Molecular systematics of the Barbirostris Subgroup and Hyrcanus Group of the genus *Anopheles* in Southeast Asia

By

Claudia Caterina Paredes-Esquivel

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

Abstract

The *Anopheles barbirostris* subgroup includes six mosquito species that are almost identical in adult morphology: *Anopheles barbirostris*, *An. campestris, An. donaldi, An. hodgkini, An. pollicaris* and *An. franciscoi*. Some of these species are implicated in the transmission of malaria and filariasis in Southeast Asia. Specimens of the Barbirostris Subgroup are also confused in the field with those from the Hyrcanus Group. Such mistakes in identification are an obstacle to the implementation of effective vector control.

A phylogenetic analysis of 756 bp of Cytochrome Oxidase I (COI) in the mitochondrial genome revealed five clades for the specimens of the Barbirostris Subgroup studied. The same clades were shown using Neighbour Joining and Maximum Parsimony trees, although internal branch points were different. A parsimony-based nested clade analysis also showed five separate networks, congruent with the phylogenetic clades, The analysis of COI showed a high level of genetic differentiation between clades (F_{ST} = 0.74-0.91, $p < 0.00001$) suggesting that each clade may represent a different species.

Analysis of the nuclear rDNA ITS2 region revealed five clades, congruent with those from the COI analysis. In all specimens of the Barbirostris Subgroup, ITS2 was >1.5kb, the largest so far recorded in any insect. The ITS2 of Clades I, IV and V differed in size. Clades II and III had an ITS2 of similar size, but they differed by 35 fixed nucleotide substitutions. Populations from Mae Hong Son (Thailand) and Kalimantan in clade I, differed by 27 fixed substitutions, this was reflected in the high level of genetic differentiation in the COI (F_{ST} = 0.63, p <0.01). However, the dataset for this clade was too small to arrive at conclusions about the species status of these allopatric populations. The extreme length of the ITS2 was a most interesting finding and resulted from the presence of four or five internal repeats of c.220bp within the ITS2, each repeat comprised of two c.110 bp sub-repeats of variable homology. Immediately 3' of these repeats was an AT-rich 26 bp region homologous to regions of transposable genetic elements (tges) found in *An. gambiae*. This transposition event must have occurred prior to the divergence of the species within the Barbirostris Subgroup. The absence of repeats and the tge-like sequence in species of the Hyrcanus group suggests that this region may have played a role in the duplication event that led to the complex internal repeats in the ITS2 of the Barbirostris Subgroup.

A morphological examination of adult specimens was carried out, to determine the relationship between morphological forms and molecular clades. For well-preserved specimens, current morphological criteria are adequate to distinguish between the Barbirostris and Hyrcanus Groups. Within the Barbirostris Subgroup, clades I, II and III were morphologically compatible with *Anopheles barbirostris* Van der Wulp,. with clades I and III found in sympatry in Thailand and clades I and II in sympatry in Java. This suggests that *Anopheles barbirostris* Van der Wulp is a species complex comprising at least 3 species (clades I, II and III). There is limited information on host preferences for clades I and II, but clade III appears to be zoophilic. Clade V was identified as the anthropophilic species *Anopheles campestris.* Clade IV is a zoophilic species with morphological characters intermediate between those of *An. campestris* and *An. barbirostris*, with which it is found in sympatry in Sa Kaeo (Thailand). Clade IV appears to be a new species.

ACKNOWLEDGMENTS

Foremost I would like to thank my supervisor Professor Harold Townson, who I truly appreciate, admire and respect. I would like to thank him for giving me the once-in-a-life time chance to learn from somebody with his wide knowledge and great passion for science, which has been an inspiration all these years. His contributions to my career and my life are invaluable.

Thanks to Martin Donnelly, for his advice and supervision and for helping me understand some difficult topics of this study. To Gwen Finnegan, for helping me in many ways all these years, always with a smile and good humour. I would also like to thank a good friend, Amy Lynd, for always being keen to help and listen, even when very busy. To all staff and students of the Liverpool School of Tropical Medicine who helped me in different ways, in particular to Dee, Lee, Helen and Sam, for their encouragement and for making this experience so enjoyable.

I am also grateful to Dr. Ralph Harbach who provided me with the specimens and gave me the opportunity to learn about the morphology of mosquitoes in the Natural History Museum, London. To the senior entomologist Kobkan Kanjanopas and to the field staff from the Ministry of Health in Thailand for providing the samples from the regions of Sa Kaeo and Trat.

I am very grateful to my sponsor, The Gunter Trust, for giving me financial support for my living expenses, and to the Leverhulme Trust and the Overseas Research Student Award for the financial support of the project. Many thanks to Gavin Laing, for invaluable advice, and to Ian Marshall and Guy Barnish for letting me work as a demonstrator in the school.

Finally I would like to thank my family. To Luis, for his strength, patience and love, and for making my life in Liverpool very special. To my dear Uchy, for being my best friend and one of the reasons I started this adventure in the UK, to my Granny ,Maria, an example of life, to my dear Mom, Carmen, for her love, encouragement and sacrifice having us so far away trying to follow our dreams, and to my Dad who I am sure would have been sharing this achievement with me.

ABBREVIATIONS

An.	*Anopheles*
bp	base pair (s)
BLAST	Basic Local Alignment and Search Tool
COOH	carboxylic Acid (functional group)
COI	Cytochrome Oxidase I
°C	degrees centigrades
DNA	Desoxyribonucleic Acid
dNTP	desoxyribonucleotide
EDTA	ethylenediaminetetraacetic acid
ITS	Internal Transcribed Spacer
IPTG	Isopropyl-β-D-galactosidase
kb	kilobase (s)
mtDNA	mitochondrial DNA
M	molar
mM	millimolar
μM	micromolar
mL	milliliter
μL	microliter
MP	Maximum Parsimony tree
NJ	Neighbour Joining tree
NOAA	National Oceanic and Atmospheric Administration (USA)
PCR	Polymerase chain reaction
rDNA	ribosomal (nuclear) DNA
RFLP	Restriction fragment length polymorphism
SDS	Sodium dodecyl sulphate
TGE(s)	Transposable Genetic Element
WHO	World Health Organization

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1 Introduction

Among the vector-borne diseases, malaria, dengue and filariasis are the most important in Southeast Asia. Mosquitoes of the genus *Anopheles* are vectors of two of these, malaria and lymphatic filariasis. Malaria is endemic in all countries of the Southeast Asian region except the Maldives. The burden of malaria in the region is second only to Sub-Saharan Africa, with over 30 000 deaths from malaria reported annually. The region also includes more than half of the global burden of lymphatic filariasis.

Vector control and clinical management of patients are the two main strategies to control mosquito-borne diseases, since there are currently no effective vaccines, except for Japanese encephalitis and yellow fever (the latter not being found in Southeast Asia) (Chadee et al., 2007). Filariasis control strategies are based on mass drug administration, which have proven successful in some regions (Weil and Ramzy, 2007) but there is a need to combine this with vector control in Southeast Asia. The increased prevalence of drug resistance in malaria parasites in various regions of the world has greatly impaired treatment. As a result, efforts to control malaria are now focused on mosquito control strategies (Hemingway, 2004).

An understanding of the distribution of vector species, their habitats and in particular their resting and feeding preferences is necessary for the development of effective control programmes. Such studies are dependent on the correct identification of species. For instance, spraying insecticides inside houses has proved very effective against *An. gambiae s.s.*, which feeds mainly indoors on humans and rests there once fed. In contrast, the degree to which *An. arabiensis* feeds on non-human hosts and rests outdoors reduces the efficacy of these strategies in Kenya (Tirados et al., 2006). In Pakistan, the density and life expectancy of *Anopheles culicifacies* and *An. stephensi* decreased notably after the application of insecticides to the surfaces of cattle, due to the zoophilic behaviour of these species (Rowland et al., 2001). Tirados et al. (2006) suggested a similar strategy to deal with

An. arabiensis in Kenya. Thus the identification of these morphologically identical species was essential to the design of new control strategies.

In Southeast Asia, vector control programmes face the problem of the high number of species involved in malaria transmission (Van Bortel et al., 2001) and the presence of species complexes (Manguin et al., 2002; Phuc et al., 2003; Surendran et al., 2006; Walton et al., 1999a). These complexes comprise species with identical or similar morphology, or overlapping morphological characters. If members of species complexes are confused, their roles in malaria transmission cannot be determined.

The *Anopheles barbirostris* Group (Reid, 1962) includes some of the less studied species in Southeast Asia; they are also difficult to identify. Thirteen species have been formally identified within this Group (Harbach, 2004). This thesis will focus on the Barbirostris Subgroup, the species of which have almost identical morphology although they vary remarkably in their behaviour. *An. campestris* (Reid, 1962) is highly anthropophilic and has been reported to be an important vector of malaria and filariasis (Harrison and Scanlon, 1975; Reid, 1968). *An. barbirostris* (Van der Wulp, 1884) is the most variable of the species of this Subgroup, with vector and non-vector forms having been reported (Lien et al., 1977). Despite the comprehensive studies carried out in the past, current morphological keys fail to distinguish vector and non-vector forms of this species (Reid, 1979). Little is known of the other four members of the Barbirostris Subgroup, *An. donaldi, An. pollicaris, An. hodgkini* and *An. franciscoi,* which are mainly forest species (Reid, 1962).

The differentiation of these species is crucial to establish effective malaria and filariasis control programmes in the Southeast Asian region. However, since the studies of Reid (Reid, 1942; Reid, 1962; Reid, 1968; Reid, 1979), there has been little research on this Group. Baskoro (2001) carried out a study of *Anopheles barbirostris* Van der Wulp in Indonesia. He recognised at least four different clades within *Anopheles barbirostris,* based on the analysis of two molecular markers, and regarded these as separate species. The research reported in this thesis was intended to extend Baskoro's work to include a wider geographical area.

Members of the Barbirostris Subgroup have become the subject of study in Thailand, where *Anopheles campestris* was implicated in the increase in malaria cases in the province of Sa Kaeo (Apiwathnasorn et al., 2002; Limrat et al., 2001). These authors reported difficulties in differentiating *Anopheles campestris* from *An. barbirostris* Van der Wulp.

This project has focused on the phylogenetic analysis of two molecular markers, the mitochondrial Cytochrome Oxidase I (COI) subunit and the nuclear rDNA Internal Transcribed Spacer 2 (ITS2). Morphological examination of specimens was carried out, where possible, to relate morphological features to clades identified from the phylogenetic studies.

1.1 Objectives

The main objectives of the research reported in this thesis were:

1. To use the molecular markers, COI and ITS2, to determine the number of distinct clades within the Barbirostris Subgroup, with a particular focus on those identifiable morphologically as *Anopheles barbirostris* and *An. campestris*.

2. To use the same molecular markers to determine if *An. barbirostris* Van der Wulp is a species complex.

3. To compare the consistency of the morphological keys used to differentiate the Barbirostris Subgroup from the closely related Hyrcanus Subgroup by reference to molecularly determined clades.

4. To determine the genetic structure of species of the *An. barbirostris* Group in Thailand, using the mitochondrial marker, COI.

5. To attempt to determine if species within the Barbirostris Subgroup are undergoing demographic expansion or are at equilibrium.

CHAPTER 2

2 Literature Review

This chapter briefly reviews the malaria problem in Southeast Asia and the species involved in its transmission. It also deals with the species concept and how cryptic or sibling species may be recognised. Finally, the methods currently used for identifying species are reviewed.

2.1 The malaria problem in Southeast Asia

In the 24th Meeting of Southeast Asian Ministers of Health, held in Bangladesh (2006), malaria was identified as a problem of high priority in Southeast Asia (WHO, 2006). Even though in general the malaria transmission in this region has diminished and the number of deaths declined, the proportion of fatal cases of *Plasmodium falciparum* has increased progressively (WHO, 2006).

The region faces serious problems related with drug resistance and increases in the number of cases of malaria due to *P. falciparum*; Southeast Asia has been described as the global epicentre of multidrug resistant malaria (Singhasivanon, 2003).

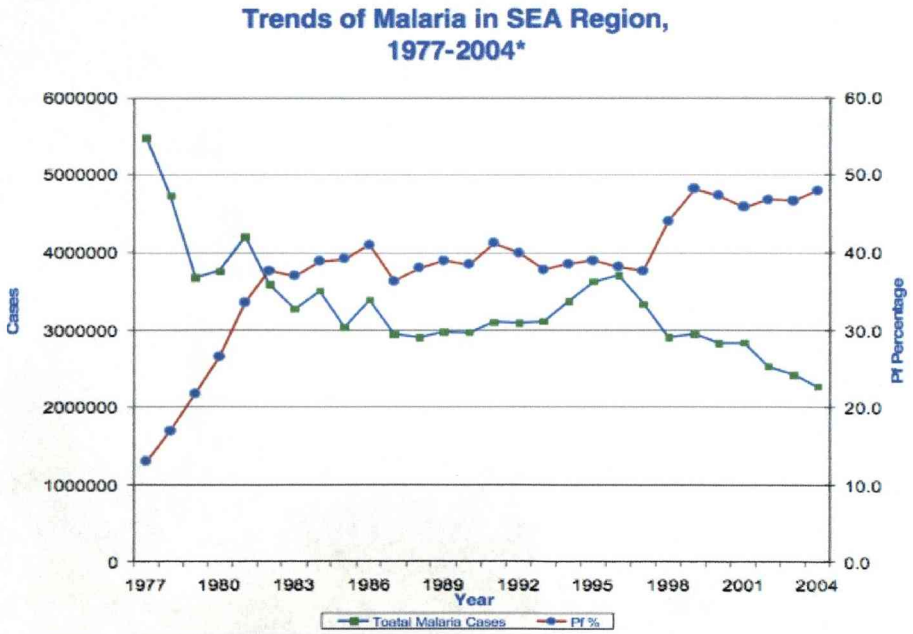
2.1.1 Incidence of malaria in Southeast Asia

Malaria is endemic in all countries of the Southeast Asian Region except the Maldives, which have remained free of malaria since 1984. In 2004 the reported incidence rate of malaria in Southeast Asia was 1.85 cases per 1000 inhabitants, with 3768 deaths per annum, the lowest annual rate reported since 1985 (WHO, 2007) (Figure 2.1). Similarly, in Vietnam, after the alarming incidence rates of malaria mortality and morbidity reached in 1991 (Hung et al., 2002), there has been more than two-fold drop in the number of cases and deaths over the period 1996-2001 (Singhasivanon, 2003). Although the morbidity and mortality has been reduced, the current level of malaria is still considered to be unsatisfactory (WHO,

2006). It is believed that the number of malaria cases in this region are actually underestimated, considering the number of patients that are treated in the private sector or, even more, those who undergo self-treatment (Singhasivanon, 2003).

In Malaysia, an increasing number of patients with malaria have tested positive for *Plasmodium knowlesi,* a malaria parasite of long-tailed macaque monkeys (Singh et al., 2004). The authors maintain that molecular methods are a prerequisite for correct identification of the malaria parasite, since *P. knowlesi* is morphologically indistinguishable from *P. malariae.* Another human case of *P. knowlesi* has also been reported in Thailand (Jongwutiwes et al., 2004). This increases the already complex malaria situation in Southeast Asia.

Figure 2.1 Malaria cases in Southeast Asia from 1999 to 2004 (WHO, 2007).



2.1.2 Drug resistance

The global epicentre of the drug resistance is the Mekong area, which comprises countries that share some part of the waters of the Mekong river: Cambodia, China (Yunnan province) Laos, Myanmar, Thailand and Vietnam (Singhasivanon, 2003).

Southeast Asia has more cases of multidrug resistance than any other region with the parasite resistant to most of the available antimalarials.

The border between Thailand and Cambodia has the most severe multidrug resistant *P. falciparum* (Looareesuwan et al., 1992). Despite intensive efforts to understand the resistance mechanisms, they are still not completely understood. It is thought that due to suboptimal levels of drugs, either from insufficient dosing regimes or due to the long half-lives of certain drugs that remain in circulation, malaria resistance is aggravated. (Rathod et al., 1997). The problem is aggravated due to the fake antimalarial drugs circulating in some countries (Dondorp et al., 2004). In Thailand, recrudescence and drug failures occur unexpectedly even in carefully controlled clinical settings (Looareesuwan et al., 1999). Drug resistance in *P. vivax* has been less studied (White, 2004).

2.1.3 Social implications

Malaria is not just a public health problem; it is also related to development as well as social, ecological and environmental changes. "Asia has the highest proportion of people below the poverty line and malaria is perpetuating the vicious cycle of poverty from which they are unable to escape" (WHO, 2006). These social implications prompted the governments of Laos, Cambodia and Vietnam to implement aggressive malaria control programmes and to declare malaria as a priority in public health programmes; as a result the malaria morbidity and mortality incidence has decreased dramatically in recent years (Trung et al., 2004).

Vietnam faced its most critical malaria situation in the period from the late 1980s to early 1990s. However, as a consequence of the application of effective vector control programmes during the last decade, malaria was declared absent in many areas of northern Vietnam. The strategy included the following main lines of attack: the inclusion of artemisin drugs for the treatment of patients, malaria diagnosis and treatment without charge, the widespread use of bednets among the population and finally an education policy, to involve the community in the prevention and control of malaria (Hung et al., 2002). Additionally, Vietnam underwent economic and

social changes that have considerably improved the quality of life of the population. The successful results were mainly focused in the northern part of the country. In central Vietnam, where the major populations are minority ethnic groups, malaria remains a major public health problem.

The effect of human migration on malaria transmission is more critical in the regions where the main vectors have not yet been eradicated (Verle et al., 1998). In East Timor, approximately 850 000 people were displaced as a consequence of the way of independence from Indonesia in 1999 and forced into refugee camps. "Many of the population were internally displaced into the jungle and mountain areas, whilst the militia and Indonesian forces pushed thousands more over the border into West Timor" (Kolaczinski and Webster, 2003). Deforestation has also become a serious problem in Southeast Asia. In Thailand the forest cover dropped from about 50% in 1960 to 18% in 1990 (Rattanarithikul et al., 2005). An increased number of breeding sites for mosquito species is a common consequence in deforested areas. However, the extensive deforestation in northern Thailand is said to have reduced the number of breeding sites of the two main vectors *Anopheles dirus* and *An. minimus.* (Singhasivanon, 2003).

2.2 *Anopheles* species complexes and the importance of their identification

Since it was first determined that human malaria was transmitted by mosquitoes of the genus *Anopheles* it became clear that malaria was not observed in some areas where all the conditions were appropriate for its establishment. This was called "Anophelism without malaria" (Fantini, 1994) and was the beginning of the study of cryptic species in *Anopheles*. Cryptic or sibling species have been defined as morphologically similar or identical populations that are reproductively isolated (Mayr, 1991). These species commonly differ in important biological characteristics, including their preference for different hosts and resting sites (White, 1977a) and hence their role in malaria transmission. The first species complex discovered in mosquitoes was the *An. maculipennis* complex (Hackett and Misiroli, 1935). Subsequently other important *Anopheles* species have been identified as species complexes.

In Africa, the first recognition of the *Anopheles gambiae* complex was based on mating barriers between its members (Davidson et al., 1967) and subsequently these were differentiated by polytene chromosome analysis (Coluzzi et al., 2002; Coluzzi et al., 1979). The complex includes seven species, the most important of which are *An. gambiae s.s.* and *An. arabiensis,* of these, *An. gambiae* is the more important due to its high anthropophily and indoor biting behaviour, whereas *An. arabiensis* feeds on humans but also extensively on cattle and can rest outdoors, particularly in man-made structures (reviewed by Service and Townson, 2002).

In the alst completed review there were 444 formally described *Anopheles* species and 40 unnamed species that are members of species complexes (Harbach, 2004). The number of species within complexes of *Anopheles* species is likely to be an underestimate (Service and Townson, 2002).

Some of the most important species complexes in the Oriental Region include: *Anopheles culicifacies* Giles, a complex of five sibling species, designated as species A, B, C, D and E. Species A, C, D and E are vectors of malaria in India. Differences in susceptibility to malaria parasites have been reported; in Sri Lanka species E has been successfully infected with *P. falciparum*, while species B has proved to be refractive to infection (Surendran et al., 2006). Species of the *Anopheles fluviatilis* James and the *Anopheles minimus* Theobald complexes comprise at least three sibling species each, closely related and important malaria vectors in the Oriental Region (Singh et al., 2006). These species are commonly confused in the field but the design of diagnostic tests, based on DNA amplification, facilitates their identification (Phuc et al., 2003). *Anopheles dirus* Peyton and Harrison is another group of species that has been found to comprise a species complex and includes major vectors of malaria in Southeast Asia (Baimai et al., 1987; Peyton, 1989; Walton et al., 1999a). Recently, a revision of this complex has revealed seven species (Sallum et al., 2005). Finally *Anopheles barbirostris* Van der Wulp, one of the species studied here was not considered to be a main malaria vector. There is evidence that this taxon comprises a species complex (Baskoro, 2001), which may include vectors of malaria and filariasis and non-vector forms.

Effective vector control programmes rely on a comprehensive knowledge of the vector species. Heterogeneities in malaria vector behaviour may result in differences in epidemiological importance (Trung et al., 2005), since these behaviours largely define vectorial status. For instance, in Southeast Asia, the late biting behaviour of *An. minimus s.s.* (Harbach et al., 2006) ensured that bed nets were an effective control measure, whilst the outdoor biting, and early feeding behaviour of *An. dirus s.s.* (Sallum et al., 2005) made insecticide-impregnated bed nets and indoor residual spraying less suitable for its control (Trung et al., 2005).

Preventive control strategies in Southeast Asia must deal with the complexity of the vector system caused by the number of species potentially involved in malaria transmission and the large number of species complexes (Trung et al., 2004; Van Bortel et al., 2001).

2.3 The concept of species

Species are units of comparison in almost all subfields of biology, from anatomy to behaviour, development, ecology, evolution, genetics, molecular biology, palaeontology, physiology and systematics (de Queiroz, 2005). Scientists have tried to define species from different perspectives, and there are currently 24 such concepts (Hey, 2001; Tibayrenc, 2006). However, a species definition that is applicable to all organisms is unlikely to become available. Some researchers even question if species actually exist in reality (Kunz, 2002). The most important concepts were summarized by Mallet et al. (2006) and are presented in Table 2.1. Four of them are described in detail in the following pages.

Table 2.1 Partial listing of species concepts and other ideas about species. Taken from Mallet et al. (2006).

Name of species concept (alphabetically arranged)	Brief definition	Reference
"Biological" or reproductive isolation species concept	Taxa possessing reproductive isolation with respect to other species. Characterized by reproductive isolating mechanisms	Poulton 1904, Mayr, 1970
Cladistic Concept	Species are unbranched segments or lineages in an organismal phylogeny	Hennig 1968, Ridley, 2004
Cohesion Concept	A taxon characterized by cohesion mechanisms, including reproductive isolation, recognition mechanisms, ecological niche, as well as by genealogical distinctness	Templeton, 1998
Darwin's morphological oncept	"Varieties" between which there are no or few morphological intermediates	Darwin, 1859
Diagnostic (phylogenetic) species concept	A species "is an irreducible (basal) cluster of organisms, diagnosably distinct from other such clusters, and within which there is a parental pattern of ancestry and descent"	Cracraft, 1989
Ecological concept	"A lineage which occupies an adaptive zone minimally different from that of any other lineage..."	Van Valen, 1976
Evolutionary concept	A lineage evolving separately and "with its own unitary evolutionary role and tendencies"	Simpson, 1951
Genealogical concept	Species are mutually monophyletic in the genealogies at all (or at a consensus of) gene genealogies in the genome	Baum and Shaw, 1995
General lineage concept	Species are independent lineages. according to de Queiroz: all other species concepts agree on this fundamental principle; conflict about species concepts refers mainly to criteria applying to different stages of lineage divergence	de Queiroz, 1998
Genotypic (genomic) cluster criterion	Sympatric species are clusters of genotypes circumscribed by gaps in the range of possible multilocus genotypes between them	Mallet, 1995, 2001
Phenetic concept	Clusters of individuals circumscribed using multivariate statistical analysis	Sokal and Crovello, 1970
Polytypic species	Taxa having many "types," i.e., geographic subspecies. Geographic populations are part of the same species if they intergrade in areas of overlap	1890 onwards, reviewed by: Mayr, 1970; Mallet, 1995, 2004
Population concept	Populations are the real units of evolution, not species, because gene flow is generally weak. Morphological and	Ehrlich and Raven, 1969

CONTINUE...

	genetic uniformity of species is explained by stabilizing selection acting separately in each population	
Recognition concept	Species possess a shared fertilization system, known as "specific-mate recognition systems"	Paterson, 1985
Taxonomy without species	Species are no more real than any other species hierarchical level in the tree of life. Species and other taxonomic ranks should be replaced either by "rank-free taxonomy" (which can name each node in a bifurcating phylogeny—Mishler), or by genotypic clusters described according to their genetic divergence from other clusters (Hendry et al.)	Mishler, 1999; Hendry et al., 2000

2.3.1 The "biological" (BSC) or reproductive isolation species concept

Edward Bagnall Poulton was the first to tackle the problem of the species definition in an evolutionary context. Poulton proposed to investigate how the meaning of "species" has changed from "that of the years before 1858, when the Darwin-Wallace conception of natural selection was launched upon the world" and laid out the research programme for speciation largely adopted today. Poulton argued that species were *syngamic* (i.e. formed reproductive communities), the individual members of which were united by *synepigony* (common descent) (Mallet, 2004).

Later, in 1942, Ernst Mayr published *Systematics and the Origins of the Species,* which was the first comprehensive conceptual analysis of biological species. All subsequent debates about the biological species concept refer to Mayr (1942). This is the species definition he gave at that time:

> "A species consists of a group of populations which replace each other geographically or ecologically and of which the neighbouring ones intergrade or hybridize wherever they are in contact or which are potentially capable of doing so (with one or more of the populations) in those cases where contact is prevented by geographical or ecological barriers "

The validity of the BSC has been under endless debate (Andersson, 1990; Noor, 2002), the principal criticism being its inability to deal with uniparentally reproducing organisms (Andersson, 1990). However because speciation is still conceptualized at the level of individual, the biological species concept (BSC) is still considered by some as the "gold standard" (Wu, 2001).

