primers were developed from polymorphism presented in internal transcribed spacer 2 (ITS2) of ribosomal DNA. The PCR reaction volume and thermocycler profiles are provided in Appendix B. Three microliters of PCR products were separated on 1.5% agarose gel in 0.5X TAE buffer (Vivantis, California) and 100 bp ladder (SibEnzyme, Russia) was used as standard marker. Agarose gel electrophoresis was performed for 25 minutes at 100 volts. Gel was strained with ethidium bromide and the PCR product bands were visualized under ultraviolet light.

Table 6 Multiplex AS-PCR primers for identification of the Minimus Complex and the closely related species (Garros *et al.*, 2004).

Species (Primer name)	Sequence (5´ to 3´)	Product size (bp)	Tm (ºC)
Universal forward primer (ITS2A)	TGT GAA CTG CAG GAC ACA T		54.5
An. minimus (MIA)	CCC GTG CGA CTT GAC GA	310	57.6
An. harrisoni (MIC)	GTT CAT TCA GCA ACA TCA GT	180	53.2
An. aconitus (ACO)	ACA GCG TGT ACG TCC AGT	200	56.0
An. varuna (VAR)	TTG ACC ACT TTC GAC GCA	260	53.7
An. pampanai (PAM)	TGT ACA TCG GCC GGG GTA	90	58.2

PCR amplification and sequencing

PCR amplification was performed in 30 µl volume. The reaction mixture and the reaction conditions for PCR amplification are described in Appendix C. The information of primers, the sequence of primers and their annealing temperature are given in Table 5. Three microliters of PCR products were separated on 1% agarose gel in 0.5X TAE buffer (Vivantis, California) and 100 bp ladder (SibEnzyme, Russia) was use as standard marker. Agarose gel electrophoresis was performed for 25 minutes at 100 volts. Gel was strained with ethidium bromide and the PCR product bands were visualized under ultraviolet light. PCR products were purified describing in Appendix D; the excess primers were digested with Exonuclease I and the dNTPs with FastAP™ Thermosensitive Alkaline Phosphatase (Fermentas, Ontario). The purified PCR products were directly sequenced in both directions using forward- and reverse-specific primers by Macrogen Inc. (Korea). Forward and reverse sequences of each gene fragment were compared and validated with *Anopheles* species using nucleotide BLAST program from the NCBI database. Further, the nucleotide sequences were manually checked for accuracy between base-calling and chromatograms before editing. Homozygotes and heterozygotes were visually analyzed from the chromatograms. One fluorescent peak pattern was defined as homozygous sequences, by the way, two peaks at a single position in the chromatograms was determined as heterozygous sequences. Nucleotide sequences were aligned and edited with BioEdit version 7.0.5.3 computer programs (Hall, 1999) using ClustalW alignment algorithm (Thompson *et al.*, 1994). Nexus file format was prepared for further population genetics analyses as implemented in DnaSP version 5.00.07 computer programs (Librado and Rozas, 2009).

Data analysis

1. Estimation of nucleotide diversity

Two parameters were used to estimate the nucleotide diversity. These are θ_w (Watterson, 1975) and π (Nei, 1987) parameters.

π is the average number of difference between all pairs of sequences sampled. This proportion is estimated by:

$$\pi = [n/(n\text{-}1)] \sum x_i x_j \pi_{ij}$$

where

n: the number of sequences examined

x_i and x_j: the observed haplotype frequencies in the sample

π_{ij}: the observed proportion of differences between the ith and jth haplotype

θ_w or Watterson estimator is the number of segregation sites, polymorphic or any nucleotide site that show two or more nucleotides among the investigated sequences (Nei and Kumar, 2000), in a sample from a population for an autosomal gene of a diploid organism. It was calculated as following:

$$\theta_w = 4N_e\mu$$

where

N_e: the effective population size

μ: the mutation rate per nucleotide site per generation under investigation

2. Test of neutrality

Neutral theory was initially assumed by Kimura (1983). This theory proposes that the interactions between neural mutations and random genetic drift are the cause of the most substitutions among species and polymorphisms within species, which they are not due to natural selection (Halliburton, 2004).

There are several statistical techniques for testing neutral expectation; i.e. The Ewens-Watterson test (1978), Tajima's test (1989), the HKA (Hudson, Kreitman and Aguadé) test (1987), the McDonald-Kreitman test (1991), the test of Fu and Li (1993), and the test of Fay and Wu (2000). Generally, a null hypothesis of these statistics proposes that the amount and pattern of genetic diversity are under the neutrality. Significant deviations might infer that these polymorphisms are under natural selection (Halliburton, 2004).

Tajima's *D* statistical method (Tajima, 1989) was used to test for deviations from the standard neutral expectation. This statistic is a test for discrimination between patterns of diversity evolving random (mutation and random genetic drift; neutrality) and non-random process (e.g. selection, demography) and requires 2 estimators for investigation; these are the number of segregating sites (θ_w) and the average number of nucleotide differences estimated from pairwise comparison (π). Under neutral evolution, Tajima's *D* value should be close to zero ($\pi \approx \theta_w$), implying the balance between genetic drift and mutation is influence of genetic variation. A negative *D* value indicates an excess of low frequency polymorphisms indicating population expansion or purifying selection (mutation-selection equilibrium) ($\pi < \theta_w$), whereas a positive *D* value indicate an excess of intermediate-frequency polymorphisms inferring balancing selection, population bottleneck or population subdivision ($\pi > \theta_w$). It was calculated as following:

$$D = \frac{\pi - \theta w}{\sqrt{\hat{V}(\pi - \theta w)}}$$

where

$\hat{V}(\pi - \theta w)$: the variance of $(\pi - \theta w)$

Fu and Li's *D* and *D** statistics (Fu and Li, 1993) were utilized to test the assumption of selective neutrality (Kimura, 1983). The distributions of the mutations between the external and internal branches in the genealogy of a random sample of genes from the population were compared. The excess of mutations in the external branches indicates the presence of purifying or negative selection, while the excess of mutations in the internal branches indicates the presence of balancing (overdominant) selection.

This Fu and Li's *D* statistic test requires the intraspecific variation and an outgroup sequences. It was calculated as following:

$$D = \frac{\eta - a_n \eta_e}{\sqrt{u_D \eta + v_D \eta^2}}$$

where

η: the total number of mutations

η_e: the number of mutations in the external branches

n: the number of sequences

k: the number of segregating sites

$$v_D = 1 + \frac{a_n^2}{b_n + a_n^2}\left(c_n - \frac{n+1}{n-1}\right)$$

$$u_D = a_n - 1 - v_D$$

$$a_n = \sum_{k=1}^{n-1} \frac{1}{k}$$

$$b_n = \sum_{k=1}^{n-1} \frac{1}{k^2}$$

$$c_n = 2\frac{na_n - 2(n-1)}{(n-1)(n-2)}$$

This Fu and Li's *D** statistic test requires only the intraspecific variation sequences. It was calculated as following:

$$D^* = \frac{\left(\frac{n}{n-1}\right)\eta - a_n\eta_s}{\sqrt{u_{D*} + v_{D*}\eta^2}}$$

where

η: the total number of mutations

η_s: the number of singletons

n: the number of sequences

k: the number of segregating sites

$$u_{D*} = \frac{\left[\left(\frac{n}{n-1}\right)^2 b_n + a_n^2\, d_n - 2\frac{na_n(a_n+1)}{(n-1)^2}\right]}{(a_n^2 + b_n)}$$

$$v_{D*} = \frac{n}{n-1}\left(a_n - \frac{n}{n-1}\right) - v_{D*}$$

$$a_n = \sum_{k=1}^{n-1}\frac{1}{k}$$

$$b_n = \sum_{k=1}^{n-1}\frac{1}{k^2}$$

$$c_n = 2\frac{na_n - 2(n-1)}{(n-1)(n-2)}$$

$$d_n = c_n + \frac{n-2}{(n-1)^2} + \frac{2}{n-1}\left(\frac{3}{2} - \frac{2a_{n+1}-3}{n-2} - \frac{1}{n}\right)$$

Fay and Wu's *H* statistic (Fay and Wu, 2000) was used to test for hitchhiking effects (positive selection). This *H* statistic measures the differences between θ_H, mostly affected by high frequency variants, and π, mostly affected by intermediate frequency variants. The estimator θ_H is calculated from the frequency of derived segregating sites. It was calculated as following:

$$\theta_H = \sum_{i=1}^{n-1}\frac{2S_i i^2}{n(n-1)}$$

Where

S_i: the number of derived variants found *i* times in the sample of *n* chromosomes

Under neutrality the difference between θ_H and π is zero, but negative *H* value indicates the excess of high frequency variants, assuming the presence of hitchhiking effects (positive selection). The coalescent simulations were used to ascertain the statistical significance of this test, using the segregating sites, no recombination, and 10,000 replications. An outgroup was used to estimate the derived state of polymorphism data.

3. Population differentiation analysis

Differentiation among subpopulations was calculated by estimation of the average degrees of gene flow from DNA sequence data which it is based on *F* coefficient; F_{ST} (Hudson *et al.*, 1992). This estimator measures the differentiation between subpopulations relate to the entire population. The definition of gene flow is the movement of individuals (or gametes) from one population to another and subsequent breeding. Gene flow causes population to homogenize; on the other hand random genetic drift tends to divergence among populations (Halliburton, 2004). F_{ST} can be calculated with the following formula:

$$F_{ST} = 1 - (H_S/H_T)$$

where:

H_S: the average of expected heterozygosity within each subpopulation (the Hardy-Weinberg expected frequency)

H_T: the expected heterozygosity in the total population (the Hardy-Weinberg expected frequency)

Normally, this estimator exhibits range from 0 to 1 because the expected heterozygosity in the total population is always greater than the expected heterozygosity within each subpopulation.

The nearest-neighbor statistic (S_{nn}) statistic (Hudson, 2000) was also utilized for estimating differentiation among subpopulations. The permutation test with 10,000 random was utilized to estimate the *p*-value of S_{nn} statistic. It was calculated as following:

$$S_{nn} = (\sum_{k=1}^{n} X_k)/n$$

where:

S_{nn}: simply the average of the X_k

X_k: the fraction of nearest neighbors of individual *k* that are from the same locality as individual *k*

Part B: Phylogenetic inference with multilocus data

Mosquito samples

The *Cellia* subgenus of *Anopheles* mosquitoes analyzed for the phylogenetic relationships were focused on malaria vectors presented in Thailand. Taxonomic information of the mosquito species is shown in Table 7.

Adult female mosquito samples of *An. maculatus*, *An. epiroticus*, *An. minimus*, and *An. harrisoni* as well as genomic DNA of *An. dirus*, *An. baimaii*, *An. aconitus*, and *An. varuna* were kindly provided by the Mosquito Insectary from Department of Entomology, Kasetsart University. Moreover, mosquito samples of *An. sawadwongporni* were kindly provided by the Vector Borne Disease Control Center 10.4, Chiang Mai province, while mosquito samples of *An. scanloni* and *An. pampanai* could not be found and collected.

From literature review indicated that *An. sundaicus* is secondary malaria vector in Thailand but this information is discordant with geographical distribution of this species, while *An. epiroticus* is predominant species transmitting *Plasmodium* spp. in the coast of Thailand (Sumruayphol *et al.*, 2010) and also occurs along the

continental coast of Myanmar, Cambodia, Vietnam and peninsular Malaysia (Dusfour *et al.*, 2007; Manguin *et al.*, 2008). Thus, *An. sundaicus* was replaced by *An. epiroticus* for this study.

Table 7 Details of analyzed *Anopheles* species (the classifications are based on Harbach [2011]).

Species	Taxonomic states	
	Species group	Series
An. dirus	Dirus Complex, Leucosphyrus Subgroup, Leucosphyrus Group	Neomyzomyia
An. scanloni	Dirus Complex, Leucosphyrus Subgroup, Leucosphyrus Group	Neomyzomyia
An. baimaii	Dirus Complex, Leucosphyrus Subgroup, Leucosphyrus Group	Neomyzomyia
An. maculatus	Maculatus Subgroup, Maculatus Group	Neocellia
An. sawadwongporni	Sawadwongporni Subgroup, Maculatus Group	Neocellia
An. epiroticus	Sundaicus Complex	Pyretophorus
An. minimus	Minimus Complex, Minimus Subgroup, Funestus Group	Myzomyia
An. harrisoni	Minimus Complex, Minimus Subgroup, Funestus Group	Myzomyia
An. aconitus	Aconitus Subgroup, Funestus Group	Myzomyia
An. varuna	Aconitus Subgroup, Funestus Group	Myzomyia
An. pampanai	Aconitus Subgroup, Funestus Group	Myzomyia

Retrieval of DNA sequence from database and primer design

Six partial DNA fragments were inclusively chosen from different genome regions, i.e. mitochondrial and nuclear genomes, and different DNA regions i.e. coding and non-coding regions, and different chromosomal locations, i.e. autosome

and X-chromosome to investigate the phylogenetic relationships of Thai malaria vectors. The details of DNA fragments used are shown in Table 8.

Table 8 Information of DNA fragments used for phylogenetic reconstruction.

Gene	Chromosome location	Coding/non-coding part
TOLL6	Nuclear genome (Autosome)	Coding
rRNA	Nuclear genome (Autosome)	Non-coding
Gbb-60A	Nuclear genome (X-chromosome)	Non-coding
CYP	Nuclear genome (X-chromosome)	Non-coding
COI	Mitochondrial genome	Coding
COII	Mitochondrial genome	Coding

Note: *rRNA*: ribosomal RNA; *Gbb-60A*: Glass bottom boat 60A (Transforming growth factor beta at 60A); *CYP*: Cytochrome P450; *COI*: Cytochrome c-oxydase I; *COII*: Cytochrome c-oxydase II.

The published partial sequences of the mitochondrial *COI*, *COII* genes and ITS2 of *rRNA* gene of *Anopheles* mosquitoes widespread in Thailand were retrieved from NCBI database (www.ncbi.nlm.nih.gov), while sequence information of mitochondrial *COI* genes of *An. sawadwongporni* and *COII* gene of *An. epiroticus* is not available, the primers were designed to amplify the partial sequences of these genes in both species. For the *COI* gene, primers were designed based on DNA sequence of *An. maculatus* (accession no: GQ259192.1, EU256336.1, and DQ267690.1), because both *An. sawadwongporni* and *An. maculatus* are belong to the members of the Maculatus Group. For mitochondrial *COII* gene, primers were designed from full length of this gene. The forward and reverse primers were located on the flanking-regions of the *COII* gene in *Anopheles* mosquitoes which they are tRNA-Leu and tRNA-Lys genes, respectively. The complete mitochondrial DNA sequence information of *Anopheles* species were download from NCBI web database. The *Anopheles* species and accession numbers of the complete mitochondrial genomes are as follow; *An. albitarsis* F: HQ335349.1, *An. janconnae*: HQ335348.1, *An. deaneorum*: HQ335347.1, *An. albitarsis* G: HQ335346.1, *An. oryzalimnetes*: HQ335345.1, *An. albitarsis*: HQ335344.1, *An. quadrimaculatus*: L04272.1, and *An.*

gambiae: L20934.1. For ITS2 of *rRNA* gene, most of published partial ITS2 sequences of *Anopheles* mosquitoes were retrieved from the NCBI database, excluding *An. scanloni*. The accession numbers of the *Anopheles* species analyzed and outgroup species are presented in Table 9.

In this study, published DNA sequences of the partial exon and complete coding sequence (cds) of *TOLL6* gene from the whole genome sequence of *An. gambiae* was retrieved and downloaded from NCBI web database for design the primers. The accession numbers of *TOLL6* gene of *An. gambiae* are as follows: AM776293.1, AM776294.1, AF444781.1, and XM_320172.2.

Moreover, primers were developed from intron flanking-region of Gbb-60A by Stump *et al.* (2005) to amplify Gbb-60A fragment, and primers for CYP fragment amplification was designed by Dixit *et al.* (personal communication) using Exon-Primed Intron-Crossing method (Bierne *et al.*, 2000). Information of analyzed DNA fragments and primer sequences are shown in Table 10.

The partial sequences of outgroup species were retrieved from NCBI web database for comparison. The information of outgroup species are shown in Table 11.

Table 9 Sequence characteristics of mitochondrial *COI*, *COII* and rDNA-ITS2.

Species	*COI*		*COII*		ITS2	
	Length (bp)	Accession no.	Length (bp)	Accession no.	Length (bp)	Accession no.
An. dirus	206	AJ877406.1	555	AJ877470.1	841	U60410.1
An. scanloni	206	AM180859.1	555	AM180873.1	**	**
An. baimaii	206	AJ877510.1	555	HQ403902.1	835	U60411.1
An. maculatus	206	DQ267690.1	555	HQ403945.1	467	FJ526582.1
An. sawadwongporni	206	newly sequenced	555	HQ404164.1	418	FJ526593.1
An. epiroticus	204	AY789200.1	555	newly sequenced	663	AY789168.1
An. minimus	206	HQ877373.1	555	FN433593.1	515	HQ228204.1
An. harrisoni	206	HQ877374.1	555	AF195048.1	461	AF194506.1
An. aconitus	206	AY423055.1	555	HQ403810.1	506	DQ000247.1
An. varuna	206	newly sequenced	555	newly sequenced	495	DQ478879.1
An. pampanai	206	AY423054.1	555	AY486112.1	491	AY737082.1

**Sequence could not be obtained from database.

Table 10 Details of studied DNA fragments and primer sequences for PCR amplification in *Anopheles* species.

Gene (Primer name)	Gene details			Details of DNA fragments		
	Chromosomal location	Putative Function	Region	Primer sequences	Product size (bp)	Annealing temperature
CYP (P9)	X-Chromosome	Metabolism of xenobiotics	Non-coding	F: TGGAAAGTAGGGGCAATCAG R: GCTGGGCAACGGACTACTTA	654	51 ºC
Gbb-60A (O8)	X-Chromosome	Signaling proteins, controls proliferation, differentiation	Non-coding	F: CGGCCATTGGGCGGACATTT R: ACGCATCTCCGATCCGTGATGATC	550	60 ºC
TOLL 6 (TOLL 6)	Autosome	Immune signaling pathways	Coding	F: CCAACTCGTTCGTCCATCTG R: CTGGCGAAAGCACACTGAT	755	62 °C
COI (COI)	Mitochondrial genome	Energy-transfer enzymes of respiratory chains	Coding	F: CTGGCTGGGCAACGGACTA R: GGAAAGTAGGGGCAATCAGTTTG	439	50 ºC
COII (COII)	Mitochondrial genome	Component of the respiratory chain and transfer of electrons from cytochrome c to oxygen	Coding	F: TGTGAACTGCAGGACACAT R: TATGCTTAAATTCAGGGGGT	773	50 ºC

Table 11 The information of outgroup species.

Fragments	Outgroup species	Length (bp)	Accession no.
P9	*Drosophila melanogaster*	506	NC_004354
TOLL6	*Drosophila melanogaster*	659	NM_079357.2
COI	*Aedes aegypti*	213	EU352212.1
COII	*Aedes aegypti*	685	EU352212.1
ITS2	*Drosophila melanogaster*	509	EU306667.1

Note: O8 fragment was excluded from analysis due to low genetic variation.

Genomic DNA extraction and molecular identification

Genomic DNA of mosquito samples was extracted by using the modification of Genomic DNA Mini Kit (Tissue) (Geneaid, Taiwan), describing in Appendix A. Each DNA sample was separated on 1.0% agarose gel (Vivantis, California) in 0.5X TAE buffer (Vivantis, California) for quality and concentration examination, and 1 Kb ladder (Vivantis, California) was use as standard marker. Agarose gel was strained with ethidium bromide and the DNA bands were visualized under ultraviolet light.

Multiplex allele-specific polymerase chain reaction (AS-PCR) assay was used to identify and confirm species of the Minimus Complex (*An. minimus* and *An. harrisoni*) and the closely related species (*An. aconitus*, and *An.varuna*), the Maculatus Group (*An. maculatus* and *An. sawadwongporni*), and the Sundaicus Complex (*An. epiroticus*), as following Garros *et al.* (2004), Walton *et al.* (2007) and Dusfour *et al.* (2007), respectively. The AS-PCR primers are shown in Table 12. The PCR reaction volume and thermocycler profiles are provided in Appendix B. PCR products were separated on 1.5% agarose gel in 0.5X TAE buffer (Vivantis, California) and 100 bp ladder (SibEnzyme, Russia) was use as standard marker. Agarose gel electrophoresis was performed for 25 minutes at 100 volts. Gel was strained with ethidium bromide and the PCR product bands were visualized under ultraviolet light.

Table 12 Multiplex AS-PCR primers for molecular identification.

Species (Primer name)	Sequence (5' to 3')	Product size (bp)
Minimus Complex and the closely related species (Garros *et al.*, 2004)		
Universal forward (ITS2A)	TGT GAA CTG CAG GAC ACA T	
An. minimus (MIA)	CCC GTG CGA CTT GAC GA	310
An. harrisoni (MIC)	GTT CAT TCA GCA ACA TCA GT	180
An. aconitus (ACO)	ACA GCG TGT ACG TCC AGT	200
An. varuna (VAR)	TTG ACC ACT TTC GAC GCA	260
An. pampanai (PAM)	TGT ACA TCG GCC GGG GTA	90
Maculatus Group (Walton *et al.*, 2007)		
Universal forward (5.8F)	ATC ACT CGG CTCGTG GAT CG	
An. maculatus (MAC)	GAC GGT CAG TCT GGT AAA GT	180
An.sawadwongporni (SAW)	ACG GTC CCG CAT CAG GTG C	242
Sundaicus Complex (Dusfour *et al.*, 2007)		
Universal forward (CBsunA)	AAT GTT ACA AGA ATT CA	
Universal forward (CBsunB)	TTA GCT ATA CAT TAT GC	575
An. sundaicus s.s. (SunSS)	TAT CAT TCT GAG GAG CC	313
An.sundaicus E (SunE)	ATG ATT TTT ACG AAT TTG C	498
Universal forward (SpCO)	GAA CGG TTT ATC CTC CT	
An.epiroticus (Epi)	TAT TCG ATC TAA AGT AAT C	167

PCR amplification, DNA sequencing, and alignment

PCR amplification was performed in 30 µl volume. The reaction mixture and the reaction conditions for PCR amplification are described in Appendix C. The primer sequences and their annealing temperature are given in Table 10. Three microliters of PCR products were separated on 1% agarose gel in 0.5X TAE buffer

(Vivantis, California) and 100 bp ladder (SibEnzyme, Russia) was use as standard marker. Agarose gel electrophoresis was performed for 25 minutes at 100 volts. Gel was strained with ethidium bromide and the PCR product bands were visualized under ultraviolet light. PCR products were purified; the excess primers were digested with Exonuclease I and the dNTPs with FastAP™ Thermosensitive Alkaline Phosphatase (Fermentas, Ontario). The purified PCR products were directly sequenced in both directions using forward- and reverse-specific primers by Macrogen Inc. (Korea). The nucleotide sequences were manually checked for accuracy between base-calling and chromatograms before editing. Homozygotes and heterozygotes were visually analyzed from the chromatograms. One fluorescent peak pattern was defined as homozygous sequences, by the way, two peaks at a single position in the chromatograms was determined as heterozygous sequences. Nucleotide sequences were aligned and manually edited using BioEdit version 7.0.5.3 computer programs (Hall, 1999). The retrieved sequences of each species were aligned and trimmed using BioEdit version 7.0.5.3 computer programs (Hall, 1999).

Phylogenetic relationship analysis

Nucleotide frequencies were obtained by MEGA version 5 computer program (Tamura *et al.*, 2011). Measure for substitution saturation of nucleotides was implemented in DAMBE version 5.2.57computer program (Xia and Xie, 2001), tested by Xia's (Xia *et al.*, 2003) and Steel's method (Steel *et al.*, 1993). Transition and transversion substitutions per site were plotted against the evolutionary distance based on Kimura's two parameter substitution model by using DAMBE version 5.2.57computer program (Xia and Xie, 2001). When both the transitions and transversions increased with increasing the evolutionary distance indicated that no saturation of substitution is presented.

For each DNA fragments, four different approaches of phylogenetic tree construction, i.e. neighbor joining (NJ), minimum evolution (ME), maximum parsimony (MP), and maximum likelihood (ML) were performed using MEGA version 5 computer program (Tamura *et al.*, 2011). The best appropriate nucleotide substitution model for maximum likelihood analysis was performed by using MEGA

version 5 computer program (Tamura *et al.*, 2011); the lowest Bayesian Information Criterion (BIC) (Schwarz, 1978) value of model tests was considered as the best substitution model. Four independent phylogenetic trees per approaches were constructed supporting with bootstrap methods (10,000 replications). Each method was performed using the heuristic search. Gap positions in aligned sequence were considered as missing data.

The mitochondrial coding regions, COI and COII sequences were translated into protein for avoidance of pseudogenes, using MEGA version 5 computer program (Tamura *et al.*, 2011) with the invertebrate mitochondrial genetic code.

Estimation of divergence time, the relative rate test (Tajima, 1993) based on the molecular evolutionary clock hypothesis (Morgan, 1998) which evolves at a constant rate of gene was firstly implemented in MEGA version 5 computer program (Tamura *et al.*, 2011). The divergence time between mosquito lineages were then estimated using the Bayesian MCMC analyses computer programs. This is mcmctree program implementing in PAML version 4.4 computer program (Yang, 2007).

RESULTS AND DISCUSSION

Part A: Genetic diversity and tests of population genetic models

Mosquito collections and morphological identification

From literature review, the distributions of *An. harrisoni* in Thailand were observed in three provinces, i.e. Chiang Mai, Tak, and Kanchanaburi. In this study, mosquito sampling sites were planned to cover the whole distribution. These localities for mosquito collections are shown in Table 13.

Table 13 List of locality for mosquito collection.

Number	Locality
1	Pong Lom Rang Village, Mae Wang district, Chiang Mai province
2	Tham Suea Village, Mae Sod district, Tak province
3	Pu Teuy Village, Sai York district, Kanchanaburi province
4	Bong Ti Noi Village, Sai Yok district, Kanchanaburi province
5	Sangkhla Buri district, Kanchaburi province
6	Thong Pha Phum district, Kanchanaburi province
7	Si Sawat district, Kanchanaburi province

For mosquito collections, samples from Kanchanaburi province (Pu Teuy and Bong Ti Noi Village) were collected, morphologically identified and provided by Mosquito Insectary of Department of Entomology, Faculty of Agriculture, Kasetsart University. Mosquito samples from Tak province could not be found and collected.

Mosquito samples from Chiang Mai province and other localities from Kanchanaburi province were collected using human-landing and cattle-bait collection methods. Adult mosquitoes were morphologically identified into the Minimus Complex using the presence or absence of humeral pale (HP) spot on the costa veins of the mosquito wings. The HP spot is usually absent on the wings of *An. minimus*,

while it is usually present on the wings of *An. harrisoni* (Rattanarithikul, 2006). The information of morphological identification of mosquitoes from each locality are shown in Table 14.

Extraction of genomic DNA

Genomic DNA of mosquitoes was isolated using the modification of Genomic DNA Mini Kit (Tissue) (Geneaid, Taiwan). The DNA bands were detected by straining the agarose gel with ethidium bromide and visualizing under ultraviolet light, as show in Figure 15.