2.3.2 Ecological species concept

"A species is a lineage (or closely related set of lineages) which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range" (Vanvalen, 1976). In other words two different organisms cannot occupy the same ecological niche at the same time. Andersson (1990) identified the ecological species concept as the most useful, in comparison with the Evolutionary and Biological species concepts. He affirmed "it has the potential to make a direct connection between environment and phenetic variation patterns and provides a useful conceptual framework for experimentation and observation at the interface between taxonomy and ecology and has the potential to explain on a universal basis variation in diversity and reproductive patterns".

2.3.3 Evolutionary species concept

This concept was first purposed by Simpson (1951) who defined species as "a lineage (an ancestral descendant sequence of populations) evolving separately from others and with its own unitary evolutionary role and tendencies". This was subsequently modified by Wiley (1978). According to Andersson (1990), this concept fails to explain why different lineages are phenetically different. It fails, therefore, to explain why species arise, and why there are differences in diversity and reproductive strategies. (Andersson, 1990).

2.3.4 Phylogenetic (Diagnostic) species concept (PSC)

This concept was first presented by Cracraft in 1989. The author maintained: "a phylogenetic species is an irreducible (basal) cluster of organisms, diagnosably distinct from other such clusters, and within which there is a parental pattern of ancestry and descent" (Cracraft, 1989). The main disadvantage of this definition is that, if these criteria are applied strictly, either small groups of individuals, or even single specimens, could be defined as separate species, leading to ''taxonomic inflation'' Mallet et al. (2006). On average, the PSC has recognized 48% more species than has the BSC (Agapow et al., 2004). This concept has been described as a "fine-grained" strategy and far from the "broad-brush" biological species concept (Isaac et al., 2004). Considering that phylogenetic tools have been widely used to solve species problems in the last years, it is important to recognize the limits to diversity under the PSC.

The Biological species concept has been the most widely used to define species complexes in *Anopheles.* However, at present, the number of new species is on the increase as a result of the use of DNA-based techniques. Harbach (2004) pointed out that one of the problems involved in identifying species by DNA sequence comparison, lies in the fact that many of the sequences submitted into databases like GenBank may be from specimens that have been wrongly identified. This author maintained that "the modern techniques must be integrated with classical morphological analysis to determine the correct identification of available names for species concepts, previously based solely on anatomical characters" (Harbach, 2003).

Regardless of the species concept used for *Anopheles*, it is indisputable that the identification of sibling species can influence the efficiency of control measures. Many control programmes are unsuccessful as they fail to target the particular *Anopheles* species involved in malaria or filariasis transmission. In the following pages, a review of the molecular methods employed in the identification of species is presented.

2.4 Molecular markers for species studies in malaria vectors

In *Anopheles* mosquitoes, many of the currently known species were originally identified using morphological characters. However, anopheline mosquitoes suffer from the limitation that many species differ in very few characteristics of adult or immature morphology (White, 1977b). Moreover, morphological variation inside species has also been observed (Hwang et al., 2004; Krzywinski and Besansky, 2003) and this has added to the confusion in the identification. The following are some of the most common molecular techniques used in the identification of species.

2.4.1 Allozymes

In the mid-1960s, protein electrophoretic techniques generated a revolution in population genetics and evolutionary biology; allozymes were the most widely used molecular markers previous to the appearance of PCR-based techniques. Allozymes of a given enzyme are heterozygous products of different alleles at a specific locus (Loxdale and Lushai, 1998). This method is based on the property of proteins to migrate at different rates in a supporting media when applying them electric current, due to differences in their net charge; as a result, different electrophoresis patterns are obtained among populations. In insects, allozymes have proven to be useful to differentiate sibling species, for instance members of the *Anopheles gambiae* complex (Mahon et al., 1976), *Drosophila* sibling species (Ayala and Powell, 1972), cryptic species of *An. albitarsis* in Brazil (Narang et al., 1993). They are still being used in population studies (Kaya and Isik, 2006; Takada et al., 2006; Uyeda and Kephart, 2006; Weisrock and Larson, 2006), although their utility has been clearly replaced by DNA-based studies.

2.4.2 DNA molecular markers

The main advantage of DNA molecular markers when compared with protein based techniques is that greater levels of polymorphism are detected. In addition DNA markers can be used with small amounts of insect material, and with stored, dry or

old samples (Loxdale and Lushai, 1998; Townson et al., 1999). The first DNA based technique to become popular was restriction enzyme analysis. This appeared in the 1970s and was widely used in population genetic studies. It was subsequently followed by fingerprinting approaches and mitochondrial DNA in the 1980s. In the early to mid-1990s, PCR primers and advances in sequencing technology greatly increased the rate of collection of DNA data, to the point where the majority of phylogenetic studies involve a molecular component (Caterino et al., 2000; Macdonald and Loxdale, 2004; Simon et al., 2006).

DNA molecular markers have been widely used to answer evolutionary questions in recent years and their contribution has enhanced the value of morphological and ecological data, making substantial contributions to the evolutionary biology of insects in the process (Caterino et al., 2000). Although the multiple molecular genetic studies have resulted in little change in the internal classification of the genus *Anopheles* (Harbach, 2004), their contribution in addressing questions about anopheline phylogenetics, biogeography, the nature of species boundaries, and the forces that have structured genetic variation within species is doubtless (Krzywinski and Besansky, 2003).

It is, however, a requisite for the success of phylogenetic studies, to select an appropriate genetic marker (Lunt et al., 1996). This should provide consistent and contrastable information about evolutionary processes and phylogenetic relationships among organisms. In the following pages a brief description of the most commonly used molecular markers is presented.

2.4.3 Mitochondrial DNA (mtDNA) markers

Mitochondrial DNA (mtDNA) is a small, single, covalently closed circular molecule (Avise et al., 1987), except in some cnidarians that seem to have one or two linear molecules (Warrior and Gall, 1985). It is usually about 16-20 kilobases long. Despite some variation in size, the coding function of the mitochondrial genome remains relatively stable in animals. In general, mitochondrial DNA codes only for genes involved in the mitochondrial translation apparatus, electron transport, and

oxidative phosphorylation (Ballard and Rand, 2005). The mitochondrial genome typically encodes for 13 protein subunits of the oxidative phosphorylation, the two rRNAs of the mitochondrial ribosome, and the 22 tRNAs necessary for the translation of the proteins encoded by mtDNA (Boore, 1999). The most distinctive feature of metazoan mitochondrial genomes is their extremely compact gene organization. Apart from the replication origin region, the genome comprises discrete genes. Introns, pseudogenes, repetitive DNA, and sizeable spacer genes are absent (Avise et al., 1987; Saccone et al., 1999).

Due to its ease of isolation, high copy number, the absence (or reduced rates) of recombination (Birky, 2001), conservation of sequence and structure across metazoa, and the range of mutational rates in different regions of the molecule (Harrison, 1989), the use of mtDNA has become standard for many phylogenetic studies (Caterino et al., 2000; Simon et al., 2006). It has been used as a valuable tool for filling the gap between population genetics and systematics (Avise et al., 1987) and is believed to be useful to address long-standing contentious issues in dipteran phylogeny (Cameron et al., 2007).

Among the many mitochondrial genes that have been studied, the protein coding regions Cytochrome Oxidase subunits I (COI) and II (COII) have been extensively used for phylogenetic inference alone or in combination with other sequences, and have proven to be phylogenetically informative in many insect groups (Chen et al., 2003). Other mitochondrial markers widely used in insect molecular systematics are: 16S and 12S rDNA subunits, followed by genes COIII, NADH dehydrogenase 5 and *cyt* b to a lesser extent and finally genes ND2, ND4 and ND1 which have received only isolated attention (Caterino et al., 2000). Further discussion of COI is presented in Chapter 4; in this section, the utility of other markers is discussed.

2.4.3.1 Cytchrome oxidase II (COII)

This is the terminal member of the mitochondrial inner membrane electron transport chain; one of three mitochondrially-encoded subunits. Its functions are to generate the adenosine triphosphate (ATP) required for cellular processes. The COII gene

region has been used to understand the phylogenetic relationships in Lepidoptera species (Silva-Brandao et al., 2005) and Diptera species (Ekrem and Willassen, 2004; Han and Ro, 2005). It has also proved useful in studying the genetic structure of Africanized honeybees in Brazil and Uruguay and to understand how they were introduced into the New World (Collet et al., 2006). Although it has not been as widely used in *Anopheles* species as COI, COII appears valuable in the identification of some *Anopheles* species (Chen et al., 2004; Goswami et al., 2005; Junkum et al., 2005; Park et al., 2003).

2.4.3.2 Subunits 12S and 16S

The molecular marker 16S has been mainly applied to phylogenetic studies in micro-organisms. In insects it is used in studies of mid-category levels, i.e. families or genera (Black and Piesman, 1994). It has not been used as extensively as other molecular markers, but there is evidence that this region can be important in phylogeny reconstruction for fossil specimens. In the termite *Mastotermes electrodominicus* preserved in Oligo-Miocene amber (25 million to 30 million years old) the analysis of the 16S together with the nuclear 18S rDNA confirmed morphological cladistic analyses of living dictyopterans (termites, cockroaches and mantids) (Desalle et al., 1992).

The 12 S gene region has proved to be highly conserved and therefore useful to infer phylogenetic relationships among higher categorical levels, phyla or subphyla (Ballard et al., 1992).

2.4.3.3 NADH Dehydrogenase subunits 4 and 5 (ND4 &ND5)

ND5 is one of the fastest-evolving mitochondrial genes (Clary and Wolstenholme, 1985). In contrast with other widely used molecular markers like *cyt* b and 12S/16S rRNA genes, the main advantages that ND5 offers are (1) ease of alignment, (2) length (ca. 3.4 kb), and (3) it contains more phylogenetically informative variation at 1st and 2nd codon positions. Moreover, the ND4/ND5 gene regions are usually easy to amplify and sequence (Miya et al., 2006).

In mosquitoes, ND5 have been used to infer the genetic structure of *Aedes albopictus* in Brazil (Birungi and Munstermann, 2002), *Anopheles farauti s.s.* in Vanuatu (Reiff et al., 2007) and it was used in the detection of population expansion of *An. arabiensis* and *An. gambiae* in Africa (Donnelly et al., 2001). In a study on phylogeny reconstruction of anopheline mosquitoes, the abundant variation at all codon positions found in the ND5 gene, allowed recovery of the basal and most of the recent relationships (Krzywinski et al., 2001).

The subunit ND4 has received only isolated attention; however, it has proved to be valuable for understanding the genetic structure of the dengue vector *Aedes aegypti* in Venezuela (Herrera et al., 2006) and it has been useful for exploring interactions within and among populations of the vector of onchocerciasis *Simulium ochraceum* s.l. in Mexico (Rodriguez-Perez et al., 2006).

2.4.3.4 Cytochrome b (cyt b)

In anophelines, this molecular marker proved to be extremely conserved at the protein level, showing also a rapid saturation at synonymous positions. This probably accounted for the lack of a meaningful phylogenetic signal in the cyt b gene (Krzywinski et al., 2001)

The high rate of substitution in the mitochondrial genome generates a rich source of variable characters, which may produce high levels of heteroplasmy and homoplasy that characterize this molecule (Engstrom et al., 2004)

2.4.3.5 Heteroplasmy

Surveys of mtDNA variation in natural populations of animals have revealed that heteroplasmy, defined as the presence of more than one type of mtDNA within cells, is a common event (Bermingham et al., 1986; Kann et al., 1998; Kijewski et al., 2006; Kmiec et al., 2006; Morel et al., 2006; Nardi et al., 2001; Shigenobu et al., 2005). It has been related to multiple diseases in humans, including cancer (Chinnery et al., 2002). Reports of heteroplasmy have come from insects of different

orders. Heteroplasmy have been detected within individuals of *Cochliomyia macellaria* (Diptera: Calliphoridae) (doValle and deAzeredoEspin, 1995) and *Cochliomyia hominivorax* (Vargas and Espin, 1995). It has also been found in coding regions in grasshoppers of the genus *Chitaura* (Walton et al., 1997). In the genus *Anopheles,* heteroplasmy has been reported in *Anopheles gambiae* (Caccone et al., 1996). This event is particularly common in the control region (Kann et al., 1998), although it has also been reported in different regions of the mitochondrial genome ND2–COIII (Kijewski et al., 2006), the cytochrome oxidase I (Steel et al., 2000; Walton et al., 1997).

Much of the variation in mtDNA haplotypes in individuals is the result of insertions, deletions and duplications in the mitochondrial genome (Kann et al., 1998). There was concern that heteroplasmy might be extensive and complicate mtDNA study, but empirical experience proved that these cases were unusual and therefore of little impact in routine study of animal mtDNA (Avise et al., 1987).

2.4.3.6 Pseudo mitochondrial genome

Nuclear mitochondrial pseudogenes are defined as nuclear copies of mitochondrial sequences that show high similarity with mtDNA sequences (Lopez et al., 1994). In some species they have been detected to occur in very high copy numbers and can be preferentially amplified with respect to mtDNA, even when using universal primers (Benesh et al., 2006). Such pseudogenes account for some instance of false mitochondrial sequence heteroplasmy (Abbott et al., 2005; Parr et al., 2006). According to Benansson *et. al.* (2000) "They complicate the use of mtDNA as a molecular marker in evolutionary studies. They have been reported in a variety of organisms, predominantly in mammals and birds (Benesh et al., 2006). Nuclear mitochondrial pseudogenes have been found in the genus *Anopheles* (Richly and Leister, 2004), however their existence cannot be underestimated since they are present in other insects (Bensasson et al., 2000). According to Benesh et al. (2006), recognition of nuclear mitochondrial pseudogenes is necessary in any study utilizing mtDNA. Heteroplasmic mtDNA can be differentiated from pseudogenes due to their

absence of stop codons, indels or radical amino acid changes (Magnacca and Danforth, 2006). These strategies were applied in this study.

2.4.3.7 Homoplasy or conserved evolution

Homoplasy occurs when the same character state evolves independently in two separate lineages (Brown, 1999). This phenomenon was first described for coincident morphological characters. In the same way, at the molecular level, we see repeated or back mutations at a specific nucleotide position. If in some specific positions in the genome rapid changes occur, by recurrent base substitutions or other events, then it is expected that some homoplasious changes will occur (Avise et al., 1987). The occurrence of homoplasy has been of concern since it can obscure the real number of evolutionary events in a sequence data and many phylogenetic studies rely on mitochondrial data only. It has been responsible for incorrect phylogeny reconstruction in some species (Naylor and Brown, 1998).

On the other hand, several approaches for dealing with this situation have been successful. Character or character-state weighting have been used to deal with homoplasy in sequence data in parsimony analysis. For instance, homoplasy is thought to occur more frequently as a result of transitions than transversions; therefore, the weight given to transitions as phylogenetic characters should be lower than that given to transversions to (Meyer, 1994).

According to Avise (1987) "typical empirical surveys of mtDNA effectively involve assays of at least several hundred base-pairs of information per individual. When viewed this way, any widespread and intricate similarities present in mtDNA are most unlikely to have arisen by convergent evolution and so they must primarily reflect phylogenetic descent rather than convergent mutation" In this respect, homoplasy is not the alarming problem in phylogenetic studies, as originally it was considered by some authors.

2.4.4 Nuclear ribosomal DNA (rDNA)

Ribosomes are the sites of protein synthesis in all organisms, a complex composed of individual ribosomal RNA (rRNA) molecules and more than 50 proteins. The functional eukaryotic ribosomes are composed of a small and a large subunit. Proteins in the two subunits differ, as do the molecules of rRNA. Within Diptera, the rDNA genes form a tandemly repeated transcriptional unit, which contains around 500 copies (Collins et al., 1987). This has contributed to making this region one of the most widely used in molecular population studies. PCR amplification and sequencing can be performed more easily than when using single copies in the genome (Caterino et al., 2000). Each transcriptional unit comprises the intergenic spacer (IGS) and genes for the 18S, 5.8S and 28S ribosomal RNA (rRNA) separated by internal transcribed spacers (ITS1 and ITS2). A more comprehensive review of the different regions of the ribosomal DNA is presented in Chapter 5.

2.3.4.1. Concerted evolution in rDNA

Concerted evolution was first observed in the spacer sequences between 18S and 28S gene regions in the rDNA of species of frogs of the genus *Xenopus.* The authors observed that the intergenic spacer (IGS) was identical in members of the same species but about 10% different between *X. laevis* and *X. mulleri*. The conclusion was that these gene regions had evolved together through a "correction" mechanism. This process later became known as concerted evolution: a pattern in which several hundred rDNA sequences within any one species show little or no genetic diversity whereas the sequences of different species diverge (Brown et al., 1972). Subsequently, the concept of concerted evolution became accepted, and the genes affected were then considered of value in phylogenetics.

According to the model presented by Brown et al. (1972), concerted evolution occurs as a consequence of unequal crossover taking place randomly among members of a gene family, a repeated pattern of this event has the effect of homogenizing the member genes. Other processes have been used to explain the random fixation of variants in species of concerted evolution: unequal chromatid (or

chromosome) exchange gene conversion and gain and loss by transposition (Dover, 1982). Gene conversion is the major mechanism contributing to the phenomenon of concerted evolution (Amstutz et al., 1985; Sugino and Innan, 2006). The main difference between gene conversion and unequal crossover is that the latter may increase or decrease the number of genes, whereas the former does not. The process of gene conversion is however not completely well understood (Nei and Rooney, 2005).

In several recent studies, the rate of homogenization has proved to be low and as a result, significant levels of intraspecific rDNA polymorphisms are found. For instance, concerted evolution appears to be extremely ineffective in the grasshopper *Podisma pedestris*; this species presents several highly divergent ribosomal DNA groups within individuals of the same species (Keller et al., 2006). Although the validity of the rDNA as a suitable molecular marker for phylogenetic studies has been questioned (Alvarez and Wendel, 2003), it is important to consider that most of these ineffective cases of concerted evolution have been reported in plants. That is why it is still extensively used in phylogenetic studies in insects.

2.5 The *Anopheles barbirostris* problem

Since first reported in 1884 by Van der Wulp *Anopheles barbirostris* has been described as a single "species." As with the Maculipennis Group in Europe, it was variations in malaria vector capacity that made researchers suspect that this was in reality a species group. An exhaustive study performed by Reid (1962) in Malaysia lead to the conclusion that in fact this was a group of "half-a-dozen or more distinct though exceedingly similar-looking species", the Barbirostris Group. This Group is placed in the Myzorhynchus Series.

Genus	Anopheles
Subgenus	Anopheles
Series	Myzorhynchus

2.5.1 Morphological Description: Barbirostris Group and Barbirostris Subgroup

The Barbirostris Group is a well marked group of 13 species (Harbach, 2004), which resemble one another very closely, especially as adults. Reid (1968) described them as follows:

> "Female with very shaggy all-black palps and a tuft of black scales on sternite VII of the abdomen. Other sternites usually with a few scattered white scales in both sexes. Sides of thorax often with some white scales associated with the groups of setae. Wing usually with a pale fringe spot at 5.2 and a few scattered pale scales on the basal half of the costa. Inner shoulder hair of larva branched from near the base, outer clypeal hairs usually bushy".

Reid (1962) divided the group into 2 subgroups based on adult, larval and egg characters. Years later, two more species were included. They have not been placed in either of these two groups.

An. freyi Meng

An. koreicus Yamada and Watanabe

Barbirostris subgroup	**Vanus subgroup**
An. barbirostris van der Wulp	*An. ahomi* Chowdhury
An. campestris Reid	*An. barbumbrosus* Strickland & Chowdhury
An. donaldi Reid	*An. manalangi* Mendoza
An. franciscoi Reid	*An. reidi* Harrison
An. hodgkini Reid	*An. vanus* Walker
An. pollicaris Reid	

Adult species from the Barbirostris and Vanus Subgroups are distinguished as follows:

> "Wing with lower apical pale fringe spot wide, at least from 3 to 4.1; no ventral pale scales on abdomen (Vanus Subgroup).
>
> Wing with lower apical fringe spot narrow, opposite 3 only; some ventral pale scales nearly always present on abdominal segments II-VI (Barbirostris Subgroup)" (Reid, 1968).

More details related to the taxonomy and structure of the Barbirostris Subgroup are presented in Chapter 6.

2.5.2 Distribution

In contrast to the Bancroftii Group, which is more Australian in its range, the Barbirostris Group is oriental (Reid, 1968). The northernmost records in the west occur in Pakistan and Nepal, while in the east, members have been collected in Kwangtung, Szechwan and Yunnan provinces of China, and Hainan and Taiwan. The center of distribution for the group appears to be Malaysia, but this may be a result of the large amount of work done by Reid in that country (Harrison and Scanlon, 1975).

In view of the difficulties of identifying *An. barbirostris* Van der Wulp with accuracy unless early stages as well as adults are available, the range of this species is uncertain at present (Reid, 1968) but it is probably present in these countries: Bangladesh, Cambodia, China, India, Indonesia, Laos, Malaysia, Myanmar, Nepal (possibly), Pakistan, Sri Lanka, Thailand, Vietnam, East Timor. It is common in Peninsular Malaysia and appears to range throughout the Indonesian archipelago from Sumatra to Sulawesi (where it is certainly present) and the Moluccas. *Anopheles barbirostris* is very abundant and widely distributed in Thailand. It is likely to be found anywhere where a thorough search is made, except at higher elevations or in dense primary forest.

In some areas (just north of Bangkok), in the rice plains along the Chao Phrya River, *An. campestris* is the predominant member of the Barbirostris Subgroup (Harrison and Scanlon, 1975). *An. campestris* appears to be largely confined to broad alluvial plains (therefore the name *campestris*) but there have been few studies using the early stages to corroborate findings in many countries. Outside Malaysia it has been positively identified only in Thailand (Harrison and Scanlon, 1975; Reid, 1968) and Cambodia (Harrison, 1975). Reid (1968) regarded *An. campestris* as absent from Sumatra and Java and the rest of the archipelago and confined to the mainland of Southeast Asia.

The distribution of other members of the Barbirostris Subgroup is also unclear for the reasons mentioned before. *An. donaldi* has been reported in Thailand, but it is probably restricted to the southern region (Harrison and Scanlon, 1975). Reid (1968) affirmed that *An. donaldi* is locally common in central and southern Peninsular Malaysia. It is also the most common member of the group in Borneo and may be present in Sumatra and Java. *An. hodgkini* and *An. pollicaris* are uncommon forest species. The former appears to be quite widespread, ranging apparently from Myanmar to Indochina, Thailand, Peninsular Malaysia and Borneo. The latter has been reported only in Malaysia and Thailand, although two pupae have been seen from Nepal (Harrison and Scanlon, 1975; Reid, 1968).

2.5.3 Biology

Members of the Barbirostris Group occupy larval habitats usually associated with still water. Of the species in the Subgroup, only *An. barbirostris* and *An. campestris* larvae are normally found closely associated with human habitation in such habitats as rice fields, ditches and open temporary ground pools. The remaining species of this group are forest dwelling species with the larvae found in shaded stream pools (Harrison and Scanlon, 1975).

2.5.4 Feeding preferences

An. barbirostris was considered a zoophilic species, harmless to humans, except in Sulawesi, where this species was seen to be highly or at least moderately anthropophilic (4180 = 5%) (Lien et al., 1977), and on the island of Java (Reid et al., 1979). In Thailand, this species had been observed biting people, but not on a large scale except in situations where the normal host animals were absent. However, in some regions of Southeast Asia "man-biting" varieties of *An. barbirostris* have been reported in the past. Harrison and Scanlon (1975) pointed out the importance of studying these varieties to understand the real structure of the species group. In India, *An. barbirostris* is generally regarded as a zoophilic species though it will bite people occasionally. Human blood has been detected in female mosquitoes from the

north of Madras. It has been reported that in the absence of cattle, *An. barbirostris* will readily bite humans (Rao, 1981).

Anopheles campestris has been ranked as the third species of *Anopheles* in Southeast Asia most attracted to humans. This is a truly anthropophilic species which under comparable conditions is more attracted to humans than to other animals. It enters houses to bite, and significant numbers may remain inside to rest by day, though probably the majority rest outside. These habits contrast with those of *An. barbirostris*, which it resembles so closely morphologically (Reid, 1968).

Some confusion exists related to the feeding preferences of *An. donaldi*. Adult behaviour has proven to be very variable and sometimes unusual for a member of the Barbirostris Group. As with some members of the Umbrosus Group, *An. donaldi* actively bites in shaded forest during the day and then enters settlements and houses at night to feed (Harrison and Scanlon, 1975). Little is known about the feeding preferences of other members of the Barbirostris Subgroup.

2.5.5 Outdoor and indoor behaviour

In a study carried out in Thailand, *Anopheles barbirostris* was the species with the highest indoor biting activity. The number of specimens collected at human bait indoors was 111 compared with over 69 found outdoors. It has not been incriminated as a malaria vector in Thailand (Abu Hassan et al., 2001).

Anopheles campestris is highly endophagic. This species have been found biting humans indoors at a ratio of 4.4 : 1 over humans outdoors (Moorhouse, 1965). Of the specimens collected by Moorhouse (1965), 85% were caught between 20:00–02:00 hours. This study also showed *An. campestris* is endophilic when pyrethrum knockdown catches revealed 51 *campestris* resting in 15 of 27 houses between 16:00-17:30 hours.

In the words of Harrison (1975), *An. donaldi* is reported to "readily bite man, but show a preference for cattle; enter houses to bite at night, but show a preference to

bite outside; leave houses before daylight; and bites in the shaded forest during daylight hours".