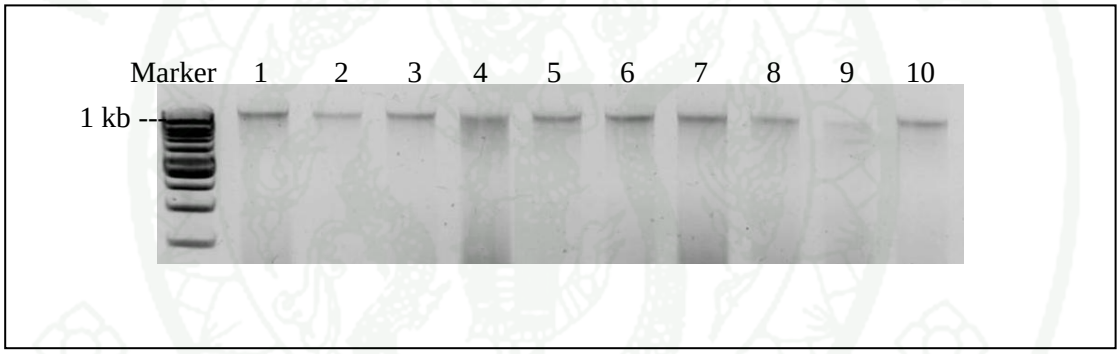


Figure 15 Genomic DNA of *An. harrisoni* from Pu Teuy Village, Sai York district, Kanchanaburi province (No. 1 to 10), comparing with 1 kb ladder.

Table 14 Information of morphological identification of mosquitoes from each locality.

Locality	**Morphological identification**		
	An. minimus	***An. harrisoni***	**Others**
(1) Tham Suea Village, Mae Sod district, Tak province **	-	-	-
(2) Pu Teuy Village, Sai York district, Kanchanaburi province	-	10 **	-
(3) Bong Ti Noi Village, Sai Yok district, Kanchanaburi province	4	51 **	1 *An. maculatus* s.l.
(4) Sangkhla Buri district, Kanchaburi province	60	5	-
(5) Thong Pha Phum district, Kanchanaburi province	80	22	-
(6) Si Sawat district, Kanchanaburi province	10	9	50 *An. dirus* s.l.
(7) Pong Lom Rang Village, Mae Wang district, Chiang Mai province	2	8	89 *An. splendidus**
			20 *An. pursati**
			10 *An. pseudojamesi**
			5 *An. kawari**
			3 *An. insulaeflorum**

** Samples were collected, identified and provided by Department of Entomology, Faculty of Agriculture, Kasetsart University.

* Samples were morphologically identified by Department of Entomology, Faculty of Agriculture, Kasetsart University.

Molecular identification

Anopheline mosquitoes belonging to the Minimus Complex were correctly identified species using diagnostic PCR assay as following Garros *et al.* (2004). The AS-PCR assay was carried out, as show in Figure 16. Lengths of amplified species-specific fragments of *An. minimus* and *An. harrisoni* were 310 bp and 180 bp, respectively.

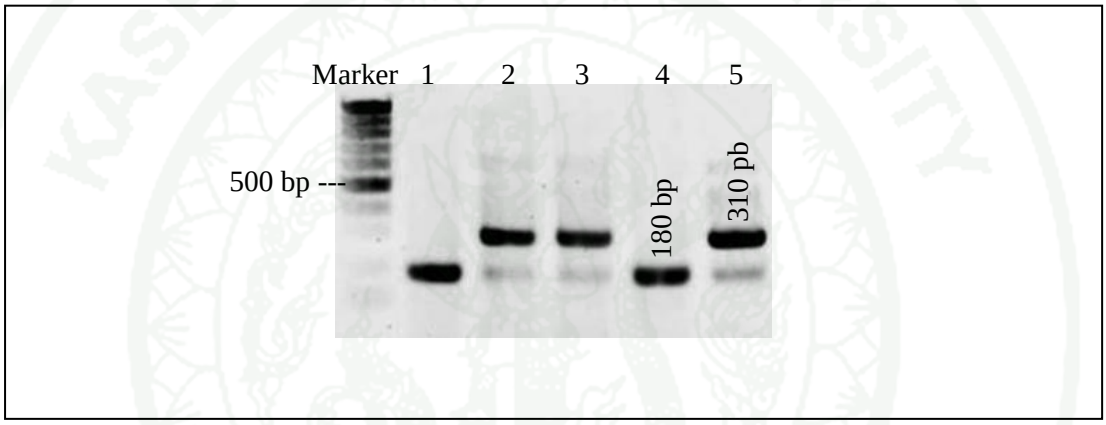


Figure 16 Amplified fragments using diagnostic AS-PCR assay of *Anopheles* mosquitoes from Bong Ti Noi Village, Sai Yok district, Kanchanaburi province (No. 1: *An. harrisoni*; No. 2 and 3: *An. minimus*; No. 4: positive control of *An. harrisoni*; No. 5: positive control of *An. minimus*), comparing with 100 bp ladder.

Molecular identification of *Anopheles* mosquitoes from each locality was performed. Samples of *An. harrisoni* were precisely distinguished from *An. minimus* based on AS-PCR technique. The results showed that most of morphological identifications of *An. harrisoni* were misidentification. Thus, molecular identification is required for precise classification of mosquito species complex. The results of comparison between morphological and molecular identification are shown in Table 15.

Table 15 Comparison between morphological and molecular identification of mosquito samples, and percentage error.

Locality	Species		% Error
	Morphological identification	**Molecular identification**	
(1) Tham Suea Village, Mae Sod district, Tak province	-	-	-
(2) Pu Teuy Village, Sai York district, Kanchanaburi province	10 *An. harrisoni*	10 *An. harrisoni*	-
(3) Bong Ti Noi Village, Sai Yok district, Kanchanaburi province	51 *An. harrisoni*	33 *An. minimus* 16 *An. varuna* 2 *An. harrisoni*	96.08
(4) Sangkhla Buri district, Kanchaburi province	5 *An. harrisoni*	5 *An. minimus*	100
(5) Thong Pha Phum district, Kanchanaburi province	22 *An. harrisoni*	22 *An. minimus*	100
(6) Si Sawat district, Kanchanaburi province	9 *An. harrisoni*	9 *An. minimus*	100
(7) Pong Lom Rang Village, Mae Wang district, Chiang Mai province	8 *An. harrisoni*	8 *An. minimus*	100

In this study, two populations of *An. harrisoni* from Kanchanaburi province were used for analysis of molecular population genetics. These populations are Pu Teuy Village, Sai York district, Kanchanaburi province and Bong Ti Noi Village, Sai Yok district, Kanchanaburi province as well as numbers of samples for analysis were 2 and 10 samples, respectively. Moreover, two samples of *An. harrisoni* from Khoi Village, Hoa Binh province, Vietnam providing by Dr. Sylvie Manguin from University Montpellier I, Faculté de Pharmacie, Montpellier, France were used to compare with mosquito populations of Thailand.

PCR amplification and sequencing

PCR amplification of interested fragments was carried out. Three populations of *An. harrisoni* were amplified using six primer pairs, as show in Table 5. The capability of PCR amplification from each locus is shown in Table 16. The results of PCR amplification are shown in Figure 17 and Figure 18.

Table 16 Capability of PCR amplification and sequencing from each locus.

Pop.	**PCR amplification**						**Sequencing**					
	P9	**P12**	**P22**	**O8**	**O10**	**ITS2**	**P9**	**P12**	**P22**	**O8**	**O10**	**ITS2**
KPT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
KBT	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓
VHB	✓	✓	×	✓	✓	×	✓	✓	×	✓	✓	×

Note: ✓ Result could be obtained; × Result could not be obtained, Pop. = Populations.

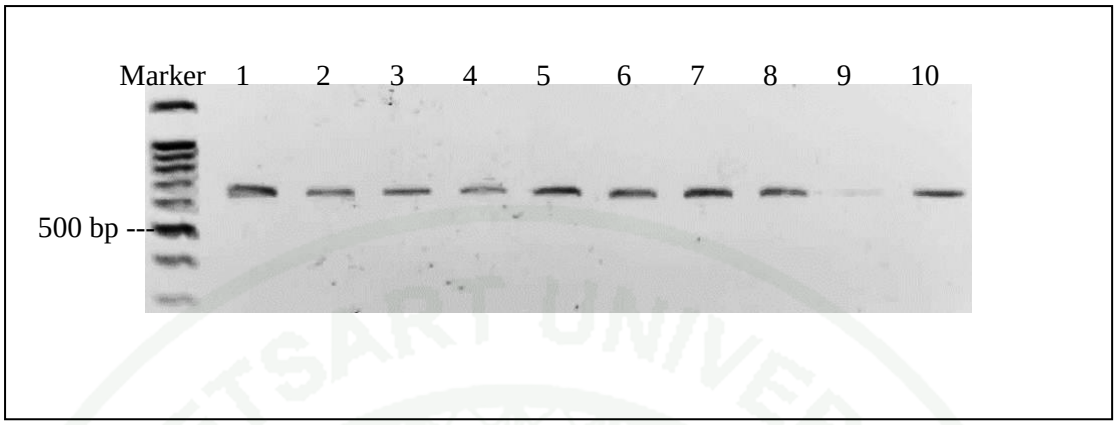


Figure 17 PCR amplification of P9 fragment of *An. harrisoni* from Pu Teuy Village, Sai Yok district, Kanchanaburi province, Thailand (No. 1 to 10), comparing with 100 bp ladder.

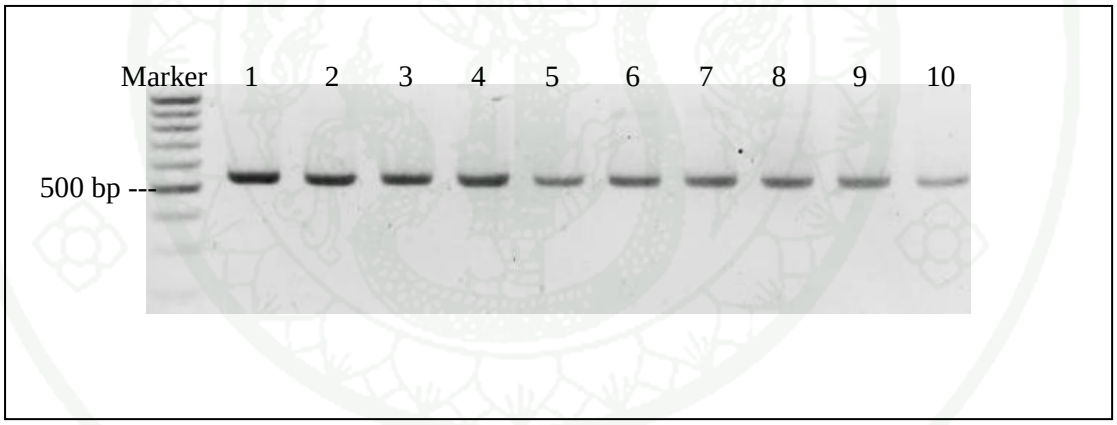


Figure 18 PCR amplification of O8 fragment of *An. harrisoni* from Pu Teuy Village, Sai Yok district, Kanchanaburi province, Thailand (No. 1 to 10), comparing with 100 bp ladder.

PCR purification and DNA sequencing were carried out. Obvious single peaks (one color at each position, and minimum noise or baseline peak) are required. Nucleotide sequences were manually checked for avoidance of mis-called nucleotides. Double or heterozygous peaks (polymorphic positions) on some nucleotide positions are common for sequences of diploid genomic DNA (The University of Michigan, n.d.). The capability of DNA sequencing from each locus is

also shown in Table 16. The chromatogram of nucleotide sequences is shown in Figure 19.

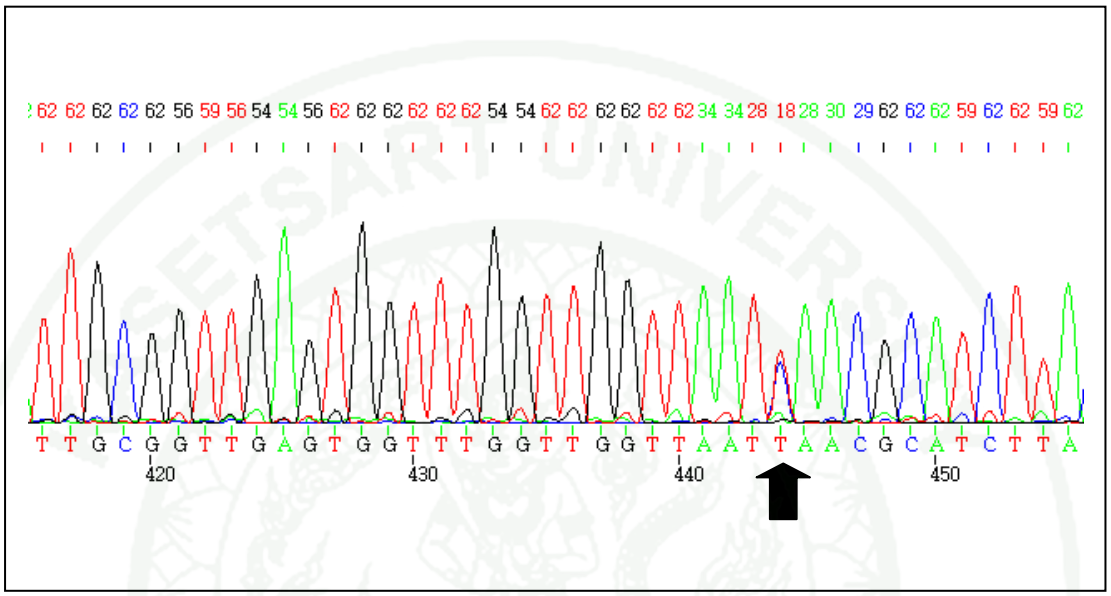


Figure 19 Chromatograms of nucleotide sequences of O8 primer was sequenced from sample No. 8 of Pu Teuy Village, Sai Yok district, Kanchanaburi province, Thailand as well as heterozygous peaks are also detected (black arrow).

Comparing nucleotide sequences using nucleotide BLAST program from the NCBI database, nucleotide sequence alignments, editing nucleotide sequences and preparing the Nexus files were carried out (BLAST results were shown in Appendix E). Nexus file format is common for further analysis by DnaSP computer programs. Sequence alignments and Nexus file are shown in Figure 20 and 21, respectively.

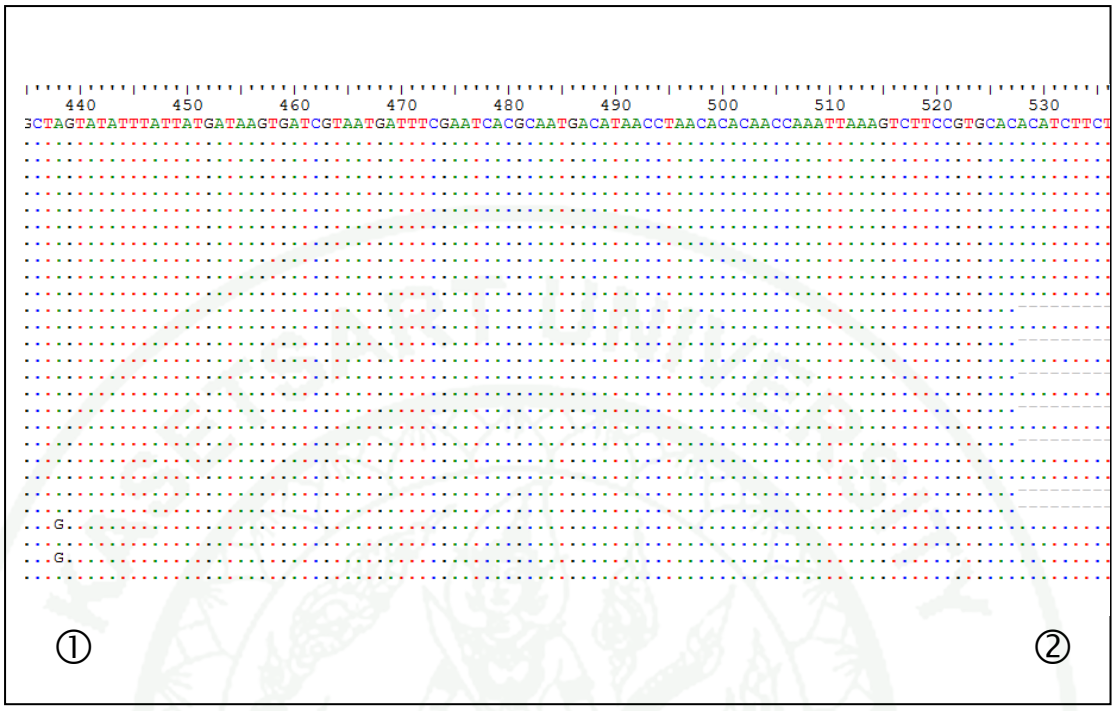
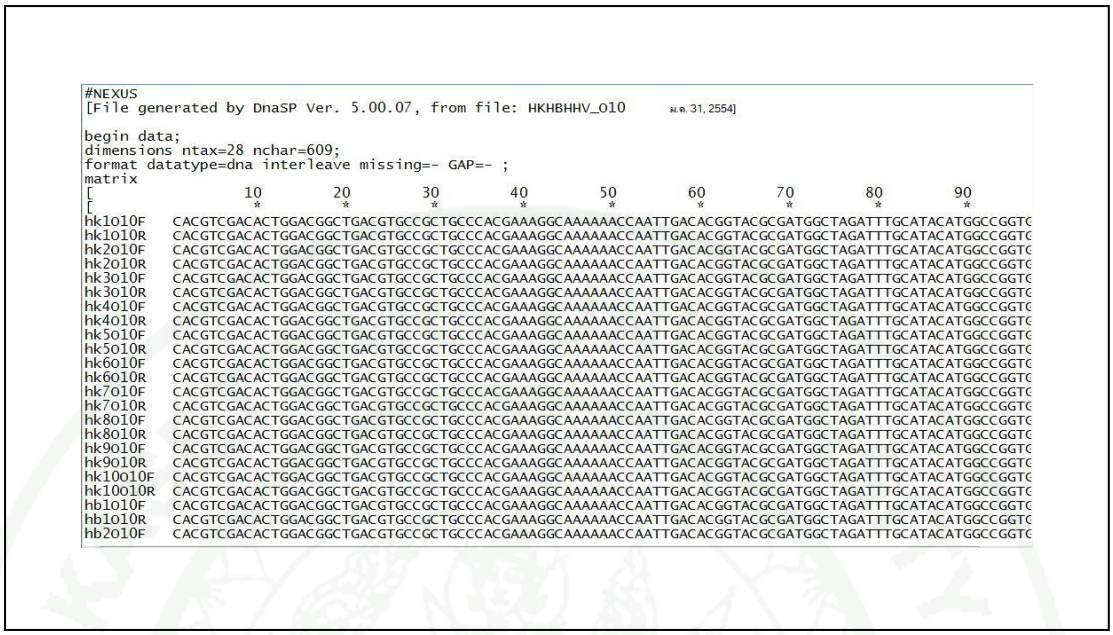


Figure 20 Nucleotide sequence alignment of O10 primer from all populations of *An. harrisoni*, position of ① is polymorphic position (Guanine/Adenine) and position of ② is indel position (dots in the sequence alignment refer to identity between a standard sequence, usually the top one, and nucleotides down a column).

Data analysis

Analyses of nucleotide sequences using DnaSP version 5.00.07 computer programs (Librado and Rozas, 2009) were performed. The intraspecific nucleotide diversity of *An. harrisoni* from KPT, KBT and VHB populations was determined from non-coding regions of Cytochrome P450, Inositol polyphosphate 4-phosphatase, Glass bottom boat 60A (Transforming growth factor beta at 60A) and Ribosomal internal transcribed spacer 2 of rRNA genes.



```
#NEXUS
[File generated by DnaSP Ver. 5.00.07, from file: HKHBHHV_O10      พ.ค. 31, 2554]

begin data;
dimensions ntax=28 nchar=609;
format datatype=dna interleave missing=- GAP=- ;
matrix
[                 10         20         30         40         50         60         70         80         90
[                  *          *          *          *          *          *          *          *          *
hk1o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk1o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk2o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk2o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk3o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk3o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk4o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk4o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk5o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk5o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk6o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk6o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk7o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk7o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk8o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk8o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk9o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk9o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk10o10F   CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hk10o10R   CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hb1o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hb1o10R    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
hb2o10F    CACGTCGACACTGGACGGCTGACGTGCCGCTGCCCACGAAAGGCAAAAAACCAATTGACACGGTACGCGATGGCTAGATTTGCATACATGGCCGGTC
```

Figure 21 Nexus file format of O10 primer from all populations of *An. harrisoni.*

Sequence Polymorphism

A total of 72,016 nucleotides across 6 non-coding fragments were analyzed. The total length of the studied loci ranges from 245 bp to 609 bp. A total of 34 polymorphic sites along 6 fragment regions were identified. The polymorphic sites including 31 single nucleotide substitutions (SNPs) and three insertion/deletion polymorphisms (indels) were observed.

Investigation of nucleotide diversity was performed. Watterson estimator (θ_w) and π were calculated as implemented in DnaSP version 5.00.07 computer programs (Librado and Rozas, 2009). Summary of nucleotide diversity from all populations and studied fragments are shown in Table 17.

Table 17 Summary of nucleotide diversity.

Population	Loci	Length of analyzed sequence (base pairs)	Nucleotide diversity	
			θ_w [a]	π [a]
KPT	P9	595	0.145	0.175
	P12	245	0.349	0.124
	P22	385	0.439	0.364
	O8	458	0.123	0.155
	O10	609	0.047	0.088
	ITS2	455	0	0
	Average		0.184	0.151
KBT	P9	595	0.187	0.228
	P12	245	0.676	0.620
	P22	**	-	-
	O8	458	0	0
	O10	609	0.182	0.195
	ITS2	455	0	0
	Average		0.209	0.209
VHB	P9	595	0.275	0.280
	P12	245	0.445	0.408
	P22	*	-	-
	O8	458	0.119	0.109
	O10	609	0.269	0.274
	ITS2	*	-	-
	Average		0.277	0.268
All	P9	595	0.264	0.232
	P12	245	0.743	0.439
	P22	385	0.439	0.364
	O8	458	0.112	0.180
	O10	609	0.215	0.183
	ITS2	455	0	0
	Average		0.296	0.233

Note: [a] Calculated in percentage scale; * No amplification; ** No sequencing; KPT: Pu Teuy Village; KBT: Bong Ti Noi Village; θ_w: Watterson estimator; π: Nei estimatior.

All studied fragments showed relatively low level of nucleotide polymorphism. The highest nucleotide variation was found at locus P12, while no variation was detected at locus ITS2. This result of locus ITS2 is consistent with a characteristic of the transcribed and non-transcribed spacer regions of rDNA sequence, holding high interspecific and low intraspecific variation (Mohanty *et al.*, 2009).

The nucleotide diversity of *An. harrisoni* populations was compared with other species. For KPT population, the nucleotide diversity at locus P9 in *An. harrisoni* (π = 0.00175) was lower than in *An. arabiensis* (π = 0.0023), and higher than in *An. gambiae* M and S forms (π = 0.0002 and 0.0002, respectively) (Stump *et al.*, 2005). Level of nucleotide variation at locus P12 in this study (π = 0.00124) was lower than in *An. gambiae* both M and S forms (π = 0.0207 and 0.0181, respectively) (Stump *et al.*, 2005). The nucleotide diversity at locus P22 of *An. harrisoni* (π = 0.00364) was lower than in *An. arabiensis* (π = 0.0081), while the polymorphism in the study was higher than in *An. gambiae* both M and S forms (π = 0.0002 and 0.0021, respectively) (Stump *et al.*, 2005). Moreover, the nucleotide diversity at locus O10 of *An. harrisoni* (π = 0.00088) was also smaller when compared with *An. minimus* from Jalpaiguri population, India (π = 0.00181) (Dixit *et al.*, 2011).

For *An. harrisoni* from KBT population, level of nucleotide variation at locus P9 in this study (π = 0.00228) was higher than in *An. gambiae* both M and S forms (π = 0.0002 and 0.0002, respectively), but lower than in *An. arabiensis* (π = 0.0023) (Stump *et al.*, 2005). The nucleotide diversity at locus P12 in *An. harrisoni* (π = 0.00228) was higher than in *An. gambiae* M and S forms (π = 0.0207 and 0.0181, respectively) (Stump *et al.*, 2005). Because nucleotide sequences of *An. harrisoni* at locus P22 could not be obtained, the nucleotide diversity could not be evaluated and compared with other species. In addition, the nucleotide variation of *An. harrisoni* at locus O8 showed no variation, while at locus O10 was higher (π = 0.00195) than in *An. minimus* from Jalpaiguri population, India (π = 0.00181) (Dixit *et al.*, 2011). For *An. harrisoni* from VHB population, the nucleotide diversity of *An. harrisoni* at locus P9 (π = 0.00280) was higher than in *An. gambiae* both M and S forms (π = 0.0002

and 0.0002, respectively) as well as *An. arabiensis* (π = 0.0023) (Stump *et al.*, 2005). At locus P12, the nucleotide diversity of *An. harrisoni* (π = 0.00408) was lower than in *An. gambiae* both M and S forms (π = 0.0207 and 0.0181, respectively). Furthermore, the nucleotide variation at locus O10 of *An. harrisoni* (π = 0.000274) was larger than of *An. minimus* (π = 0.00181) from Jalpaiguri population, India (Dixit *et al.*, 2011). While the nucleotide diversity at locus P22 and ITS2 could not be examined because PCR amplification could not be obtained.

The fragments P9 and P22 locate near the centromere on the X chromosome (cytogenetic map division 6, Figure 22), which they could imply low recombination in this region. Nucleotide diversity at locus P12 was small, in spite of this locus locates distal the centromere on the X chromosome (cytogenetic map division 4B, Figure 22). Moreover, low genetic diversity was also detected at fragment O8 and O10 which they locate distal and near the centromere on the X chromosome, respectively (cytogenetic map division 1C, Figure 22 and cytogenetic map division 6, Figure 22, respectively). This result suggested that the low genetic diversity in *An. harrisoni* might due to low recombination event and perhaps small population size since they have been found in restricted area in Thailand. Moreover, a nature barrier, the Khwae Noi River, might have effects on the genetic structure of KPT and KBT populations. This barrier may cause gene flow restriction between these populations.

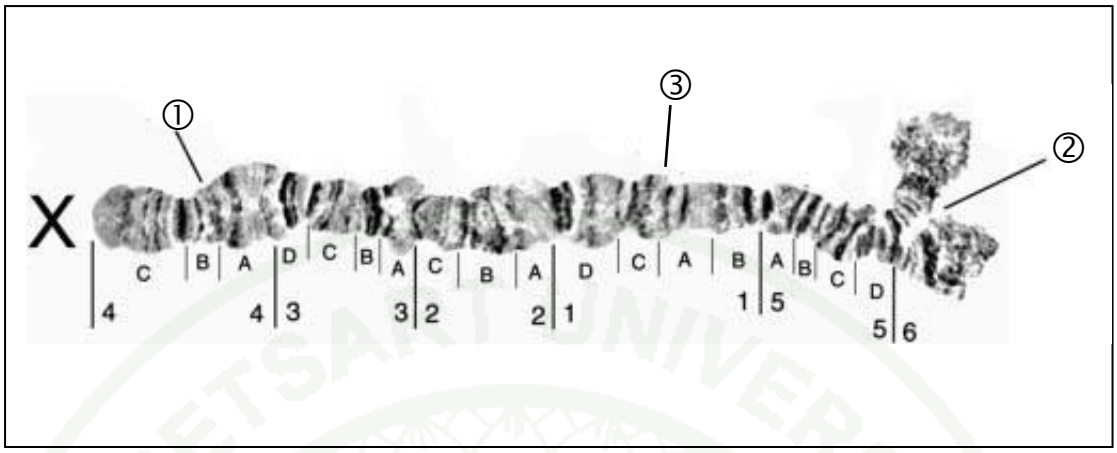


Figure 22 Location of loci on cytogenetic map of the *An. gambiae* X chromosome; ① related to fragment P12; ② related to fragment P9, P22 and O10; ③ related to fragment O8 (the centromere oriented on the right).

Source: Modified from Stump *et al.* (2005)

The mean nucleotide polymorphism at each locus from whole population was compared with other species. The nucleotide diversity at locus P9 in *An. harrisoni* (π = 0.00232) was higher than in *An. arabiensis* (π = 0.0023), and also in *An. gambiae* M and S forms (π = 0.0002 and 0.0002, respectively) (Stump *et al.*, 2005). Level of nucleotide variation at locus P12 in *An. harrisoni* (π = 0.00439) was lower than in *An. gambiae* M and S forms (π = 0.0207 and 0.0181, respectively) (Stump *et al.*, 2005). The nucleotide diversity at locus P22 of *An. harrisoni* (π = 0.00364) was lower than in *An. arabiensis* (π = 0.0081), while the polymorphism in the study was higher than in *An. gambiae* both M and S forms (π = 0.0002 and 0.0021, respectively) (Stump *et al.*, 2005). Besides, the level of nucleotide variation at locus O10 (π = 0.000183) of *An. harrisoni* was higher than in fragments O10 (π = 0.00181) of *An. minimus* from Jalpaiguri population, India, respectively (Dixit *et al.*, 2011). While the nucleotide diversity at locus ITS2 showed no variation and locus O8 showed low variation (π = 0.000180).

Other data analyses of *An. harrisoni* for each fragment in whole populations are also shown in Table 18.

Table 18 Data analysis of *An. harrisoni* for each fragment in all population.

Population	Loci	No. of samples	No. of SNPs	No. of haplotypes	Haplotype diversity (Hd)	Divergence (K)
KPT	P9	10	3	4	0.679	1.021
	P12	10	3	3	0.195	0.3
	P22	10	6	7	0.689	1.4
	O8	10	2	4	0.595	0.711
	O10	10	1	2	0.526	0.526
KBT	P9	2	2	4	1	1.333
	P12	2	3	3	0.833	1.5
	O8	2	0	1	0	0
	O10	2	2	3	0.833	1.167
VHB	P9	2	3	4	1	1.667
	P12	2	2	2	0.5	1
	O8	2	1	1	0.5	0.5
	O10	2	3	2	1	1.667

Note: Data could not be obtained from locus ITS2 of KPT, loci P22 and ITS2 of KBT, and loci P22 and ITS2 of VHB.

Test of neutrality

Tajima's *D* statistic was utilized to test the deviations from the neutral evolution. Fu and Li's *D* and *D** and Fay and Wu's *H* statistics were also used to test the selective neutrality. These tests were performed for each gene in all populations using DnaSP version 5.00.07 computer programs (Librado and Rozas, 2009). For an outgroup, the homologous sequences of *An. gambiae* were retrieved from the NCBI database, the accession no. of fragments from sex chromosome (P9, P12, P22, O8, and O10) and autosome (ITS2) are AAAB01008963.1 and EU104640.1, respectively. Results of the neutrality tests are shown in Table 19.

Table 19 Summary of neutrality tests for each gene in all populations.

Population	Analyzed fragment	D_T	D_{FL}	D^*_{FL}	H
KPT	P9	0.554 [ns, a]	1.012 [ns, a]	-0.124 [ns, a]	-2.589 [ns, c]
	P12	-1.723 [ns, a]	-2.177 [ns, b]	-2.385 [ns, b]	0.189 [ns, c]
	P22	-0.548 [ns, a]	0.128 [ns, a]	0.547 [ns, a]	-0.357 [ns, c]
	O8	0.610 [ns, a]	0.854 [ns, a]	0.866 [ns, a]	0.473 [ns, c]
	O10	1.565 [ns, a]	0.626 [ns, a]	0.649 [ns, a]	0.000 [ns, c]
KBT	P9	1.893 [ns, a]	1.440 [ns, a]	1.893 [ns, b]	0.000 [ns, c]
	P12	-0.754 [ns, a]	-1.200 [ns, a]	-0.754 [ns, a]	0.666 [ns, c]
	O10	0.591 [ns, a]	1.440 [ns, a]	0.591 [ns, a]	-1.000 [ns, c]
VHB	P9	-0.065 [ns, a]	-0.912 [ns, a]	0.167 [ns, a]	0.333 [ns, c]
	P12	-0.709 [ns, a]	-1.200 [ns, a]	-0.709 [ns, a]	0.666 [ns, c]
	O8	-0.612 [ns, a]	1.095 [ns, a]	-0.612 [ns, a]	-1.000 [ns, c]
	O10	0.167 [ns, a]	-0.368 [ns, a]	0.167 [ns, a]	0.666 [ns, c]

Note: D_T: Tajima's D statistic; D_{FL}: Fu and Li's D statistic; D_{FL}*: Fu and Li's D* statistic; H: Fay and Wu's H statistic; [a] p-value > 0.10; [b] 0.10 > p-value > 0.05; [c] p-value > 0.05; ns, not significant; **, significant; No data at locus ITS2 of KPT, loci P22, O8 and ITS2 of KBT, loci P22 and ITS2 of VHB.

In this study, fragments O8, O10 and P9 from KPT population, fragments P9 and O10 from KBT population, and fragment O10 from VHB population showed positive *D* values, while fragment P12 and P22 from KPT population, fragment P12 from KBT, and fragment P9, P12 and O8 from VHB showed a negative *D* value. However, none of *D* value from each locus showed statistically significant Tajima's *D* value (*p*- value > 0.01), suggesting these studied fragments are under a standard neutral equilibrium. These results were consistent with the neutrality tests of *An. gambiae* M and S forms (Stump *et al.*, 2005) as well as *An. minimus* (Dixit *et al.*, 2011). Tajima's *D* values for *An. gambiae* M and S forms at locus P9 were -1.1644 and -1.1639, at locus P12 were -1.1198 and -1.2741, and at locus P22 were -1.1610 and -1.2234, respectively, assuming these loci are under neutrality (*p*-value > 0.05) (Stump *et al.*, 2005). Negative *D* value for *An. minimus* from Jalpaiguri population, India at locus O10 was -1.78132, assuming this locus is under neutrality (*p*-value < 0.05) (Dixit *et al.*, 2011).

For other tests of neutrality, Fu and Li' *D* and *D** statistic tests of all populations and loci do not show significant departures from neutral equilibrium expectations, as well as the *H* test of Fay and Wu do not show significant departures from neutral evolutions, suggesting no signal of selective sweep in all populations.

Population differentiation

Estimations of the standard F_{st} statistic based on Hudson (1992) and Hudson's S_{nn} statistic (2000) across 6 loci in 2 (KPT and KBT, excluding VHB) and 3 populations were performed. Probability of S_{nn} statistic was obtained by the permutation test with 10,000 replicates. Results of the standard F_{st} statistic and S_{nn} statistic are shown in Table 20 and 21.

Table 20 Population differentiation across 5 loci (2 populations, excluding VHB) of *An. harrisoni.*

Locus	F_{st} [a]	S_{nn} [b]	***p*-value of S_{nn}**
P9	0	0.705	0.572^{ns}
P12	0	0.675	1.000^{ns}
O8	0.210	0.733	0.263^{ns}
O10	0	0.727	0.239^{ns}
ITS2	0	0.704	1.000^{ns}
Average	0.042	0.709	

Note: negative F_{st} were set to zero.

[a] F_{st} statistic based on Hudson (1992).

[b] S_{nn} statistic based on Hudson *et al.* (2000).

ns, not significant; No data for locus P22.

Genetic differentiation estimates were performed only 2 populations from Thailand (excluding VHB population), the result showed that no significant genetic differentiations at all loci were observed. The average F_{st} estimate was 0.04210, indicating low genetic variation among these populations.

Table 21 Population differentiation across 5 loci (3 populations) of *An. harrisoni*.

Locus	F_{st} [a]	S_{nn} [b]	*p*-value of S_{nn}
P9	0.248	0.629	0.016*
P12	0.583	0.721	0.000***
O8	0.661	0.628	0.027*
O10	0.389	0.697	0.001*
ITS2	0.000	0.704	1.000^{ns}
Average	0.376	0.676	

Note: [a] F_{st} statistic based on Hudson (1992).

[b] S_{nn} statistic based on Hudson *et al.* (2000).

ns, not significant; *, 0.01< P <0.05; ***, P <0.001.

For comparison with other *Anopheles* species, the F_{st} value of *An. minimus* s.l. studying in Kanchaburi province, Thailand (F_{st} = 0.061) (Poolprasert *et al.*, 2008) was higher than in *An. harrisoni* (F_{st} = 0.04210). The F_{st} value of *An. minimus* in Vietnam (F_{st} = 0.075) (Van Bortel *et al.*, 2003) was higher than in *An. harrisoni* in Thailand (F_{st} = 0.04210) and lower than in *An. harrisoni* in Thailand and Vietnam (F_{st} = 0.37651). Level of genetic differentiation among *An. maculatus* populations in Thailand (F_{st} = 0.0406) (Rongnoparut *et al.*, 1999) was lower than in *An. harrisoni* in Thailand (F_{st} = 0.04210) and *An. harrisoni* in Thailand and Vietnam (F_{st} = 0.37651). F_{st} value of *An. gambiae* across loci in division 1-4 on X chromosome (F_{st} = 0.018) (Stump *et al.*, 2005) was lower than in *An. harrisoni* in Thailand (F_{st} = 0.04210) and *An. harrisoni* in Thailand and Vietnam (F_{st} = 0.37651), F_{st} value of *An. gambiae* across loci in division 5 on X chromosome (F_{st} = 0.273) (Stump *et al.*, 2005) was higher than in *An. harrisoni* in Thailand (F_{st} = 0.04210) and lower than in *An. harrisoni* in Thailand and Vietnam (F_{st} = 0.37651) as well as F_{st} value of *An. gambiae* across loci in division 6 on X chromosome (F_{st} = 0.874) (Stump *et al.*, 2005) was higher than in *An. harrisoni* in Thailand (F_{st} = 0.04210) and *An. harrisoni* in Thailand and Vietnam (F_{st} = 0.37651). Genetic differentiation level of *An. arabiensis* (F_{st} range from 0.75 to 0.99) (Stump *et al.*, 2005) was higher than in *An. harrisoni* in Thailand (F_{st} = 0.04210) and *An. harrisoni* in Thailand and Vietnam (F_{st} = 0.37651).

In addition, average pairwise estimated population differentiation across 5 loci in 3 populations was also calculated and results are shown in Table 22.

Table 22 Average pairwise estimated population differentiation across 5 loci in 3 populations of *An. harrisoni*.

Population	KPT	KBT	VHB
KPT	-	0	0.382
KBT	-	-	0.349
VHB	-	-	-

Note: negative F_{st} were set to zero; KPT: Pu Teuy Village; KBT: Bong Ti Noi Village; VHB: Hoa Binh province.

The result showed that no genetic differentiation was observed among 2 populations from Kanchanaburi province, Thailand (KPT and KBT). Genetic differentiations between VHB and KPT, as well as VHB and KBT were 0.382 and 0.349, respectively. Genetic differentiation between Vietnam and Thailand population are high, because VHB population is geographically isolated from KPT and KBT populations.

Part B: Phylogenetic inference with multilocus data

Phylogenetic relationships

Most of the nucleotide sequences of COI, COII, and ITS2 fragments were retrieved from GenBank and a few sequences were newly sequenced. The accession no. of these samples are shown in Table 9. Nucleotide frequencies of each DNA fragment were carried out and shown in Table 23, 24, 25, 26, 27, and 28.

Table 23 Nucleotide compositions of *Anopheles* mosquitoes from COI fragment.

Species	Base compositions (%)			
	Thymine	Cytosine	Adenine	Guanine
An. dirus	42.2	16.5	27.7	13.6
An. baimaii	42.2	16.5	28.2	13.1
An. scanloni	42.2	17.5	26.7	13.6
An. maculatus	40.8	17.0	28.2	14.1
An. sawadwongporni	43.2	15.0	27.2	14.6
An. epiroticus	43.6	16.2	27.0	13.2
An. minimus	44.2	16.0	25.7	14.1
An. harrisoni	44.7	15.5	25.7	14.1
An. aconitus	43.7	16.0	26.2	14.1
An. varuna	41.3	17.0	26.7	15.0
An. pampanai	42.7	14.1	28.2	15.0
Ae. aegypti	41.3	19.9	25.2	13.6
Avg.	42.8	16.1	27.0	14.0

Table 24 Nucleotide compositions of *Anopheles* mosquitoes from COII fragment.

Species	Base compositions (%)			
	Thymine	Cytosine	Adenine	Guanine
An. dirus	38.4	13.5	35.9	12.3
An. baimaii	38.4	13.5	36.0	12.1
An. scanloni	38.4	13.7	35.5	12.4
An. maculatus	39.3	13.0	35.0	12.8
An. sawadwongporni	39.3	13.0	35.0	12.8
An. epiroticus	38.7	13.0	35.5	12.8
An. minimus	38.4	13.5	36.2	11.9
An. harrisoni	38.6	13.7	35.9	11.9
An. aconitus	40.2	12.4	35.7	11.7
An. varuna	38.6	14.6	34.4	12.4
An. pampanai	40.0	13.0	35.1	11.9
Ae. aegypti	39.6	15.1	34.8	10.5
Avg.	38.9	13.4	35.5	12.3

Table 25 Nucleotide compositions of *Anopheles* mosquitoes from O8 fragment.

Species	Base compositions (%)			
	Thymine	Cytosine	Adenine	Guanine
An. dirus	26.0	21.3	27.8	24.8
An. baimaii	26.9	20.6	26.7	25.8
An. sawadwongporni	26.5	20.9	27.8	24.8
An. epiroticus	26.5	20.9	27.8	24.8
An. minimus	26.6	21.0	27.3	25.2
An. harrisoni	26.7	20.9	27.6	24.8
An. aconitus	26.5	20.9	27.8	24.8
An. varuna	26.5	20.9	27.8	24.8
Avg.	26.5	20.9	27.6	25.0

Table 26 Nucleotide compositions of *Anopheles* mosquitoes from P9 fragment.

Species	Base compositions (%)			
	Thymine	Cytosine	Adenine	Guanine
An. dirus	29.3	19.2	26.5	25.1
An. baimaii	30.0	19.3	26.1	24.5
An. maculatus	28.1	22.0	23.4	26.5
An. sawadwongporni	28.0	22.6	21.7	27.8
An. minimus	28.0	20.8	23.6	27.6
An. harrisoni	27.4	21.0	23.7	27.9
An. aconitus	21.7	23.4	31.4	23.4
D. melanogaster	32.8	19.4	30.0	17.8
Avg.	27.5	21.2	25.2	26.1

Table 27 Nucleotide compositions of *Anopheles* mosquitoes from TOLL6 fragment.

Species	Base compositions (%)			
	Thymine	Cytosine	Adenine	Guanine
An. dirus	20.7	27.4	21.1	30.8
An. baimaii	20.9	27.4	21.1	30.7
An. maculatus	23.7	24.8	22.1	29.4
An. sawadwongporni	22.6	25.1	21.5	30.8
An. epiroticus	23.0	26.9	22.8	27.4
An. minimus	23.8	24.9	21.9	29.4
An. harrisoni	23.8	24.7	22.2	29.2
An. aconitus	21.9	26.5	21.4	30.1
An. varuna	24.0	25.1	21.9	29.0
D. melanogaster	20.2	28.3	24.3	27.1
Avg.	22.7	25.9	21.8	29.6

Table 28 Nucleotide compositions of *Anopheles* mosquitoes from ITS2 fragment.

Species	Base compositions (%)			
	Thymine	Cytosine	Adenine	Guanine
An. dirus	18.0	32.9	16.4	32.7
An. baimaii	17.2	34.0	16.2	32.6
An. maculatus	20.3	28.5	23.8	27.4
An. sawadwongporni	19.1	29.4	23.7	27.8
An. epiroticus	22.8	26.2	21.6	29.4
An. minimus	21.2	27.6	24.1	27.2
An. harrisoni	22.1	28.0	22.8	27.1
An. aconitus	20.2	28.1	26.1	25.7
An. varuna	21.0	26.7	25.5	26.9
An. pampanai	21.2	28.9	25.5	24.4
D. melanogaster	35.6	10.0	39.3	15.1
Avg.	20.2	29.0	22.2	28.5

Test of substitution saturation

Strong substitution saturation (too diverged sequence) is mainly effects for the accuracy of phylogenetic reconstruction, reducing phylogentic information contained in sequences (Xia and Lemey, 2009). Substitution saturation of sequences was assessed by using Xia's method (Xia *et al.*, 2003) and Steel's method (Steel *et al.*, 1993) with DAMBE version 5.2.57computer program (Xia and Xie, 2001). With the expectation of nucleotide substitution rates among coding and noncoding sequences is different. Thus, these sequences were separately considered. In coding sequences, nucleotide positions are evolved at different rate; third codon positions are more variable than at first or second positions (Halliburton, 2004). The results of test of substitution saturation indicate that the sequences of each fragment showed little substitution saturation of nucleotides which these fragments can be utilized for phylogenetic relationships analysis. The details of those tests are presented in Appendix F.

Test of the nucleotide substitution model

The most appropriate substitution model was carried out. The information of selected models from each fragment is provided in Table 29.

Table 29 Appropriate nucleotide substitution model for likelihood analysis.

Locus	BIC	AICc	*lnL*	Selected model
P9	5459.5	5367.8	-2668.5	T92
TOLL6	5275.1	5142.7	-2551.3	T92+G
COI	1749.9	1593.8	-769.6	TN93+G
COII	4120.9	3917	-1928.4	GTR+G
ITS2	6888.4	6756	-3356.9	K2+G

Note: O8 fragment was excluded from analysis due to low genetic variation; BIC: Bayesian Information Criterion; AICc: Akaike Information Criterion, corrected; Maximum Likelihood value (*lnL*); T92: Tamura 3-parameter; TN93: Tamura-Nei; K2: Kimura 2-parameter; GTR: General Time Reversible; +G: a discrete Gamma distribution.

Phylogenetic inference with multilocus DNA sequences

Phylogenetic trees were inferred with four different methods based on five partial fragments by heuristic searches using MEGA version 5 computer program (Tamura *et al.*, 2011).

Phylogenetic relationships based on morphological characters were utilized to compare with molecular relationships in this study. Because species classifications based on morphological are not covered species in this study, to examine relationships was performed in taxonomic scale of the series. But these relationships data are incongruent (Foley *et al.*, 1998; Anthony *et al.*, 1999; Sallum *et al.*, 2000).

Phylogenetic trees based on four methods from each fragment showed incongruent. When tree topologies of each fragment were considered arrangements between ingroup and outgroup species, appropriate MP analysis was chosen for considering.

Phylogenetic analysis of mitochondrial genes (COI and COII)

For the COI gene, nucleotide sequences of 9 *Anopheles* species were retrieved from Genbank and 2 species were newly sequenced. *Aedes aegypti* was used as outgroup. Nucleotide sequences of the COII fragment were retrieved from the GenBank with 9 *Anopheles* species and compared with newly sequenced 2 *Anopheles* species. Sequence of *Aedes aegypti* was used as outgroup. Results of MP tree topologies are shown in Figure 23 and details of other methods are showed in Appendix G.

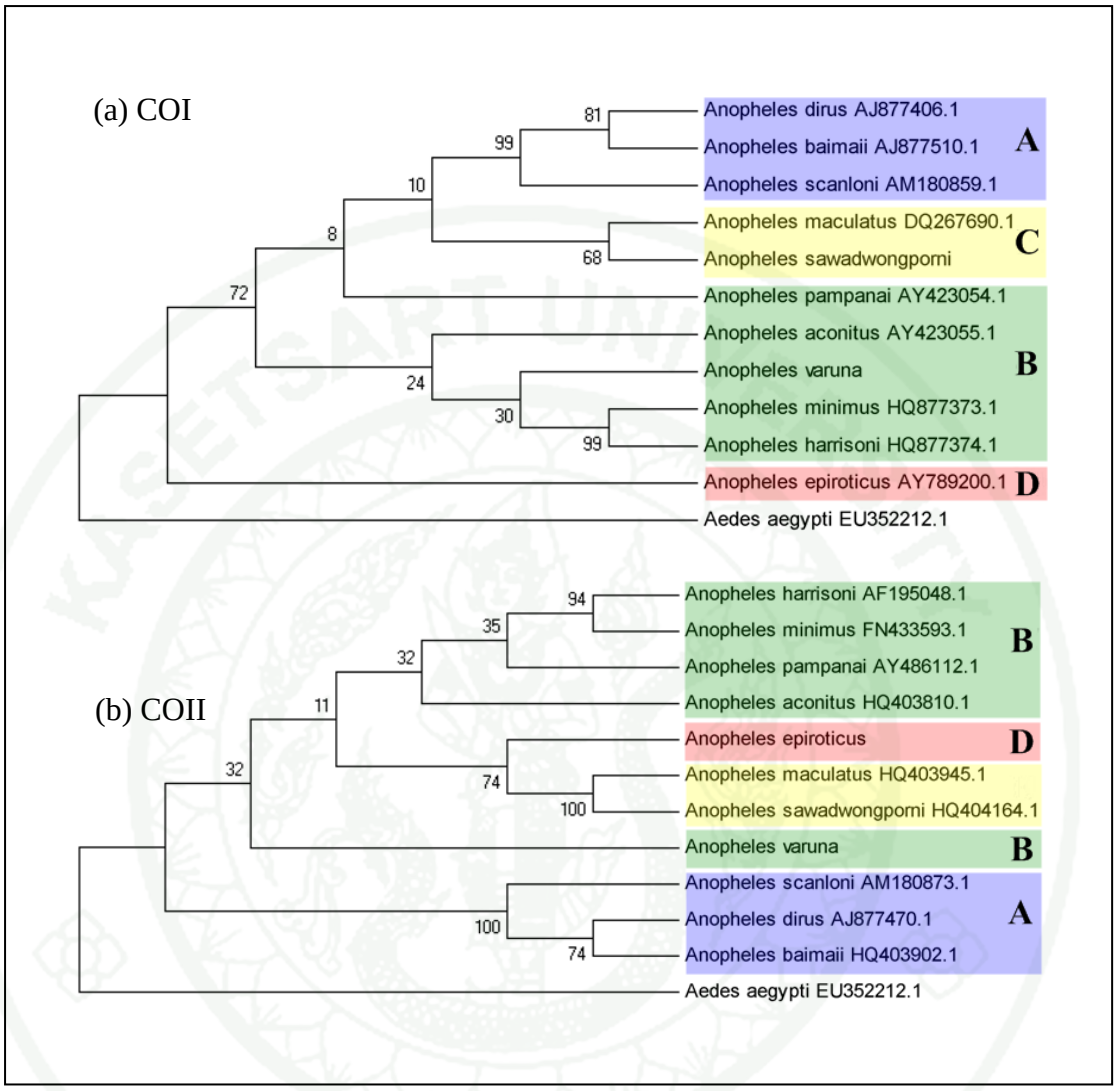


Figure 23 Phylogenetic tree based on the COI (a) and COII (b) fragments (MP reconstruction) with bootstrap values; A: Neomyzomyia Series, B: Myzomyia Series, C: Neocellia Series, and D: Pyretophorus Series. The name of retrieved mosquito species were mentioned with reference accession no.

MP analysis of the COI fragment (Figure 23, (a)) divided into three main clusters. First cluster is monophyly of the Neomyzomyia cluster (*An. dirus*, *An. baimaii* and *An. scanloni*). Second cluster is paraphyletic of Neocellia (*An. maculatus* and *An. sawadwongporni*) cluster relating to the Neomyzomyia cluster, but branch support is poor. Third cluster is the Myzomyia cluster. *An. pampanai* showed improper arrangement from the Myzomyia cluster, the relationship between *An. minimus* and *An. harrisoni* is consistent with the result of Garros *et al.* (2005b), and *An. epiroticus* is the basal species of this topology. MP analysis of the COII fragment (Figure 23, (b)) separated groups into three clusters. First cluster contains monophyletic group of the Myzomyia Series (*An. minimus*, *An. harrisoni* and *An. pampanai* and *An. aconitus*), the relationship between *An. minimus* and *An. harrisoni* is consistent with the results of Sharpe *et al.* (2000) and Garros *et al.* (2005a). Second cluster contains the Pyretophorus (*An. epiroticus*) and Neocellia (*An. maculatus* and *An. sawadwongporni*) Series. Third cluster is basal cluster of the Neomyzomyia cluster (*An. dirus*, *An. baimaii* and *An. scanloni*). These congruent clusters are supported by morphological relationships of Anthony *et al.* (1999). However, branch supports among each cluster are poor and unreliable.

The phylogenetic relationships based on the mitochondrial DNA data showed that branch supports of species complexs are moderate to strong, while poor branch supports are observed among groups of species complexs and closely related species. This is consistent with studying phylogeny of subfamily Anophelinae (Sallum *et al.*, 2002). Sallum *et al.* (2002) suggested that the mitochondrial DNA data hold little phylogenetic signal, but utilizing for relationships of very recently derived species are exception. Howland and Hewitt (1995) investigated phylogenetic relationships of the family Coleoptera (Ground Beetles) based on the mitochondrial COI sequence data. They suggested that the COI sequence data may be suitable for molecular systematics at lower taxonomic levels of the family. On the other hand, the mitochondrial COI and COII data was not likely suitable to resolve phylogenetic relationships at higher taxonomic level of species complexs.

Phylogenetic analysis of nuclear fragments (P9, ITS2, and TOLL6)

Nucleotide sequences of the P9 fragment were newly sequenced from 7 *Anopheles* species. Sequence of *Drosophila melanogaster* was used as outgroup. Nucleotide sequences of the TOLL6 fragment were newly sequenced from 9 *Anopheles* species. Sequence of *Drosophila melanogaster* was used as outgroup. 10 nucleotide sequence of *Anopheles* species from the ITS2 fragment were retrieved from GenBank and sequence of *Drosophila melanogaster* was used as outgroup. Results of MP tree topologies are shown in Figure 24 and details of other methods are showed in Appendix G.