2.5.6 Malaria vector status

In the past, *An. barbirostris* had been incriminated as a malaria vector in different regions of Southeast Asia; however, since the recognition of the Barbirostris Group (Reid 1962) some of these reports were corrected. For instance the "*An. barbirostris*" reported as a malaria vector from the west side of the Malaysian peninsula and that from Borneo are now recognized as *An. campestris* and *An. donaldi*, respectively (Harrison and Scanlon, 1975). In Sulawesi this species has been confirmed to be a malaria vector (Lien et al., 1977). Initial evidence indicated that this species could be composed by two different forms: the ordinary *An. barbirostris* and a vector form (Reid, 1968). However a careful morphological analysis by Reid et al. (1979) concluded all varieties are conspecific and belong to a single species: *Anopheles barbirostris.*

Anopheles campestris was incriminated as a vector of malaria in Malaysia (Reid, 1968). In a more recent paper, *An. campestris* has been catalogued as an important malaria vector in Thailand, related to an outbreak in the province of Sa Kaeo. (Apiwathnasorn et al., 2002).

2.5.7 Filaria vector status

It was not until the discovery of a new filarial parasite of man in West Timor, Indonesia, that *An. barbirostris* was recognized as the vector of filaria (Atmosoedjono et al., 1977a). This species is also likely to be the main vector in East Timor due to its widespread occurrence and anthropophilic behaviour, showing exophagic and endophagic biting habits (WHO, 2000). *Anopheles campestris* and *An. donaldi* have been incriminated as vectors of filariasis. *Anopheles donaldi* is the vector of periodic *Brugia malayi* in the inland hills areas of Peninsular Malaysia and is probably a vector of this parasite in Borneo (Reid 1968). Peninsular Malaysian *An. donaldi* were found to have a very low level of susceptibility to *Wuchereria*

bancrofti (Wharton et al., 1963); however, de Zulueta (1957) considered this as a vector of this parasite in Borneo. It has been also described as a filarial vector in Flores (Atmosoedjono et al., 1977b).

Since the research for this thesis was completed, Saeung et al. (2007) have carried out a molecular phylogenetic study with laboratory reared specimens of the Barbirostris Subgroup. A discussion of their findings in relation to those from this thesis is given in Chapter 8.

CHAPTER 3

3 Material and Methods

3.1 Specimens received

Specimens were obtained either through Dr. Ralph Harbach of the Natural History Museum, London or from Ms. Kobkan Kanjanopas, a senior entomologist of the Ministry of Health of Thailand. In addition two specimens were obtained from Vietnam. The list of localities where mosquitoes were collected is given in Table 3.1. Field workers were given the following guidelines for killing and storage of specimens.

1. Once collected, leave mosquitoes in a tube in sunlight. This procedure is used to kill and dry mosquitoes at the same time. Alternatively they can be left in a freezer but care has to be taken to avoid condensed moisture inside.

2. Place dead mosquitoes in 1.5 mL eppendorf tubes (a maximum of 5 mosquitoes per tube).

3. Put tubes containing mosquitoes in plastic bags containing silica gel and these bags in plastic containers provided. This is to absorb all the water and preserve them dry. Please observe that silica gel should always be blue. If it appears pink, please proceed to change it.

4. Mosquitoes were kept in the lab at room temperature. Silica gel colour was regularly checked.

Unfortunately, the field staff did not follow instructions 2 and 3 correctly. Instead of using eppendorf tubes, dozens of specimens were placed in plastic tubes. Moreover, the silica gel came pink. As a result many of the specimens had poor quality DNA, which was reflected in the difficulties found in PCR amplification.

Table 3.1 List of dates and sites of collection of this study, collection methods and specimen codes.

Collection site	Date of collection	Number collected	No sequenced (No processed)	Collection method	Specimen codes
Indonesia					
Palu, Central Sulawesi. Sulawesi.	March, 1993	9	0 (9)	Na	pal (n)
Padang, Cermin-Lampung. Sumatra.	September, 1993	8	6 (8)	Na	l (n)
Aebubu-Paga-Flores	March-April, 1994	18	0 (18)	Na	f (n)
Lengkong Sukabumi. Western Java	March, 1994	10	1 (10)	Na	s2 (n), s7 (n)
Pelabuhan, Ratu-Sukabumi. W.Java	February, 2003	3	0 (3)	Na	s7 (n)
Jambu, Central Java.	April, 1994	13	0 (12)	Na	j (n)
Lasung, Kalimantan	May, 2003	13	9 (13)	Na	a(n)
Kalimantan	No data	4	4 (4)	Na	k (n)
Thailand					
Aranyaprathet. Sakaeo	May, 2006	113	57 (85)	Animal bait and resting	bsk(n), csk (n), S(n), **
Mueng, Trat	April, 2006	296*	37 (84)	Animal bait and resting	btr (n), ctr (n), T(n) **
Mae Sariang, Mae Hong Son	August, 2000	60	40 (24)	larvae as progeny broods and no data for adults	th(n)
Mae Sod. Tak	2006	2	2 (2)	adults as progeny broods	ta(n)
Maeramat, Tak	September, 2006	22	7 (14)	Na	ta(n)
Sangkhlaburi Kanchanaburi	September, 2006	22	5 (11)	Na	kh(n)
Vietnam					
Phuoc Long, Binh Phuoc	May, 2002	1	1 (0)	Na	v1
Tuy Phong, Binh Thuan	October, 2003	1	1 (1)	Na	v2

* Specimens were in bad condition as many were placed in the same tube.

** Specimen codes starting with "b" were identified as *An. barbirostris* in the field. Those with starting with "c" as *An. campestris*. In capital letters, those morphologically identified.

Na: information not available

3.2 Collection sites

3.2.1 Sampling sites in Indonesia

The Indonesian archipelago is located between the Indian and the Pacific Oceans, close to the equatorial line and bridges two continents. It is composed of 17,508 islands with approximately 6000 of those inhabited. The major and most populated islands are Kalimantan (the Indonesian portion of the island of Borneo), Java, Sumatra, Sulawesi and Western New Guinea. Approximately 80% of the total population live in rural areas (Suroso et al., 2006). Indonesia is mainly mountainous with about 400 volcanoes. The land area is generally covered by thick tropical rain forests, where fertile soils are continuously replenished by volcanic eruptions. The weather and climate of Indonesia is typical of equatorial regions. Rainfall is heavy and well distributed around the year almost everywhere. Most places receive 180-280 mm per month throughout the year (NOAA, 2007). Although collections of specimens were carried out in six different regions of Indonesia, we could only obtain results with specimens from Sumatra, Java and Kalimantan. Details of these collection sites are given below.

Lampung, Sumatra

This province is located in the southern tip of the island of Sumatra, the largest volcanic island in the Indonesian archipelago (Bellon et al., 2004). Lampung has undergone the largest changes in land use when compared with other regions of Indonesia due to a governmental colonization programme which led the population to increase more than 10-fold, rising from 376,000 inhabitants in 1930 to 5,250,000 in 1986 (Imbernon, 1999).

Pelabuhan, Ratu-Sukabumi, Western Java

Java is a densely populated island in the Republic of Indonesia . Malaria has all but vanished from the populous cities but is still commonly found in some rural areas, with widespread rice cultivation contributing to the problem. Climate is warm (21-28 °C) and humid (70-95°C) with average rainfall up to 1500 mm. The main rainy season is in November to April.

Figure 3.1 Map showing the collection sites of this study

Collection sites in Southeast Asia.

1) Mae Hong Son, 2) Tak, 3) Kanchanaburi, 4) Sa Kaeo, 5) Trat, 6) Bin Phuoc, 7) Bhin Thuan, 8) Sumatra, 9) Western Java, 10) Central Java, 11) Flores, 12) Kalimantan, 13) Sulawesi

Lasung (Lasoen, Lasoeng), South Kalimantan

Kalimantan is the Indonesian part of the island of Borneo. Being crossed by the Equator, Kalimantan climate is warm all year, rainy seasons occur from January to March. The average temperature is 33.1 °C and the lowest is 22.0 °C. The average humidity ranges from 79% to 86%. Collections were carried out in Lasung place, situated in Kota Baru, South Kalimantan.

3.2.2 Sampling sites in Thailand

The kingdom of Thailand is situated in the centre of the Southeast Asian mainland. Laos and Cambodia border it in the east, Mynamar in the west and Malaysia in the south. The country is divided into 76 Provinces, which are the first order administrative units.

Thailand has two types of climate. A savanna type climate, found in the north, north-east and central region; have a climate with three distinct seasons: rainy, from June to October; cool, from November to February; and hot of highest temperatures and sunny weather from March to May. The southern region has a characteristic tropical rainforest climate, a number of microclimates can be found. In this region, temperature is almost the same all the year, this is on average 28 °C, with March and April normally the driest months of the year.

Collection sites in this country are all located in border regions, where malaria is still prevalent.

Aranyaprathet, Sakaeo (Sa Kaeo, Sakaew)

This is one of the six districts of Sakaeo province, located in the east of the country, on the border with Cambodia. The north of the province is covered with the forested mountains and the south is a mostly deforested foothill plains region.
Rainy season occurs from June to October. Main crops are rice, followed by maize, cassava and sugarcane (Apiwathnasorn et al., 2002). Sakaeo was selected as asampling site, taking into consideration the dramatic increase of malaria cases in the recent years. It was regarded as a low-risk area until 1996, and it currently

reports one of the 10 provinces with highest malaria indices in Thailand (Thimasarn et al., 1995).

Trat

This province is located in the southeastern subregion (subregion 4), in the border with Cambodia, close to the Gulf of Thailand. This province has reported the highest malaria incidence and the highest number of imported cases (Thimasarn et al., 1995), with most of these cases contracted in Cambodia. This province includes some hilly forested regions where malaria control is difficult to achieve.

Mae Sariang, Mae Hong Son

Mae Sariang is a small mountainous district located in the Mae Hong Son province, alongside the Yuam river in the northern subregion in Thailand (Rattanarithikul et al., 2005), in the border with Myanmar. This province is one of the poorest of Thailand, mainly populated by hill tribe minorities, dedicated to growing rice. The area is commonly affected by floods.

Tak province

Tak province is located in the western second subregion (Rattanarithikul et al., 2005), in hidden mountainous area at the border with Mynamar, with the Moei river as a natural boundary. In this province, two collection sites are located:

Mae Sod (Mae Sot) district, Tak

This area is of particular interest for malaria control programmes due to having the highest malaria incidence in the country (Thimasarn et al., 1995) and high levels of resistance levels of malaria by *Plasmodium falciparum* (Tasanor et al., 2006).

Maeramat (Mae Ramat) district, Tak

It is one of the eight districts of the Tak province, population are mainly dedicated to field crops.

Sangkhlaburi district, Kanchanaburi

Kanchanaburi is Thailand's largest province, located contiguous to Tak province, in the western subregion, also in the Thai-Mynamar border. Sangkhlaburi is a rural district located 230 km to the north of the province. This region includes large valleys, plains, caves and waterfalls and many protected areas (Rattanarithikul et al., 2005). This province contains a number of malaria-endemic remote villages along the Thai-Mynamar border (Thimasarn et al., 1995)

3.2.3 Sampling sites in Vietnam

Officially known as the Socialist Republic of Vietnam, is the easternmost nation on the Indochinese Peninsula. It borders China to the north, Laos to the northeast, and Cambodia to the southeast. Population is composed of various ethnic minority groups, living from subsistence farming, mainly rice and forest labour. Cases of *Plasmodium vivax* and *P. falciparum* occur in this region, with the majority of the cases taking place from October through December-January, during and after the rainy season (Hung et al., 2002).

3.3 DNA extraction

Prior to DNA extraction mosquitoes were observed in the microscope. Spare segments of body like legs, proboscis, antennae, etc were separated to avoid cross contamination among specimens. Presence of fungus was also observed.

The DNA extraction method was based on that described by Ballinger-Crabtree et al. (1992) with the modification used by Townson et al. (1999). This involved a more through processing of DNA through phenol/chloroform/isoamyl alcohol (Ballinger-Crabtree et al., 1992; Townson et al., 1999).

Reagents:

Lysis buffer, which was prepared using the following reagents:

100 mM Tris-HCl

- 50 mM of Sodium Chloride
- 50 mM Ethylenediaminetetraacetic Acid (EDTA)
- 0.15 mM Spermine
- 0.5 mM Spermidine

5 µL of a solution of Proteinase K

10% Sodium dodecyl sulphate (SDS)

Phenol:chlorophorm:isoamyl alcohol (25:24:1)

Chlorophorm:isoamyl alcohol. Ratio 24:1

Ethanol 70% and100%

10 M Ammonium acetate

Other solutions:

TE buffer (*pH 8.0)*

10 mM Tris –HCl (*pH 8.*0)

1 mM EDTA (*pH 8.0*)

3.3.1 Procedure:

- Grind individual mosquitoes/larvae/mosquito legs with a pestle in an eppendorf tube containing 270 µL lysis buffer, 5 µL of Proteinase K and 30 µL of 10% SDS.
- Incubate overnight at 50 °C. In case of larvae maintained in alcohol, they were cleaned, before extraction, in double distilled water per 24 hours changing the water twice, in the process.
- Spin tubes briefly.
- Add equal volumes of phenol: chlorophorm:isoamyl alcohol. Ratio 25:24:1. Components of this reagent form phases; it was shaken before use.
- Rotate gently for 20 minutes.
- Spin for 10 minutes at 14000 rpm at room temperature.
- Transfer upper phase to a fresh eppendorf and add equal volumes of chlorophorm:isoamyl alcohol. Ratio 24:1.

- Mix gently and spin for 5 minutes at 14000 rpm room temperature.
- Remove upper phase and transfer to a fresh eppendorf.
- Add 0.2 volumes of 10M ammonium acetate, mix very well by inverting the tube slowly several times.
- Add 2 volumes of 100% ethanol (stored at –20 °C); again mix well but gently. Place at –20 °C x 15 minutes.
- Spin at 4 °C for 30 minutes at 14000 rpm.
- Carefully discard the ethanol, avoid dislodging the pellet of DNA and wash with 200 µL of 70% ethanol (stored at –20 °C); the pellet should be visible either glassy, grey or pink.
- Spin at 4 °C x 5 minutes at 14000 rpm.
- Discard 70% ethanol (stored at –20 °C) carefully and allow the pellet to air dry for approximately 20-30 minutes at room temperature.
- Resuspend pellet in 200 µL of water or TE buffer (for whole adult or larval mosquitoes) or 20-30 µL (for mosquito legs).
- Samples were storage at 4 °C (a 20 µL aliquot) and - 80°C (remaining sample).

This is a variation of the method of Ballinger-Crabtree et al. (1992), this method incorporates the use of chloroform, taking advantage of the fact that proteins can be more efficiently denatured when two different solvents are used. It is also used to inhibit RNase activity. In addition, isoamyl alcohol reduces foaming during extraction (Maniatis et al., 1982).

DNA extraction method "DNA easy Blood and tissue kit" by QUIAGEN was also tested, however no clear advantages were observed over the method of Ballinger-Crabtree modified; therefore, the latter was selected for all DNA extractions.

3.4 Estimation of DNA concentration

DNA was then measured with the NanoDrop® ND-1000 Spectrophotometer.
1 µL of DNA sample is pipetted onto the end of a fiber optic cable. A second fiber optic cable is then brought into contact with the liquid sample causing the liquid to

bridge the gap between the fibre optic ends. Then the spectrometer utilizing a linear CCD array, analyze the light after passing through the sample. The instrument is controlled by special software run from a PC.

The reading at 260 nm allows calculation of the concentration of nucleic acid in the sample meanwhile; proteins show an absorption peak of 280 nm.

260/280: This ratio of absorbance was used to assess the purity of DNA and RNA. A ratio of ~1.8 is generally accepted as "pure" for DNA. If the ratio is appreciably lower, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm.

260/230: This was a secondary measure of nucleic acid purity. The 260/230 values for "pure" nucleic acid are often higher than the respective 260/280 values. They are commonly in the range of 1.8-2.2. If the ratio is appreciably lower, this may indicate the presence of co-purified contaminants.

Measure of DNA was used to evaluate the efficiency of the extraction and purification methods.

3.5 PCR amplification of the Cytochrome Oxidase I (COI) mitochondrial gene region

A range of different primers was needed to amplify this mitochondrial region. Primers bound with different efficiencies due to presence of nucleotide substitutions. Primers designed and PCR conditions are described below:

3.5.1 Amplification of fragment COI-CULR

COI (Forward)	5' (TTGATTTTTTGGTCATCCAGAAGT)3' (Tm 64.3 °C)
CULR (Reverse)	5' (TAGAGCTTAAATTCATTGCACTAATC)3'(Tm 60.5 °C)

Primers COI and CULR (Crozier et al., 1989; Roehrdanz, 1993) modified by Townson (personal communication) amplified a fragment of about 880 bp. Reaction mixture (50 µL) contained QIAGEN reagents as follow: 1x reaction buffer, containing Tris-Cl, KCl, and $(NH4)_2SO_4$; 0.2 mM of dNTPs; 125 nM of each primer; 1 unit of DNA polymerase; 1.5 – 3 mM of $MgCl_2$ (depending on DNA quality) and finally 1-2 µL of DNA template. Thermal profile was as follows: 94 °C for 5 min, followed by 35- 40 cycles of 95 °C x 40 s., 50 °C x 1 min of annealing temperature 72 °C x 1 min and a final extension temperature of 72 °C x 10 minutes.

3.5.2 Amplification of fragment UEA3-UEA4

UEA3 (Forward) 5' TATAGCATTCCCACGAATAAATAA (Tm 60.1 °C)
UEA4 (Reverse) 5' AATTTCGGTCAGTTAATAATATAG (Tm 54.2 °C)

Primers UEA3 and UEA4 (Lunt et al., 1996) were located upstream primers COI-CULR (See primers map). The 25 µL reaction mixture contained: 0.2 mM of dNTPs, 125 nM of each primer, 2 units of DNA polymerase and 1.95 mM $MgCl_2$. The following thermal profile was used: an initial denaturation temperature of 94 °C for 5 min, followed by 28-35 cycles of 95 °C x 40 s, 45-50°C x 1 min and 72 °C x 1 min.

3.5.3 Designed primers

The following primers were designed using programme Primer3, v. 0.3.0 (available http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Basic primer properties, according to theoretical information, were also considered in their design. They were used with specimens of difficult (low yield PCR) or impossible amplification (not amplified to date).

SEAF (Forward) CTAGAAGTATAGTAGAAAATGGGGC (Tm 59.2 °C)
SEAR (Reverse) TCATACAATAAATCCTAATAAACCAA (Tm 58.4 °C)
Based in DNA sequences of specimens collected from Southeast Asia.

Kal COI (Reverse) TGGAAATGGGCAACAACATA (Tm 63.6 °C)
Based in sequences of specimens from Lasung, Kalimantan, Indonesia (1rst batch).

BarF1 (Forward) CACATATTATTACTCAAGAAAGAGGA (Tm 58.4 °C)
BarR1 (Reverse) CAGGAAAATCAGAATATCGTCGAG (Tm 64.6 °C)
Based in sequences of specimens morphologically identified as *An. barbirostris.*

Optimal conditions of a PCR of these primers were similar to those defined for primers COI-CULR: A 25-50 µL reaction mixture containing, 0.2 mM of dNTPs, 125 nM of each primer, 1 unit of DNA polymerase, 1.5-3 mM $MgCl_2$ and 1-2 µL of DNA template. The PCR amplification conditions were defined as follows: Initial denaturation temperature of 94 °C for 5 min, followed by 30 - 35 cycles of 95 °C x 40 s., 50-55 °C x 1 min of annealing temperature 72 °C x 1 min and a final extension temperature of 72 °C x 10 minutes.

3.5.4 Degenerate Primers

Degenerate primers were used in a last attempt to sequence specimens that could not be amplified. Sequence data obtained in this project and available data from Genbank from *An. coustani, An. pullus* and *An. sinensis* (Hyrcanus Group) and *An. barbirostris s.s.* were retrieved and aligned. Substitutions were used to design degenerate sites.

KAL3 TGAGTCCCRTGTATTGTWGC (Tm 60.1 °C)
KAL2 TACATARTGGAAATGRGCWAC (Tm 61.0°C)
Being W: A/T and R: G/A.

Since these primers were not useful in the amplification of difficult samples, they were not used in routine experiments. For this reason, PCR conditions and thermal profiles are not described in detail, however they were similar to those used for primers COI-CULR.

3.6 DNA purification

After PCR amplification, products were concentrated and purified using QIAquick PCR Purification Kit Protocol (QIAGEN). This method is used to purify single- or double- stranded DNA fragments from PCR and other enzymatic reactions. In this project, it was employed to purify PCR products prior to being sent for sequencing. It was only used for COI amplification products.

This method takes advantage of the fact that DNA adsorbs to the silica-membrane (in the column) in the presence of high salt while contaminants pass through it. Impurities are efficiently washed away, and the pure DNA is eluted with Tris buffer or water. DNA fragments ranging 100 bp-10 kb are separated from primer-dimers, nucleotides, polymerase and other contaminants.

Procedure was follow according to QIAquickSpin Handbook, available at: http://www1.qiagen.com/literature/handbooks/PDF/DNACleanupAndConcentration/QQ_Spin/1043789_HB.pdf. An exact description of the procedure followed is detailed below:

Protocol

1. Add 5 volumes of Buffer PB to 1 volume of the PCR sample in an eppendorf tube and mix.
2. To bind DNA, apply the mix to the QIAquick column and centrifuge for 60 s. Discard flow-through.

3. To wash, add 0.75 ml of Buffer PE to the same column and centrifuge for 60 s.

4. Discard flow-through and place the column back in the same tube. Centrifuge the column for an additional 1 min.

5. Place QIAquick column in a clean 1.5 ml microcentrifuge tube.

6. To elute DNA add 30 -50 µl of water to the centre of the membrane inside the column; let the column stand for 1 min and then centrifuge. We used 30 µl to obtain higher concentrations.

7. Finally DNA concentration in purified products was quantified in NanoDrop® ND-1000 Spectrophotometer.

3.7 PCR amplification of the Internal Transcribed Spacer 2 (ITS2) gene region

PCR amplification was based on the protocol used to amplify ITS2 in *Anopheles minimus* (Phuc et al., 2003). The fragment of interest, ITS2 is amplified using primers aligning in regions 5.8S and 28S, which are highly conserved regions (Paskewitz and Collins, 1990).

5,8S (Forward) 5' (ATCACTCGGCTCATGGATCG) 3'
28s (Reverse) 5' (ATGCTTAAATTTAGGGGGTAGTC) 3'

In addition, another forward primer 5.8S (2) was used to confirm initial sequences obtained:

5.8S (Forward) 5' (TGTGAACTGCAGGACACATGGAAC) 3'

Reaction mixture components are, 1X reaction buffer, containing 1.5 mM $MgCl_2$; 0.2 mM dNTPs, 25 ng of each primers, 0.5 U of taq polymerase and 20-50 ng DNA polymerase. Thermal profile conditions were: initial denaturation temperature of 94 °C x 5 min, 30-35 cycles of 94°C x 1 min, 55°C x 2 min annealing temperature, an extension temperature of 72°C x 2 min and a final extension of 72°C x 10 min.

The ITS2 region in species of the Barbirostris Subgroup is large and internal primers were needed to obtain the entire sequence.

BIF (Forward primer) 5' GTTTCGCCCTCGGTATATCA 3' (Tm 63.6 °C)
BIR (Reverse primer) 5' CAACACACGCCCTTCAACAT 3' (Tm 65.9 °C)

Optimal conditions of a PCR were similar to those defined for primers 5.8S and 28S: 25-50 µL reaction mixture containing, 0.2 mM of dNTPs, 25 ng of each primers, 1 unit of DNA polymerase, 1.5 mM $MgCl_2$ and 1 µL of DNA template. The PCR amplification conditions were defined as follows: Initial denaturation temperature of 94 °C for 5 min, followed by 27 - 30 cycles of 94 °C x 1 min, 55 °C x 1 min of annealing temperature 72 °C x 1 min and a final extension temperature of 72 °C x 10 minutes.

3.7.1 Touchdown PCR

This technique was used to improve the quality of the PCR amplification, reducing the presence of the small spurious bands that appear together with the amplification of the region of interest. This was the case of the amplification of the ITS2 in specimens of the Barbirostris Subgroup, due to its length and the presence of internal repeats (see Chapter 5).

The reaction mixture was the same as that described for ITS2. The PCR protocol was as follows:

Initial denaturation temperature of 94 °C for 5 min, followed by 3 cycles of 94 °C x 30 s, 68 – 65 °C x 30 s and 72 °C x 1 min. A second time with 3 cycles of 94 °C x 30 s, 64 – 60 °C x 30 s and 72 °C x 1 min and a third time 3 cycles of 94 °C x 30 s 59 – 56 °C x 30 s and 72 °C x 1 min. An annealing temperature of 72 °C x 1 min and a final extension temperature of 72 °C x 10 minutes.

3.7.2 Cloning of Internal transcribed Spacer 2 (ITS2) gene region

All ITS2 (including partial fragments of 5.8S and 28S genes) fragments amplified were cloned prior sequencing. Direct sequence was not possible; probably due to the presence internal repeats (see details in Chapter 5).

System used was pGEM®-T Vector System (PROMEGA). It is based on the pGEM-5Zf(+) Vector(g) backbone. Sequence includes T7 and SP6 promoters to serve as sequencing primer binding sites or for in vitro transcription of either strand of the insert with the appropriate RNA polymerase.

Components

- 2X Rapid Ligation Buffer, T4 DNA ligase
- pGEM®-T Easy Vector (50 ng)
- T4 DNA ligase (3Weiss units/μL)
- *Escherichia coli* JM109 High Efficiency Competent Cells. Kept at -80 °C
- SOC Media: 2 g Tryptone, 0.5 g Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 5 mM $MgCl_2$, 5 mM MgSO4 and 97 mL deionized water. Preparation was autoclaved and 20mM glucose added when cool. Media was then placed in 15 mL tubes and kept at 4 °C.
- LB (Luria-Bertani) broth:1g bactotryptone, 0.5g yeast extract, 1g sodium chloride, 100 mL H2O, LB broth was autoclaved and kept at room temperature. To avoid contaminations 0.1 mL of ampicillin (100 μg/μL) was alternatively added once autoclaved and cooled.
- LB (Luria-Bertani) agar: 1L LB broth, 15 g agar. In addition 1mL of ampicillin (100 μg/μL) was added after autoclaving. After homogenization, a volume of ~ 10-15 mL was placed on plastic Petri dishes.
- X-gal 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside is a substrate for beta-galactosidase, an enzyme that promotes lactose utilization. This enzyme hydrolyzes X-Gal into a colourless galactose and 4-chloro-3-brom-indigo, which forms an intense blue precipitate.
- IPTG (isopropyl-beta-D-thiogalactopyranoside) is an analogue of lactose. It induces synthesis of beta-galactosidase by inactivating repressor *lac.*
- Primers M13F 5' (CGCCAGGGTTTTCCCAGTCACGAC) 3'. Located at position 2941 -2957 and M13R 5'(TCACACAGGAAACAGCTATGAC) 3' at 176 –197 of the T-easy vector sequence.