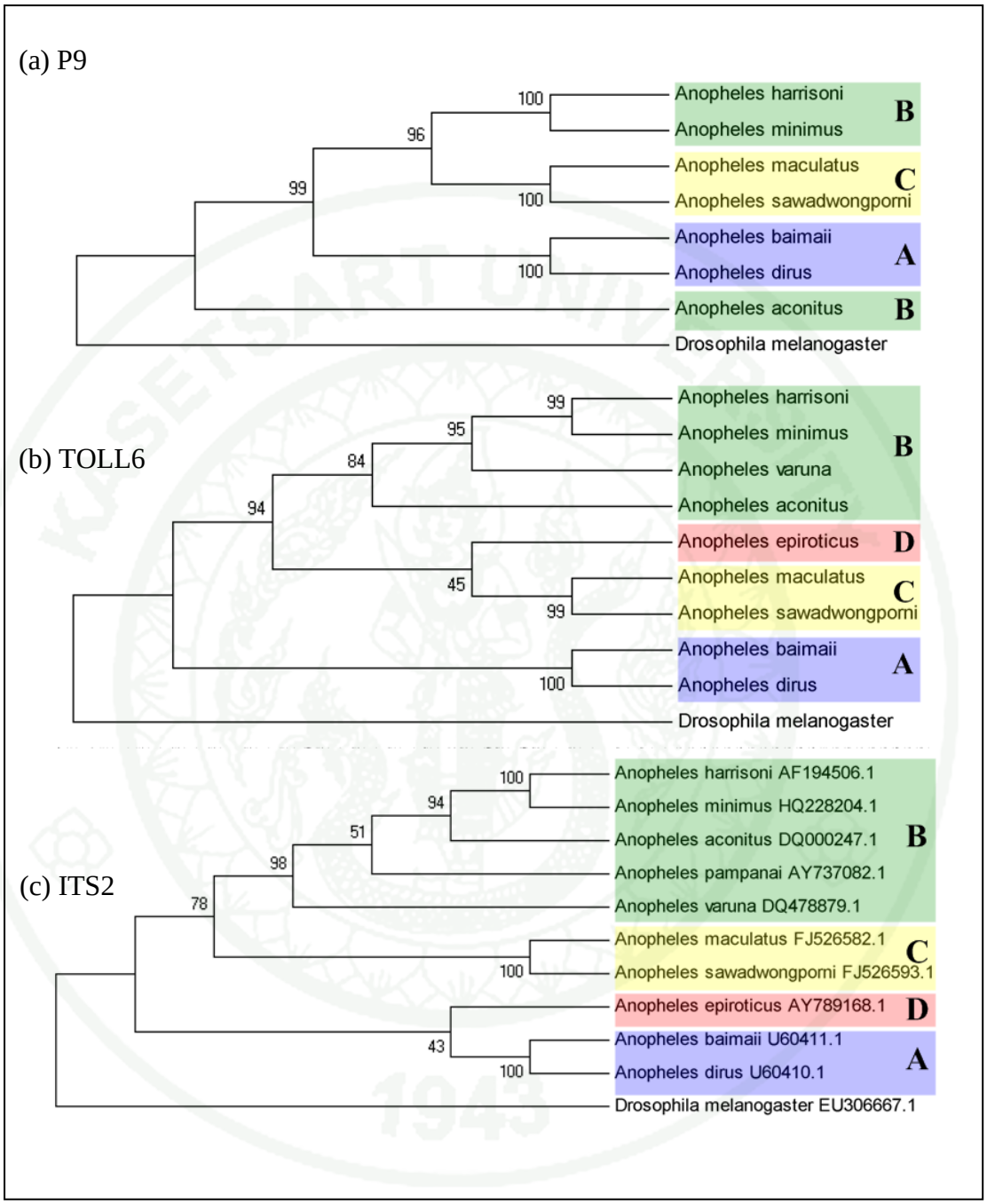


Figure 24 Phylogenetic tree based on the P9 (a), TOLL6 (b) and ITS2 (c) fragments (MP reconstruction) with bootstrap values; A: Neomyzomyia Series, B: Myzomyia Series, C: Neocellia Series, and D: Pyretophorus Series. The name of retrieved mosquito species were mentioned with reference accession no.

MP analysis results of the P9 fragment (Figure 24, (a)) separated investigated species into three clusters. First cluster is the Myzomyia cluster (*An. minimus*, *An. harrisoni*). This cluster formed monophyly with second cluster of the Neocellia Series (*An. maculatus* and *An. sawadwongporni*). Third cluster contains the Neomyzomyia Series (*An. dirus* and *An. baimaii*). These clusters are supported by morphological relationships of Anthony *et al.* (1999) (Figure 25 (B)), excepting *An. varuna*; it showed improper basal position because it belongs to member of the Myzomyia Series. For this fragment, more species are required for test the phylogenetic signal in higher taxonomic levels. Topologies from all analyses of the TOLL6 fragment (Figure 24, (b)) are consistent. Investigated species were separated into three main clusters. First cluster contains monophyly of the Myzomyia cluster (*An. minimus*, *An. harrisoni*, *An. varuana* and *An. pampanai*). Second cluster contains the Pyretophorus (*An. epiroticus*) and Neocellia (*An. maculatus* and *An. sawadwongporni*) Series. The Neomyzomyia Series (*An. dirus* and *An. baimaii*) is paraphyly and basal cluster of these topologies. Relationships of these species are consistent with morphological relationships of Foley *et al.* (1998) (Figure 25 (A)). Results suggested that the TOLL6 fragment data may be an informative gene containing phylogenetic signal. More species from several taxonomic levels are required to confirm. Results of NJ, ME, MP and ML analyses of the ITS2 fragment (Figure 24, (c)) generated similar topologies (figures of NJ, Me, and ML phylogenetic trees are also presented in Appendix G). Three main clusters were constructed. First cluster is monophyly of the Myzomyia cluster (*An. minimus*, *An. harrisoni*, *An. varuana*, *An. pampanai*, and *An. aconitus*). Second cluster is the Neocellia (*An. maculatus* and *An. sawadwongporni*) cluster and basal cluster contains the Pyretophorus (*An. epiroticus*) and Neomyzomyia Series (*An. dirus* and *An. baimaii*). Most of branch support showed moderate to strong supporting, except branch between the Pyretophorus and Neomyzomyia Series in MP analysis and most of branchs in ML analysis.

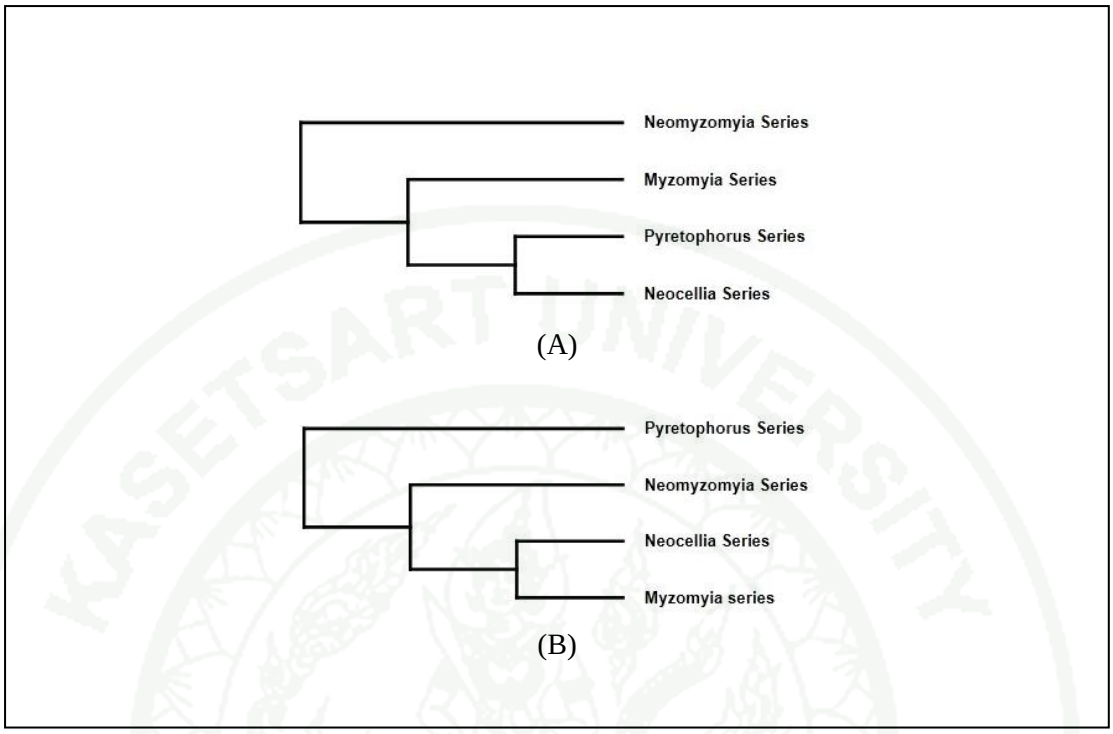


Figure 25 Illustration of relationships among the Series of mosquito taxonomic level based on morphological characters; (A) Foley *et al.* (1998), (B) Anthony *et al*. (1999).

In this study, the results of phylogenetic suggested that TOLL6 (Toll-like receptor) fragment appears much more informative and should be useful in phylogenetic reconstruction within these samples of Anopheline mosquitoes because the phylogenetic trees from all approaches were similar tree topology and high bootstrap supports, although bootstrap support for some clade dropped. Additionally, P9 fragment also appears informative signal for phylogenetic reconstruction, but phylogenetic positions of some species were incorrect. However, the observation of Mardulyn and Whitfield (1999) suggested that informative and useful gene for phylogenetic reconstruction in one taxon properly might not resolve in other taxa, even closely related. Thus, more *Anopheles* species are required for testing phylogenetic information of these regions.

In addition, the consistent relationships of the Series taxonomic level were observed in the multilocus phylogenetic analysis. Relationships among main species, the Series of mosquitoes, and interested fragments were considered and depicted in Figure 26. This result showed that arrangements of the Series taxonomic levels between the TOLL6 (nuclear gene) and COII (mitochondrial gene) fragments were consistent. Additionally, relationships between the P9 and ITS2 fragments were quite consistent, but unclear due to lack of amplification result of *An. epiroticus* in P9 fragment while arrangement of COI fragment was unique.

Similar phylogenetic status between the TOLL6 and COII fragments presented that the Neocellia and Pyretophorus Series are monophyletic group and the Myzomyia Series has a close affinity with monophyletic group, supporting with morphological character analysis of Foley *et al.* (1998). Moreover, in this study, the phylogenetic positions of COII are consistent to results of Dixit *et al.* (2010) that the Pyretophorus Series has a close affinity with the Neocellia Series and this clade was clustered together with the Myzomyia Series.

Dixit *et al.* (2010) suggested that horizontal gene transfer and gene duplication may have resulted in detecting discordant phylogenetic status of member of gene families which it is consistent to TOLL6 and P9 (non-coding region of CYP) fragments.

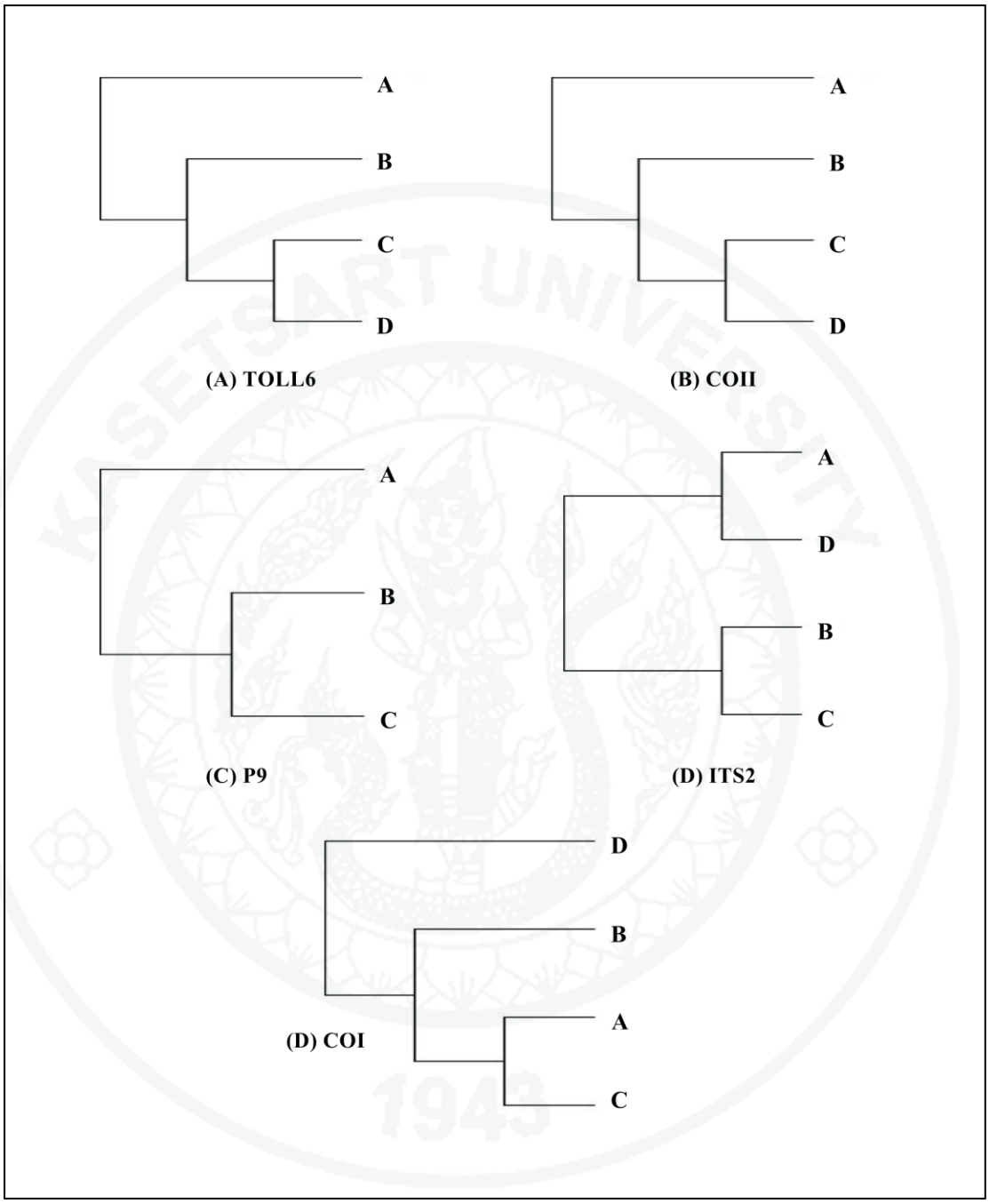


Figure 26 Illustration of relationships between interested fragments and the Series of mosquito taxonomic level based on MP phylogenetic trees; A: Neomyzomyia Series, B: Myzomyia Series, C: Neocellia Series, D: Pyretophorus Series.

Estimation of divergence time

The molecular evolutionary clock hypothesis was firstly proposed by Zuckerkandl and Pauling in 1965 - the rate of evolution in protein or DNA molecule is approximately constant over time and among evolutionary lineages (Morgan, 1998). This hypothesis is useful for estimating the evolutionary history and divergence time of organisms having poor fossil information record (Hoy, 2003).

For testing the molecular evolutionary clock hypothesis, simple statistical methods (relative rate test) based on the chi-square test are developed by Tajima (1993). This test was implemented in MEGA version 5 computer program (Tamura *et al.*, 2011). *P*-value less than 0.05 is considered to reject the null hypothesis of the equality of evolutionary rate between lineages. Firstly, no fragment was found to follow the molecular evolutionary clock hypothesis. The results were then reconsidered. Most of problematic species observing in the COI, P9, TOLL6 and ITS2 fragments were main malaria vectors (data not shown), but the COII gene fragment, only *An. aconitus* considering secondary vector was problematic and *An. varuna* is differently positioned from the species groups of the Myzomyia Series. Thus, these species were excluded from analysis. Using the mitochondrial COII fragment as genetic marker, Gissi *et al.* (2000) suggested that unique properties of the mitochondrial genome, such as the presence of strictly orthologous genes, the lack of recombination, and an appropriate substitution rate, make it become a common tool of studying phylogenetic relationships and estimation of divergence times between species. Results of relative rate test between Anopheline mosquitoes at the COII gene fragment are shown in Table 30.

Table 30 Relative rate tests between Anopheline mosquitoes at the COII gene fragment.

Species	*Dir*	*Sca*	*Bai*	*Mac*	*Saw*	*Epi*	*Har*	*Min*	*Pam*
Dir	-	0.14,0.705	1.00,0.317	0.16,0.685	0.16,0.685	0.02,0.896	0.02,0.890	0.02,0.888	0.17,0.680
Sca	*Pam*	-	0.00,1.000	0.07,0.796	0.27,0.605	0.15,0.696	0.07,0.785	0.72,0.396	1.19,0.276
Bai	*Har*	*Pam*	-	0.02,0.894	0.07,0.789	0.00,1.000	0.07,0.785	0.08,0.781	0.30,0.586
Mac	*Min*	*Har*	*Pam*	-	0.00,1.000	0.09,0.763	0.47,0.492	1.85,0.173	0.17,0.680
Saw	*Har*	*Min*	*Har*	*Har*	-	0.20,0.654	0.47,0.492	1.85,0.173	0.17,0.680
Epi	*Min*	*Pam*	*Min*	*Min*	*Har*	-	0.83,0.362	1.19,0.276	0.44,0.507
Har	*Epi*	*Epi*	*Epi*	*Dir*	*Dir*	*Dir*	-	0.06,0.808	0.03,0.872
Min	*Mac*	*Mac*	*Mac*	*Sca*	*Sca*	*Sca*	*Dir*	-	0.10,0.751
Pam	*Saw*	*Saw*	*Saw*	*Bai*	*Bai*	*Bai*	*Mac*	*Saw*	-

Note: Above-diagonal: The chi-square test statistic and *P*-value, respectively; below-diagonal: the species considering as an outgroup between species pair; *Dir*: *An. dirus*, *Sca*: *An. scanloni*, *Bai*: *An. baimaii*, *Mac*: *An. maculatus*, *Saw*: *An. sawadwongporni*, *Epi*: *An. epiroticus*, *Har*: *An. harrisoni*, *Min*: *An. minimus*, *Pam*: *An. pampanai*.

For analysis with mcmctree computer program (a part of PAML version 4.4 computer program), sequence data file and tree file formats are required. Thus, phylogenetic trees were newly generated based on NJ, MP and ML methods and the trees are shown in Figure 27, 28 and 29.

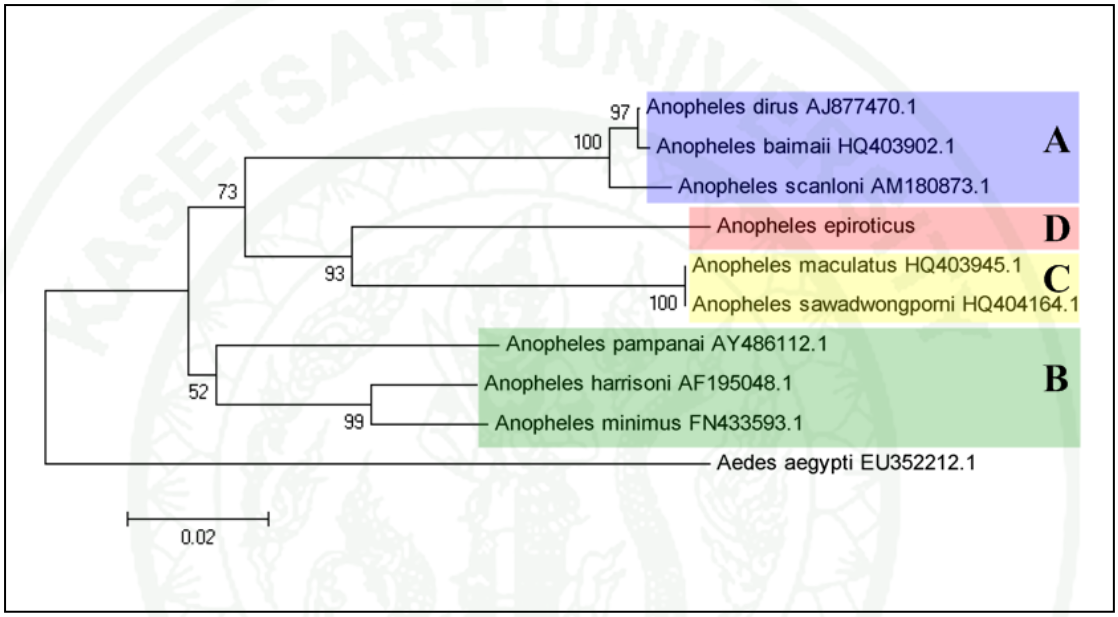


Figure 27 NJ phylogenetic tree with the COII gene fragment in nine *Anopheles* mosquitoes; A: Neomyzomyia Series, B: Myzomyia Series, C: Neocellia Series, and D: Pyretophorus Series. The name of retrieved mosquito species were mentioned with reference accession no.

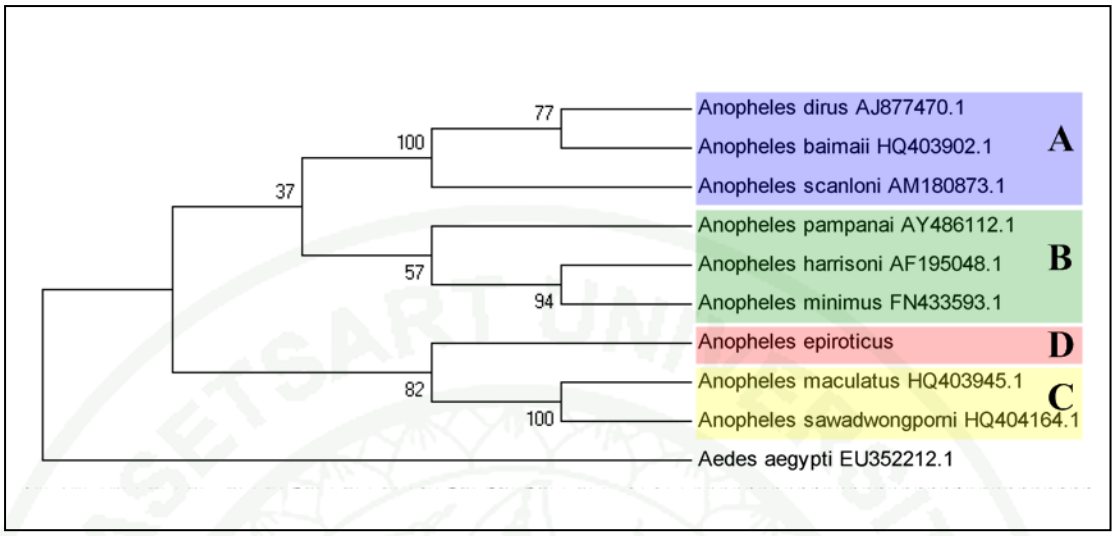


Figure 28 MP phylogenetic tree with the COII gene fragment in nine *Anopheles* mosquitoes; A: Neomyzomyia Series, B: Myzomyia Series, C: Neocellia Series, and D: Pyretophorus Series. The name of retrieved mosquito species were mentioned with reference accession no.

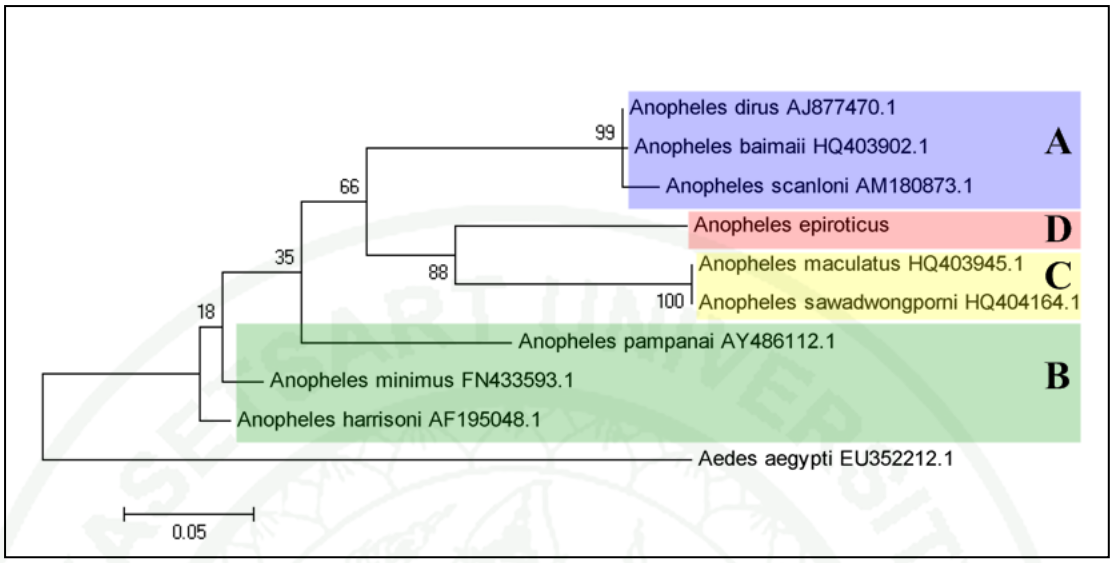


Figure 29 ML phylogenetic tree with the COII gene fragment in nine *Anopheles* mosquitoes; A: Neomyzomyia Series, B: Myzomyia Series, C: Neocellia Series, and D: Pyretophorus Series. The name of retrieved mosquito species were mentioned with reference accession no.

Each topology of different methods of phylogenetic tree reconstruction was considered. Branch pattern supporting with boostrap test less than 50% majority rule was considered as questionable branch. Thus, phylogenetic tree of NJ method supporting with moderate to high branch support was chosen for estimating divergence time. The fossil calibration information is required, but the fossil record is so poor for estimating the divergence and actual ages of the taxa in family Culicidae (Poinar *et al.*, 2000; Harbach, 2007). The problem of using one calibration point from previous study (Dixit *et al.*, 2010) was addressed in this study by using two calibration points (Garros *et al.*, 2005b; Krzywinski *et al.*, 2006). The estimation of divergence time points of mosquito mtDNA genomes between *Anopheles* and *Aedes* (approximately 145 to 200 mya) (Krzywinski *et al.*, 2006) and between *An. minimus* and *An. harrisoni* (approximately 1.1 to 1.3 mya) (Garros *et al.*, 2005b) were utilized to estimate the splits between lineages within subgenus *Cellia* (Dixit *et al.*, 2010). The most complex nucleotide substitution model of HKY 85, the 10,000 burning periods, the 100,000 numbers of samples, and the 5 sample frequencies were set for computer simulations and performed with four independent analyses. Phylogeny and

divergence times with their associated 95% confidence intervals in *Anopheles* mosquitoes based on the COII gene fragment are shown in Figure 30.

The estimated divergence time within this subgenus *Cellia* (late 41 mya) is consistent with divergence between subgenus *Anopheles* and *Cellia*, i.e. approximately 90 to 100 mya (Krzywinski *et al.*, 2006) and after the break-up of Gondwana (approximately 100 mya at the late Cretaceous Period) (Garros *et al.*, 2005b). Lineages of the Myzomyia Series (*An. minimus*, *An. harrisoni*, and *An. pampanai*) were firstly splitted from the Pyretophorus Series (*An. epiroticus*), Neomyzomyia Series (*An. dirus*, *An. scanloni*, and *An. baimaii*), and Neocellia Series (*An. maculatus* and *An. sawadwongporni*). Speciation between the Aconitus Subgroup (*An. pampanai*) and the Minimus Subgroup (*An. minimus* and *An. harrisoni*) within the Funestus Group of the Myzomyia Series was estimated to have occurred during the late Eocene to late Miocene (Eocene epoch, approximately 34 to 56 mya; Miocene epoch, approximately 5 to 23 mya). The estimated divergence time of the Pyretophorus Series (*An. epiroticus*) and the Neocellia Series occurred during the Miocene (Miocene epoch, approximately 5 to 23 mya). The divergence time estimating between *An. maculatus* and *An. sawadwongporni* (the Neocellia Series) (approximately 2.1 mya; early Pleistocene) is earlier than the previous studies (based on the combined ND5 and COII), occurred during the Pliocene to early Pleistocene (Morgan *et al.*, 2009).

The species divergence of the Dirus Complex within the Neomyzomyia Series (*An. dirus*, *An. baimaii* and *An. scanloni*) estimated approximately 0.051 to 7.614 mya. The divergence time of *An. scanloni* from *An. dirus* and *An. baimaii* is approximately 0.271 to 7.614 mya, and the divergence time between *An. baimaii* and *An. dirus* is approximately 0.051 to 2.338 mya. Thus, the speciation events are most likely due to Pleistocene forest fragmentation (Pleistocene epoch, approximately 0.0117 to 2.588 mya). During the glacial period of the effects of Pleistocene climatic change, forest-dependent species survived in tropical forest refugia and remained in some areas of mainland Southeast Asia. Then, the allopatric species of the Dirus Complex would have expanded with the extent of forest cover across Southeast Asia in post-gracial recolonization, according to the habitats of these mosquitoes which are

only found in tropical forests (O'Loughlin *et al.*, 2008; Morgan *et al.*, 2010; Morgan *et al.*, 2011; Loaiza *et al.*, 2012). This result of divergence times is similar to the allopatric divergence in the speciation of *An. baimaii*, *An. dirus* and *An. scanloni* within the last 1.5 mya: approximately 192-877 thousand years ago (kya) between *An. baimaii* and *An. scanloni*, approximately 204-932 kya between *An. dirus* and *An. scanloni* and approximately 163 kya to 1.53 mya between *An. dirus* and *An. baimaii* (Morgan *et al.*, 2010).

When the estimations of divergence time of mosquito species in this study were considered overall, speciation of these species was estimated to have occurred during the late Miocene to early Plesitocene. This period was considered as an important time for diversification in other forest-dependent taxa across mainland and island Southeast Asia (Morgan *et al.*, 2009). Pliocene climatic change according to increasing forest fragmentation is likely the driving force for this diversification, precipitation levels were high and consequently tropical forest had extensively covered across Southeast Asia (Chandler *et al.*, 1994; Cronin *et al.*, 1994; Ravelo *et al.*, 2004). These data are consistent with habitats of these mosquito species.

Using a single locus may be problematic for estimating divergence events. Thus, the multiple loci following molecular clock hypothesis are required for the estimation of divergence time for further study. However, this estimation of divergence time is the initial step for basic understanding for Anopheline (*Cellia*) mosquitoes. Multiple gene loci, in particular nuclear genes, and more *Cellia* mosquito samples are required for further study in order to understanding the evolution of *Anopheles* (*Cellia*) mosquitoes.

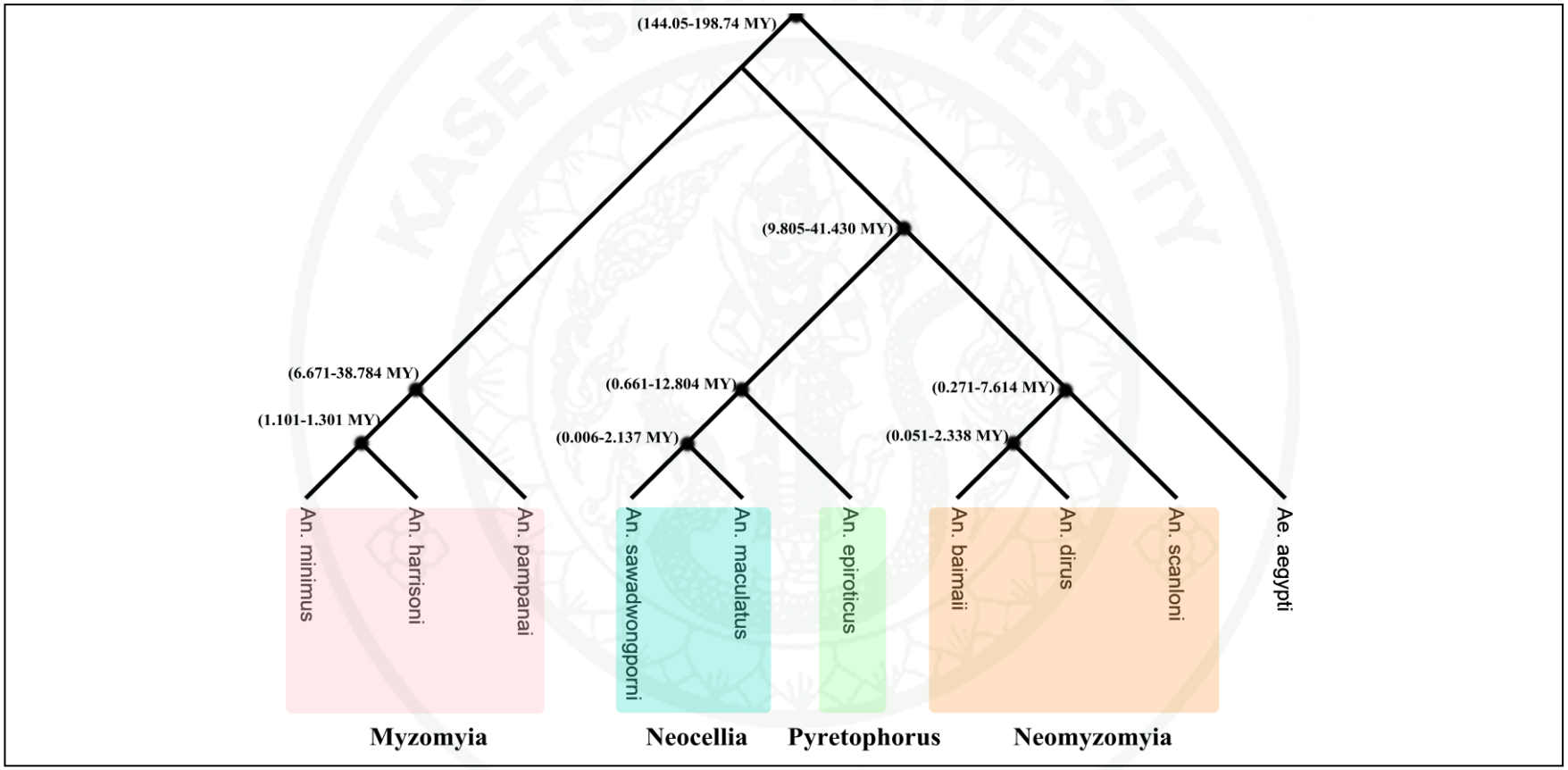


Figure 30 Phylogeny and divergence times in *Anopheles* mosquitoes from the COII gene fragment based on PAML version 4.4 computer program; MY: million year.

CONCLUSION

Part A: Genetic diversity and tests of population genetic models

Two *An. harrisoni* populations from Kanchanaburi province and one population from north-western Vietnam were used to study pattern of genetic variation and level of population differentiation.

The total length of six studied loci ranges from 245 bp to 609 bp. Nucleotide diversity was estimated based on SNPs using π and Watterson's θ_w estimators. The average level of nucleotide variation (π) across all loci varies from 0.151 – 0.268 %. For neutrality, Tajima's *D* statisctic test did not show significant deviation from standard neutral expectation for all populations as well as, Fu and Li's *D* statistic sand the *H* test of Fay and Wu showing no departure from neutral equilibrium expectations. Furthermore, no genetic differentiation was observed among each population in Kanchanaburi province. However genetic differentiation was observed among populations from Vietnam and Kanchanaburi province, due to geographical isolation among these populations.

This preliminary study exhibited low level of nucleotide diversity for all three *An. harrisoni* populations, and they are under mutation-random genetic drift equilibrium. For further study, more populations may be required for better understanding of evolutionary history of this malaria vector species.

Part B: Phylogenetic inference with multilocus data

Phylogenetic relationships of *Anopheles* species presented in Thailand were analyzed using multilocus DNA sequence information. The results based on four different reconstructed algorithms as well as different gene regions revealed congruent phylogenetic status of malaria vector species especially at the Series taxonomic level. The nuclear TOLL6 fragment may be an informative gene containing phylogenetic signals to open a new perspective for reconstructing relationships in *Anopheles* mosquitoes. However, more samples of different taxonomic levels are required to confirm the taxonomic positions. Moreover, estimation of the divergence time among the *Anopheles* species studied was performed based on the COII fragment inferring that *Anopheles* species probably existed around the Eocene and Miocene periods (later than 41 million years ago).

LITERATURE CITED

Andersen, D.H., E. Fabbri, A. Santini, S. Paget, E. Cadieu, F. Galibert, C. André and E. Randi. 2006. Characterization of 59 canine single nucleotide polymorphisms in the Italian wolf (*Canis lupus*) population. **Mol. Ecol. Notes** 6: 1184–1187.

Anonymous. 2009. **Malaria**. Wikipedia, the free encyclopedia. Available source: http://en.wikipedia.org/wiki/Malaria, May 2, 2009.

Anthony, T.G., R.E. Harbach and I.J. Kitching. 1999. Phylogeny of the Pyretophorus Series of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). **Syst. Entomol.** 24(2): 193–205.

Arbogast, B.S., S.V. Edwards, J. Wakeley, P. Beerli and J.B. Slowinski. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. **Annu. Rev. Ecol. Syst.** 33: 707-740.

Bierne, N., S. A. Lehnert, E. Bédier, F. Bonhomme and S.S. Moore. 2000. Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. **Mol. Ecol.** 9(2): 233-235.

Black IV, W.C. and J.G. Vontas. 2007. Affordable assays for genotyping single nucleotide polymorphisms in insects. **Insect Mol. Biol.** 16(4): 377-387.

Boonsuepsakul, S., R. Sithiprasasna, V. Baimai and P. Rongnoparut. 2005. Microsatellite DNA polymorphism in *Anopheles minimus*, a major malaria vector in Thailand. **31st Congress of Science and Technology of Thailand**. Suranaree University of Technology, Nakhon Ratchasima, Thailand.

Booy, G., R.J.J. Hendriks, M.J.M. Smulders, J.M. Van Groenendael and B. Vosman. 2000. Genetic Diversity and the Survival of Populations. **Plant biol.** 2(4): 379-395.

Brumfield, R.T., P. Beerli, D.A. Nickerson and S.V. Edwards. 2003. The utility of Single nucleotide polymorphisms in inferences of population history. **Trends Ecol. Evol.** 18(5): 249-256.

Cann, A.J. 1996. **Biology of *Plasmodium* Parasites and *Anopheles* Mosquitos**. Introduction. Available source: http://www.wehi.edu.au/other_domains/MalDB/intro.html, April 25, 2009.

Castelloe, J. and A.R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. **Mol. Phylogenet. Evol.** 3(2): 102-113.

Chandler, M.A., D. Rind and R.S. Thompson. 1994. Joint investigations of the middle Pliocene climate II: GISS GCM Northern Hemisphere results. **Global Planet. Change** 9: 197-219.

Chareonviriyaphap, T., M.J. Bangs and S. Ratanatham. 2000. Review of malaria in Thailand. **Southeast Asian J. Trop. Med. Public Health**. 31(2): 225-237.

Chareonviriyaphap, T., A. Prabaripai, M.J. Bangs and B. Aum-Aung. 2003. Seasonal abundance and blood feeding activity of *Anopheles minimus* Theobald (Diptera: Culicidae) in Thailand. **J. Med. Entomol.** 40(6): 876-881.

Chen, B., R.K. Butlin and R.E. Harbach. 2003. Molecular phylogenetics of the Oriental members of the Myzomyia Series of *Anopheles* subgenus *Cellia* (Diptera: Culicidae) inferred from nuclear and mitochondrial DNA sequences. **Syst. Entomol.** 28: 57-69.

Cohuet, A., S. Krishnakumar, F. Simard, I. Morlais, A. Koutsos, D. Fontenille, M. Mindrinos and F.C. Kafatos. 2008. SNP discovery and molecular evolution in *Anopheles gambiae*, with special emphasis on innate immune system. **BMC Genomics** 9(227).

Cramer, E.R.A., L. Stenzler, A.L. Talaba, C.A. Makarewich, S.L. Vehrencamp and I.J. Lovette. 2008. Isolation and characterization of SNP variation at 90 anonymous Loci in the banded wren (*Thryothorus pleurostictus*). **Conserv. Genet.** 9(6): 1657-1660.

Cronin, T.M., A. Kitamura, N. Ikeya, M. Watanabe and T. Kamiya. 1994. Late Pliocene climate change 3.4-2.3 Ma: paleoceanographic record from the Yabuta Formation, Sea of Japan. **Palaeogeogr. Palaeoclimatol. Palaeoecol.** 108(3–4): 437-455.

Dixit, J., H. Srivastava, M. Sharma, M.K. Das, O.P. Singh, K. Raghavendra, N. Nanda, A.P. Dash, D.N. Saksena and A. Das. 2010. Phylogenetic inference of Indian malaria vectors from multilocus DNA sequences. **Infect. Genet. Evol.** 10: 755–763.

Dixit, J., H. Srivastava, O.P. Singh, D.N. Saksena and A. Das. 2011. Multilocus nuclear DNA markers and genetic parameters in an Indian *Anopheles minimus* population. **Infect. Genet. Evol.** 11(3): 572-579.

Drummond, A.J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. **BMC Evol. Biol.** 7(214).

Dusfour, I., J.R. Michaux, R.E. Harbach and S. Manguin. 2007. Speciation and phylogeography of the Southeast Asian *Anopheles sundaicus* complex. **Infect. Genet. Evol.** 7: 484–493.

Fay, J.C. and C.-I. Wu. 2000. Hitchhiking under positive Darwinian selection. **Genetics** 155: 1405-1413.

Foley, D.H., J.H. Bryan, D. Yeates and A. Saul. 1998. Evolution and systematics of *Anopheles*: insights from a molecular phylogeny of Australasian mosquitoes. **Mol. Phylogenet. Evol.** 9(2): 262-75.

Fu, Y.-X. and W.-H. Li. 1993. Statistical tests of neutrality of mutations. **Genetics** 133: 693-709.

Galinski, M.R. and J.W. Barnwell. 2009. Monkey malaria kills four humans. **Trends. Parasitol.** 25(5): 200-204.

Garros, C., L.L. Koekemoer, M. Coetzee, M. Coosemans and S. Manguin. 2004. A single multiplex assay to identify major malaria vectors within the African *Anopheles funestus* and the Oriental *Anopheles minimus* groups. **Am. J. Trop. Med. Hyg.** 70: 583-590.

Garros, C., R.E. Harbach and S. Manguin. 2005a. Morphological assessment and molecular phylogenetics of the Funestus and Minimus groups of *Anopheles* (*Cellia*). **J. Med. Entomol.** 42(4): 522-536.

Garros, C., R.E. Harbach and S. Manguin. 2005b. Systematics and biogeographical implications of the phylogenetic relationships between members of the funestus and minimus groups of *Anopheles* (Diptera: Culicidae). **J. Med. Entomol.** 42(1): 7-18.

Garros, C., W. Van Bortel, H.D. Trung, M. Coosemans and S. Manguin. 2006. Review of the Minimus Complex of *Anopheles*, main malaria vector in Southeast Asia: from taxonomic issues to vector control strategies. **Trop. Med. Int. Health** 11(1): 102-114.

Garros, C., W. Van Bortel, H.D. Trung, M. Coosemans and S. Manguin. 2006. Review of the Minimus Complex of *Anopheles*, main malaria vector in Southeast Asia: from taxonomic issues to vector control strategies. **Trop. Med. Int. Health** 11(1): 102-114. *Cited* R.E. Harbach, J.B. Gingrich and L.W. Pang. 1987. Some entomological observations on malaria transmission in a remote village in northwestern Thailand. **J. Am. Mosq. Control Assoc.** 3: 296-301.

Garros, C., W. Van Bortel, H.D. Trung, M. Coosemans and S. Manguin. 2006. Review of the Minimus Complex of *Anopheles*, main malaria vector in Southeast Asia: from taxonomic issues to vector control strategies. **Trop. Med. Int. Health** 11(1): 102-114. *Cited* S.R. Meek. 1995. Vector control in some countries of Southeast of Asia: comparing the vectors and the strategies. **Ann. Trop. Med. Parasit.** 89: 135–147.

Genetic Health. 2005. **What is a Single nucleotide polymorphisms SNP?**. Doctors Area. Available Source: http://www.genetic-health.co.uk/doctors/doctors-area- snp.htm, February 14, 2011.

Gissi, C., A. Reyes, G. Pesole and C. Saccone. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. **Mol. Biol. Evol.** 17(7): 1022-1031.

GPSFileDepot. 2007. **GPS Sway**. GPSFileDepot. Available Source: http://www.gpsfiledepot.com/tools/gpssway.php, August 21, 2011.

Green, C.A., R.F. Gass, L.E. Munstermann and V. Baimai. 1990. Population-genetic evidence for two species in *Anopheles minimus* in Thailand. **Med. Vet. Entomol.** 4: 25-34.

Guo, Y., and D.C. Jamison. 2005. The distribution of SNPs in human gene regulatory regions. **BMC Genomics** 6(140).

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucl. Acids. Symp. Ser.** 41: 95-98.

Halliburton, R. 2004. **Introduction to Population Genetics**. Pearson Education, Inc., United States of America.

Harbach, R.E. 2004. The classification of genus *Anopheles* (Diptera: Culicidae) : a working hypothesis of phylogenetic relationships. **Bull. Entomol. Res.** 94: 537-553.

Harbach, R.E. 2007. The Culicidae (Diptera): a review of taxonomy, classification and phylogeny, pp. 591–638. *In* Z.-Q. Zhang and W.A. Shear, eds. **Zootaxa**. Magnolia Press, Auckland.

Harbach, R.E. 2011. **Culicidae (Diptera)**. Mosquito Taxonomic Inventory. Available Source: http://mosquito-taxonomic-inventory.info/, October 9, 2011.

Harbach, R.E, C. Garros, N.D. Manh and S. Manguin. 2007. Formal taxonomy of species C of the *Anopheles minimus* sibling species (Diptera: Culicidae). **Zootaxa** 1654: 41-54.

Hare, M.P. 2001. Prospects for nuclear gene phylogeography. **Trends Ecol. Evol.** 16: 700-706.

Heaton, M.P., G.P. Harhay, G.L. Bennett, R.T. Stone, W.M. Grosse, E. Casas, J.W. Keele, T.P.L. Smith, C.G. Chitko-McKown and W.W. Laegreid. 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. **Mamm. Genome** 13: 272-281.

Howland, D.E. and G.M. Hewitt. 1995. Phylogeny of the Coleoptera based on mitochondrial cytochrome oxidase I sequence data. **Insect. Mol. Biol.** 4(3): 203-215.

Hoy, M.A. 2003 **Insect Molecular Genetics: An Introduction to Principles and Application**. 2nd ed. Academic Press, California.

Hudson, R.R. 2000. A new statistic for detecting genetic differentiation. **Genetics** 155: 2011–2014.

Hudson, R.R, M. Slatkin and W.P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. **Genetics** 132: 583–589.

Jarukamjorn, K. and W. Jatupornprasert. 2007. SNP: Basic knowledge to applications. **Thai Pharm. Health Sci. J.** 2(2): 166-174.

Kengluecha, A., P. Rongnoparut, S. Boonsuepsakul, R. Sithiprasasna, P. Rodpradit and V. Baimai. 2005. Geographical distribution of *Anopheles minimus* species A and C in western Thailand. **J. Vector Ecol.** 30(2): 225-230.

Kimura, M. 1983. **The neutral theory of Molecular Evolution**. Cambridge University Press, Cambridge, Massachusetts.

Kohn, M.H., W.J. Murphy, E.A. Ostrander and R.K. Wayne. 2006. Genomics and conservation genetics. **Trends Ecol. Evol.** 21(11): 629-637.

Krzywinski, J. and N.J. Besansky. 2003. Molecular systematic of *Anopheles*: From subgenera to subpopulations. **Annu. Rev. Entomol.** 48: 111-139.

Krzywinski, J, O.G. Grushko and N.J. Besansky. 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: An improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. **Mol. Phylogenet. Evol.** 39: 417-423.

Lai, E. 2001. Application of SNP technologies in medicine: lessons learned and future challenges. **Genome Res.** 11: 927-929.

Li, Y.B., Z. Hu, Q. Zhang, X. Deng and H.X. Wang. 2000. **Genetic variation in wild soybean (Glycine soja Sieb. & Zucc.) populations revealed with simple sequence repeat markers**. Soybean Genetics Newsletter 27 (Online journal). Available source: http://www.soygenetics.org/previewArticle.php?autoID=19&view=long, February 14, 2011.

Librado, P. and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. **Bioinformatics** 25: 1451-1452.

Lister, R., B.D. Gregory and J.R. Ecker. 2008. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. **COPLBI.** 12: 1-12.

Loaiza, J.R., E. Bermingham, O.I. Sanjur, M.E. Scott, S.A. Bickersmith and J.E. Conn. 2012. Review of genetic diversity in malaria vectors (Culicidae: Anophelinae). **Infect. Genet. Evol.** 12(1): 1-12.

Lowe, B.S., M. Mosobo and P.C. Bull. 1998. All four species of human malaria parasites form rosettes. **Trans. R. Soc. Trop. Med. Hyg.** 92(5): 526.

Manguin, S., C. Garros, I. Dusfour, R.E. Harbach and M. Coosemans. 2008b. Bionomics, taxonomy, and distribution of the major malaria vector taxa of *Anopheles* subgenus *Cellia* in Southeast Asia: An updated review. **Infect. Genet. Evol.** 8: 489-503.

Manguin, S., P. Carnevale and J. Mouchet. 2008a. **Biodiversity of Malaria in the World**. John Libbey Eurotext, France.

Mardulyn, P. and J.B. Whitfield. 1999. Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera; Braconidae): evidence of a high diversification rate in this group of parasitoids. **Mol. Phylogenet. Evol.** 12(3): 282-294.

Mehlotra, R.K. and P.A. Zimmerman. 2006. Resistance to antimalarial drugs: parasite and host genetic factors, pp. 81-124. *In* K.R. Dronamraju and P. Arese, eds. **Malaria: Genetic and Evolutionary Aspects**. Springer, New York.

Morgan, G.J. 1998. Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. **J. Hist. Biol.** 31: 155–178.

Morgan, K., S.M. O'Loughlin, F. Mun-Yik, Y.M. Linton, P. Somboon, S. Min, P.T. Htun, S. Nambanya, I. Weerasinghe, T. Sochantha, A. Prakash and C. Walton. 2009. Molecular phylogenetics and biogeography of the Neocellia Series of *Anopheles* mosquitoes in the Oriental Region. **Mol. Phylogenet. Evol.** 52(3): 588-601.

Morgan, K., S.M. O'Loughlin, B. Chen, Y.M. Linton, D. Thongwat , P. Somboon , M.Y. Fong, R. Butlin, R. Verity, A. Prakash, P.T. Htun, T. Hlaing, S. Nambanya, D. Socheat , T.H. Dinh and C. Walton. 2011. Comparative phylogeography reveals a shared impact of pleistocene environmental change in shaping genetic diversity within nine *Anopheles* mosquito species across the Indo-Burma biodiversity hotspot. **Mol. Ecol.** 20(21): 4533-4549.

Mohanty, A., S. Swain, S.K. Kar and R.K. Hazra. 2009. Analysis of the phylogenetic relationship of *Anopheles* species, subgenus *Cellia* (Diptera: Culicidae) and using it to define the relationship of morphologically similar species. **Infect. Genet. Evol.** 9: 1204-1224.

Morin, P.A., G. Luikart, R.K. Wayne and the SNP workshop group. 2004. SNPs in ecology, evolution and conversation. **Trends Ecol. Evol.** 19(4): 208-216.

Narain, J.P. 2008. Malaria in South-East Asia Region: Myth & the reality. **Indian J. Med. Res.** 128: 1-3.

National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services. 2009. **Life cycle of the malaria Parasite**. Malaria Available Source: http://www3.niaid.nih.gov/topics/Malaria/ lifecycle.htm, May 2, 2009.

Nei, M. 1987. **Molecular Evolutionary Genetics**. Columbia University Press, New York.

Nei, M and S. Kumar. 2000. **Molecular Evolution and Phylogenetics**. Oxford University Press, Inc., New York.

Novelli, V.M., M.A. Takita and M.A. Machado. 2004. Identification and analysis of single nucleotide polymorphisms (SNPs) in citrus. **Euphytica** 138: 227-237.

O'Loughlin, S.M., T. Okabayashi, M. Honda, Y. Kitazoe, H. Kishino, P. Somboon, T. Sochantha, S. Nambanya, P.K. Saikia, V. Dev and C. Walton. 2008. Complex population history of two *Anopheles dirus* mosquito species in Southeast Asia suggests the influence of Pleistocene climate change rather than human-mediated effects. **J. Evol. Biol.** 21(6): 1555-1569.

Petkovski, E., C. Keyser, B. Ludes and R. Hienne. 2003. Validation of SNPs markers for individual identification. **Int. Congress Series** 1239: 33-36.

Phuc, H.K., A.J. Ball, L. Son, N.V. Hanh, N.D. Tu, N.G. Lien, A. Verardi and H. Townson. 2003. Multiplex PCR assay for malaria vector *Anopheles minimus* and four related species in the Myzomyia Series from Southeast Asia. **Med. Vet. Entomol.** 17: 423-428.

Poinar, G.O.Jr., T.J. ZavorTink, T. Pike and P.A. Johnston. 2000. *Paleoculicis minutus* (Diptera: Culicidae) n. gen., n. sp., from Cretaceous Canadian amber, with a summary of described fossil mosquitoes. **Acta. Geol. Hisp.** 35: 119-128.

Polzin, T. and S.V. Daneschmand. 2003. On Steiner trees and minimum spanning trees in hypergraphs. **Oper. Res. Lett.** 31: 12-20.

Poolprasert, P., S. Manguin, M.J. Bangs, S. Sukhontabhirom, S. Poolsomboon, P. Akaratanakul and T. Chareonviriyaphap. 2008. Genetic structure and gene flow of *Anopheles minimus* and *Anopheles harrisoni* in Kanchanaburi Province, Thailand. **J. Vector Ecol.** 33(1): 158-165.

Pouniotis, D.S., O. Proudfoot, G. Minigo, J.L. Hanley and M. Plebanski. 2004. Malaria parasite interactions with the human host. **J. Postgrad. Med.** 50(1): 30-34.

Rafalski, A. 2002. Application of single nucleotide polymorphisms in crop genetics. **Curr. Opin. Plant Biol.** 5: 94-100.

Rattanarithikul, R., B.A. Harrison, R.E. Harbach, P. Panthusiri and R.E. Coleman. 2006. Illustrated keys to the mosquitoes of Thailand IV. *Anopheles*. **SE Asian J. Trop. Med. Publ. Hlth.** 37 (Supplement 2): 1-128.

Rich, S.M. and F.J. Ayala. 2006. Evolutionary origins of human malaria parasites, pp. 125-146. *In* K.R. Dronamraju and P. Arese, eds. **Malaria: Genetic and Evolutionary Aspects**. Springer, New York.