3.7.3 Step1: Ligation

In this process, the enzyme DNA ligase will stick together two fragments of DNA, in this case the T-easy vector (plasmid) with the ITS2 fragment amplified.

Procedure

1. Briefly centrifuge pGEM®-T Easy Vector and Control.

2. Set up ligation reactions as described below.

Component	Standard Reaction	Positive Control
PCR product	1 µL (or more)	-
2X Rapid Ligation Buffer, T4 DNA ligase	5 µL	5 µL
pGEM®-T Easy Vector (50 ng)	0.5 µL	0.5 µL
Control Insert DNA		2 µL
T4 DNA ligase (3Weiss units/µl)	1 µL	1 µL
Deionized water to a final volume of	10 µL	10 µL

Ligation mixture was then incubated at 4 °C overnight.

When initial conditions are suboptimal, e.g. with low concentration of PCR product, a ratio optimization is important to calculate the amount of PCR product (insert) to be used in ligation reaction. The formula below was used with this purpose:

$$\frac{\text{Ng of vector x kb size of insert}}{\text{kb size of vector}} \text{ x insert:molar ratio = ng of insert}$$

T-easy vectors are approximately 3 kb (size of vector) and are supplied at 50 ng/µl (ng of vector). In addition a insert:molar ratio of 3/1 provide good initial parameters,

hence established as ratio standard. Then, for a 1.5 kb ITS2 insert, the amount of insert required is

$$\frac{50\text{ng} \times 1.5\ \cancel{\text{kb}}}{3.0\ \cancel{\text{kb}}} \times \frac{3}{1} = 75\text{ ng}$$

3.7.4 Step 2: Transformation

Into the galactosidase gene of the plasmid vector a region called polylinker has been inserted, which contains many restriction sites, useful for inserting donor fragments. This region does not interfere with the enzyme translation. If donor fragment is being inserted to the polylinker, translation of the enzyme is disrupted in the vector and it cannot degrade substrate X-gal. Then, colonies observed in successful transformations are colourless.

Procedure

1. Spin ligation reaction and add 1-2 μL in a 0.5 mL tube on ice.
2. Add 25 μL of JM109 High Efficiency Competent Cells, keeping reaction mixture on ice.
3. Mix gently by flicking, avoiding excessive pippeting.
4. Live mixture in ice for 25 μL.
5. Heat shock for 45 seconds at 42 °C. Place on ice for 2 minutes.
6. Add 450 μL of SOC media.
7. Incubate at 37 °C with shaking for 1.5 hours.
8. Using a drigalski spatula distribute homogeneously 10 μL of IPTG and XGal. This has to be done at least one hour prior inoculating cells.
9. Plate 50-100 μL of transformation culture into LB agar /amplicillin/IPTG, X-gal plates. Incubate overnight.

3.7.5 Step 3: PCR amplification

White colonies, those supposed to have the inserted fragment, were selected for screening; one portion streaked on new LB/amplicillin agar plates and another portion of the same colony placed in 0.5 µL PCR tubes containing reaction mixture, which composition is detailed below:

The reaction mixture (15 µL) containing 1x reaction buffer (Bioline), 1.5 µL of Bioline taq polymerase, 0.03 mM dNTPs, 0.6mM of $MgCl_2$, 0.3µM of each primer. Thermal profile was set as follows: 95 °C for 10 min, followed by 35 cycles of 94 °C x 30 s, 55 °C x 30 s of annealing temperature, 72 °C x 1.5 min and a final extension temperature of 72 °C x 10 minutes.

Those colonies containing the insert are recognized by their size after PCR amplification. Those with size of the fragment amplified (1.5-1.7) + fragment of vector amplified (~200 bp) = 1.7-1.9 kb were selected for purification and sequencing. Cultures were transferred from agar to 2 mL LB broth containing 2 µL amplicillin and growth overnight at 37 °C.

3.7.6 Step 4: DNA purification from cultures

System used was QIAprep Spin Miniprep Kit. This is the final step of the cloning process. This protocol has been designed for purification of plasmids DNA from 1-5 mL of cultures of *E. coli* in LB medium and is based on alkaline lysis of bacterial cells followed by adsorption of DNA onto silica in the presence of high salt. More details can be found at: http://www1.qiagen.com/literature/handbooks/PDF/PlasmidDNAPurification/PLS_QP_Miniprep/1043788_HB_QIAprep_122006.pdf. In this project, protocol was followed as described below:

1. Centrifuge bacterial cells for 10 min. Discard supernatant and resuspend pelleted bacterial cells in 250 µl Buffer P1 (containing RNase). No cell clumps should be visible after resuspension of cells.

2. Add 250 µl Buffer P2 and mix thoroughly by inverting the tube 4–6 times, until the solution becomes viscous and slightly clear. Do not allow the lysis reaction to proceed for more than 5 min.
3. Add 350 µl Buffer N3 and mix immediately and thoroughly by inverting the tube 4–6 times.
4. Centrifuge for 10 min at 13,000 rpm
5. Apply the supernatants from step 4 to the QIAprep spin column by pipetting.
6. Centrifuge for 60 s. Discard the flow-through.
7. Add 0.5 ml of buffer PB to QIAprep spin column by adding and centrifuging for 60 s.
8. Wash QIAprep spin column by adding 0.75 ml Buffer PE and centrifuging for 60 s.
9. Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.
10. Place the QIAprep column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50 µl Buffer of water to the centre of each QIAprep spin column, let stand for 1 min and centrifuge for 1 min. Samples were then sent for sequencing.

3.8 Sequencing of PCR products

PCR reactions were carried out with Big Dye Terminator Version 3.1 supplied by Applied Biosystems. In this method for automated fluorescent sequencing, fluorescent dye labels were incorporated into DNA extension products using 3' dye labeled dideoxynucleotide triphosphates (dye terminators). Applied Biosystems DNA sequencers detect fluorescence from four different dyes that are used to identify the A, C, G and T extension reactions. Each dye emits light at a different wavelength when excited by an argon ion laser.

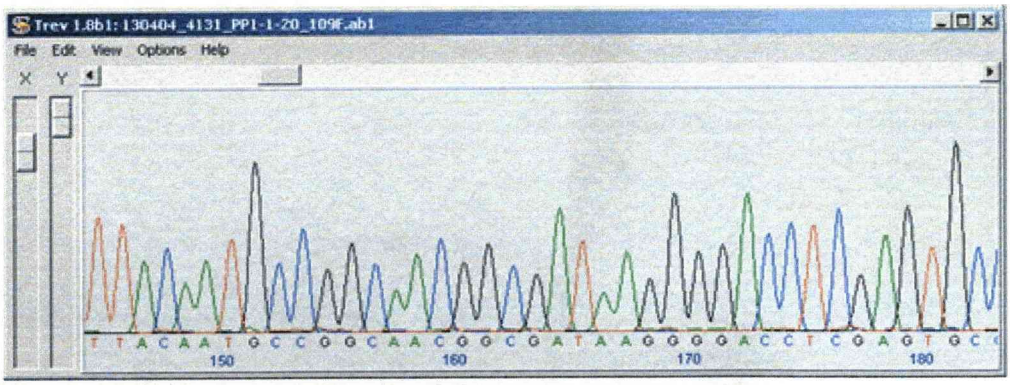
The products were precipitated and washed twice with 70% isopropanol. The dried pellets were then resuspended in Applied Biosystems Hi-Di Formamide and run on an Applied Biosystems 3130xl Genetic Analyzer using a 50cm capillary array

containing POP-7 polymer. For each specimen, forward and reverse sequences were obtained.

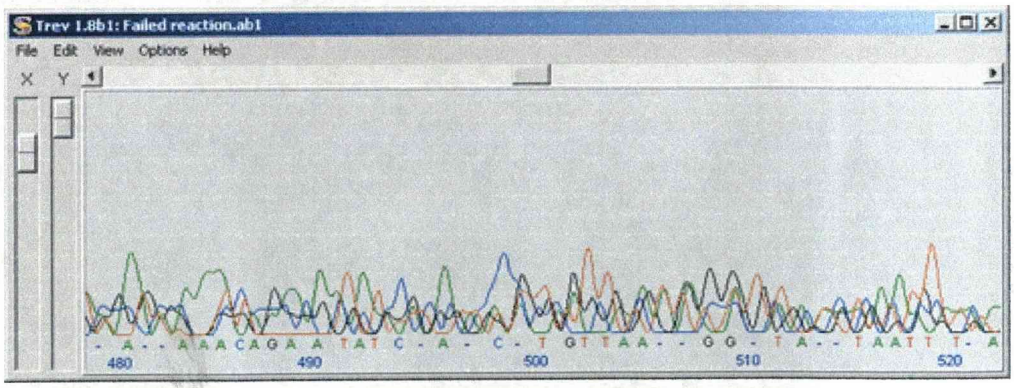
3.9 Sequence analysis

Prior to analysis, chromatogram was observed to validate sequence information. For this purpose, I used programme Bioedit version 7.0.5.3. (This is a biological sequence editor that runs in Windows 95/98/NT/2000 and is intended to provide basic functions for protein and nucleic sequence editing, alignment, manipulation and analysis). It is available at http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Peaks in chromatogram had to be clear and without background (more than one peak in the same position). This programme allowed me to judge the quality of the sequence and manually assign bases that the software was been unable to call. Those sequences with ambiguous unordered chromatogram were not employed in the analysis.



Example of an optimal chromatogram of nucleotide sequences



Example of chromatogram not used in analysis

3.9.1 Sequence alignment

Using Bioedit, forward sequence was aligned with the "reverse complement" of reverse sequence and then compared to obtain a single consensus sequence.

A multiple alignment was then carried out with all consensus sequences obtained, followed by a protein sequence alignment. As mentioned previously, Bioedit was the programme selected to perform these tasks.

3.9.2 Measuring genetic distance

The simplest measure of distance between two pair of sequences is to count the number of nucleotide sites at which the two sequences differ (Page and Holmes, 1998) With this purpose, we used programme DnaSP, DNA Sequence Polymorphism, available online in the address: http://www.ub.es/dnasp/. Transition and transversion rates were also calculated with this programme.

To calculate the number of nucleotide substitutions and therefore evolutionary distances, mathematical models were used. The efficiency of models Jukes Cantor (JC) and Kimura Two-parameter was tested with data obtained.

3.10 Phylogenetic inference

The following were the methods employed in the analysis of sequence data.

3.10.1 Distance methods

- Cluster analysis (UPGMA and WPGMA) *Unweighted-pair group method with arithmetic means* and the *weighted- pair method with arithmetic means*
- Minimum evolution (ME)
- Neighbour joining (NJ)

3.10.2 Discrete methods

- Maximum parsimony
- Maximum likelihood

The program MEGA 3.1 http://www.megasoftware.net/mega.html was used to computationally obtain phylogenetic trees using these methods. The evaluation of their applicability is discussed in following chapters.

3.10.3 TCS analysis

The TCS software package is distributed freely and is available at http://bioag.byu.edu/zoology/crandall_lab/programs.htm. This program collapses identical sequences into haplotypes and calculates frequencies of the haplotypes in the sample. A more detailed explanation is presented in Chapter 4.

3.11 Morphological examination of specimens

Using keys by Harrison and Scanlon (1975) and Reid et al. (1968), a morphological examination was carried out, using a MEIJI techno binocular stereo microscope. Specimens were not mounted, they were observed in Petri dishes, using thin forceps for this purpose. Photographs of the specimens were taken to keep a record of the most important details observed. The morphological keys and other procedures are presented in detail in Chapter 6.

CHAPTER 4

4 Phylogenetic analysis and population structure of species of the Barbirostris Subgroup based on the Cytochrome Oxidase I gene region

4.1 Introduction

This chapter describes studies in which the mtDNA subunit Cytochrome Oxidase I (COI) has been used to determine the phylogenetic relationships between members of the Barbirostris Subgroup and to determine the populations structure. A further objective has been to use the DNA sequences (both COI and rDNA ITS2).

The Barbirostris Subgroup includes six species of almost identical morphology (*An. barbirostris*, *An. campestris*, *An. donaldi*, *An. franciscoi*, *An. hodgkini* and *An. pollicaris* (Reid, 1962). The precise identification of these species requires the examination of the immature stages of associated reared adults. Since a morphological examination of progeny broods can be tedious and time-consuming, this approach is not widely used in field studies. The consequence is that reports of the distribution of species of the Barbirostris Subgroup have not been confirmed for several regions in Southeast Asia (Apiwathnasorn et al., 2002; Harrison and Scanlon, 1975; Limrat et al., 2001).

These findings will subsequently be combined with the analysis of ITS2 (Chapter 5) and the results of the morphological examination of specimens from Thailand (Chapter 6).

The mtDNA COI subunit possesses characteristics that make it particularly suitable as a molecular marker for evolutionary studies. It has been relatively well studied at the biochemical level, as it is the terminal catalyst in the mitochondrial respiratory chain (Lunt et al., 1996). There is little variation in the gene content (Harrison, 1989) and its size and structure appear to be conserved across all aerobic organisms investigated (Saraste, 1990), with the exception of some Cnidaria, where it consists of one or two linear molecules (Warrior and Gall, 1985). Based on these attributes,

the COI was chosen in the "Barcoding of life" project as a preferred marker to delimit species in the animal kingdom (Hebert et al., 2003; Roe and Sperling, 2007). In *Anopheles* species, it has been widely used in molecular systematics and phylogenetic studies (Dusfour et al., 2004; Foley et al., 2007; Lehr et al., 2005; Linton et al., 2003; Sallum et al., 2007; Torres et al., 2006).

The COI amino acid sequence has been divided into twenty-five regions comprising five structural classes (Lunt et al., 1996). These regions evolve at different rates, and the patterns of sequence variability seem associated with functional constraints on the protein. The COOH-terminal was found to be significantly more variable than internal loops (I), external loops (E), transmembrane helices (M) or the NH2 terminal. The central region of the COI (M5-M8) has lower levels of variability, which is related to several important functional domains in this region (Lunt et al., 1996).

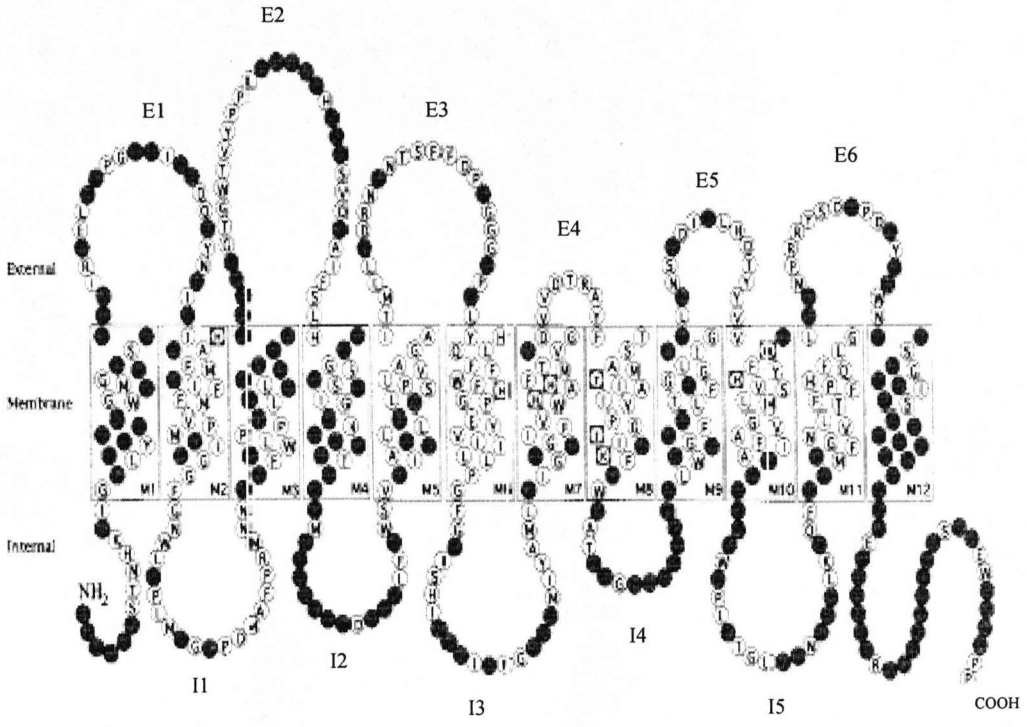
The variability of the different regions of the COI Sub-unit needs to be considered in inferring phylogenetic relationships among organisms. The most variable regions are more suitable for low-level analyses such as intraspecific variation or studies in closely related species; on the other hand, regions with lower levels of sequence variability among organisms would be used for higher-level evolutionary studies, for example, at genus or family level.

In this study, the region amplified comprises a fragment from internal loop 3 to the COOH terminal region (see fig. 4.1). This fragment includes some of the most variable regions of the COI according to Lunt (1996). To determine if the phylogenetic signal in this region is suitable for distinguishing *Anopheles* species from the Barbirostris Group, an analysis of the evolutionary rates in different structural regions were carried out and are described in this chapter.

Although Bazin and collaborators doubted the value of ecological and biodiversity studies (Bazin et al., 2006), COI has proved the most useful molecular marker in studies of the population structure of *Anopheles* species (Mirabello and Conn, 2006a; O'Loughlin et al., 2007; Walton et al., 2000). Furthermore, in earlier studies

of the Barbirostris Subgroup (Baskoro, 2001), COI sequences were found to be useful in defining clades within the Subgroup.

Figure 4.1 Schematic representation of the structural regions of the COI internal loops (I), external loops (E) and transmembrane helices (M). Taken from Lunt et al. (1996)



This chapter reports the following results

1. The species composition of the Barbirostris Subgroup based on the analysis of the COI gene fragment.

2. The genetic structure of species and populations within the Barbirostris Subgroup.

3. To demographic history of species within the Barbirostris Subgroup

4. Whether the nucleotide substitutions in COI in the Subgroup followed a neutral or non-neutral model of evolution.

4.2 Material and Methods

4.2.1 Specimens

Specimens morphologically identified as *Anopheles barbirostris* were sent from Kalimantan, Sumatra and Java in Indonesia and Mae Hong Son, Kanchanaburi and Tak in Thailand. Samples from Sa Kaeo and Trat included *An. barbirostris and An. campestris* specimens.

4.2.2 Measuring DNA polymorphism

The procedure for DNA extraction, PCR amplification and sequence analysis are detailed in chapter III (Material and methods). This section focuses on the approaches used to analyse the data; including tests of neutrality, phylogenetics and population genetic structure.

For sequence data, the extent of polymorphism is determined by calculating the average number of pairwise nucleotide differences Π. Since the Π value is likely to be larger for long sequences than for short, a better measure for comparisons of DNA sequence variation among loci is to calculate the nucleotide diversity π (Nei and Li, 1979), because this measure considers the length of the sequence. This is estimated as $\pi = \Pi/L$, where L is the length of the sequence studied.

Tajima (1989) used π value to estimate θ, the main parameter of the neutral theory of evolution (Mousset et al., 2004). θ represents the amount of variation of a DNA sequence in a population if evolution is entirely neutral. This parameter is estimated as $4N_e\mu$, for diploid organisms and in haploid genomes, like mitochondrial DNA, θ is estimated as $2N_e\mu$, where μ is the mutation rate and *Ne,* the effective population size. This parameter can be used to test neutrality (see Tajima's test). In this study, π and θ estimates were calculated using DnaSP (Rozas and Rozas, 1999).

4.2.3 Testing the neutrality hypothesis

The neutral theory of evolution was first proposed by Kimura (1968). This theory assumes that the majority of mutations appear as a consequence of a stochastic process, genetic drift. These mutations will not affect fitness of the organism. Although this theory was first based on the consistency in rates of amino acid changes, this has also been supported by the study of degenerate genetic code, and the observation of synonymous and non-synonymous substitutions. The following tests were used to test neutrality in Barbirostris Group populations. All tests were performed using the DnaSP program.

4.2.3.1 The MacDonald-Kreitman test

MacDonald and Kreitman (1991) proposed a statistical method based on the study of polymorphisms in nucleotide sequences. If the observed substitutions are neutral, the ratios of non-synonymous to synonymous substitutions are expected to be the same within and among species because both are the result of the same evolutionary processes. Differences between species are identified as *fixed sites*. If we have species A and B, a fixed site has the same nucleotide in all sequences from species A, but a different one in all sequences from species B. Differences within species are tested through observation of *polymorphic sites*, where a nucleotide substitution can be found at any variable site present within or between species. If the number of fixed replacements is high, it suggests that these excess replacement substitutions are the result of positive selection (McDonald and Kreitman, 1991) and as a consequence, we reject the hypothesis of neutrality.

4.2.3.2 Tajima's D test

The value of θ can be calculated using the number of segregating sites or S or using π, the nucleotide diversity as explained above. According to Tajima (1989), if evolution is neutral, both parameters will give the same result. The most important difference between both values is that they are affected in different ways by natural selection. Deleterious mutants are maintained at low frequencies in a population. If

present in a population, they will strongly affect the number of segregating sites, and the value of θ_s will be more affected than the value of θ_π, since the latter is not affected by low frequency alleles. On the other hand, if the population is subject to balancing selection, an increment of allele frequencies is expected and as a result the value of π should also be increased (Page and Holmes, 1998). The advantage of this test is that it only requires polymorphism data from one species, in contrast to Mac Donald-Kreitman, which requires data from at least two species. In addition, if a population goes through a bottleneck, *D* may lead to erroneous conclusions (Tajima, 1989).

4.2.3.3 Fu and Li (1993) tests

Fu and Li's approach maintains that under negative selection, there will be high numbers of mutations in the external, younger branches, because deleterious alleles are present in low frequencies; a similar situation would occur if a new advantageous mutation is starting to be fixed in the population. Conversely if a balancing selection is operating at the locus, then the mutations would appear present in the internal branches and the external ones would be fewer in number. Thus, under positive selection the value expected for D* and F* would be negative values (Fu and Li, 1993).

4.2.4 Gene flow and genetic divergence

F-statistics was originally based in the determination heterozygosity (*H*) at different levels of population structure (Wright, 1965). F-statistics is calculated in three hierarchical levels: F_{IS} (interbreeding coefficient), F_{IT} (overall fixation index), and F_{ST} (fixation index). The F_{ST} estimate has also been extrapolated to analyse nucleotide sequences using π, nucleotide diversity values (Nei, 1973). It immediately provides an estimate of the degree of subpopulation structure in any organism. In mitochondrial DNA, F_{ST} is expected to diverge four times faster than at a nuclear locus, considering it is maternally inherited and haploid. Hence, the values of this parameter are normally four times larger for mtDNA (Page and Holmes, 1998).

The Bonferroni correction was used when several pairwise comparisons were performed simultaneously, to avoid type I errors (the error of rejecting a null hypothesis when it is actually true). In this study, it was used when comparing F_{ST} between populations. If *n* hypotheses are being tested, the statistical significance (α) used for each hypothesis separately should be α/n. Both F_{ST} and Bonferroni corrections were performed using Arlequin 3.1.

4.2.5 Nested clade analysis

Nested analysis was performed using the TCS program (Clement et al., 2000), which employs the cladogram estimation algorithm proposed by Templeton (1998). This program presents important advantages compared to traditional methods to infer phylogeny. For example, the possibility of finding an ancestral haplotype in the population is considered and it also incorporates the possibility of recombination, but this is not needed in mtDNA. These variables are not included in the traditional methods like Maximum parsimony, Maximum likelihood and Neighbour-joining.

The program calculates the network based on the most parsimonious connections of haplotypes in the following way. It collapses all sequences in haplotypes and calculates each haplotype frequency in the sample. These frequencies are used to estimate the outgroup probabilities, which correlate with haplotype age. Following a pairwise comparison of haplotypes, it evaluates the limits of parsimony (Templeton et al., 1992) using a probability of 95%. The number of mutational differences associated with this probability just before 95% cut-off is defined as the maximum number of mutational steps within a network (Clement et al., 2000). The program TCS 1.21 was used to construct the haplotype network of the specimens of the Barbirostris Subgroup.

4.2.6 Mismatch distribution

The mismatch distribution involves the analysis of the distribution of pairwise differences in a population. Populations under expansion commonly "leave a wave" in the mismatch distribution. Therefore, a smooth and unimodal distribution is

characteristic of a bottleneck or population expansion; on the other hand, if the mismatch distribution is bimodal, the population is likely to be under equilibrium, and have a constant population size (Rogers and Harpending, 1992). The mismatch distribution has been used to infer the demographic history of populations, particularly when using mitochondrial DNA, in which recombination events are rare or absent (Mousset et al., 2004). Values of the mismatch distribution were obtained with DnaSP 4.10.9 and significant deviation from the Poisson distribution was tested with a goodness-of-fit Chi square test.

4.2.7 Phylogenetic analyses

Two methods were used to infer phylogenetic trees, the Neighbour-joining tree (NJ) (Saitou and Nei, 1987) and the Maximum parsimony (MP).

The NJ tree is a clustering distance method that operates on the basis of pairwise distances between two taxa; this distance represents the estimate of the amount of divergence of these taxa. To determine the rate of substitution per site, the Kimura 2-parameter model was used (Kimura, 1980). This model takes into consideration the fact that in nature, the rate of transition per site is greater than the rate of transversions. This rate appears particularly common in mitochondrial genes (Page and Holmes, 1998). The total rate of substitution per site is given by: $\alpha + 2\beta$, where α and β are the different probabilities for transitions and transversions.