Rongnoparut, P., N. Sirichotpakorn, R. Rattanarithikul, S. Yaicharoen and K.J. Linthicum. 1999. Estimates of gene flow among *Anopheles maculatus* populations in Thailand using microsatellite analysis. **Am. J. Trop. Med. Hyg.** 60(3): 508-515.

Rwegoshora, R.T., R.G. Sharpe, K.J. Baisley and P. Kittayapong. 2002. Biting behavior and seasonal variation in the abundance of *Anopheles minimus* species A and C in Thailand. **Southeast Asian J. Trop. Med. Public Health** 33: 694-701.

Saccone, C., C. Lanave and A. De Grassi. 2006. Metazoan OXPHOS gene families: evolutionary forces at the level of mitochondrial and nuclear genomes. **Biochim. Biophys. Acta.** 1757: 1171-1178.

Sallum, M.A.M., T.R. Schultz, P.G. Foster, K. Aronstein, R.A. Wirtz and R.C. Wilkerson. 2002. Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. **Syst. Entomol.** 27: 361–382.

Sallum, M.A.M., T.R. Schultz and R.C. Wilkerson. 2000. Phylogeny of Anophelinae (Diptera Culicidae) based on morphological characters. **Ann. Entomol. Soc. Am.** 93: 745–775.

Sanofi-aventis. 2008. **Dispersion and Oviposition: Life-Cycle of *Anopheles***. Impact Malaria. Available Source: http://en.impact-malaria.com/iml/cx/en/layout.jsp?cnt=0A0E4961-70BD-450A-963A-3D349756C4D7, August 20, 2011.

Schwarz, G. 1978. Estimating the Dimension of a Model. **Ann. Statist.** 6(2): 461-464.

Seddon, J.M., H.G. Parker, E.A. Ostrander and H. Ellegren. 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. **Mol. Ecol.** 14: 503-511.

Sharpe, R.G., M.M. Hims, R.E. Harbach and R.K. Butlin. 1999. PCR-based methods for identification of species of the *Anopheles minimus* group: allele-specific amplification and single-strand conformation polymorphism. **Med. Vet. Entomol.** 13: 265-273.

Sharpe, R.G., R. Harbach and R.K. Butlin. 2000. Molecular variation and phylogeny of Member of the Minimus Group of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). **Syst. Entomol.** 25: 263-272.

Silvie, O., M.M. Mota, K. Matuschewski and M. Prudêncio. 2008. Interaction of the malaria parasite and its mammalian host. **Curr. Opin. Microbiol.** 11: 352-359.

Smith, K. 2002. **Genetic Polymorphism and SNPs**. Genetic Polymorphism and SNPs. Available Source: www.cs.mcgill.ca/~kaleigh/compbio/snp/snp_summary.html, January 6, 2009.

Steel, M.A., P. Lockhart and D. Penny. 1993. Confidence in evolutionary trees from biological sequence data. **Nature** 364: 440-442.

Stump, A.D., M.C. Fitzpatrick, N.F. Lobo, S. Traore´, N. Sagnon, C. Costantini, F.H. Collins and N.J. Besansky. 2005. Centromere-proximal differentiation and speciation in *Anopheles gambiae*. **PNAS.** 102 (44): 15930–15935.

Sungvornyothin, S., C. Garros, T. Chareonviriyaphap and S. Manguin. 2006b. How reliable is the humeral pale spot for identification of cryptic species of the Minimus Complex?. **J. Am. Mosq. Control Assoc.** 22(2):185–191.

Sungvornyothin, S., V. Muenvorn, C. Garros, S. Manguin, A. Prabaripai, M.J. Bangs and T. Chareonviriyaphap. 2006a. Tropical behavior and biting activity of the two sibling species of the *Anopeles minimus* complex in western Thailand. **J. Vector Ecol.** 31(2): 252-261.

Sumruayphol, S., C. Apiwathnasorn, N. Komalamisra, J. Ruangsittichai, Y. Samung and P. Chavalitshewinkoon-Petmitr. 2010. Bionomic status of *Anopheles epiroticus* Linton & Harbach, a coastal malaria vector, in Rayong Province, Thailand. **SE. Asian J. Trop. Med.** 41(3): 541-547.

Swain, S., A. Mohanty, H.K. Tripathy, N. Mahapatra, S.K. Kar and R.K. Hazra. 2010. Molecular identification and phylogeny of Myzomyia and Neocellia series of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). **Infect. Genet. Evol.** 10(7): 931-939.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics** 123: 585-595.

Tajima, F. 1993. Simple methods for testing molecular evolutionary clock hypothesis. **Genetics** 135: 599-607.

The office of Diseases Prevention and Control, Department of Disease Control, Ministry of Public Health. 2008. **Information of Malaria**. Vector Borne Diseases. Available Source: http://www.dpc1.in.th/insect/index.php?option=com_content&task=view&id=14&Itemid=28, February 7, 2011.

The University of Michigan. n.d. **Interpretation of Sequencing Chromatograms**. DNA Sequencing Core. Available Source: http://seqcore.brcf.med.umich.edu/doc/dnaseq/interpret.html, August 14, 2011.

Thompson, J.D., D.G. Higgins and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res.** 22: 4673–4680.

Trung, H.D., W. Van Bortel, T. Sochantha, K. Keokenchanh, N.T. Quang, L.D. Cong and M. Coosemans. 2004. Malaria transmission and major malaria vectors in different geographical areas of Southeast Asia. **Trop. Med. Int. Health** 9(2): 230-237.

Tsuda, Y., M. Takagi, T. Toma, A. Sugiyama and I. Miyagi. 1999. Mark-release-recapture experiment with adult *Anopheles minimus* (Diptera: Culicidae) on Ishigaki Island, Ryukyu Archipelago, Japan. **J. Med. Entomol.** 36(5): 601-604.

Valgimigli, M.C. 2004. **Genetic variability in Italian populations of *Cupressus sempervirens* L. assessed by SSR and RAPD markers**. Ph.D. Thesis, Università degli Studi di Verona.

Van Bortel, W., H.D. Trung, N.D. Manh, P. Roelants, P. Verlé and M. Coosemans. 1999. Identification of two species within the *Anopheles minimus* complex in northern Vietnam and their behavioural divergences. **Trop. Med. Int. Health** 4(4): 257-265.

Van Bortel, W., H.D. Trung, P. Roelants, R.E. Harbach, T. Backeljau and M. Coosemans. 2000. Molecular identification of *Anopheles minimus s.l.* beyond distinguishing the members of the species complex. **Insect Mol. Biol.** 9(3): 335-340.

Van Bortel, W., H.D. Trung, P.Roelants, T. Backeljau and M. Coosemans. 2003. Population genetic structure of the malaria vector *Anopheles minimus* A in Vietnam. **Heredity** 91: 487-493.

Vignal, A., D. Milan, M.S. Christobal and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. **Genet. Sel. Evol.** 34: 275-305.

Walton, C., P. Somboon, S. M. O'Loughlin, S. Zhang, R. E. Harbach, Y. M. Linton, B. Chen, K. Nolan, S. Duong, M. Y. Fong, I. Vythilingum, Z. D. Mohammed, Ho Dinh Trung and R. K. Butlin. 2007. Genetic diversity and molecular identification of mosquito species in the *Anopheles maculatus* group using the ITS2 region of rDNA. **Infec. Genet. Evol.** 7(1): 92-102.

Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. **Theor. Popul. Biol.** 7: 256-276.

World Health Organization. 2007a. **What is Malaria?**. Malaria. Available Source: http://www.searo.who.int/ EN/Section10/Section21/Section334.htm, January 6, 2009.

World Health Organization. 2007b. **Anopheline Species Complexes in South and South-East Asia**. WHO Regional Office for South-East Asia, New Delhi.

World Health Organization. 2008. **World malaria report 2008**. WHO Press, Switzerland.

World Health Organization. 2009. **World malaria report 2009**. WHO Press, Switzerland.

World Health Organization. 2010. **Malaria in the greater mekong subregion: Regional and country profiles**. WHO publications, India.

Xia, X. and P. Lemey. 2009. Assessing substitution saturation with DAMBE, pp. 615-630. *In* P. Lemey, M. Salemi and A.-M. Vandamme, eds. **The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny**. Cambridge University Press.

Xia, X. and Z. Xie. 2001. DAMBE: software package for data analysis in molecular biology and evolution. **J. Hered.** 92(4): 371-373.

Xia, X., Z. Xie, M. Salemi, L. Chen and Y. Wang. 2003. An index of substitution saturation and its application. **Mol. Phylogenet. Evol.** 26(1): 1-7.

Xu, J., H. Guo and Z.L. Yang. 2007. Single nucleotide polymorphisms in the ectomycorrhizal mushroom *Tricholoma matsutake*. **Microbiology** 153: 2002-2012.

Yang, Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. **Mol. Biol. Evol.** 24: 1586-1591.

Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young and P.B. Cregan. 2003. Single-Nucleotide Polymorphism in soybean. **Genetics** 163: 1123-1134.

APPENDICES

Appendix A

Genomic DNA Mini Kit (Tissue) Protocol (modification)

1. Transfer mosquito to a 1.5 ml microcentrifuge tube.
2. Add 20 µl of GT Buffer to the tube and homogenize the sample tissue by using the micropestle to grind the tissue to a pulp.
3. Add 80 µl of GT Buffer to the tube and continue to homogenize the sample tissue by grinding.
4. Add 100 µl of GT Buffer to the tube for washing the micropestle.
5. Add 20 µl of Proteinase K to the sample mixture and mix by vortex.
6. Incubate at 60ºC for 40 minutes to lyse the sample and invert the tube every 5 minutes during incubation.
7. Add 200 µl of GBT Buffer and mix by vortex for 5 seconds.
8. Incubate at 70ºC for 40 minutes and invert the tube every 5 minutes during incubation.
9. For DNA elution step, preheat the Elution Buffer (50 µl per sample) to 70ºC.
10. Add 200 µl of absolute ethanol to the sample and vortex immediately for 10 seconds.
11. Place a GD Column in a 2 ml Collection Tube and transfer all of the mixture to the GD Column.
12. Centrifuge at 12,000 rpm for 2 minutes.
13. Discard the 2 ml Collection Tube containing the flow-through and transfer the GD Column to a new 2 ml Collection Tube.
14. Add 400 µl of W1 Buffer to the GD Column.
15. Centrifuge at 12,000 rpm for 1 minutes.
16. Discard the flow-through and place the GD Column back in the 2 ml Collection Tube.
17. Add 600 µl of Wash Buffer (ethanol added) to the GD Column.
18. Centrifuge at 12,000 rpm for 1 minutes.
19. Discard the flow-through and place the GD Column back in the 2 ml Collection Tube.
20. For drying the column matrix, centrifuge again for 3 minutes at 12,000 rpm.
21. Transfer the dried GD Column to a new 1.5 ml microcentrifuge tube.

22. For first elution, add 20 µl of preheated Elution Buffer to the center of the column matrix and let stand for 10 minutes.
23. Centrifuge at 12,000 rpm for 1 minutes to elute the purified DNA.
24. For second elution, add 30 µl of preheated Elution Buffer to the center of the column matrix and let stand for 10 minutes.
25. Centrifuge at 12,000 rpm for 1 minutes to elute the purified DNA.

Appendix B

Amplification with multiplex allele-specific polymerase chain reaction (AS-PCR)

PCR mixture:

0.5	µl	DNA template
0.8	µl	10x PCR buffer ($MgCl_2$ free)
0.4	µl	25mM $MgCl_2$
0.4	µl	8mM dNTPs (2 mM each)
0.4	µl	ITS2A primer (10 µM)
0.4	µl	MIA primer (10 µM)
0.4	µl	MIC primer (10 µM)
0.4	µl	ACO primer (10 µM)
0.4	µl	PAM primer (10 µM)
0.4	µl	VAR primer (10 µM)
0.1	µl	*Taq* DNA polymerase (5 U/µl)
5.4	µl	Sterilized distilled water
10.0	**µl**	**Total volume**

PCR programme:

Step 1: 94ºC	2	min
Step 2: 94ºC	30	sec
Step 3: 50ºC	30	sec
Step 4: 72ºC	40	sec
Repeat step 2 to 4 for 35 cycles		
Step 5: 72ºC	10	min

Appendix C

PCR Amplification of interested fragments

Reaction mixture:

1.0	µl	DNA template
2.5	µl	10x PCR buffer ($MgCl_2$ free)
1.0	µl	25mM $MgCl_2$
1.5	µl	8mM dNTPs (2 mM each)
1.25	µl	Forward primer (10 µM)
1.25	µl	Reverse primer (10 µM)
0.1	µl	DNA polymerase (5 U/µl)
21.4	µl	Sterilized distilled water
30.0	**µl**	**Total volume**

PCR programme:

Step 1: 95ºC	5	min
Step 2: 95ºC	30	sec
Step 3: 50-60ºC	40	sec
Step 4: 72ºC	1	min
Repeat step 2 to 4 for 30 cycles		
Step 5: 72ºC	8	min

Appendix D

Reaction mix and condition require for PCR product purification

Reaction mix:

1.4	µl	Exonuclease I (20 U/µl)
5.6	µl	Themosensitive Alkaline Phosphatase (1 U/µl)
7.0	**µl**	**Total volume**
27	µl	PCR product

The reaction mixtures were mixed and incubated at 37ºC for 15 min. Then, the enzymes were inactivated at 80ºC for 15 min.

Appendix E

BLAST results

Part A: Genetic diversity and tests of population genetic models

Fragments	Accession	Max score	E value	Max ident
ITS2	DQ665848.1	821	0.0	100%
O8*	-	-	-	-
O10	DQ102065.1	170	9e-39	79%
P9	DQ102065.1	170	9e-39	79%
P12	JF264228.1	545	5e-152	99%
P22	XM_310912.5	187	7e-44	82%

Note: * could not be obtained from BLAST.

Part B: Phylogenetic inference with multilocus data

Fragments	Accession	Max score	E value	Max ident
TOLL6	XM_320172.2	673	0.0	84%
P9	DQ102065.1	170	9e-39	79%
ITS2	GQ500119.1	913	0.0	100%
COI	DQ267690.1	473	3e-130	93%
COII	AF417748.1	1076	0.0	99%

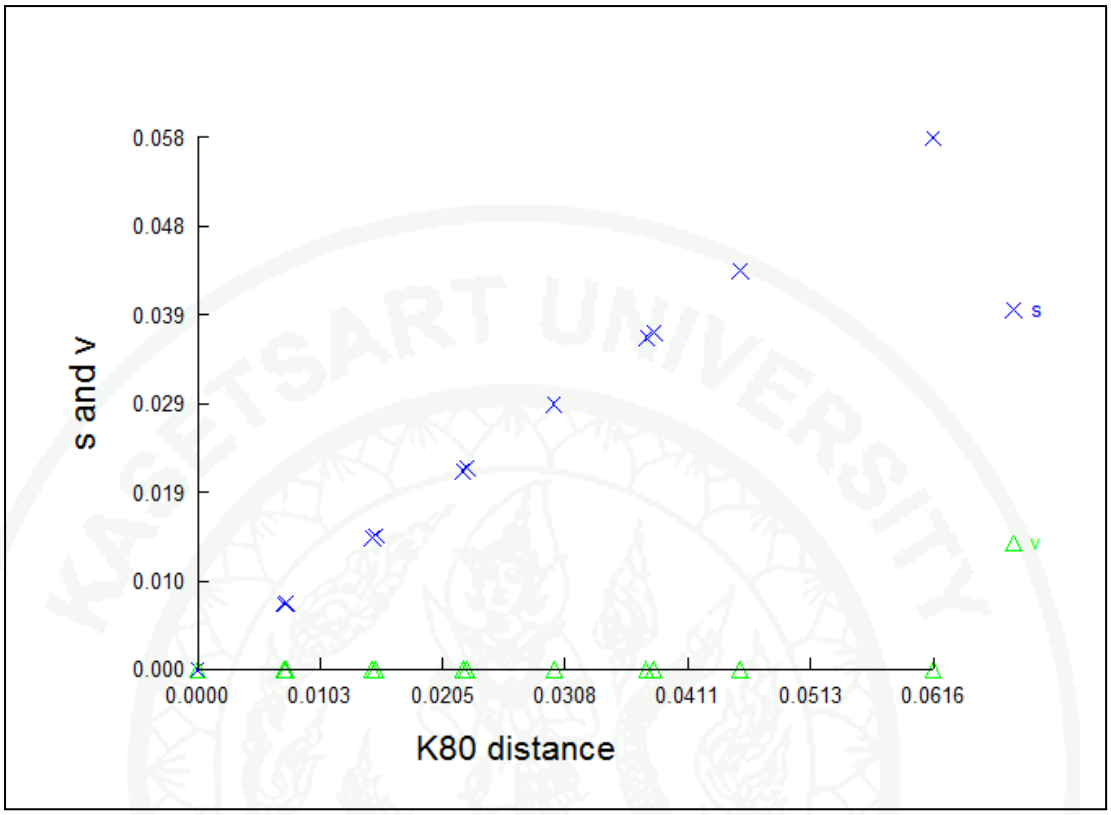
Appendix F

Test of substitution saturation and nucleotide substitution pattern

Test of substitution saturation and nucleotide substitution pattern of different portions of datasets were plotted against evolutionary distances of each fragments

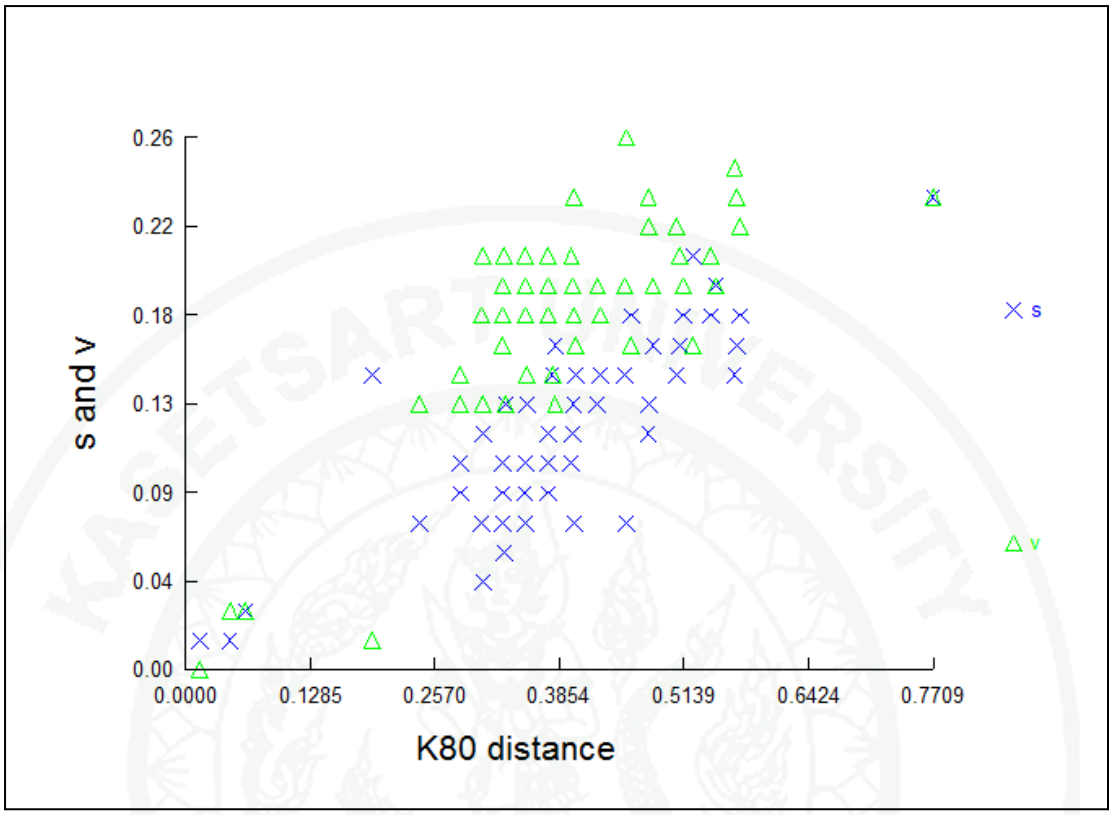
Analysis of cytochrome c oxidase I (COI) fragment

The first and second codon positions were firstly observed. For Xia's test, the proportion of invariant sites P_{inv} can be set at default value (zero) because no invariable sites are observed. The observed I_{ss} value of 0.0251 is significantly smaller than the $I_{ss.c}$ value of 0.6823 for the symmetrical topology and 0.5557 for the asymmetrical topology, assuming the sequences are observed little substitution saturation. For Steel's test, the mean ϕ values range from 0.2226 to 0.2738. *An. maculatus* has the smallest mean ϕ value and no tests that fail to reject the null model. Both transitions and transversions were plotted against evolutionary distances based on Kimura 2-parameter model (Appendix Figure 31). This result showed transversions occur less frequently than transitions, and substitutions increased with increasing evolutionary distance, representing rejection of substitution saturation.



Appendix Figure 1 Transition (s) and transversion (v) substitutions of the first and second codon positions of COI fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).

The third codon position of the COI sequences was then observed. For Xia's test, the proportion of invariant sites (P_{inv}) was estimated equal 0.01718. The resulting I_{ss} = 0.5019, significantly smaller than the $I_{ss.c}$ value of 0.7553 for the symmetrical topology and the $I_{ss.c}$ value of 0.7295 for the asymmetrical topology, assuming these sequences are little substitution saturation. Steel's test, the mean ϕ values range from 0.1317 to 0.1989. The mean ϕ value of *An. varuna* shows the smallest and no tests that fail to reject the null hypothesis. In Appendix Figure 2, both transitions and transversions were plotted against evolutionary distances based on Kimura 2-parameter model and increased with increasing evolutionary distance, no saturation of substitution is recognized.

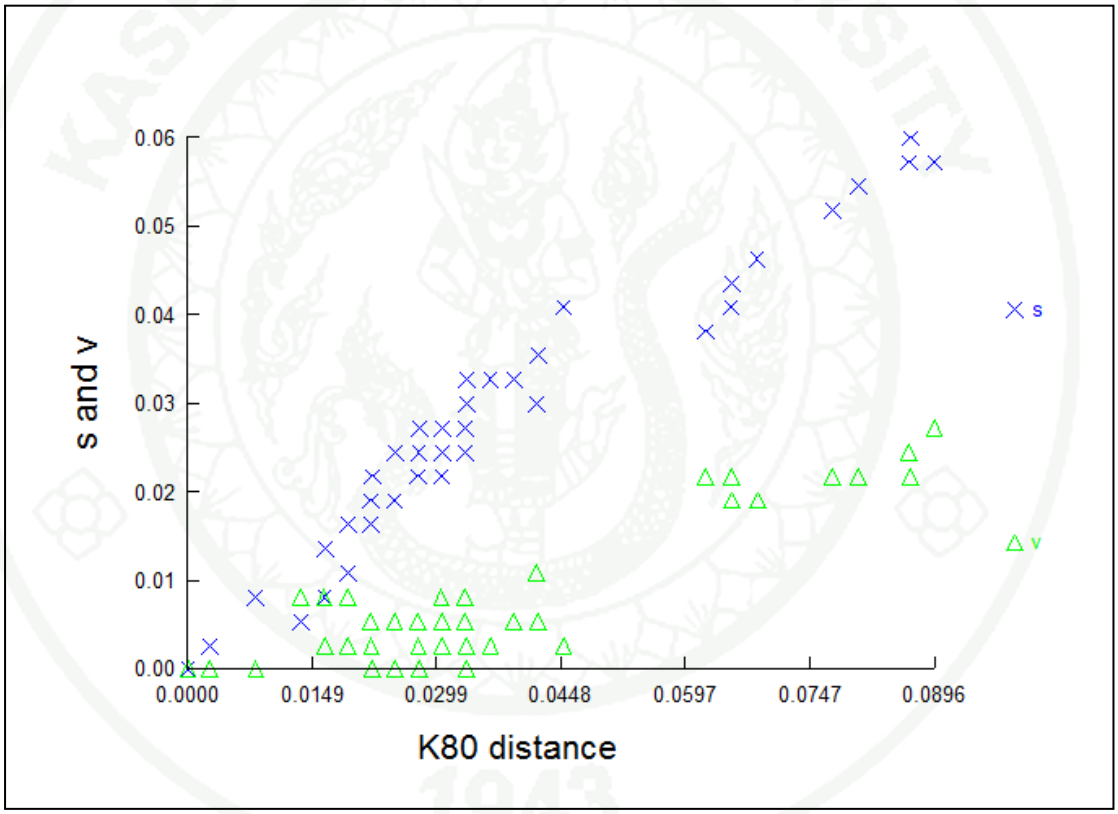


Appendix Figure 2 Transition (s) and transversion (v) substitutions of the third codon positions of COI fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).

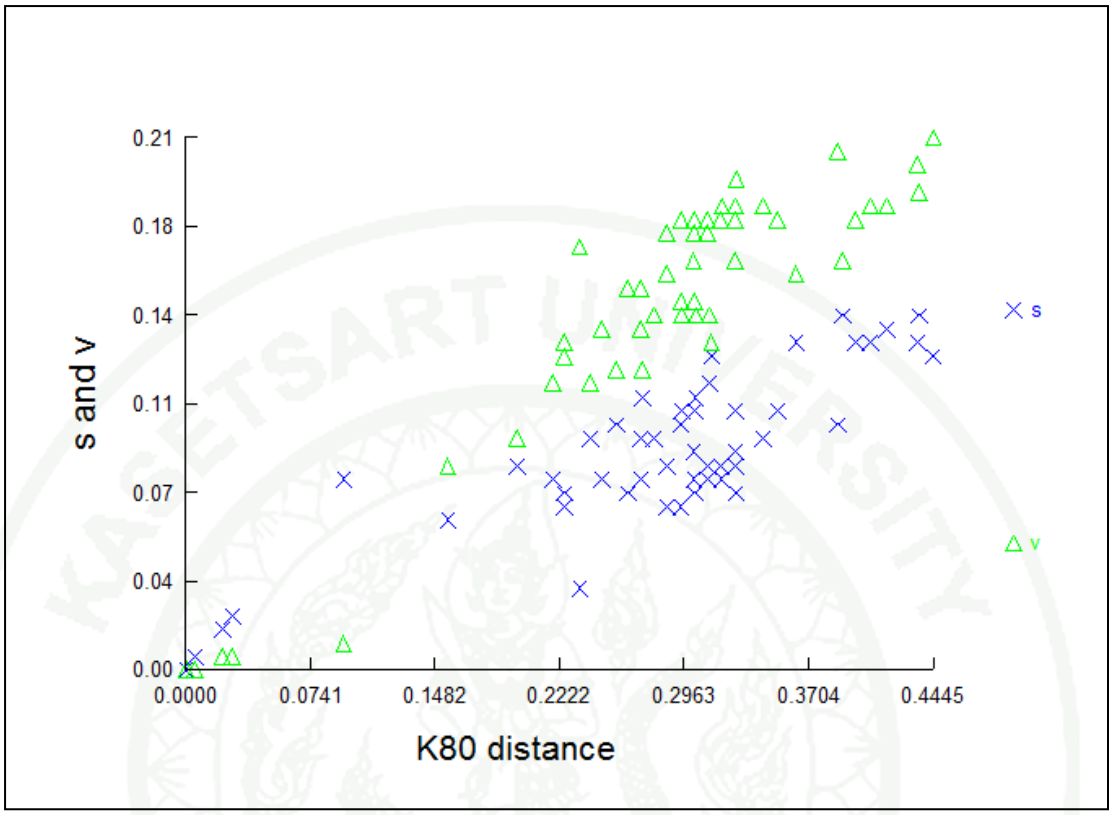
Analysis of cytochrome c oxidase II (COII) fragment

The first and second codon positions were observed. For Xia's test, the proportion of invariant sites (P_{inv}) was estimated equal 0.00116. The observed I_{ss} value of 0.0417 is significantly smaller than the $I_{ss.c}$ value of 0.7035 for the symmetrical topology and 0.5371 for the asymmetrical topology, assuming the sequences are observed little substitution saturation. Steel's test, the mean ϕ values range from 0.2052 to 0.2436. *An. aconitus* has the smallest mean ϕ value and no tests that fail to reject the null model. The scatter plot between substitutions and evolutionary distances are shown in Appendix Figure 3. The third codon position of the COII sequences was observed. Xia's test, the P_{inv} was estimated equal 0.28819. The resulting I_{ss} = 0.5881, significantly smaller than the $I_{ss.c}$ value of 0.6790 for the

symmetrical topology and the $I_{ss.c}$ value of 0.5310 for the asymmetrical topology, assuming these sequences are little substitution saturation. Steel's test, the mean ϕ values range from 0.1160 to 0.2025. The mean ϕ value of *Aedes aegypti* (an outgroup) shows the smallest and the largest number of tests that fail to reject the null hypothesis $N_{insignificant}$ = 115, suggesting it might be too divergent from the others. The scatter plot between substitutions and evolutionary distances are also shown in Appendix Figure 4.



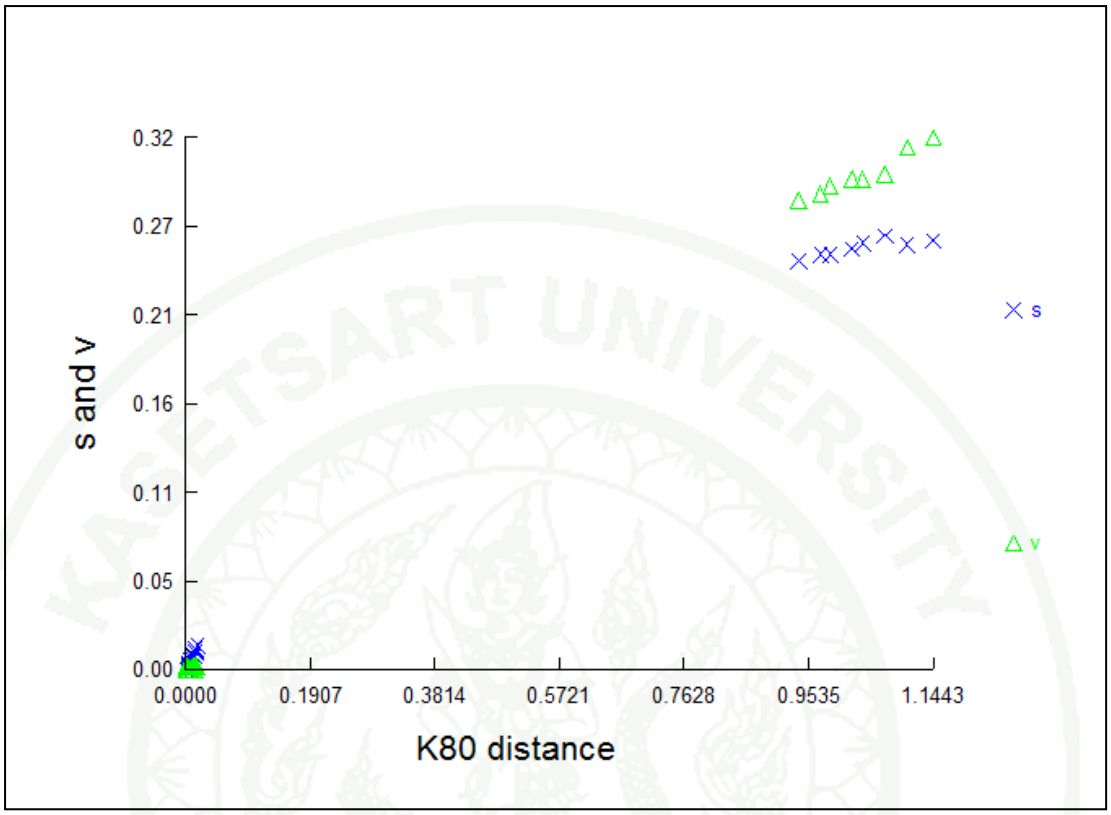
Appendix Figure 3 Transition (s) and transversion (v) substitutions of the first and second codon positions of COII fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).



Appendix Figure 4 Transition (s) and transversion (v) substitutions of the third codon positions of COII fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).

Analysis of O8 fragment

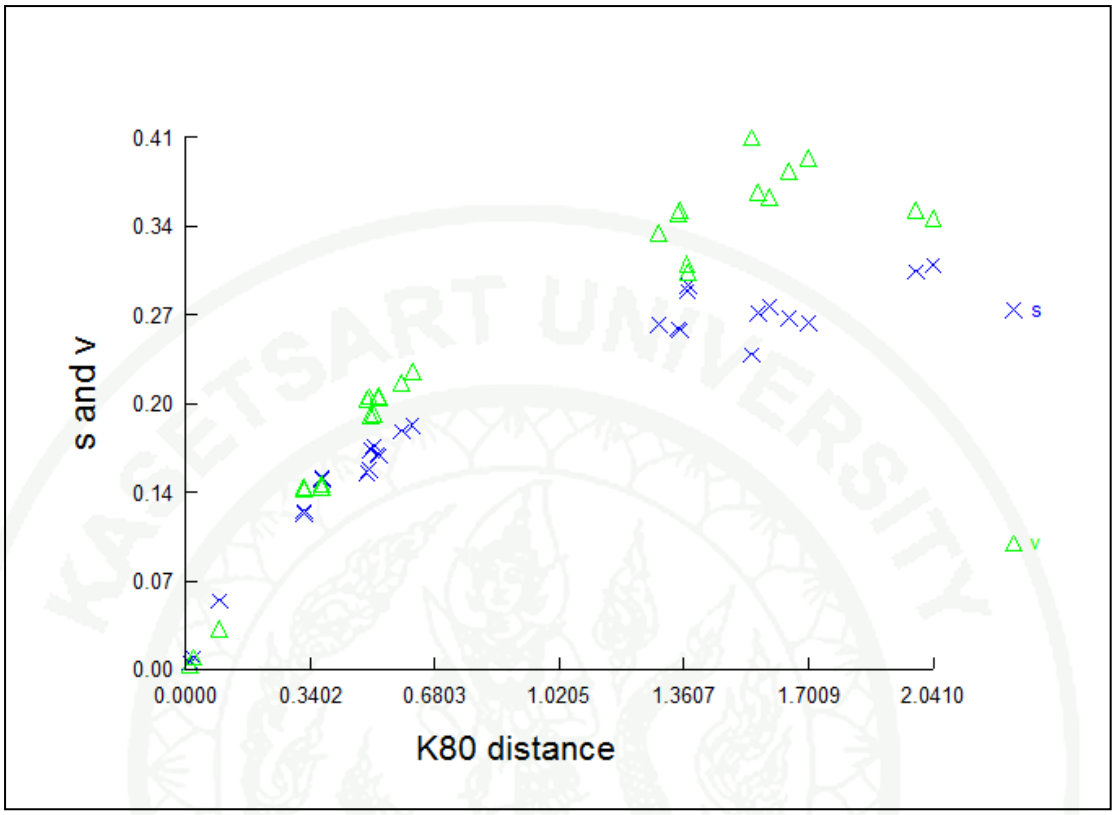
For Xia's test, the proportion of invariant sites (P_{inv}) is 0.00174. The resulting I_{ss} = 0.1725, much smaller than $I_{ss.c}$ (= 0.7273 assuming a symmetrical topology and 0.6012), indicating little substitution saturation. Steel's test, the mean ϕ values range from 0.2795 to 0.3496. *An. maculatus* has the smallest mean ϕ value and the largest number of tests that fail to reject the null hypothesis ($N_{insignificant}$ = 6). The test showed the failure to reject the null hypothesis is due to lack of sequence variation, not due to substitution saturation. The scatter plot between substitutions and evolutionary distances are shown in Appendix Figure 5.



Appendix Figure 5 Transition (s) and transversion (v) substitutions of O8 fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).

Analysis of P9 fragment

For Xia's test, the proportion of invariant sites (P_{inv}) was estimated equal 0.00436. The observed I_{ss} value of 0.5223 is significantly smaller than the $I_{ss.c}$ value of 0.7458 for the symmetrical topology and 0.6342 for the asymmetrical topology, assuming the sequences have experienced little substitution saturation. Steel's test, the mean ϕ values range from 0.2983 to 0.3498. *An. aconitus* has the smallest mean ϕ value and the largest number of tests that fail to reject the null hypothesis ($N_{insignificant}$ = 8), suggesting it might be too divergent from the others. The scatter plot between substitutions and evolutionary distances are shown in Appendix Figure 6.

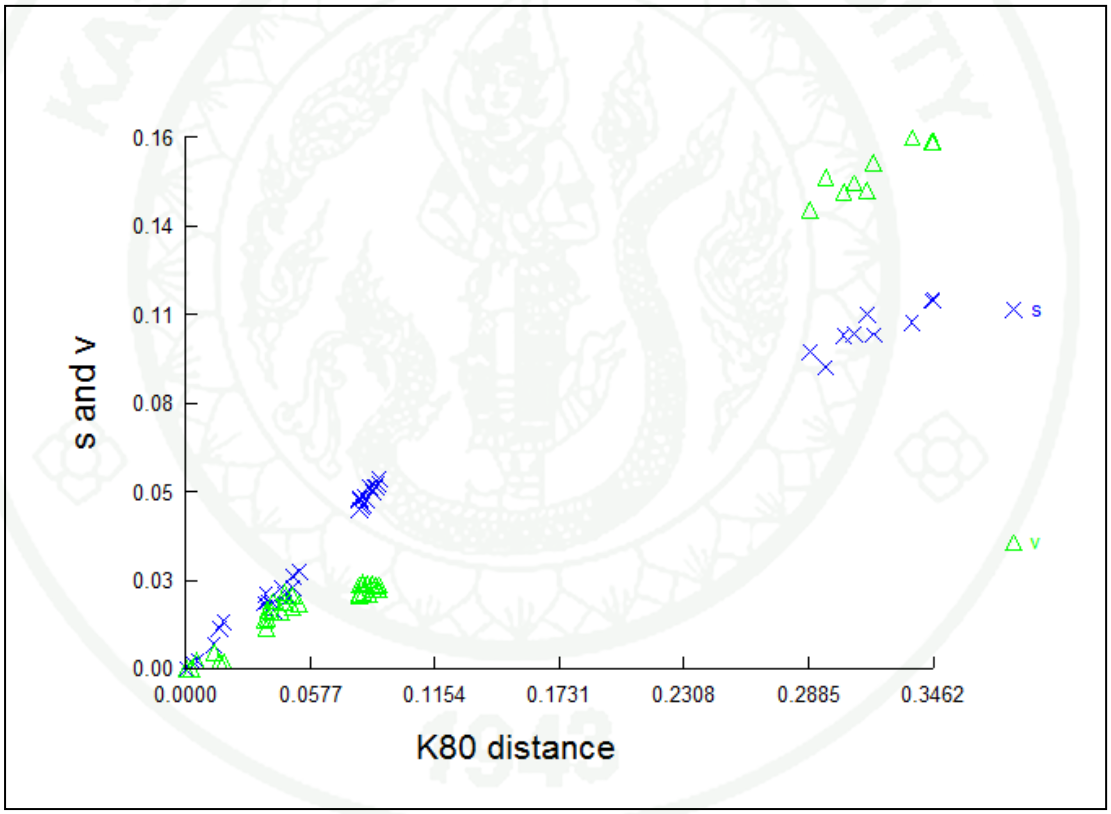


Appendix Figure 6 Transition (s) and transversion (v) substitutions of P9 fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).

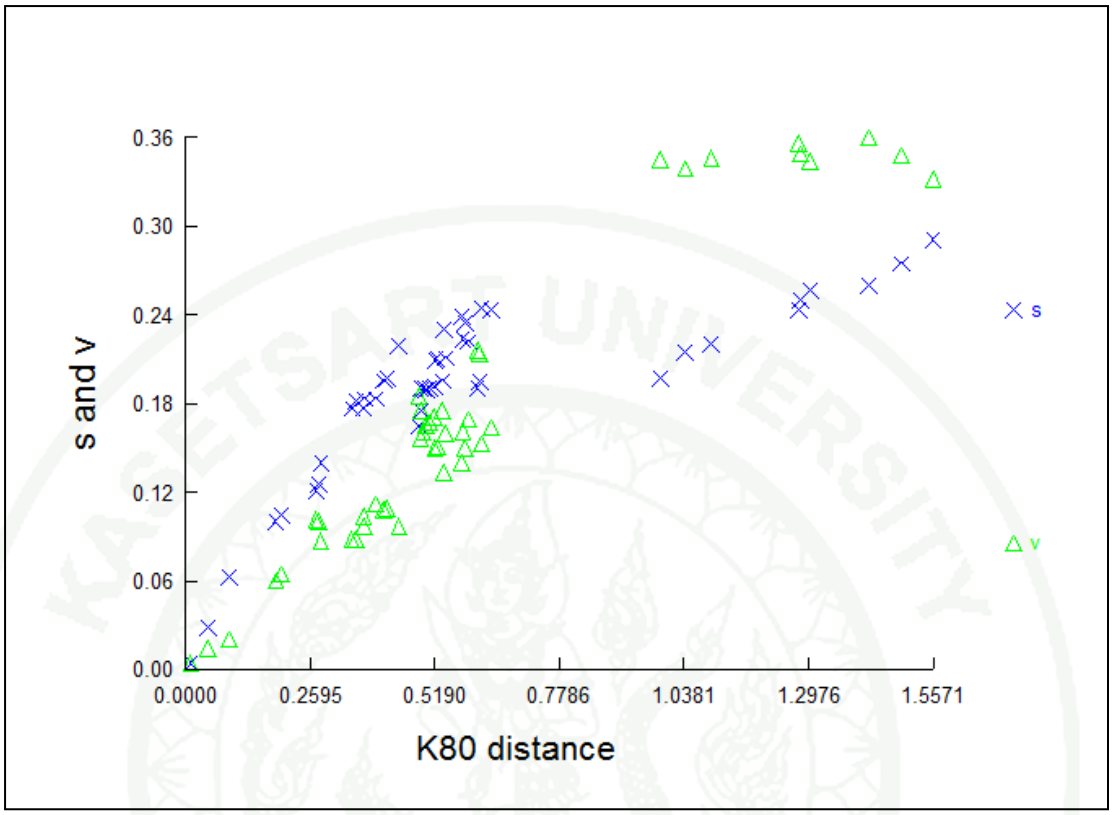
Analysis of TOLL 6 fragment

The highly conserved first and second codon positions were observed. For Xia's test, the proportion of invariant sites (P_{inv}) was estimated equal 0.39461. The observed I_{ss} value of 0.1886 is significantly smaller than the $I_{ss.c}$ value of 0.7160 for the symmetrical topology and 0.5755 for the asymmetrical topology, assuming the sequences are observed little substitution saturation. Steel's test, the mean ϕ values range from 0.2083 to 0.2628. *An. epiroticus* has the smallest mean ϕ value and the largest number of tests that fail to reject the null hypothesis ($N_{insignificant}$ = 8), suggesting it might be too divergent from the others. The scatter plot between substitutions and evolutionary distances are shown in Appendix Figure 7. Then, the highly third codon position was observed. Xia's test, the P_{inv} can be set at default value (zero) because the invariable sites are smaller. The resulting I_{ss} = 0.4843,

significantly smaller than the $I_{ss.c}$ value of 0.7048 for the symmetrical topology and not significantly smaller than the $I_{ss.c}$ value of 0.5805 (p = 0.0083, two-tailed t test) for the asymmetrical topology, assuming these sequences are not substantial saturation. Steel's test, the mean ϕ values range from 0.1518 to 0.2111. The mean ϕ value of *An. epiroticus* shows the smallest and the largest number of tests that fail to reject the null hypothesis $N_{insignificant}$ = 43, suggesting it might be too divergent from the others. The scatter plot between substitutions and evolutionary distances are shown in Appendix Figure 8.



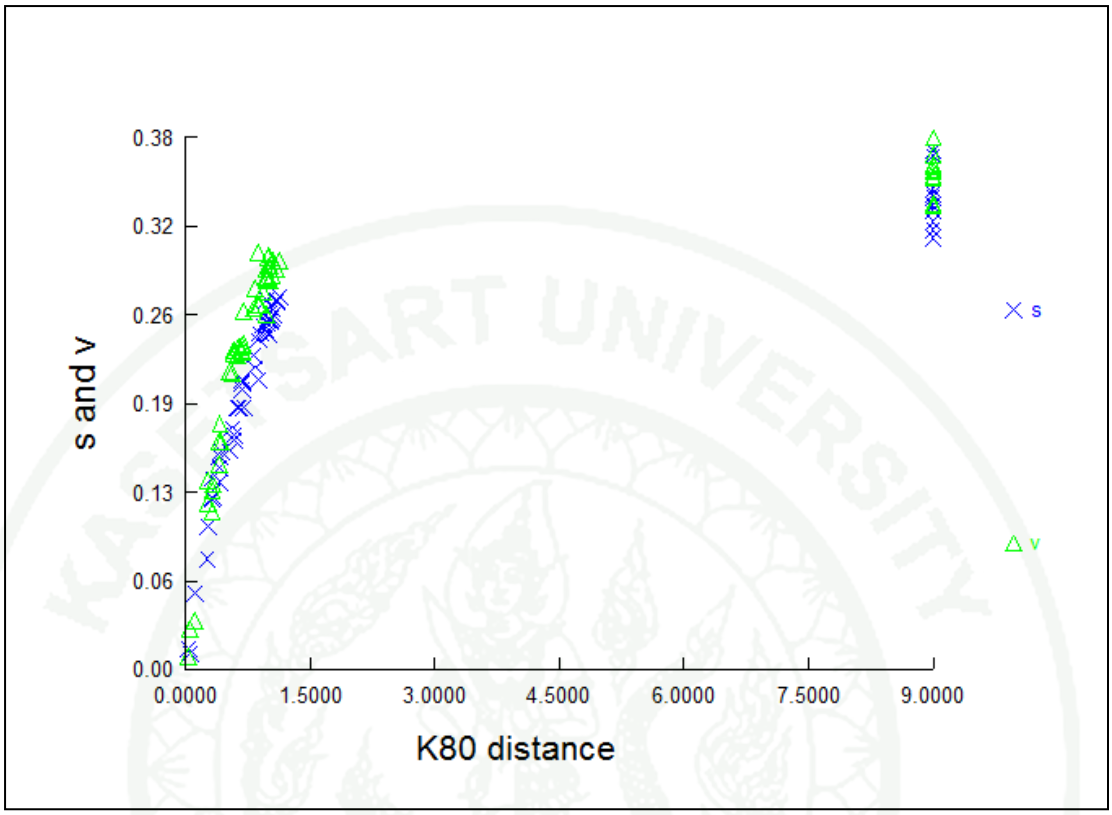
Appendix Figure 7 Transition (s) and transversion (v) substitutions of the first and second codon positions of TOLL6 fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).



Appendix Figure 8 Transition (s) and transversion (v) substitutions of the third codon positions of TOLL6 fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).

Analysis of the Internal Transcribed Spacer 2 (ITS2) fragment

For Xia's test, the proportion of invariant sites (P_{inv}) was estimated equal 0.03266. The observed I_{ss} value of 0.6156 is significantly smaller than the $I_{ss.c}$ value of 0.7095 for the symmetrical topology and substantially larger than $I_{ss.c}$ (0.5755) assuming an asymmetrical topology, indicating the sequences may still be useful if the true topology is not very asymmetrical. Steel's test, the mean ϕ values range from 0.1447 to 0.2115. *Drosophila melanogaster* (an outgroup) has the smallest mean ϕ value and *An. varuna* is involved in the largest number of tests that fail to reject the null model ($N_{insignificant}$ = 27). The scatter plot between substitutions and evolutionary distances are shown in Appendix Figure 9.



Appendix Figure 9 Transition (s) and transversion (v) substitutions of ITS2 fragment were plotted against pairwise evolutionary distances (K80 model; Kimura 2-parameter model).

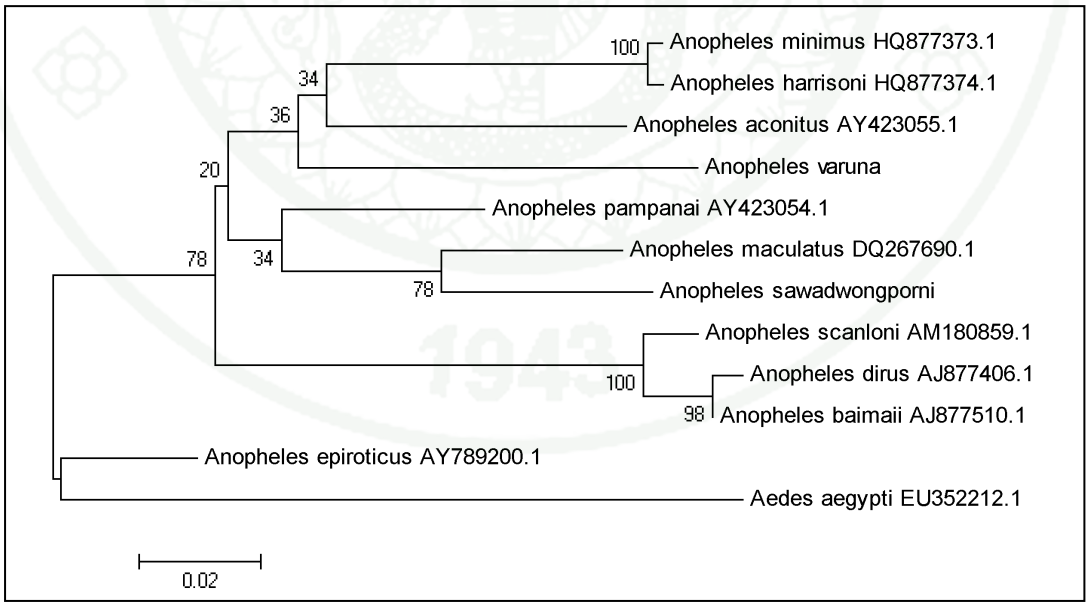
Appendix G

Phylogenetic trees with different gene fragments

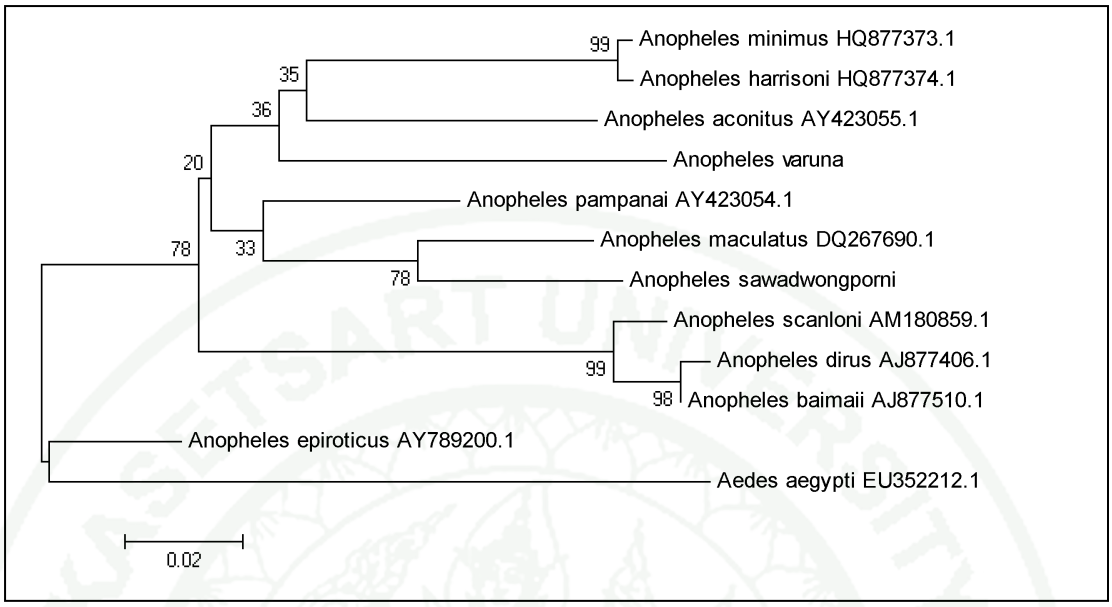
Phylogenetic analysis of cytochrome c oxidase I (COI) gene

NJ and ME analyses (Appendix Figure 10 and 11) showed a similar topology. These species were mainly divided into three clusters, i.e. clusters of Myzomyia (*An. minimus*, *An. harrisoni*, *An. aconitus* and *An. varuna*), Neocellia (*An. maculatus* and *An. sawadwongporni*) and Neomyzomyia (*An. dirus*, *An. baimaii* and *An. scanloni*) Series. *An. pampanai* was located in the Neocellia cluster with poor branch support, in spite of it should be located in Myzomyia cluster and *An. epiroticus* was located with outgroup.

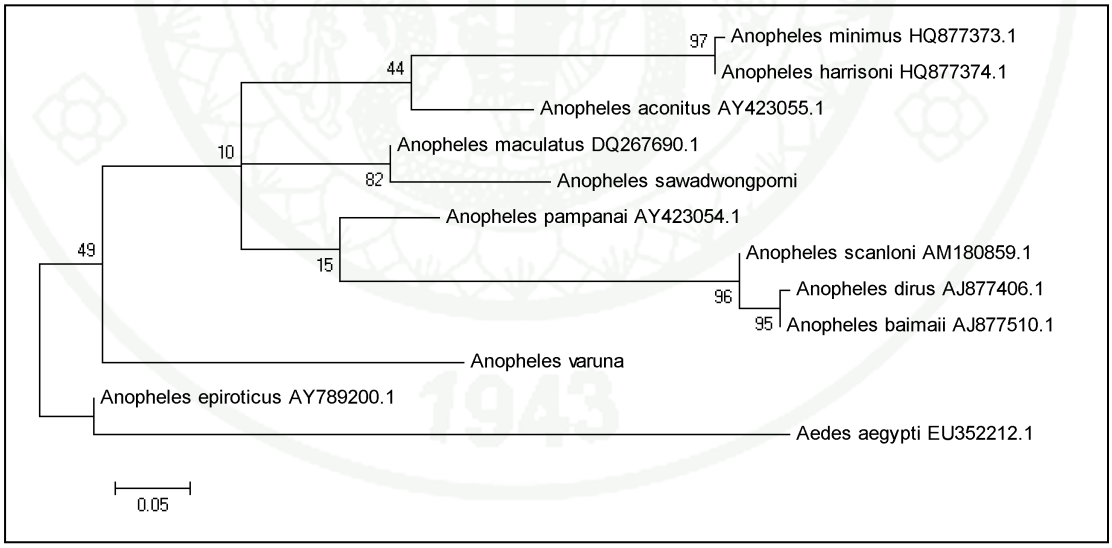
ML analysis (Appendix Figure 12) resulted in three main clusters, i.e. clusters of Myzomyia (*An. minimus*, *An. harrisoni* and *An. aconitus*), Neocellia (*An. maculatus* and *An. sawadwongporni*) and Neomyzomyia (*An. dirus*, *An. baimaii* and *An. scanloni*) Series. *An. varuna* and *An. pampanai* was separately placed from the Myzomyia cluster and *An. epiroticus* was formed a cluster with outgroup.