The Maximum Parsimony method operates using a matrix of discrete characters, assigning one or more characters to each taxon (i.e., sequences in this study). This method finds the tree topology for a set of aligned sequences with the smaller number of character changes, i.e. mutations (Salemi and Vandamme, 2003).

To evaluate the robustness of these methods, a bootstrap test (Felsenstein, 1985) was included in the analysis. This is a statistical method for estimating the statistical error, based on a re-sampling of the data. This procedure can be repeated hundreds of times, but it is recommended that these should be between 200 and 2000 re-samplings (Salemi and Vandamme, 2003). In this present study 500 re-samplings are used. Considerable confidence can be given if groups are supported by 70% or

more; branches supported by less than 70% should be treated with caution (Zharkikh and Li, 1992). This method has improved significantly the precision of distance methods. In this study we have carried out a bootstrap analysis in both Neighbour-joining and Maximum parsimony trees. The program used with this purpose was Mega 3.1.

4.3 Results

4.3.1 Phylogenetic analysis

4.3.1.1 Differentiation between Barbirostris and Hyrcanus Groups

Phylogenetic relationships among individuals were inferred using Neighbour-joining based on Kimura 2-parameter method for estimating nucleotide substitutions per site and Maximum parsimony to infer trees topologies. A total of 109 haplotypes of 756 bp length were obtained from 163 specimens, morphologically identified in the field as members of the Barbirostris Subgroup. The *An. gambiae* COI sequence was included as an outgroup. A bootstrap test was performed with 500 replications to determine the robustness of the tree obtained (Figure 4.2).

Neighbour-joining (NJ) and Maximum parsimony (MP) trees show clear division of specimens in two main clades (Figure 4.2). Those with yellow labels were identified as members of the Hyrcanus Group based on Blast comparisons with sequences of Hyrcanus Group species. This assumption was subsequently confirmed when the morphology of mosquitoes was examined in the Natural History Museum in London. This revealed that those individuals involved at sampling mosquitoes in the field had made mistakes with identification in sampling sites in Thailand and in Indonesia, thereby emphasising the need for clear and unequivocal methods of identification.

Based on the phylogenetic analysis of COI from the 163 specimens received as members of the Barbirostris Subgroup, 19 were identified as members of the Hyrcanus Group and 142 as members of the Barbirostris Subgroup. Further analyses

are centred on the Barbirostris Subgroup specimens. A phylogenetic analysis of the Hyrcanus Group will be described in Chapter 7.

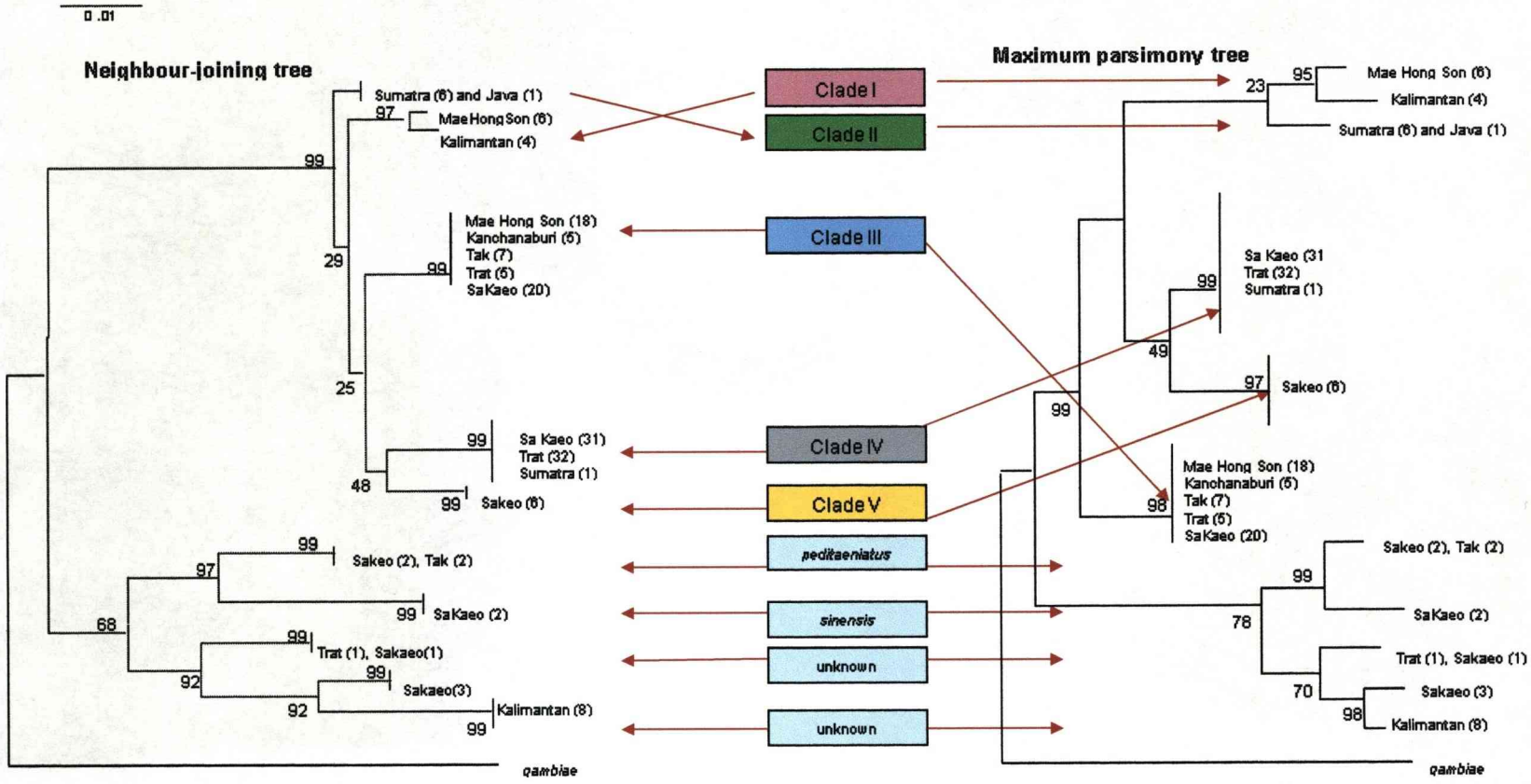
4.3.1.2 Phylogenetic relations inside the Barbirostris Group

Within the Barbirostris Group, 142 specimens were located in 5 distinct clades (named I-V) based on NJ and MP analyses. In both trees some of the branches show low bootstrap values (Figure 4.2); as a result relation among clades are poorly resolved.

The Maximum Parsimony (MP) analysis produced 105 possible topologies. Although clades I and II appear in the same cluster in MP tree, this is poorly supported (bootstrap value=23%). These clades appear separated in the NJ tree. In a similar way, the relation of clade III with clades IV and V is different in the two methods of tree reconstruction and a bootstrap test again indicates that the relationships are poorly resolved. The majority of specimens were assigned to clades III (55) and IV (64).

Considering that clade III included specimens morphologically identified as *Anopheles barbirostris* in the Natural History Museum as progeny broods, it is almost certain that these specimens are *Anopheles barbirostris*. The remaining clades comprise specimens only identified as adults and are therefore of doubtful identity. Among them, specimens from clades I and II were identified in the field as *An. barbirostris*, clade IV had a mixture of specimens morphologically identified as *An. barbirostris* and *An. campestris* and finally clade V, comprised 6 individuals identified in SaKaeo (Thailand) as *An. campestris.* A more exhaustive analysis of the relation between morphological and molecular analysis is presented in Chapter 8.

Figure 4.2 NJ (K 2-parameter) and MP trees using a 756 bp fragment from Cytochrome Oxidase I.



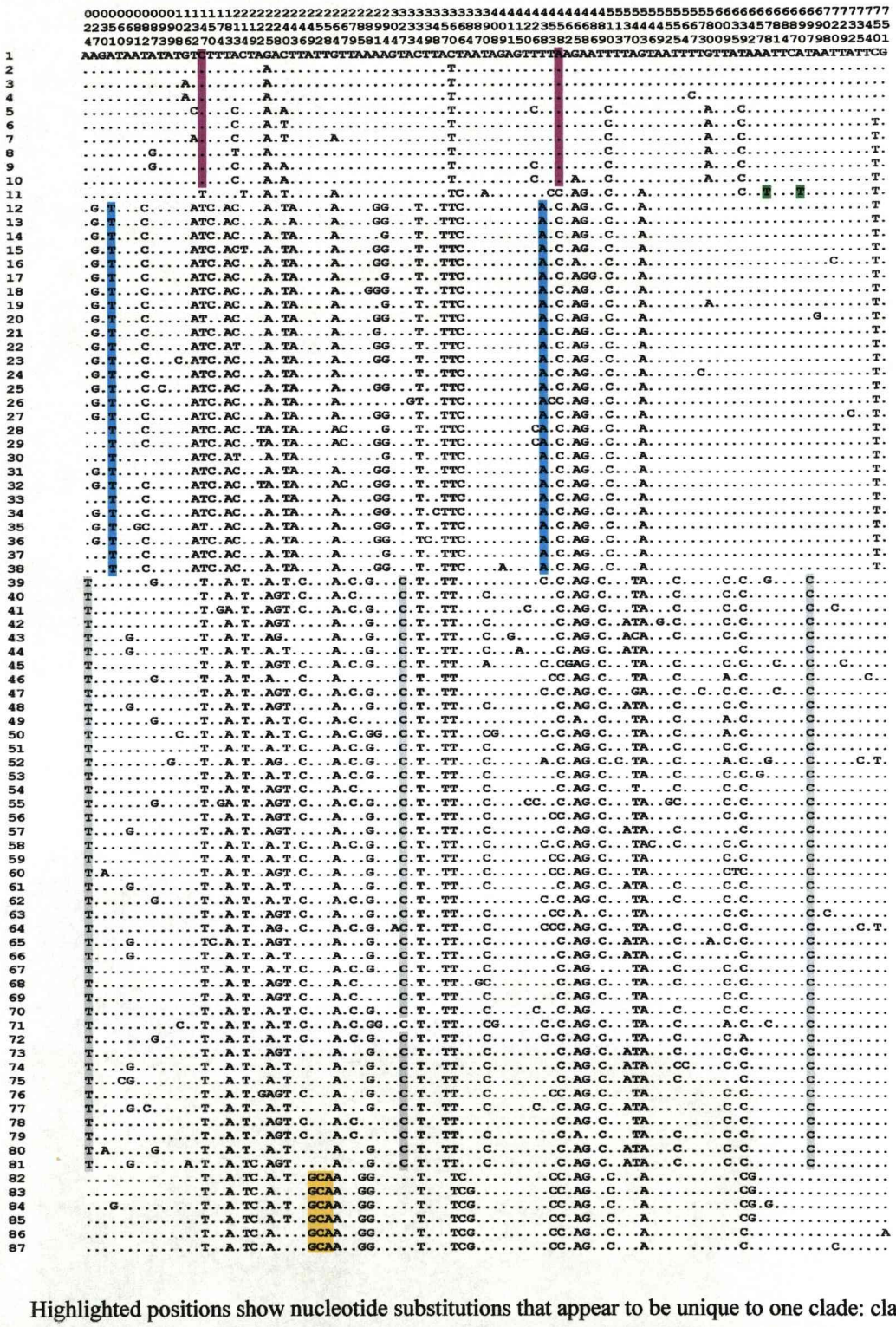
A bootstrap test of 500 replicates was used and values obtained are shown next to internal nodes. Dataset includes specimens from the Hyrcanus and Barbirostris Groups. Only haplotypes (109) were used in tree reconstructions. Two main clades are observed with both trees, one corresponding to specimens within the Barbirostris Group, organized in 5 clades and the other to Hyrcanus specimens (blue labels). The number of specimens in each clade is shown in brackets.

4.3.2 Sequence variation and genetic diversity

In this section, the genetic diversity of specimens from the Barbirostris Group is measured to address the issue of the genetic structure in this group of species. The analysis was performed on the 5 clades obtained from the phylogenetic analysis. With this objective, the estimates of S (number of segregating sites), π (the average number of nucleotide substitutions per site) and haplotype diversity (frequency of a haplotype in the dataset) are analysed.

Analysis was based on 142 sequences. All specimens were AT rich, the average content of G+C was 29.1% and A+T 70.9%. As expected for mitochondrial coding genes, the most variable codon position was the third. Eighty-eight of the substitutions were located at this position, 9.6% in the first and 2% in the second. In addition, the number of transitions was higher than transversions in all clades (Table 4.1).

Figure 4.3 Haplotype distributions in specimens from the Barbirostris Subgroup. Of the 109 haplotypes found in the dataset, 87 correspond to the Barbirostris Subgroup.



Highlighted positions show nucleotide substitutions that appear to be unique to one clade: clade I: pink, clade II: green, clade III: blue, clade IV: grey and clade V: yellow. Nucleotide position numbers are defined according to the amplified fragment of 756 bp and are shown in the top row. Haplotypes numbers are shown in the left column,

Table 4.1 Statistics of polymorphism parameters in COI. Analysis was performed for the 5 clades from the Barbirostris Group obtained from the phylogenetic analysis

Phylogenetic clades	*n*	S	*h*	Hd	k	π	st/sv
Clade I	10	15	10	1 ± 0.045	6.04	0.008 ± 0.0008	15/3
Clade II	7	0	1	0	0	0	0
Clade III	49	28	27	0.92 ± 0.03	2.4	0.0032 ± 0.0004	26/2
Clade IV	64	54	43	0.98 ± 0.007	6.63	0.0088 ± 0.0005	49/10
Clade V	6	7	6	1 ± 0.096	2.8	0.0037 ± 0.0007	7/0

n = number of sequences, S =number of segregating sites, *h*=number of haplotypes, Hd= haplotype diversity ± SD, *k* = average number of mutations, π = nucleotide diversity ± SD, sv = number of transversion, st number of transitions. DNA polymorphism analysis was performed with DnaSP 4.0; transition/ transversion ratios were calculated with Arlequin 3.1

Polymorphic sites include for 94 of the 756 sites, of which 69 were parsimony informative (substitutions present in more than one specimen) and 25 were single substitutions (Table 4.2). The distinction between clades is evident from the nucleotide substitutions shown in Figure 4.3. Twelve substitutions, located at positions 24, 51, 147, 249, 252, 258, 327, 438, 458, 678, 690 and 697, appear fixed within and exclusive to a particular clade, whereas others appear to be shared among clades.

Eighty-seven haplotypes were found in the 142 specimens of the Barbirostris Group studied in this project, using a 756 bp fragment of the cytochrome oxidase gene region. These included 10 larva obtained as progeny from 4 adult females captured in the field. Only one larva of each of the 4 adults were included in further analysis since including the entire dataset would affect predictions on their population structure.

General haplotype diversity value was 0.984 ± 0.005. This value is high, but expected considering that the entire dataset includes specimens identified as *An. barbirostris* and *An. campestris*. A detailed analysis of the DNA polymorphism and further inferences of population structure were performed, taking into consideration the results obtained in the phylogenetic analysis, which show five distinctive clades.

Of the 87 haplotypes found, 64 are singleton, that is represented by a single individual and 6 were widespread in different geographic populations, 4 in clade III and 2 in clade IV (Table 4.2 and 4.3). High haplotype diversity values were obtained in all clades (Table 4.1), with the exception of clade II that shows a single haplotype (Hap11) in seven specimens from the islands of Sumatra (6) and Java (1) in Indonesia. Conversely, clade I comprised 10 different haplotypes in 10 specimens from Mae Hong Son (Thailand) and Kalimantan (Indonesia). In clade III, some haplotypes were widely distributed; particularly Hap12, which appeared in all collections sites in Thailand, from the northernmost province of Mae Hong Son, through Kanchanaburi and Tak provinces, the western region bordering Myanmar, and into the eastern provinces of Sa Kaeo and Trat bordering Cambodia (Figure 4.4)

Table 4.2 List of haplotypes within the Barbirostris Group and the localities where they are found.

Phylogenetic clades are separated by lines. Clades I, II and III are shown in this table. Clade II has one single haplotype. In bold, haplotypes distributed in various localities; underlined, samples obtained as progeny broods.

Haplotype number	**N**	Specimens source						
		Kal	Mae	Sum	Jav	SaK	Tra	Tak	Kan
Hap 1	1	k1							
Hap 2	1	k2							
Hap 3	1	k3							
Hap 4	1	k4							
Hap 5	1		th1-1						
Hap 6	1		th1-3						
Hap 7	1		th1-12						
Hap 8	1		th1-13						
Hap 9	1		th1-4						
Hap 10	1		th1-7						
Hap 11	**7**			**111,112 113,115 133,131 132**	**s71**				
Hap 12	**14**		**th39.4 th39.3 th1.6**			**bsk33 csk5 S9.1**	**T32.1 T37.1 T29.3**	**ta19 ta21 ta25**	**kh7, kh9**
Hap 13	3					csk3 csk7 S15.2			
Hap 14	**5**		**th73.2 th73.4**			**bsk26 S11.2**		**ta24**	
Hap 15	2					bsk1 S13.1			
Hap 16	1					bsk20			
Hap 17	1		th1.2						
Hap 18	1					bsk5			
Hap 19	1		th1.11						
Hap 20	1		th1.14						
Hap 21	1		th1.5						
Hap 22	1		th1.8						
Hap 23	1		th1.9						
Hap 24	1					bsk15			
Hap 25	2					bsk13 S13.2			
Hap 26	1						btr14		
Hap 27	1					bsk25			
Hap 28	4		th41.4 th41.5 th41.2 th41.3						
Hap 29	2		th69.2 th69.3						
Hap 30	**2**		**th1.10**						**kh10**
Hap 31	2								Kh3, kh4
Hap 32	1							ta10	
Hap 33	2							ta5, ta6	
Hap 34	1					S13.3			
Hap 35	1					S15.4			
Hap 36	1					S17.1			
Hap 37	1						T31.2		
Hap 38	1					bsk17			

Kal=Kalimantan, Mae=Mae Hong Son, Sum=Sumatra, Jav=Java, SaK=Sa Kaeo, Tra= Trat, Tak=Tak, Kan=Kanchanaburi.

Table 4.3 List of haplotypes in clades IV and V and the localities where they are found.

The clades are separated by lines. In bold, haplotypes distributed in various localities.

Haplotype number	**N**	Specimens source		
		SaKaeo	Trat	Sumatra
Hap 39	1	bbs8		
Hap 40	1	bsk22		
Hap 41	2	bsk23,S24.5		
Hap 42	2	bsk24,S10.1		
Hap 43	1	bsk27		
Hap 44	2	bsk31,S27.3		
Hap 45	1		btr13	
Hap 46	1		btr23	
Hap 47	2		btr24,T38.1	
Hap 48	**5**	**S23.2**	**btr25,T35.1,T32.2,T29.2**	
Hap 49	1	csk6		
Hap 50	1		ctr4	
Hap 51	3	bsk2,bsk6,S27.5		
Hap 52	1			114
Hap 53	4	bsk4,bsk8,S11.1,S30.1		
Hap 54	1	bbs6		
Hap 55	1	bsk3		
Hap 56	2		ctr10,T30.1	
Hap 57	1		ctr8	
Hap 58	2	bsk7, S40.1		
Hap 59	1	bsk9		
Hap 60	1	bsk10		
Hap 61	1		btr8	
Hap 62	**2**	**bsk11**	**btr11**	
Hap 63	1		btr15	
Hap 64	1		btr16	
Hap 65	1		btr17	
Hap 66	4		btr18,btr22,T13.1,T36.1	
Hap 67	3		btr19,btr21,T12.3	
Hap 68	1	csk2		
Hap 69	1		ctr2	
Hap 70	1		T35.2	
Hap 71	1		T35.3	
Hap 72	1	S1		
Hap 73	1	S15.1		
Hap 74	1	S19.1		
Hap 75	1	S24.1		
Hap 76	1	S24.4		
Hap 77	1		T37.2	
Hap 78	1		T22.1	
Hap 79	1		T36.2	
Hap 80	1		T7.1	
Hap 81	1		T33.1	
Hap 82	1	S25.1		
Hap 83	1	S27.4		
Hap 84	1	S24.2		
Hap 85	1	csk11		
Hap 86	1	csk10		
Hap 87	1	csk1		

Sum=Sumatra, SaK=Sa Kaeo, Tra= Trat.

The haplotype diversity value of clade IV = 0.98 ± 0.007 reflects the numerous haplotypes found within this clade (43 in 64 specimens). This included specimens from Sa Kaeo and Trat in Thailand and a single specimen obtained from Sumatra (Table 4.3). Finally, clade V comprised only 6 specimens, all collected in Sa Kaeo.

Table 4.1 show the results of genetic diversity for COI (mtDNA) among clades of the Barbirostris Group. Values of nucleotide diversity obtained varied from $\pi = 0$ in clade II, in which the COI fragment is the same in all specimens, to higher values in other clades: clade III (0.32%), clade V (0.37%) clade IV (0.88%) and clade I (0.80%).

4.3.3 Nested clade analysis

To infer the genealogical relationships among the 87 haplotypes in the Barbirostris Subgroup sampled in this study, a nested clade analysis (NCA), was used. This is an alternative to the traditional methods for inferring tree topologies (Templeton, 1998). Traditional methods make assumptions that are not valid at a population level; such as ancestral haplotypes are no longer in a population. The NCA accept the existence of ancestral haplotypes, considering that the coalescent theory assumes that these would be the most frequent in a population (Crandall and Templeton, 1993). The program used to perform analysis is TCS (Clement et al., 2000) and results are contrasted with those obtained from the phylogenetic analysis.

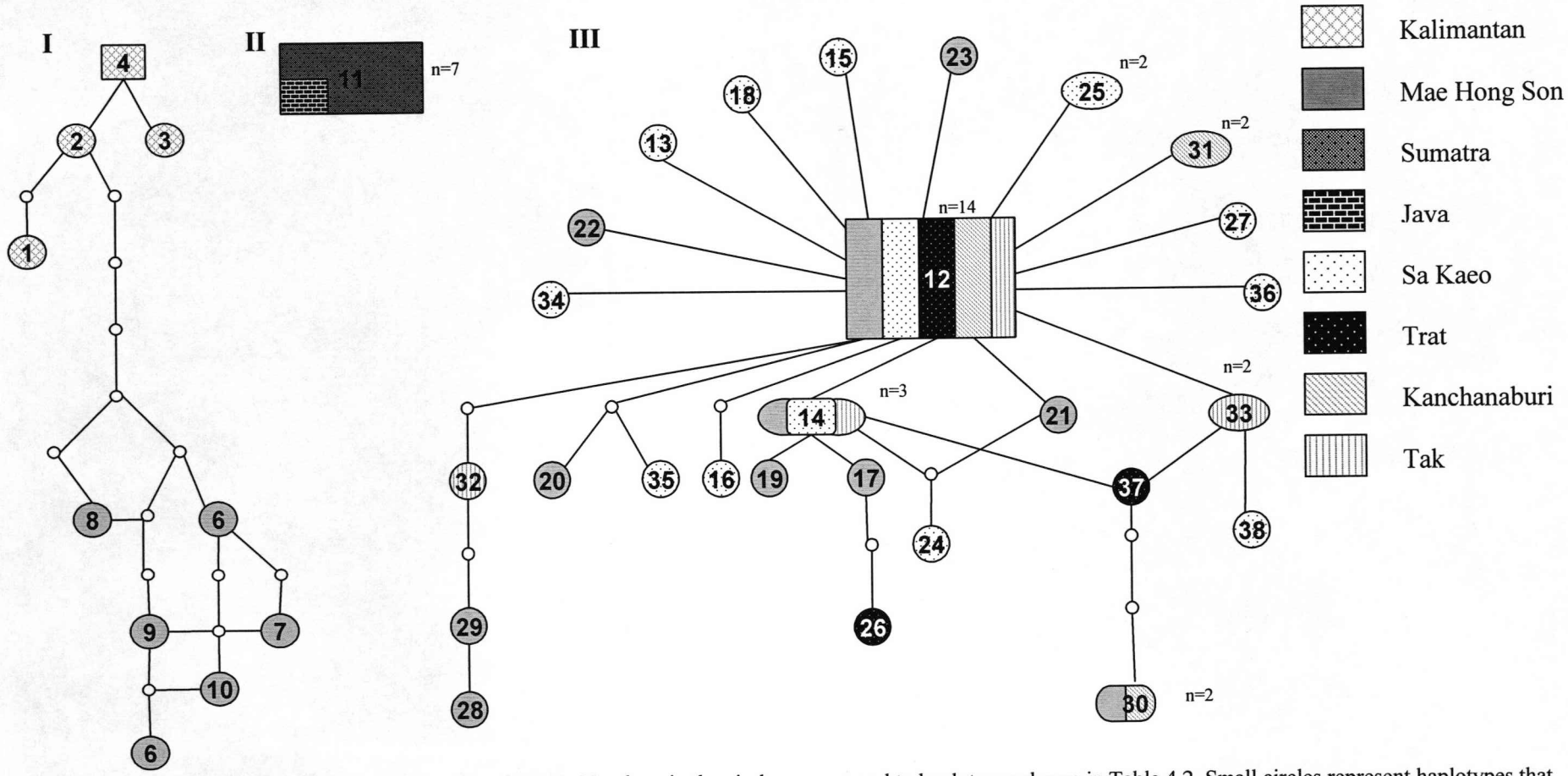
TCS analysis was carried out on sequences of specimens from the Barbirostris Group (including *An. barbirostris* and *An. campestris*). The TCS programme calculated 11 as the maximum number of mutational steps to form a network (see section 1.2.5 for further explanation). This was obtained from testing the probability of parsimony in each pairwise comparison until the probability exceeds the 95% level. Then, haplotypes which exceed more than 11 mutational steps were separated in a different network. TCS collapsed the haplotypes from the Barbirostris Subgroup into 5 different networks or clades (Figures 4.3 and 4.4). These networks were represented by the same specimens found in the 5 clades obtained in NJ and MP trees.