Appendix Figure 10 Phylogenetic tree based on the COI fragment (NJ reconstruction).



Appendix Figure 11 Phylogenetic tree based on the COI fragment (ME reconstruction).



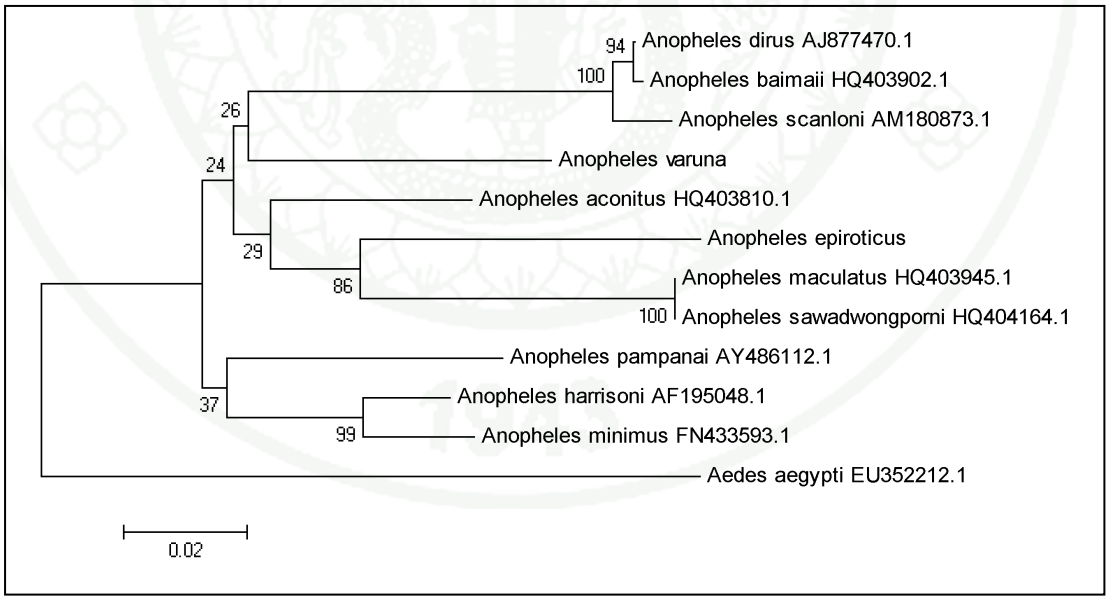
Appendix Figure 12 Phylogenetic tree based on the COI fragment (ML reconstruction).

Phylogenetic analysis of cytochrome c oxidase II (COII) gene

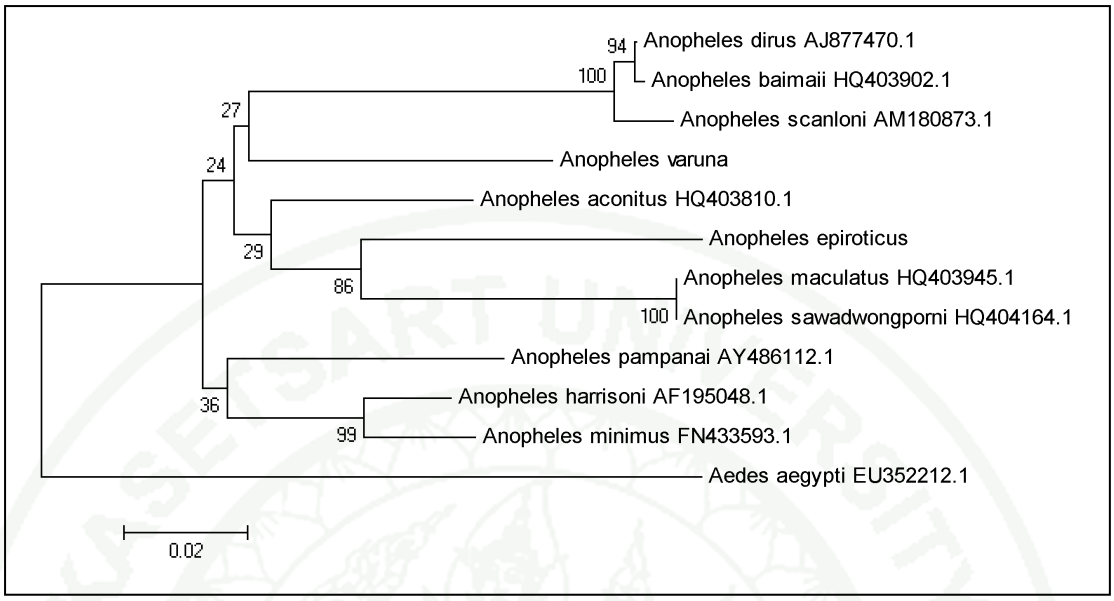
Similar topology was observed from NJ and ME analyses (Appendix Figure 13 and 14). These topologies separated three main clusters. First cluster is monophyly

of the Neomyzomyia cluster (*An. dirus*, *An. baimaii* and *An. scanloni*) relating to improper position of *An. varuna* with poor branch support. Second cluster contains the Pyretophorus (*An. epiroticus*) and Neocellia (*An. maculatus* and *An. sawadwongporni*) Series with strong branch support conforming to morphological phylogeny of Foley *et al.* (1998) and this cluster also contains improper position of *An. aconitus* with poor branch support. Third cluster is the Myzomyia cluster (*An. minimus*, *An. harrisoni* and *An. pampanai*) and basal cluster of these topologies

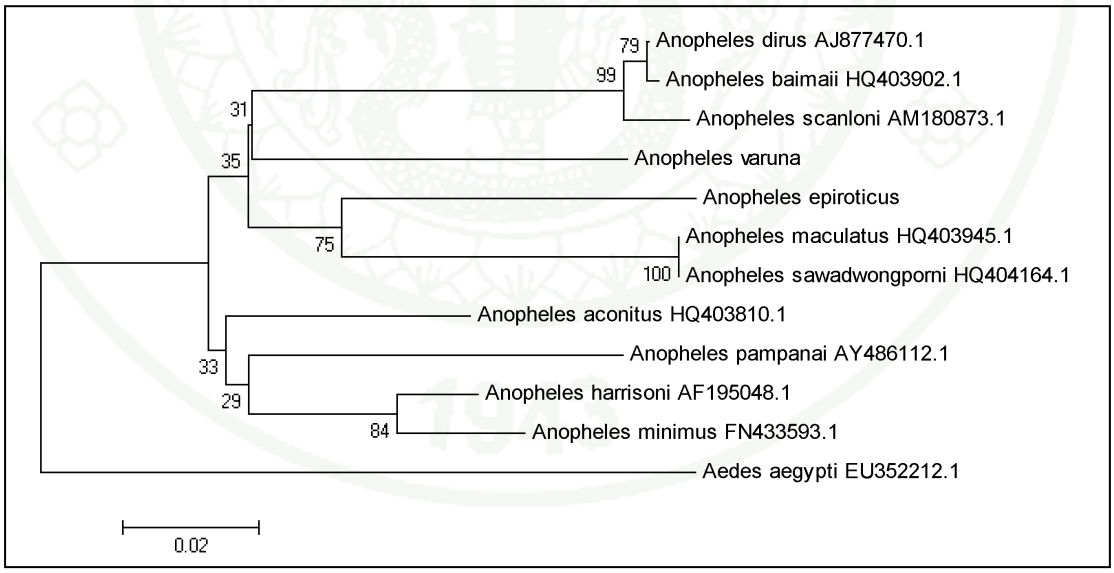
Topology of ML analysis (Appendix Figure 15) is similar to topologies of NJ and ME analyses. Three main clusters were separated. First cluster contains the Neomyzomyia cluster (*An. dirus*, *An. baimaii* and *An. scanloni*) relating to improper position of *An. varuna*. Second cluster contains the Pyretophorus (*An. epiroticus*) and Neocellia (*An. maculatus* and *An. sawadwongporni*) Series and third cluster contains the Myzomyia cluster (*An. minimus*, *An. harrisoni*, *An. pampanai* and *An. aconitus*).



Appendix Figure 13 Phylogenetic tree based on the COII fragment (NJ reconstruction).



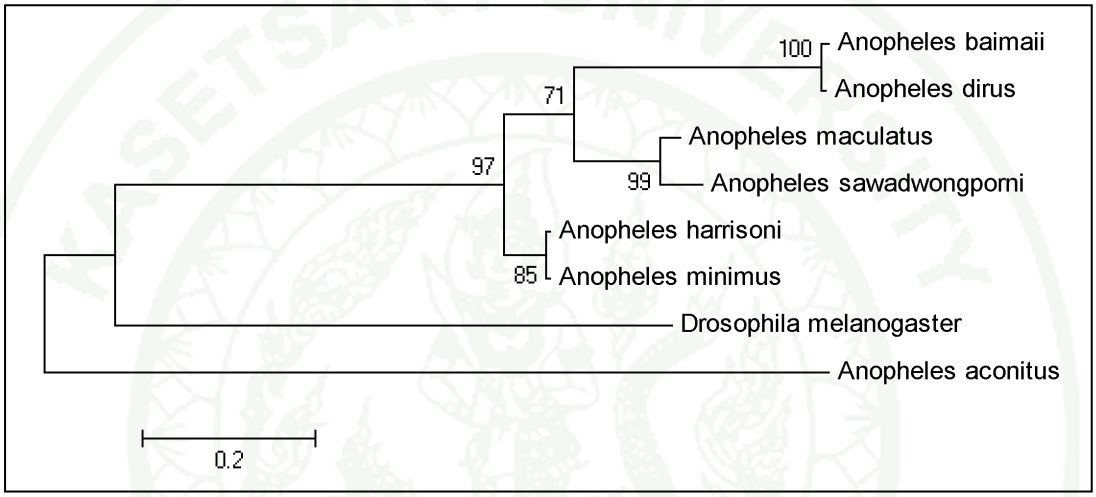
Appendix Figure 14 Phylogenetic tree based on the COII fragment (ME reconstruction).



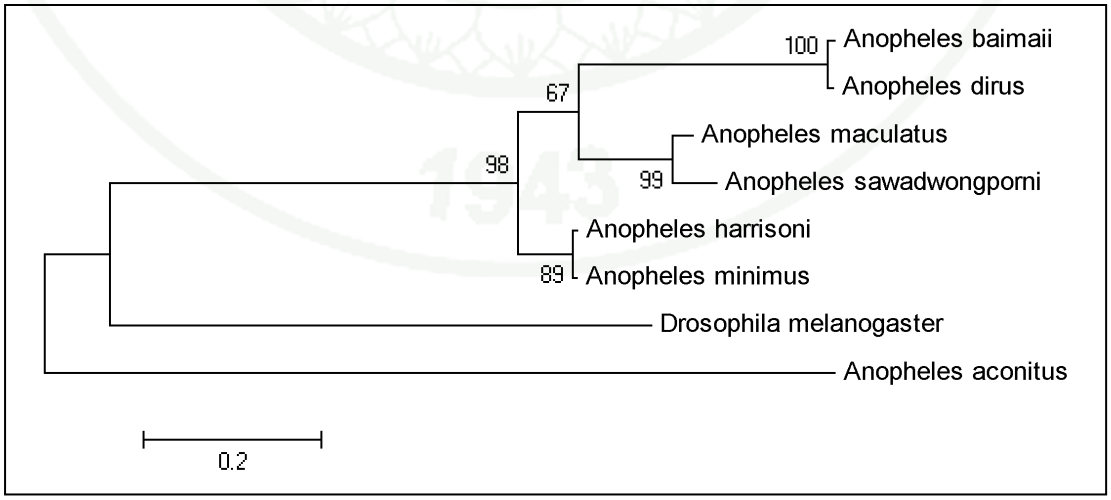
Appendix Figure 15 Phylogenetic tree based on the COII fragment (ML reconstruction).

Phylogenetic analysis of the P9 fragment

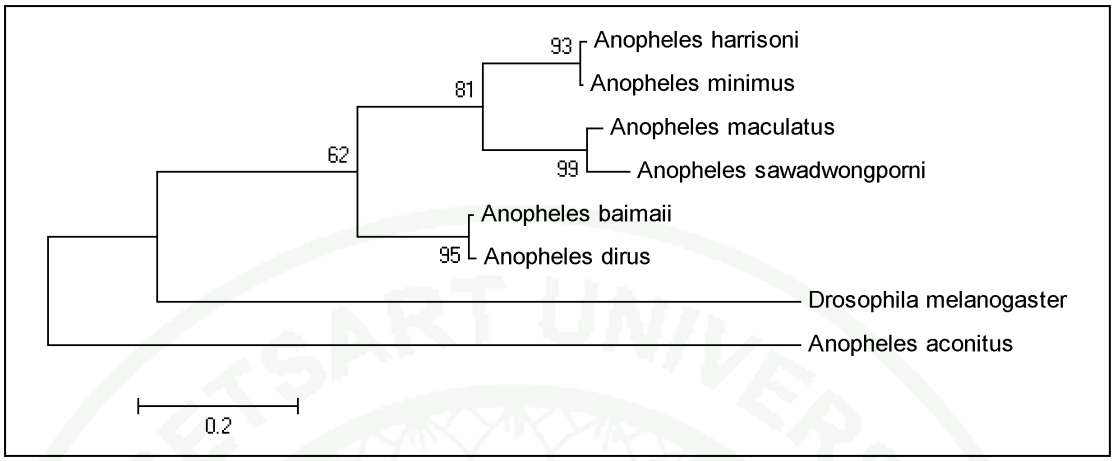
Results of NJ, ME and ML analyzed showed that sequence of outgroup were included in the phylogenetic relationships of ingroup species. This makes unreliable phylogeny of these species.



Appendix Figure 16 Phylogenetic tree based on the P9 fragment (NJ reconstruction).

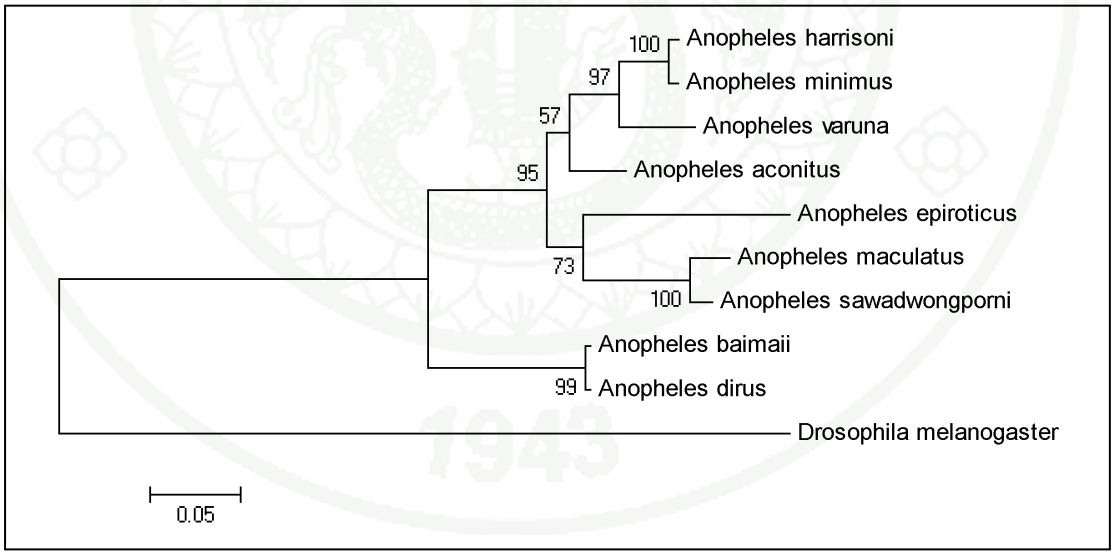


Appendix Figure 17 Phylogenetic tree based on the P9 fragment (ME reconstruction).

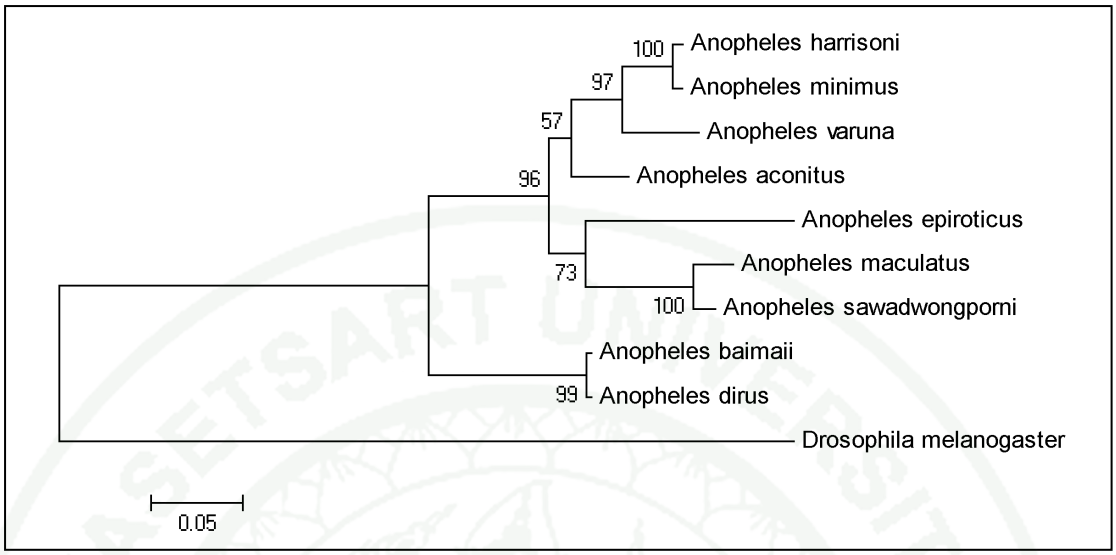


Appendix Figure 18 Phylogenetic tree based on the P9 fragment (ML reconstruction).

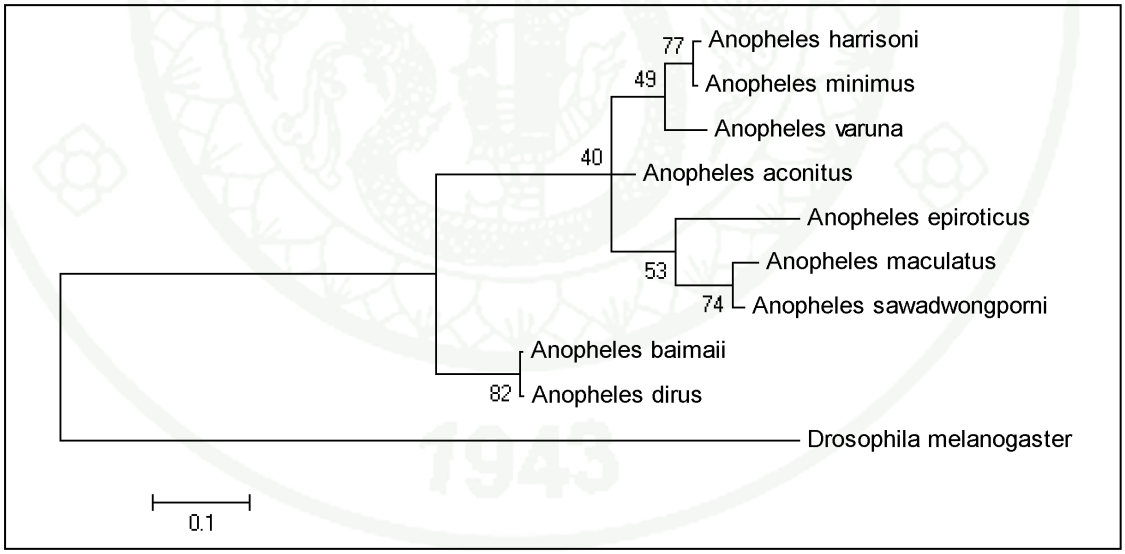
Phylogenetic analysis of the TOLL6 fragment



Appendix Figure 19 Phylogenetic tree based on the TOLL6 fragment (NJ reconstruction).

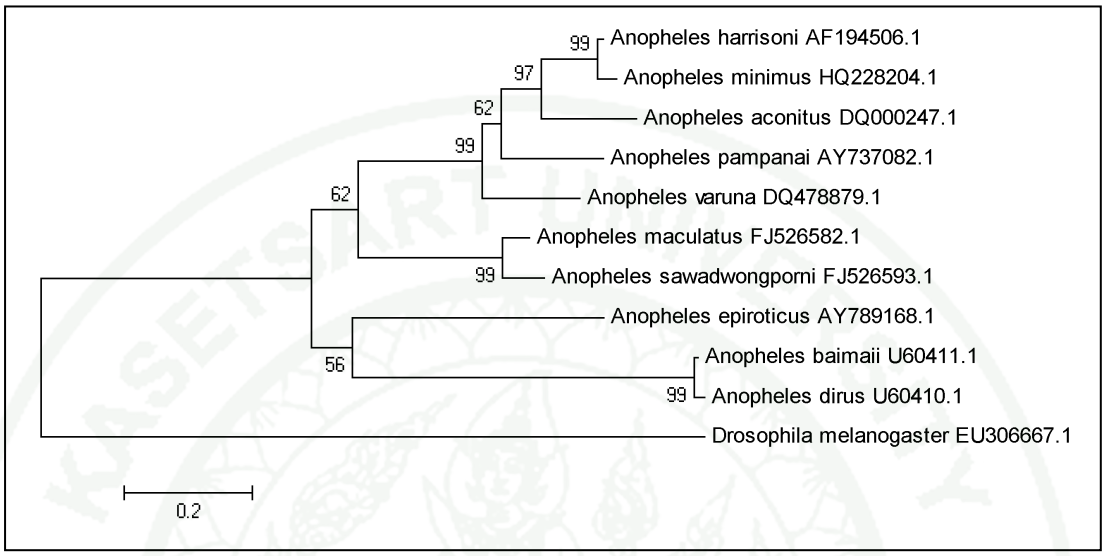


Appendix Figure 20 Phylogenetic tree based on the TOLL6 fragment (ME reconstruction).

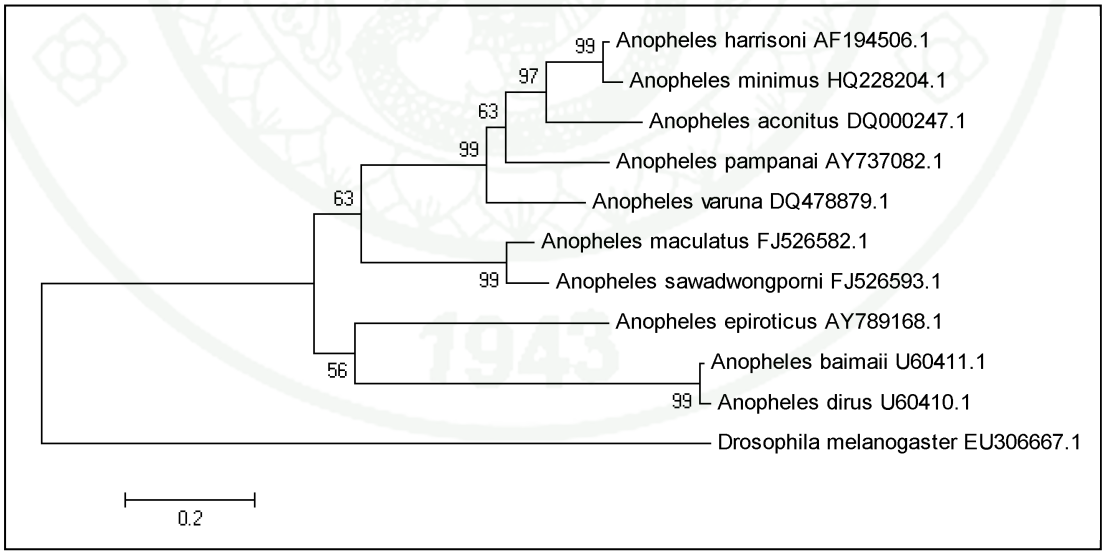


Appendix Figure 21 Phylogenetic tree based on the TOLL6 fragment (ML reconstruction).

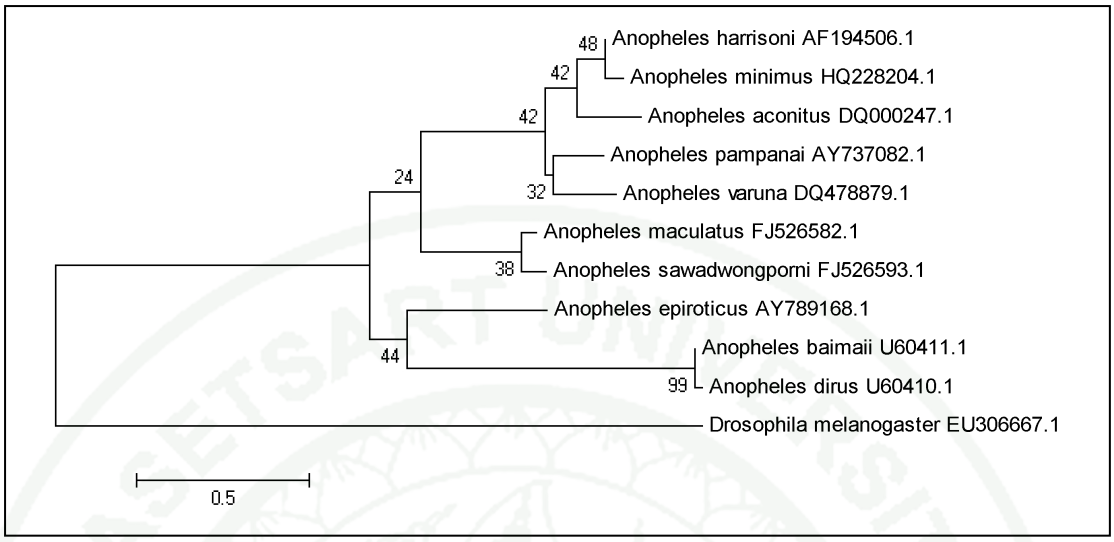
Phylogenetic analysis of the ITS2 fragment



Appendix Figure 22 Phylogenetic tree based on the ITS2 fragment (NJ reconstruction).



Appendix Figure 23 Phylogenetic tree based on the ITS2 fragment (ME reconstruction).



Appendix Figure 24 Phylogenetic tree based on the ITS2 fragment (ML reconstruction).

CIRRICULUM VITAE

NAME : Mr. Prin PHUNNGAM

BIRTH DATE : November 10, 1983

BIRTH PLACE : Bangkok, Thailand

EDUCATION :

YEAR	**INSTITUTE**	**DEGREE**
2005	Kasetsart University	B.Sc.

GRANTS : The Thailand Research Fund and Faculty of Science, Kasetsart University (TRF-MAG-WII)

: The Graduate School of Kasetsart University