Specimens from clade I comprise individuals from Kalimantan and Mae Hong Son. These two populations are separated by at least 6 mutational steps. Haplotype 4, which corresponds to a specimen from Kalimantan was identified by the program as the ancestral haplotype (Figure 4.4). Mae Hong Son individuals show one ambiguous connection as evidenced by the presence of loops (Figure 4.4). This separation of geographically separate populations is consistent with the NJ and MP trees. Clade II comprised a single haplotype (11) and formed a distinctive clade (Figure 4.4). Specimens in this clade were obtained from Sumatra and Java (Indonesia).

In clade III (see Figure 4.4), the most widely distributed haplotype (12) was identified as the ancestral haplotype. Thirteen of the 26 haplotypes differ from the ancestral form by just 1 mutational step, and 10 lineages in this clade are represented by a single haplotype (see haplotypes 34, 22, 13, 18, 15, 23, 25, 31, 27 and 36 in Figure 4.4). In general, 5 was the maximum number of mutational steps observed within this clade. As a result, clade III shows a starbust pattern which may be an indication of population expansion. This is discussed further below.

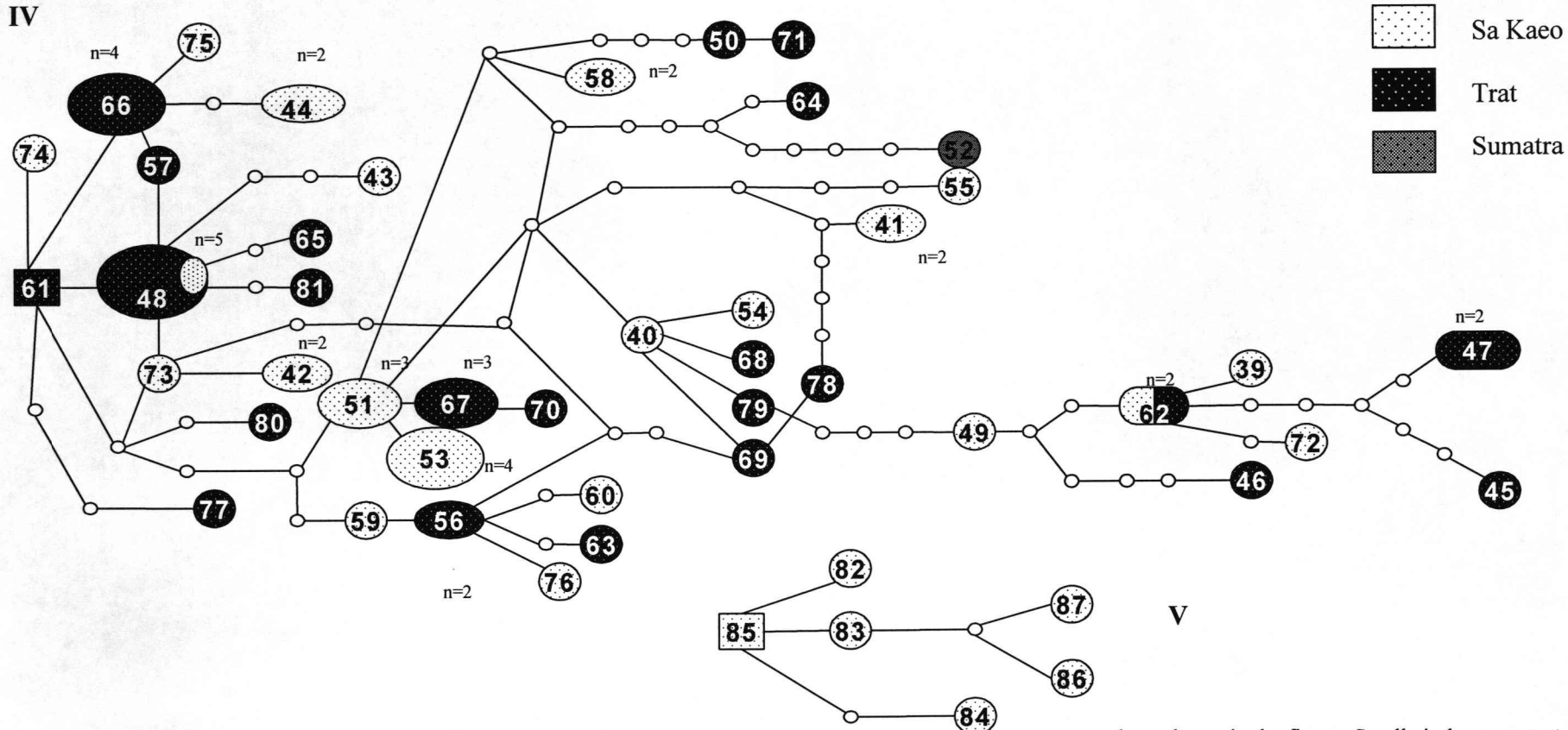
In contrast to clade III, in clade IV the ancestral haplotype lies on a node represented by a single specimen (haplotype 61) and none of the haplotypes are appreciably more abundant than the others (Figure 4.5). Lineages in this clade are made up of multiple haplotypes, with many missing nodes being inferred. Most haplotypes were found in a single geographic population, with the exception of haplotype 62, which comprised one specimen from Sa Kaeo and one from Trat. The different lineages observed in clade IV contain specimens from more than one collection site, indicating that no geographic clustering of haplotypes was observed. Numerous ambiguous connections were also observed in this network and no geographic clustering of haplotypes was observed (Figure 4.5). Finally, a parsimony network of clade V is also presented (Figure 4.5). Although the number of specimens is too small to arrive at a firm conclusion, these specimens, all collected in Sa Kaeo, were distinctive enough to be located in a separate clade.

Figure 4.4 Network analysis of Clades I , II and III estimated by TCS (Clement et al., 2000).



Rectangles in networks symbolize ancestral haplotypes. Numbers in the circles correspond to haplotypes shown in Table 4.2. Small circles represent haplotypes that have not been observed but are assumed to be intermediates. Size of circles, ellipses and rectangles represent frequency. n=1 except where shown in the figure. Links between nodes are all single mutational steps regardless of length. Localities where haplotypes were found are represented by different filling patterns (see legend at the

Figure 4.5 Network analysis of Clades IV and V estimated by TCS.



Rectangles represent ancestral haplotypes. Size of circles, ellipses and rectangles represent frequency. n=1 except where shown in the figure. Small circles represent haplotypes that have not been observed but are assumed to be intermediates. Haplotype numbers are shown within nodes; they correspond to haplotype list in Table 4.3. Links between nodes are all single mutational steps regardless of length. Geographic location of specimens is represented by different fill patterns in the nodes (see legend at the right)

4.3.4 Genetic differentiation and population structure

4.3.4.1 Estimate of genetic differentiation using F_{ST} values

In this section, the genetic differentiation among different clades and within is evaluated. With this purpose F_{ST} values (Nei, 1973) are calculated. F_{ST} values range from 0 to 1; an F_{ST} value of one would indicate complete genetic differentiation and a value of 0, no genetic differentiation. The significance of the estimated genetic distances is tested using a permutation analysis (Hudson et al., 1992) of 1000 replicates. To test the statistical significance, *p*-values are determined for a significance level (α) of 0.05. Where necessary, significant estimates were subject to Bonferroni correction for multiple comparisons to avoid type I errors.

The F_{ST} values obtained from the pairwise comparison among five clades obtained from the phylogenetic analysis are shown in Table 4.4. High levels of differentiation, as shown by high F_{ST} values, indicate that clades are clearly distinct. The lowest F_{ST} value obtained was 0.74, between clades IV and V and between II and IV. The highest value was 0.86 between clades I and III (Table 4.4). To evaluate the statistical significance of these values, a Bonferroni correction for multiple hypotheses was applied. As a result all F_{ST} obtained are statistically significant ($p << 0.00001$.

Table 4.4 F_{ST} values among clades. P-value was calculated using 1000 permutations

	Clade I	Clade II	Clade III	Clade IV	Clade V
Clade I	-				
Clade II	0.78***	-			
Clade III	0.86***	0.89***	-		
Clade IV	0.75***	0.74***	0.82***	-	
Clade V	0.79***	0.91***	0.89***	0.74***	-

*** P < 0.001

The modified α value with a Bonferroni correction for 10 pairwise comparisons is $\alpha = 0.005$, calculated p-values were all << 0.00001.

The three most populated clades were analysed separately to determine the genetic structure among subpopulations. That is the geographic localities where specimens were collected. Clade I, with 10 specimens (subpopulations Mae Hong Son and Kalimantan, Clade III with 49 specimens (subpopulations Mae Hong Son, Sa Kaeo, Trat, Kanchanaburi and Tak) and Clade IV formed by 64 specimens (Sa Kaeo, Trat and Sumatra).

Clade I show the highest F_{ST} value = 0.63 between specimens from Kalimantan and Mae Hong Son. These subpopulations appear also clearly differentiated in the nested clade analysis (Figure 4.4) and phylogenetic analysis (Figure 4.2). The *p*-value obtained from the permutation analysis (1000 replicates) for populations in this clade was $p < 0.001$ (Table 4.5), indicating that specimens from clade I differ significantly between localities.

Table 4.5 F_{ST} values obtained from the comparison between populations Mae Hong Son and Kalimantan in clade I.

	I-Mae Hong Son	I-Kalimantan
I-Mae Hong Son	0.00	
I-Kalimantan	0.63**	0.00

** p < 0.01
α = 0.05

In Table 4.6, F_{ST} estimates obtained from the comparison among the five geographic populations in clade III (Mae Hong Son, Sa Kaeo, Trat, Kanchanaburi, and Tak) are shown. F_{ST} estimates were not significantly different between any of the populations, using a Bonferroni corrected $\alpha = 0.005$; permutation test=1000 replicates. In fact, estimated F_{ST} values were not significant even without the Bonferroni correction. F_{ST} values were in general very low. This suggests that the different populations within this clade are genetically similar.

Table 4.6 F_{ST} values obtained from the comparison between five different subpopulations of clade III. Mae=Mae Hong Son, SaK=Sakaeo, Tra=Trat, Kan=Kanchanaburi and Tak=Tak.

	III-Mae	III-SaK	III-Tra	III-Kan	III-Tak
III-Mae	0.00				
III-Sak	0.056	0.00			
III-Tra	-0.064	-0.005	0.00		
III-Kan	0.003	0.112	0.067	0.00	
III-Tak	0.045	-0.01	-0.033	0.108	0.00

Bonferroni corrected α:
α = 0.05/10= 0.005

Finally in clade IV, a low level of genetic differentiation was found between populations of Sa Kaeo and Trat. F_{ST} was not significant (see below) for an α = 0.025. As in previous cases, statistical significance was calculated using 1000 permutations.

Table 4.7 F_{ST} values between geographic populations within clade IV.

	IV-Sa Kaeo	IV-Trat
IV-Sa Kaeo	0.00	
IV-Trat	0.011	0.00

Bonferroni corrected
α: = 0.05/2= 0.025

4.3.4.2 Estimate of genetic differentiation using the Kimura 2-parameter distance method (K 2-P)

The levels of sequence divergence among the 5 clades were also tested using the Kimura 2-parameter distance method (Kimura, 1980). This model calculates

distances between nucleotide sequences taking into consideration that the probability of a transition is greater than the probability of a transversion. It is expected that estimates from comparisons within species would give lower K 2-P than comparison between species. In this study, the Kimura 2-parameter was calculated for the 5 clades in the Barbirostris Group and is presented in Table 4.8.

The average K 2-P distance among clades was 2.93%, with the highest value between clades I+IV (3.6%) and the lowest between clades II and V (1.9%). Conversely, the values obtained from the comparison within clades were much lower, the highest value was obtained in clade IV: K 2-P= 0.9%.

Table 4.8 K 2-parameter values between and within 5 clades of the Barbirostris Group.

Below diagonal, the comparison between clade. Above diagonal, estimates within clades.

	Clade I	Clade II	Clade III	CladeIV	Clade V
Clade I	0.008				
Clade II	0.022	0			
Clade III	0.031	0.026	0.004		
Clade IV	0.036	0.028	0.036	0.009	
Clade V	0.032	0.019	0.032	0.031	0.004

The two genetic divergence estimators used in this project, F_{ST} and Kimura 2-parameter, agree in showing high levels of genetic differentiation between clades than within clades (Tables 4.4 and 4.8).

4.3.5 Mismatch distribution. Testing recent demographic history

A mismatch distribution is the frequency distribution of pairwise differences in a population. If this is smooth and unimodal, represented by a Poisson distribution, it is indicative of a model of sudden expansion. On the contrary, a multimodal distribution is characteristic of a population under mutation drift equilibrium (MDE)

(Rogers and Harpending, 1992). The frequency of segregating sites was compared with frequencies expected under a model of population expansion. Simulated (expected) estimates were obtained with DnaSP.

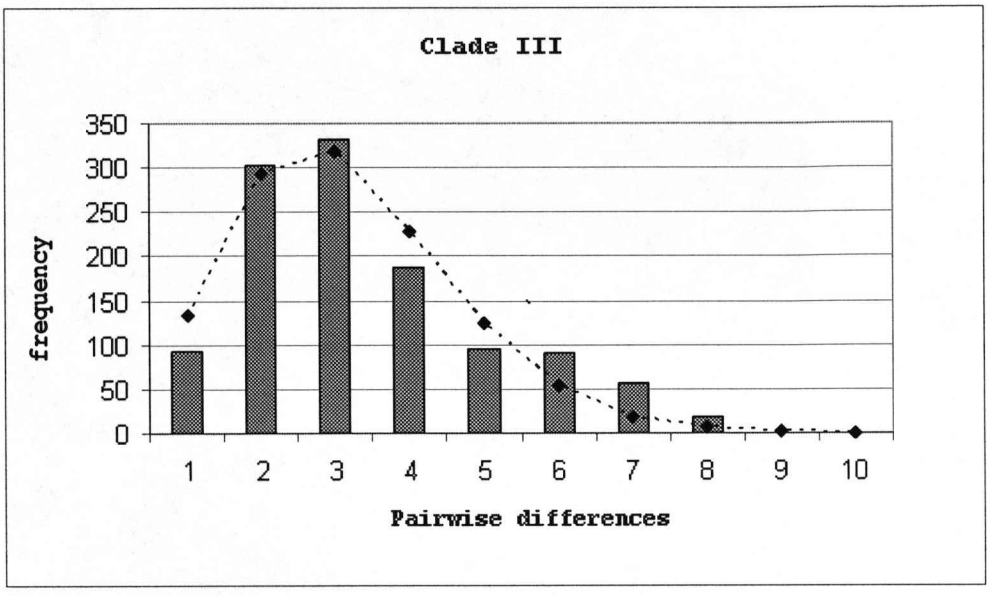
The mismatch distribution was only applied to clades III and IV because there were too few specimens in the other clades. In each case, the distribution seems to fit well with its expected distribution under population expansion (Figure 4.6), even though there was a significant deviation from a Poisson distribution (χ^2 $P<0.05$). The pattern obtained does not resemble that seen for a population in equilibrium (see fig. 5 in Rogers and Harpending, 1992).

The Raggedness Index (r) was calculated to evaluate the demographic expansion of both populations tested. This index takes larger values if the distribution is multimodal, representing a stationary population, whereas lower values are obtained if the distribution is smooth and unimodal, typical of an expanding population (Harpending et al., 1993). Raggedness indices were low, supporting the hypothesis of population expansion in both clades. In clade III the value was higher ($r < 0.056$) than in clade IV ($r < 0.0082$). Results are consistent with visual observations (Figure 4.6). Clade I showed a more ragged, bimodal distribution, but since the data set was limited in this clade (only ten specimens), further conclusions can not be drawn.

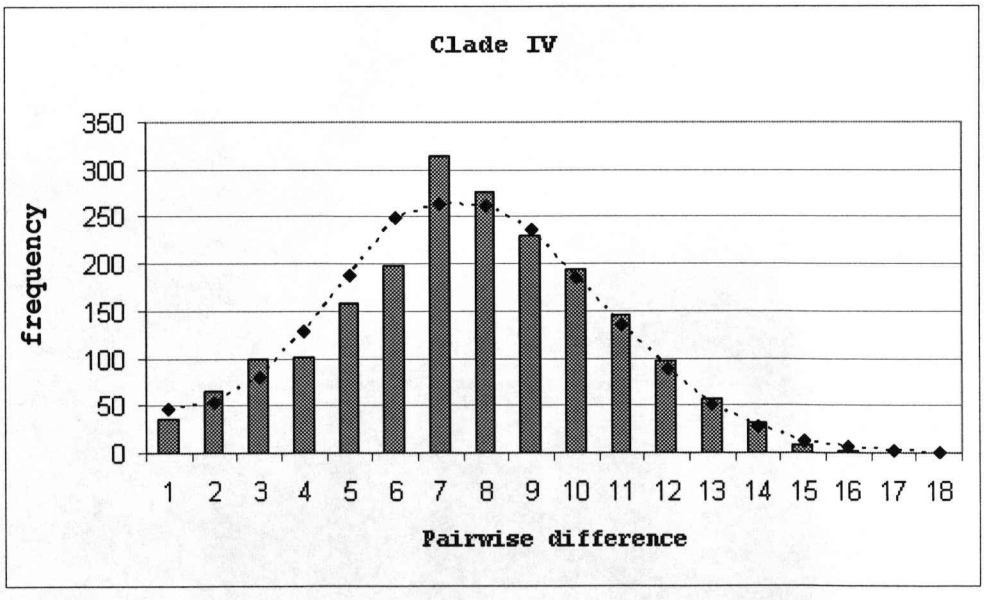
Figure 4.6 Mismatch distribution of haplotypes in clade III (A) and clade IV (B).

Solid lines correspond to expected distribution under sudden population expansion models, whereas dashed lines represent the observed values. Raggedness indexes were calculated with DnaSP. $r < 0.056$ (clade III) and $r < 0.0082$ (clade IV).

A



B



4.3.6 Neutrality tests

According to the Mc Donald-Kreitman test (M-K), under neutrality, the ratio of non-synonymous (NS) to synonymous (S) substitutions should be the same within and between species. Any significant difference between them is considered to be the result of deviation from neutrality, where a ratio <1 indicates purifying or stabilizing selection and a ratio of >1 suggest positive selection (McDonald and Kreitman, 1991). As shown in Table 4.9, very few non-synonymous substitutions were observed in both polymorphic and fixed substitutions, indicating that the different clades in *An. barbirostris* do not deviate significantly from neutrality. The excess of non - synonymous substitutions may indicate a purifying selection in the COI gene region.

Table 4.9 The number of synonymous (S) and non- synonymous (NS) substitution P-values obtained using Fisher's exact test.

	Fixed		Polymorphic		
Clade	S	NS	S	NS	*p*-value
I - II	9	1	17	1	ns
I - III	9	1	39	2	ns
I - IV	9	1	66	5	ns
I - V	14	1	22	2	ns
II - III	12	0	26	2	ns
II - IV	11	0	54	5	ns
II - V	14	1	22	2	ns
III - IV	11	0	69	6	ns
III - V	13	0	30	3	ns
IV - V	11	0	57	6	ns

The other tests of neutrality used in this study are Tajima's D test (Tajima, 1989) and Fu and Li's (1993) D* and F* tests. Tajima's test assumes that under neutrality, the number of segregating sites (θ_S) should be the same as the average number of nucleotide differences (θ_π) (Tajima, 1989), whereas Fu and Li's test is based on

comparison of mutations in the internal and external branches of a rooted phylogenetic tree. In contrast to the Mac Donald-Kreitman test, these tests have the advantage of requiring only one population to test the neutrality hypothesis, whereas M-K requires two populations to be compared.

As shown in table 4.10, all values obtained were negative. However the majority of these results was non-significant with the exception of clade III, for which values were negative and statistically significant at the 5% significance level ($p < 0.05$). This is the result of an excess of segregating sites with alleles at low frequencies. These results agree with Fu and Li's F* test (Table 4.10), in which negative values were also significant for clade III. In contrast the value of the D* test was not significant. Negative values indicate deviation of neutrality or they could be the result of the fixation of an advantageous mutation, the presence of deleterious mutations or as a result of a recent hitchhiking event or changes in the population size. The main disadvantages of these tests are that they depend on the hypothesis that populations have been under mutation-drift balance for a long time, which is unlikely to occur in most populations. Furthermore they assume that that all nucleotides are equally mutable, which is incorrect because the first, second and third codon positions mutate at different rates (Nei and Kumar, 2000). These pitfalls indicate that results have to be interpreted cautiously.

Table 4.10 Tests of neutrality of the 5 clades in the Barbirostris Subgroup are shown. Result of Tajima's D test and Fu and Li's F and D tests are shown. Results were tested at the statistical significance $\alpha=0.05$.

Populations tested	Tajima's D	Fu and Li's	
		D*	F*
Clade I	-0.23 ns	-0.20 ns	-0.24 ns
Clade II	-	-	-
Clade III	**-2.04 ***	**-2.42 ns**	**-2.72 ***
Clade IV	-1.58 ns	-2.04 ns	-2.23 ns
Clade V	-0.50 ns	-0.57 ns	-0.59 ns

- $P < 0.05$

4.3.7 Estimation of the time of coalescence

The parameter τ was used to estimate the time since population growth in clade III, which shows evidence of population expansion. This parameter is calculated as $\tau = 2\mu t$ in populations under expansion (Slatkin and Hudson, 1991), where μ is the mutational rate, estimated for mtDNA in *Drosophila* as 10^{-8} substitutions/site/year (Powell et al., 1986) and t is time in generations. The estimation of τ was calculated (using DnaSP) as 1.827. Solving $t = \tau/2\ \mu$ = 120 833 generations. Assuming 10 generations per year (Powell et al., 1986), an estimate used by Walton et al. (2000), for *Anopheles dirus*, the time of expansion of clade III occurred approximately 12 083 years ago. These data have to be interpreted cautiously considering that clade III may be distributed in a much larger geographic area than Thailand.

4.3.8 Analysis of the evolutionary rates in different regions within cytochrome oxidase I

The identification of regions of maximum divergence should ensure a successful delineation of taxa, particularly in sister species, in which divergence is expected to be low (Roe and Sperling, 2007). In this section, the evolutionary rates of the COI in clades III and IV and in other *Anopheles* mosquitoes are studied, to ensure this mtDNA region is informative enough to delimit species.

Differences in genetic divergence within the different regions within the COI fragment amplified in this study were compared with those found by Lunt (1996). This author found differences in evolutionary rates within this gene region, based on comparisons of the amino acid substitutions. In this section, the general assumptions about the evolutionary rates within the COI obtained by Lunt (1996) are compared with the results obtained with Barbirostris Subgroup specimens and in other *Anopheles* species.

The 756-bp region studied was located from the internal loop I3, to the COOH terminal in the Cytochrome Oxidase I (see Figure 4.1); from position 764 to 1520, defined according to the data presented by Lunt (1996). For comparisons, nucleotide

diversity values obtained for clades III and IV (i.e. those containing the most specimens) were used and nucleotide diversity values obtained from the alignment of seven other *Anopheles* species: *An. oswaldoi* (Scarpassa and Conn, 2006), *An. albimanus* (Sallum et al., 2002), *An. darlingi* (Sallum et al., 2002), *An. stephensi* (Oshaghi et al., 2006) , *An. gambiae* (Sallum et al., 2002), *An. quadrimaculatus* (Cockburn et al., 1990) and *An. minimus A* (Sallum et al., 2002).

As shown in Figure 4.7, region M8 was highly conserved among all *Anopheles* species included in the analysis, including those from clades III and IV of the Barbirostris Group. These results are consistent with those of Lunt, in that central regions M5-M8 (positions 559-967) had a low level of divergence, which Lunt postulated was due to functional constrains (Figure 4.7). Similarly, internal loops 4 (967-1009) and M9 (1010-1069) are highly variable regions, results that are consistent with those presented by Lunt (1996). Inferences were based on the comparison of nucleotide diversity values (π).

Considerable differences have been observed when comparing COI sequences from *Anopheles* species and those obtained by Lunt. For instance, the transmembrane region M11 (1241-1297) has been reported to be highly conserved in different insect taxa (Lunt et al., 1996). The levels of nucleotide diversity in clade III are consistent with this, whereas those for clade IV are not (Figure 4.7). Moreover, Lunt defined the COOH as the most variable region; however in this study, this terminal region appear to be one of the most conserved in clades III and IV and in the *Anopheles* species included in this analysis; in which only the last portion of these regions appear to be variable (Figure 4.7).

Figure 4.7 Nucleotide diversity values in clades III (pink line), IV (blue line) and several *Anopheles* species (green line) are compared with those studied by Lunt in 1996 (red line).

Regions covered by this sequence analysis are situated from sites 764 (situated in internal loop 3) to position 1520 (region COOH). π values in the internal loop 4 are high in both comparison groups; conversely COOH is not the most variable region as defined by Lunt. Nucleotide positions are defined according to Lunt (1996). Notice differences in scale of axis for (A) and (B).

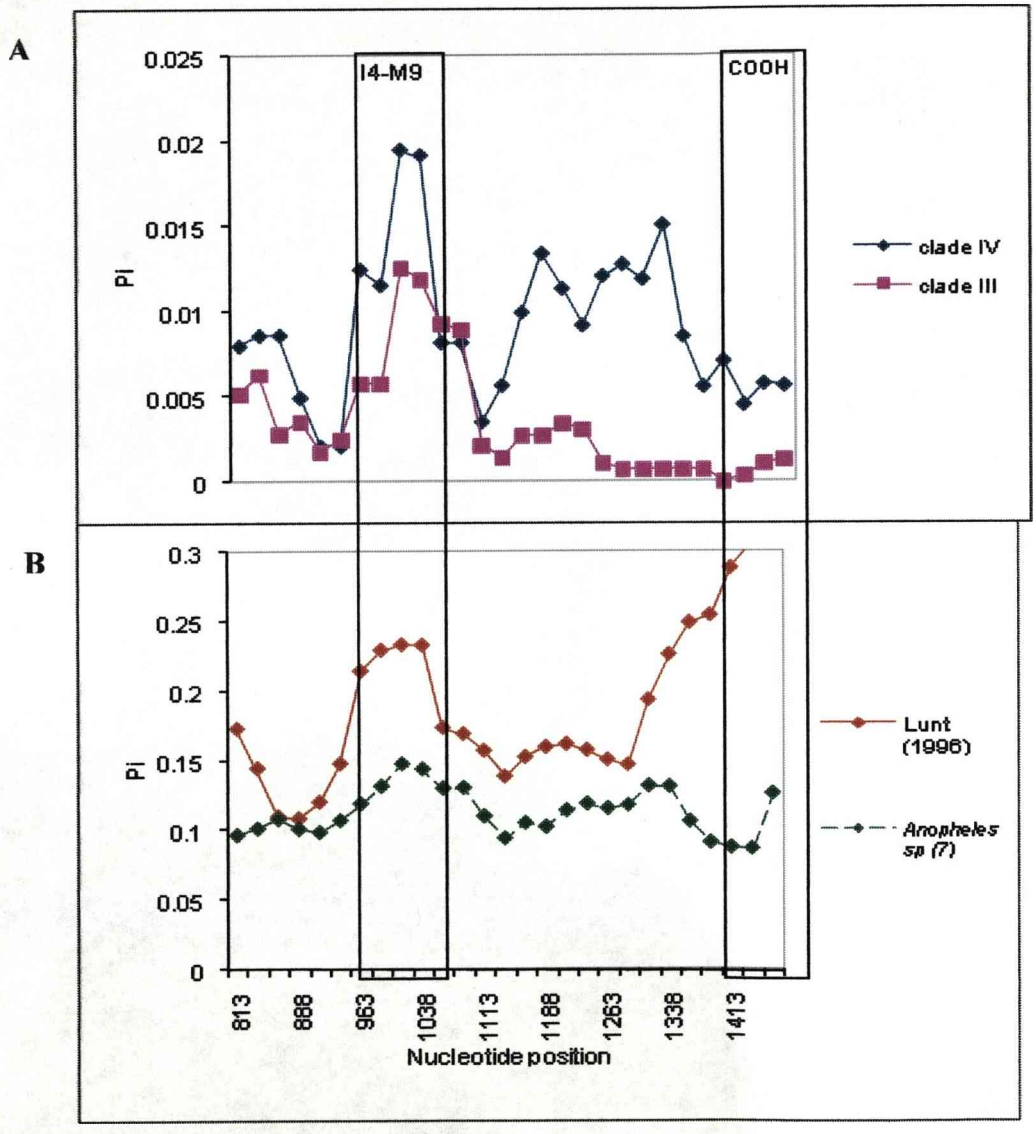
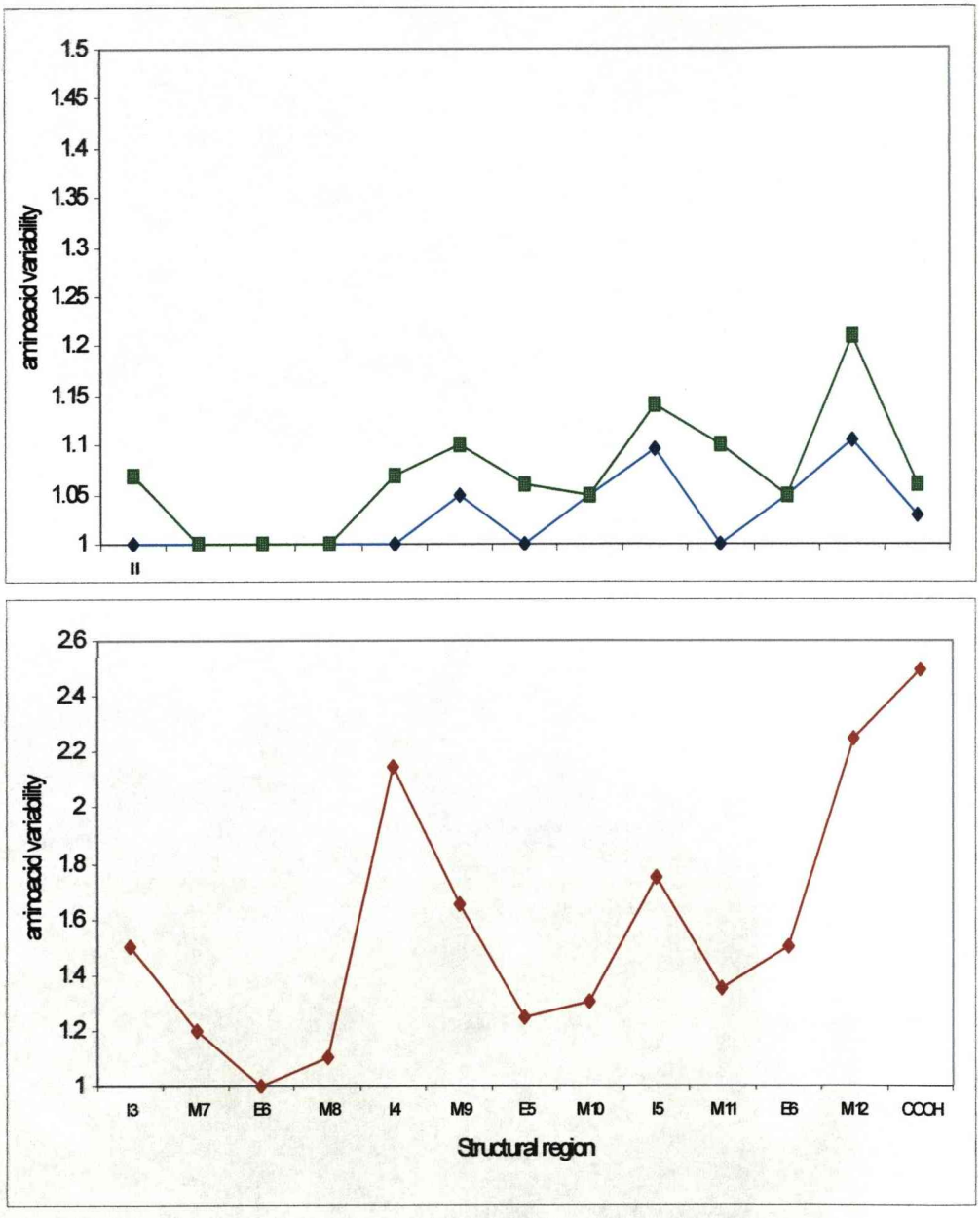


Figure 4.8 Amino acid variations defined by structural regions in the 756-bp COI region from region I3 to COOH. Amino acid variation is defined as the average number of substitutions per site.

Results for clades III and IV of Barbirostris Subgroup members are in blue, for the seven other *Anopheles* species (see text) in green line and those of Lunt et al. in red. Amino acid variability is similar in all *Anopheles* species. There are noticeable differences with the results of Lunt et al. (1996).



Of the 512 amino acids coded for COI (Lunt et al., 1996), 252 amino acids are coded for the 756-bp fragment amplified in this study. Only two fixed amino acid substitutions were identified in this fragment. The first is a Tyrosine (TAT) located at position 1220-1222 (407 of amino acid sequence) unique to specimens of clade I replacing Serine (TCT), present in all remaining specimens (Fig. 4.9) and the second fixed amino acid substitution was a Threonine (ACT), in positions 1343-1345 (448 of amino acid sequence) present only in those specimens from clade I collected in Mae Hong Son. Specimens of this clade collected from Kalimantan and specimens from other clades of the Barbirostris Group showed a (GCT) Alanine at this position with the exception of two specimens from clade III, one collected in Mae Hong Son and the other in Trat (Figure 4.9). All positions are numbered as in Lunt, 1996.

The substitution resulting in Tyrosine replacing Serine is located in the internal loop I5, which Lunt et al. (1996) regarded as a region of relative variability. Conversely, the second fixed substitution is situated in external loop 6 (E6), previously identified as a conserved region. Other substitutions (non-fixed) were located at amino acid positions 356 (in transmembrane helix M9), 391 (M10), 393 (Internal loop I5), 466 (M12), 467 (M12) and 505 (COOH). These substitutions were all located at positions defined by Lunt et al. (1996) as relatively or highly variable regions, with the exception of the transmembrane helix M10, which he considered to be highly conserved (Figure 4.8).

Figure 4.9 Amino acid substitutions within the 152 aa sequence coded for the 756-bp fragment amplified in specimens from the Barbirostris Group.

Sites are defined according to Lunt et al. (1996) amino acid sequence.

	3 5 6	3 9 1	3 9 3	4 0 7	4 4 8	4 6 6	4 6 7	5 0 5
H1	V	G	V	Y	A	Y	F	A
H2	.	.	.	.	.	.	.	.
H3	.	.	.	.	.	.	.	.
H4	.	.	.	.	.	.	.	.
H5	.	.	.	.	T	.	.	.
H6	.	.	.	.	T	.	.	.
H7	.	.	.	.	T	.	.	.
H8	.	.	.	.	T	.	.	.
H9	.	.	.	.	T	.	.	.
H10	.	.	.	.	T	.	.	.
H11	.	.	.	S	.	.	.	.
H12	.	.	.	S	.	.	.	.
H13	.	.	.	S	.	.	.	.
H14	.	.	.	S	.	.	.	.
H15	.	.	.	S	.	.	.	.
H16	.	.	.	S	.	.	.	.
H17	.	.	.	S	.	.	.	.
H18	.	.	.	S	.	.	.	.
H19	.	.	.	S	T	.	.	.
H20	.	.	.	S	.	.	.	.
H21	.	.	.	S	.	.	.	.
H22	.	.	.	S	.	.	.	.
H23	.	.	.	S	.	.	.	.
H24	.	.	.	S	.	.	.	.
H25	.	.	.	S	.	.	.	.
H26	.	.	.	S	.	.	.	.
H27	.	.	.	S	.	.	.	.
H28	.	.	.	S	.	.	.	.
H29	.	.	.	S	.	.	.	.
H30	.	.	.	S	.	.	.	.
H31	.	.	.	S	.	.	.	.
H32	.	.	.	S	.	.	.	.
H33	.	.	.	S	.	.	.	.
H34	.	.	.	S	.	.	.	.
H35	.	.	.	S	.	.	.	.
H36	.	.	.	S	.	.	.	.
H37	.	.	.	S	.	.	.	.
H38	.	S	.	S	.	.	.	.
H39	.	.	.	S	.	.	.	.
H40	.	.	.	S	.	.	.	.
H41	.	.	.	S	.	.	.	.
H42	.	.	.	S	.	.	.	.
H43	.	.	.	S	.	.	.	.
H44	.	.	I	S	.	.	.	.
H45	.	.	.	S	.	.	.	.
H46	.	.	.	S	.	.	.	.
H47	.	.	.	S	.	.	.	.
H48	.	.	.	S	.	.	.	.
H49	.	.	.	S	.	.	.	.
H50	.	.	.	S	.	.	.	.
H51	.	.	.	S	.	.	.	.
H52	.	.	.	S	.	.	.	.
H53	.	.	.	S	.	.	.	.
H54	.	.	.	S	.	.	.	.
H55	.	.	.	S	.	.	.	.
H56	.	.	.	S	.	.	.	.
H57	.	.	.	S	.	.	.	.
H58	.	.	.	S	.	.	.	.
H59	.	.	.	S	.	.	.	.
H60	.	.	.	S	.	F	.	.
H61	.	.	.	S	.	.	.	.
H62	.	.	.	S	.	.	.	.
H63	.	.	.	S	.	.	.	.
H64	I	.	.	S	.	.	.	.
H65	.	.	.	S	T	.	.	.
H66	.	.	.	S	.	.	.	.
H67	.	.	.	S	.	.	.	.
H68	.	.	.	S	.	.	.	.
H69	.	.	.	S	.	.	.	.
H70	.	.	.	S	.	.	.	.
H71	.	.	.	S	.	.	.	.
H72	.	.	.	S	.	.	L	.
H73	.	.	.	S	.	.	.	.
H74	.	.	.	S	.	.	.	.
H75	.	.	.	S	.	.	.	.
H76	.	.	.	S	.	.	.	.
H77	.	.	.	S	.	.	.	.
H78	.	.	.	S	.	.	.	.
H79	.	.	.	S	.	.	.	.
H80	.	.	.	S	.	.	.	.
H81	.	.	.	S	.	.	.	.
H82	.	.	.	S	.	.	.	.
H83	.	.	.	S	.	.	.	.
H84	.	.	.	S	.	.	.	.
H85	.	.	.	S	.	.	.	.
H86	.	.	.	S	.	.	.	T
H87	.	.	.	S	.	.	.	.

4.4 Discussion

4.4.1 Phylogenetic analysis

Cytochrome oxidase I is an efficient molecular marker for the differentiation of members of the Barbirostris and Hyrcanus Groups, as shown in the Maximum parsimony (MP) and Neighbour-joining (NJ) trees (Figure 4.2). Members of these groups are commonly confused by field workers as seen in this study: 19 specimens from the Hyrcanus group were sent as specimens from the Barbirostris Group. This was surprising since there are reliable morphological characters to distinguish species of the Barbirostris and Hyrcanus Groups. The relationship between the morphological and molecular results will be explained in detail in Chapter 6.

Tree topologies gave consistent results in identifying 5 distinctive clades within the Barbirostris Group, which were also supported by the five parsimony networks obtained with the nested clade analysis. However, the inferred phylogenetic relationships between clades were different between the trees. Clades II and III were grouped in the same cluster in the MP tree but not in the NJ tree and clade III appeared closely related to clades IV and V in the NJ tree but not in the MP tree. However, clades IV and V appear in the same cluster in both NJ and MP trees, although this relation is not well supported (Bootstrap values 48% and 49% respectively). One interesting finding was the division of clade I into two clusters, one comprising specimens from Mae Hong Son and the other specimens from Kalimantan, suggesting the influence of isolation by distance. No further conclusions about the phylogenetic relationships among clades can be made since low bootstrap values were obtained with both NJ and MP trees.

There are various explanations for low bootstrap values. They may arise due to differences in evolutionary rates within the cytochrome oxidase (Saito et al., 2000), when there is an excess in the number of taxa employed in the analysis (Rokas and Carroll, 2005) and appear to be common in cases where closely related species have not diverged extensively (Soltis and Soltis, 2003). To test if low bootstrap values were related to differences in evolutionary rates of the COI, the fragment amplified

was split into regions of high and low variability (Figure 4.7) and these regions were then used to construct phylogenetic trees. Trees obtained from the comparisons of these regions separately did not improve on the results obtained from the entire amplicon (data not shown). The explanation that best fits these results is that the clades contain closely related species that have only recently diverged.

The interpretation of bootstrap values is open to discussion; nevertheless this test is widely used to support inferred phylogenetic relationships among taxa. The monophyly of the Barbirostris Group is strongly supported (100% NJ and 99% MP), as indicated by the high bootstrap values obtained for the distinction between this clade and the Hyrcanus Group (Figure 4.2).

4.4.2 Sequence variation and haplotype diversity

The amplification of "pseudogenes" (mitochondrial-like sequences in the nuclear genome) represent a problem since they can confound phylogenetic and population genetic studies (Zhang and Hewitt, 1996). In this study no stop codons were found within fragments amplified, so the presence of pseudogenes is disccounted. Sequences of the 5 clades were A+T rich (70.9%), and the percentage of these nucleotides are typical of values found in the mitochondrial genes of insects (O'Loughlin et al., 2007; Simon et al., 1994). *Haematobia irritans* 71.4%, *Stomoxys calcitrans*, 70.6%, *Musca domestica* 69.6% (De Oliveira et al., 2005) and in *Anopheles* species: 72.1% in *An. darlingi* (Mirabello and Conn, 2006a), 74.6 % for *An. jeyporiensis* (Chen et al., 2004), 71.0% in *An. dirus* (Walton et al., 2000) and 72.0% in *An. scanloni* (O'Loughlin et al., 2007).

The haplotype diversity values were high within all clades. Such high values are common in *Anopheles* species. For instance, 50 haplotypes from 76 individuals in a fragment of the COII gene have been reported for *An. jeyporiensis* (Chen et al., 2004), 70 in 84 specimens in the same COI gene in *An. dirus* (Walton et al., 2000); 33 in 65 and 32 in 54 in *ND5* sequences of *An. gambiae* and *An. arabiensis* respectively (Besansky et al., 1997).

4.4.3 Genetic structure

The parameter F-statistic (F_{ST}), employed to detect genetic differentiation among clades within the Barbirostris Subgroup, provided compelling evidence that the 5 clades within this subgroup were distinct, since estimates were high in all comparisons, 0.74 -0.91 ($p < 0.001$, see Fig 4.4). A lower F_{ST} estimate = 0.093 has been reported between *An. arabiensis* and *An. gambiae* in Africa. In Southeast Asia, a high F_{ST} = 0.680 was obtained from the comparison between species C and D of *Anopheles dirus* (Walton et al., 2000). These species were subsequently recognized as *An. scanloni* and *An baimaii*, respectively. Moreover, in *An. scanloni,* high levels of genetic differentiation (F_{ST} =0.47-0.63) have been reported between populations in Thailand, suggesting that *An. scanloni* could be more than one species (O'Loughlin et al., 2007) In *Lutzomyia longipalpis* the high F_{ST} values observed in sympatric populations = 0.395 were evidence that suggest that this taxon could be a species complex (Bauzer et al., 2007). If we compare these results on genetic differentiation to the studies in mosquitoes mentioned, it is likely that these clades are different species.

4.4.3.1 Genetic structure in clade I

The F_{ST} values obtained from the comparison of populations within the clades were all very low with the exception of clade I. A high F_{ST} value = 0.645 ($p < 0.01$) was obtained when comparing specimens from Kalimantan (4 individuals) with Mae Hong Son (6 individuals) in this clade. Although the dataset for this clade is very limited, these two geographic populations appear distinct. This is also supported by high bootstrap values in the NJ and MP trees that support their separation (Figure 4.2) and in the nested clade analysis in which these two populations appear separated by at least 6 mutational steps (Figure 4.4). Taking into account that these collection sites are separated by ~ 3000 km and have the sea as geographic barrier, high levels of genetic diversity were expected. It is not possible to determine if these populations represent distinctive species since the dataset is too small to arrive at further conclusions.

4.4.3.2 Genetic structure in clade II

The presence of a single haplotype among individuals from clade II was surprising, since specimens were collected from localities of Sumatra and Java, that are located ~ 400 km from each other and are separated by the sea. The expansion of this clade may have occurred when these islands were connected by the Sunda shelf, which appeared as an extension of mainland Southeast Asia. Within a species, higher genetic diversity is observed in older populations than in their younger counterparts as observed in *An. darlingi* (Mirabello and Conn, 2006a) and *An. albimanus* (Molina-Cruz et al., 2004). Therefore it may be possible that the seven specimens studied in clade II (6 from Sumatra and 1 from Java) are a younger portion of a larger population. However it is difficult to arrive to any conclusion due to the reduce data set in this clade.

4.4.3.3 Genetic structure in clade III

Specimens from this clade were identified as *An. barbirostris*, based on morphological identification of immature stages of 10 of its individuals. A shallow genetic structure in clade III was seen in the nested clade analysis (Figure 4.4) and a weak effect of distance on differentiation was observed in Thailand. For instance, haplotype 12 appeared from the network analysis to be the ancestral haplotype. This was the most widely distributed haplotype in clade III, comprising specimens from the provinces of Mae Hong Son, Tak and Kanchanaburi to Sa Kaeo and Trat. Although several mountains are found in these provinces, they have relative low elevation (Rattanarithikul et al., 2005). The low level of genetic differentiation could be explained by the fact that these provinces are separated by a broad and relatively flat land (Figure 4.9) that does not represent a significant geographic barrier, not withstanding the 800 km between the most widely separated sites, Mae Hong Son and Trat.

The effects of distance on the genetic diversity of *Anopheles* species appear to be less important than the presence of geographic barriers: levels of genetic diversity in populations of *An. gambiae* in Africa are low, even over distances of 6000 km

(Besansky et al., 1997). In contrast, the highest level of differentiation within this species was reported from populations < 700 km apart but separated by the Rift Valley as geographic barrier (Lehmann et al., 1999).

The demographic expansion of clade III may be inferred from all the tests used: 1) it shows a starburst pattern, which appears to be typical of populations under expansion. Walton et al. (2000) found a similar star-like genealogy of the haplotypes in *An. dirus A* and *D* of the *Anopheles dirus* complex in Asia and concluded that these populations of these species were undergoing expansion. Similar results were also found in *An. jeyporiensis* in Asia, again suggesting recent range expansion (Chen et al., 2004). 2) haplotypes in clade III of the Barbirostris Subgroup are only separated by one or two mutational steps, which is consistent with this hypothesis (Slatkin and Hudson, 1991). 3) The pattern of mismatch distribution seen in this clade is typical of a population expansion, since this type of mismatch profile is almost never produce under equilibrium (Rogers and Harpending, 1992).

Based on these observations, it is believed that clade III has undergone population expansion. This is consistent with significant negative values obtained in the Tajima's D and Fu & Li's F* neutrality tests, obtained in this clade, as they appear associated to a recent population expansion (Kreitman, 2000), rather than to deviation from neutrality (see section 4.4).

4.4.3.4 Genetic structure in clades IV and V

Although specimens in clade IV appear collapsed into one single network, they do not show the starburst pattern, which in clade III indicated a recent population expansion. However the type of mismatch distribution observed in clade IV, which fits a Poisson distribution model (Rogers and Harpending, 1992), is an indication that this population may also be expanding. However, it is important to consider that specimens of this clade have only been found in Sa Kaeo and Trat in Thailand in addition to a single specimen from Sumatra. Sa Kaeo and Trat are only separated by

a distance < 200 km. In this respect, it is difficult to make any generalization about the genetic structure of this clade.

No conclusions can be drawn about the genetic structure of clade V, since it comprised just six specimens from Sa Kaeo. However it is clearly separated from the other four clades.

4.4.3.5 Structural analysis of the Cytochrome Oxidase I gene region. Comments on the "Barcoding of life"

The relevance of the work of Lunt (1996) in insect phylogeny studies has increased since the introduction of the proposal for "barcoding of life" (Hebert et al., 2003). His findings have been cited in several studies (Erpenbeck et al., 2006; Roe and Sperling, 2007); and even taken as reference study in Barcoding websites http://www.dnabarcoding.ca/primer/COIProtein.html.

Some of the most widely used primers for the amplification of COI are based on Lunt (1996) (Zhang and Hewitt, 1997). For instance, primer set UEA9-UEA10 has been employed to distinguish sibling species of members of the *Anopheles annulipes* group (Foley et al., 2007), and to define the species status of *An. flavirostris* (Torres et al., 2006). These primers amplify the terminal COOH region, defined by Lunt as the most variable in the COI enzyme. However, results in this present study show that COOH is highly conserved in clades III and IV, which are likely to be sibling species. In addition this region appears relatively conserved in the seven *Anopheles* species studied (Figure 4.7). Therefore, conclusions related to speciation based on this region should be interpreted with care. The main problem in drawing generalized conclusions from the work of Lunt (1996) is the limited number of taxa used in his analysis. He included only nine species: two from the Orthoptera, six Diptera and one Hymenoptera, these are surely insufficient to describe evolution patterns in all species of insects.

Supporters of the Barcoding project proposed that the "Folmer partition", which is a ~ 640 bp fragment located at the 5' end of the COI subunit, be used in the

identification of species of the animal kingdom. This was based on the belief that sequence divergence in the different regions within cytochrome oxidase I were very similar (Hebert et al., 2003). However a recent paper by Roe and Sperling (2006) examined the pattern of evolution of the cytochrome oxidase I and II in several species within the Lepidoptera and Diptera. These authors found that regions of maximum diversity were highly variable among the taxa surveyed and no single region of maximal divergence was found among 114 sequences examined. These results are consistent with those presented in this present study. The success of the use of the COI to distinguish species within the Barbirostris Subgroup will be discussed in detail in Chapter 8.

CHAPTER 5

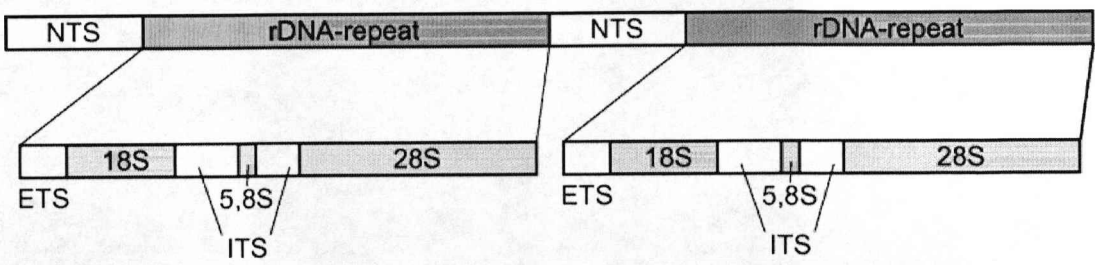
5 Phylogenetic analysis within the Barbirostris Subgroup based on the Internal Transcribed Spacer 2

5.1 Introduction

This chapter describes the analysis of the rDNA Internal Transcribed Spacer 2 (ITS2) in specimens from the Barbirostris Subgroup. The use of this marker to identify species within the Barbirostris Subgroup and to resolve their phylogenetic relationships is presented. All members of the Barbirostris Subgroup were found to have a long ITS2 (>1.5 kb). The length of the spacer in this Subgroup is due to the presence of repeat sequences within the ITS2, which vary in copy numbers and organization. A comprehensive analysis of these repeats and a discussion of their possible origin is also presented.

The ribosomal DNA (rDNA) is a family of genes that occur as tandemly repeated units of many copies per cell (Gerbi, 1985). Each repeat consists of the 18S, 5.8S, and 28S RNA genes, external transcribed spacers (ETS1 and ETS2), internal transcribed spacers (ITS1and ITS2) and an intergenic spacer (IGS), formerly known as non-transcribed spacer (NTS) (Figure 5.1). In eukaryotes, the number of repeats varies from 30 to 3000 (Prokopowich et al., 2003). In mosquitoes there are typically ≤ 500 copies per genome (Collins et al., 1987).

Figure 5.1 Schematic representation of the rDNA genes separated by transcribed spacers. Each repeat unit is separated by an IGS (or NTS).



The functional regions (18S, 5.8S and 28S genes) that produce ribosomal RNA are highly conserved whereas the spacer regions have higher mutational rates and hence

a high interspecific and low intraspecific variability, which make them useful in systematics studies (Williams et al., 1988). The low intraspecific variation is the result of the process of concerted evolution (Dover, 1982), in which all the repeats evolve as a unit. Hence, if a mutation occurs in a repeat, it spreads through the entire family of genes. In the hypothetical case of absence of mutation, this process may eventually lead to complete homogenization (Nei and Rooney, 2005). A recent study based on whole-genome shotgun cloning has shown that repeat sequence variation is at very low level, confirming that rDNA does evolve via a process of concerted evolution (Ganley and Kobayashi, 2007). This model of evolution is explained in detail in the literature review (see Chapter 2).

Ribosomal DNA is of considerable utility in the identification of *Anopheles* species (Collins and Paskewitz, 1996). The Internal Transcribed Spacer 2 (ITS2) has proved a valuable marker in molecular systematics in part due to the fact that it is flanked by highly conserved regions (5.8S and 28S), which allow the design of universal primers of use in a wide range of organisms. For example, the primers designed for the amplification of the ITS2 region of *Anopheles gambiae* (Paskewitz et al., 1993) have proved useful in numerous other *Anopheles* species.

The ITS2 has been used to identify cryptic species within *An. quadrimaculatus* (Cornel et al., 1996), species within the *A. dirus* Group in Thailand (Walton et al., 1999a), species of the Hyrcanus Group in Southeast Asia (Hwang, 2007; Ma and Xu, 2005), and to identify Palearctic members of the *Anopheles maculipennis* complex in northern Iran (Djadid et al., 2007). Differences in feeding preferences within the *An. fluviatilis* complex have been associated with individual species based on sequence variation of the ITS2 region (Manonmani et al., 2001). Moreover, the results from the analysis of the spacer region appear consistent with results obtained from the analysis of mitochondrial markers. Three closely related members of the *Anopheles moucheti* complex, were differentiated with the combined use of the cytochrome oxidase b and spacers ITS1 and ITS2. Species within the Maculatus Group of species were also differentiated using ITS2 and COII (Ma et al., 2006).

The results obtained from the ITS2 analysis will be compared with those from analysis of COI and with previous results obtained in *Anopheles barbirostris* from Indonesia (Baskoro, 2001). Recently, Seung et al. (2007) published ITS2 sequences for members of the Barbirostris Subgroup, which will also be used for comparative purposes.

Objectives

The main objective of this present chapter is to investigate the value of ITS2 for differentiating species within the Barbirostris Subgroup. With this purpose, the following questions will be addressed:

1. Are the results with ITS2 consistent with those obtained from the analysis of the COI?

2. Are there advantages in the use of the ITS2 to differentiate species within the Barbirostris Subgroup, relative to COI?

3. Are the transposable genetic elements within the ITS2 responsible for its length?

5.2 Material and methods

In this chapter, 51 specimens were included in the analysis of the ITS2 (Table 5.3). They were obtained from the following localities: Kalimantan (n=2) and Sumatra (n=5) in Indonesia, and Mae Hong Son (n=10), Trat (n=14), Tak (n=6), Sa Kaeo (n=9) and Kanchanaburi (n=5) in Thailand. The method for DNA extraction, PCR amplification, cloning and sequencing are given in Chapter 3 (Material and Methods). Direct sequencing of PCR products from amplification with 5.8 and 28S primers was not possible because of the large size of the amplicon (up to 1862 bp). As a consequence, PCR products were cloned and subsequently sequenced using newly designed internal primers (see Material and Methods).

5.3 Results

5.3.1 PCR amplification

Following the extraction procedure explained in Chapter 3, a PCR amplification of the ITS2 was carried out using primers in the 5.8S and 28S regions. The resulting amplicon comprised: 5.8S (partial)+ ITS2 (complete) + 28S (partial). All PCR products obtained were cloned and then sequenced.

All specimens identified as members of the Barbirostris Subgroup with the mitochondrial marker COI, showed an exceptionally large amplicon >1.5 kb, larger than any other *Anopheles* species. In contrast, specimens from the Hyrcanus Group showed a ~700 bp PCR product (Figure 5.2), a size more typical of *Anopheles* species (Table 5.1). This difference in size was sufficient to differentiate these two groups. In this chapter results are centred on the Barbirostris Subgroup. Data for the Hyrcanus Group of species are presented in Chapter 7.

The PCR amplification of the 5,8S-ITS2-28S fragment was complicated due to a variety of factors. These included 1) poor quality of DNA in some specimens, 2) inhibition of PCR, a not infrequent problem with mosquitoes of the Barbirostris Group, 3) the large size of the ITS2, and 4) the presence within the ITS2 of an internal repeat structure. Even when some specimens were successfully amplified, the product was sometimes lost in the cloning process. Significant effort and time was spent in the amplification of this fragment.

Problems with secondary bands in PCR amplification were particularly common in specimens identified in the mitochondrial analysis as members of clade I. Results were improved with the use of a touchdown PCR, starting with an annealing temperature of 68 °C, which was slowly reduced to 50 °C (see material and methods). With this technique, results were improved thereby, facilitating subsequent steps (Figure 5.3). In other specimens, the quality of the amplification was enhanced only by increasing the annealing temperature to reduce non-specific PCR amplification.

Figure 5.2 Showing PCR products from amplification of ITS2 (plus flanking 5.8S and 28S regions) run in 1% agarose gel with EtBr and visualized under UV.

Product size differences reflect differences in the length of ITS2. Clades are those defined by COI sequencing (Chapter 4).

Lane 1: 100bp molecular ladder

Lane 2: unsuccessful amplification

Lane 3: unsuccessful amplification

Lane 4: amplicon for Barbirostris Subgroup (~1.6kb)

Lane 5: unsuccessful amplification

Lane 6: amplicon of Hyrcanus Group (~700bp)

Lane 7: amplicon of Hyrcanus Group (~700bp)

Lane 8: unsuccessful amplification

Lane 9: amplicon of Barbirostris Subgroup Clade II (~1.8kb)

Lane 10: amplicon of Barbirostris Subgroup Clade III (~1.8kb)

Lane 11: amplicon of Barbirostris Subgroup Clade I (~1.6kb)

Lane 12: negative control

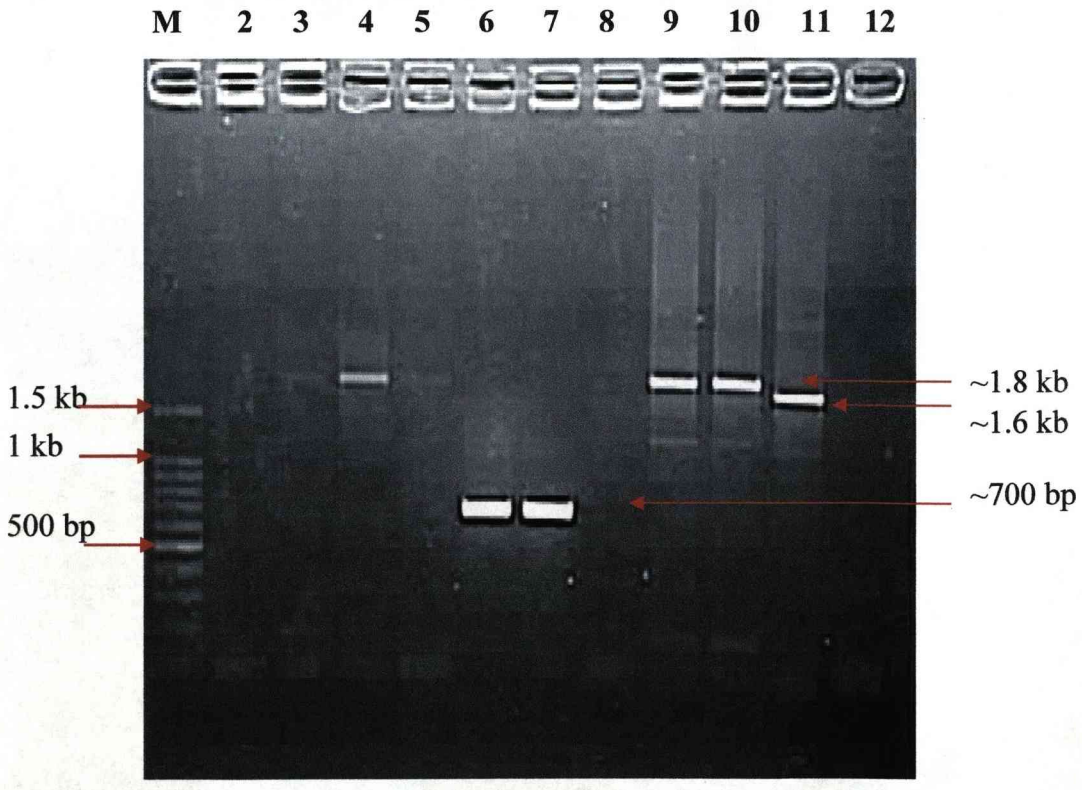


Figure 5.3 Comparison of products obtained with a standard PCR (A) and touchdown PCR (B) in the amplification of the fragment 5.8S (partial) + ITS2 (complete) + 28S (partial)

Bands obtained in the electrophoresis agarose gel show

A

Lanes 2-7: specimens k3, k4, th1-1, th1-3, th14 and th1-12, respectively

lanes 8: Positive control

Lane 9: Negative control

B

Lanes 2-4: specimens k3, k4 and th1-1, respectively

Lanes 6: Positive control

Lane 8: negative control

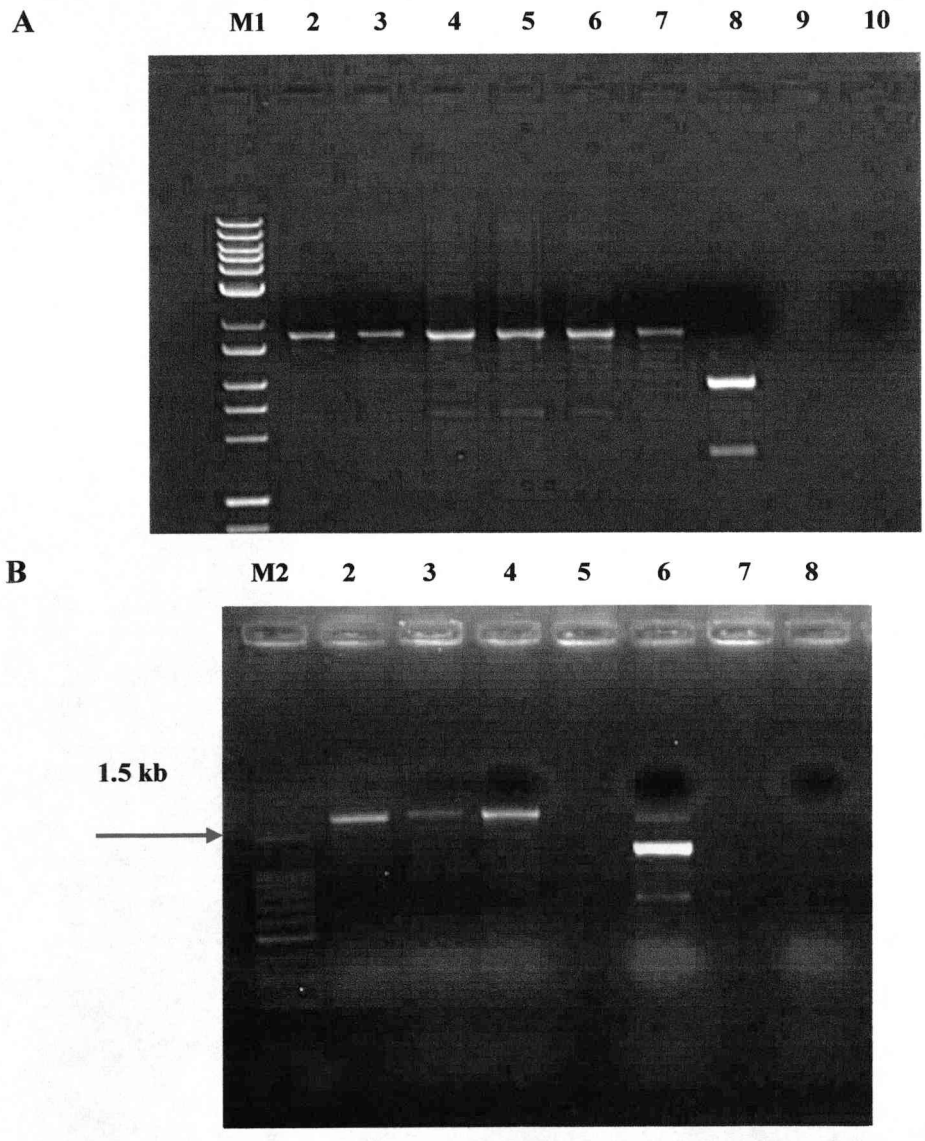


Table 5.1 List of Internal Transcribed Spacer 2 sizes (bp) and the G+C content in several species of *Anopheles* mosquitoes. *An. beklemishevi, An. crucians B* and *An. fluminensis* also show long ITS2 fragments.

Species	ITS2	GC	Reference
Clade I	1545	52.3%	This study
Clade II	1727	55.6%	This study
Clade III	1730	55.9%	This study
Clade IV	1583	54.3%	This study
Clade V	1519	55.1%	This study
An. gambiae complex	426-427	55.00%	(Paskewitz et al., 1993)
An. freeborni	305-310	52.00%	(Porter and Collins, 1991)
An. fluviatilis (species X)	372	ND	(Manonmani et al., 2001)
An. fluviatilis (species Y)	374	ND	(Manonmani et al., 2001)
An. beklemishevi	638	46.6%	(Kampen, 2005)
An. punctulatus	ND	61.3-70.9%	(Beebe et al., 1999)
An. crucians A	461	54.00%	(Wilkerson et al., 2004)
An. crucians B	872-1021	51.00%	(Wilkerson et al., 2004)
An. crucians C	204	54.00%	(Wilkerson et al., 2004)
An. crucians D	293	54.00%	(Wilkerson et al., 2004)
An. crucians E	195	56.00%	(Wilkerson et al., 2004)
An. bradleyi	208	54.00%	(Wilkerson et al., 2004)
An. fluminensis	596	56.5%	(Brelsfoard et al., 2006)
An. minimus A	373	ND	(Phuc et al., 2003)
An. minimus C	375	ND	(Phuc et al., 2003)
An. maculipennis	280	54.10%	(Marinucci et al., 1999)
An. sacharovi	300	49.00%	(Marinucci et al., 1999)
An. nuneztovari	363-369	55.3-55.9%	(Fritz et al., 1994)
An. hyrcanus	436	44.90%	(Ma and Xu, 2005)
An. crawfordi	469	46.80%	(Ma and Xu, 2005)
An. bancroftii	430	ND	(Beebe et al., 2001)

In total it was possible to amplify, clone and sequence a fragment of interest in the following specimens: Clade I: 2 specimens from Kalimantan and 4 from Mae Hong Son; clade II: 4 specimens from Sumatra; clade III: 5 from Sakaeo, 6 from Mae Hong Son, 6 from Tak and 5 from Kanchanaburi; clade IV: 14 from Trat, 2 from Sa Kaeo and 1 from Sumatra and finally clade V: 2 specimens from Sa Kaeo (The numbering of clades follows that in the COI analysis).

Different size PCR products were observed within the Barbirostris Subgroup. The fragments varied from ~1.6 kb in specimens from clades I, IV and V to ~1.8 in specimens from clades II and III (Figure 5.2).

5.3.2 Sequence analysis.

Forward and reverse sequencing of the inserts from cloned PCR products was carried out to obtain entire sequence of ITS2. However, typically this would yield a maximum product of ~850 bp in each direction (forward and reverse), i.e. a total of ~1.7 kb. This was insufficient in many cases and internal primers successfully complete the whole 5.8S- ITS2-28S sequence (see Chapter 3). The alignment of sequences required considerable care due to the presence of internal repeats (see below).

The sequences obtained were subjected to a BLAST (Basic Local Alignment Search Tool) analysis. This permits comparison of a query sequence with DNA database. The results of this BLAST search showed that fragments of the 5.8S and 28S regions were amplified in specimens from all clades since these regions are highly similar in all *Anopheles* mosquitoes. The highest score using the conserved 5.8S and 28S were obtained from comparison with members of the Hyrcanus Group (Figure 5.4). However, in contrast the ITS2 region was not similar to any dataset in Gen Bank (In proof: See subsequent references to Saeung *et al.*, 2007).

Figure 5.4 BLAST analysis of the whole amplicon 5.8S (partial)-ITS2- 28S (partial).

This revealed that the ITS2 sequence did not match any in the data base as of October 2007. For the conserved 5.8S and 28S regions, the highest score corresponded to *An. sinensis* a member of the Hyrcanus Group, with 88% identity in the 5.8S and 84% in the 28S (see below).

BLAST OUTPUT

gb|AY339278.1| Anopheles sinensis 5.8S ribosomal RNA gene, partial sequence, internal transcribed spacer 2, complete sequence, and 28S ribosomal RNA gene, partial sequence
Length=682

Score = 178 bits (196), Expect = 9e-41
Identities = 128/144 (88%), Gaps = 8/144 (5%)
Strand=Plus/Plus

5.8 S

```
Query  1    ATCACTCGGCTCATGGATCGATGAAGACCGCAGCTAAATGCGCGTCATAATGTGAACTGC  60
            ||||||||||||||||||||||||||||||||||||| |||||||  ||||||||||||
Sbjct  11   ATCACTCGGCTCATGGATCGATGAAGACCGCAGCTAAATGCGCGTCATAATGTGAACTGC  70

Query  61   AGGACACATGGGAACATTGATAAGTTGAACGCATATTACACGTCGTG---CGTCACGGCA  116
            |||||||| ||||||||||||||||||||||||| |||||||||     ||| ||  ||
Sbjct  71   AGGACACAT-GAACATTGATAAGTTGAACGCATATTGCACGTCGTGGGAACGT-AC--CA  126

Query  117  TGATGTACACATGGTTGAGTGCTT  140
             ||| ||||||||  ||||| ||||
Sbjct  127  TGACGTACACATACTTGAGCGCTT  150
```

28 S

```
Score = 44.6 bits (48),  Expect = 1.4
Identities = 33/39 (84%), Gaps = 0/39 (0%)
Strand=Plus/Plus

Query  1689  GCCTCAACTCATGGGAGACTACCCCCTAAATTTAAGCAT  1727
              ||||||| | ||| | ||| ||||||| ||||||||||
Sbjct  616   GCCTCAAGTTATGTGTGACAACCCCCTGAATTTAAGCAT  654
```

Interestingly, slight discrepancies in relation to the exact start point of the ITS2 with respect to 5.8S were seen in the GenBank sequences, reflecting a degree of uncertainty as to where the 5.8S ends at the ITS2 begins. In this study, the limits of ITS2 were based on the *Anopheles gambiae* sequence (Paskewitz et al., 1993).

The sizes of the complete sizes 5.8S (partial) - ITS2 (complete) - 28S (partial) amplicon are shown in Table 5.2. This comprised a ~90 bp fragment from the 5.8S region, a complete ITS2 fragment and a 42 bp fragment from the 28S rDNA. A full list of specimens and the localities from which they were collected are in Table 5.3.

Table 5.2 Size of the rDNA regions amplified (bp) using primers 5.8S and 28S. Clade numbers are based on the results from the COI analysis. Note the size variation in ITS2.

Region amplified	Clade I	Clade II	Clade III	Clade IV	Clade V
5.8S	91	90	91	90	91
ITS2	~1545	~1727	~1730	~1583	~1519
28S	42	42	42	42	42
Total size	1674-1677	1858	1859-1862	1713-1718	1652

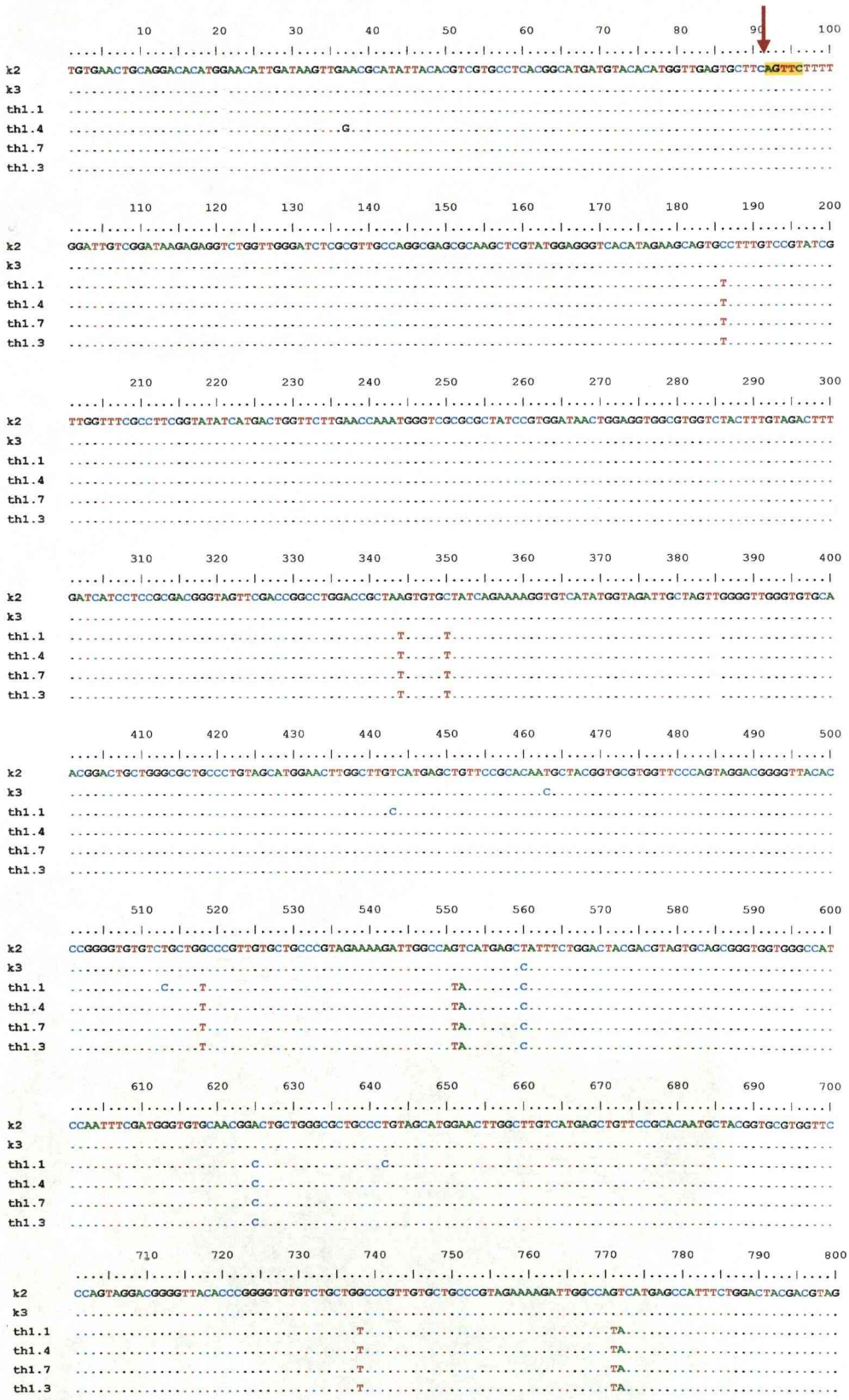
Alignment within clades was relatively easy due to the high sequence similarity. In contrast, sequences from different clades were "unalignable", with the exception of those from clades II and III (Figure 5.6), which show a similar size. Only a few gaps and substitutions were observed among sequences from the same clade. Entire alignments are shown at the end of this chapter (Figures 5.5-5.8)

The two populations (Kalimantan + Mae Hong Son) that comprise clade I, differ by 27 fixed substitutions and one fixed insertion/deletion event in the ITS2, together with one fixed substitution in the normally conserved region 5.8S (Figure 5.5).

Table 5.3 List of specimens and size of ITS2 (bp).

	Specimen	ITS2 size (bp)	Locality
Clade I	k2	1677	Kalimantan (Indonesia)
	k3	1676	"
	th1.1	1676	Mae Hong Son (Thailand)
	th1.3	1674	"
	th1.4	1674	"
	th1.7	1674	"
Clade II	l12	1858	Sumatra (Indonesia)
	l13	1858	"
	l15	1858	"
	l33	1858	"
Clade III	th39.3	1860	Mae Hong Son (Thailand)
	th1.6	1860	"
	th1.2	1861	"
	th1.8	1861	"
	th1.9	1860	"
	th1.10	1860	"
	bsk33	1861	Sa Kaeo (Thailand)
	bsk5	1861	"
	S17.1	1861	"
	S11.2	1861	"
	S24.3	1859	"
	ta10	1861	Tak (Thailand)
	ta19	1861	"
	ta21	1862	"
	ta22	1861	"
	ta23	1861	"
	ta24	1860	"
	kh3	1860	Kanchanaburi (Thailand)
	kh4	1861	"
	kh7	1860	"
	kh9	1860	"
	kh10	1860	"
Clade IV	l14	1713	Sumatra (Indonesia)
	btr7	1715	Trat (Thailand)
	btr8	1715	"
	btr10	1714	"
	btr11	1715	"
	btr16	1715	"
	btr17	1714	"
	btr18	1715	"
	btr19	1716	"
	btr22	1716	"
	btr23	1715	"
	ctr2	1716	"
	ctr4	1716	"
	T35.1	1715	"
	T35.2	1718	"
	bsk3	1716	Sa Kaeo (Thailand)
	S24.1	1716	"
Clade V	csk10	1652	Sa Kaeo (Thailand)
	bsk34	1652	"

Figure 5.5 Clade I: alignment of the fragment 5.8S (partial) + ITS2 (complete) + 28S (partial). Arrows indicate start and end of the internal transcribed spacer 2 in the sequence.



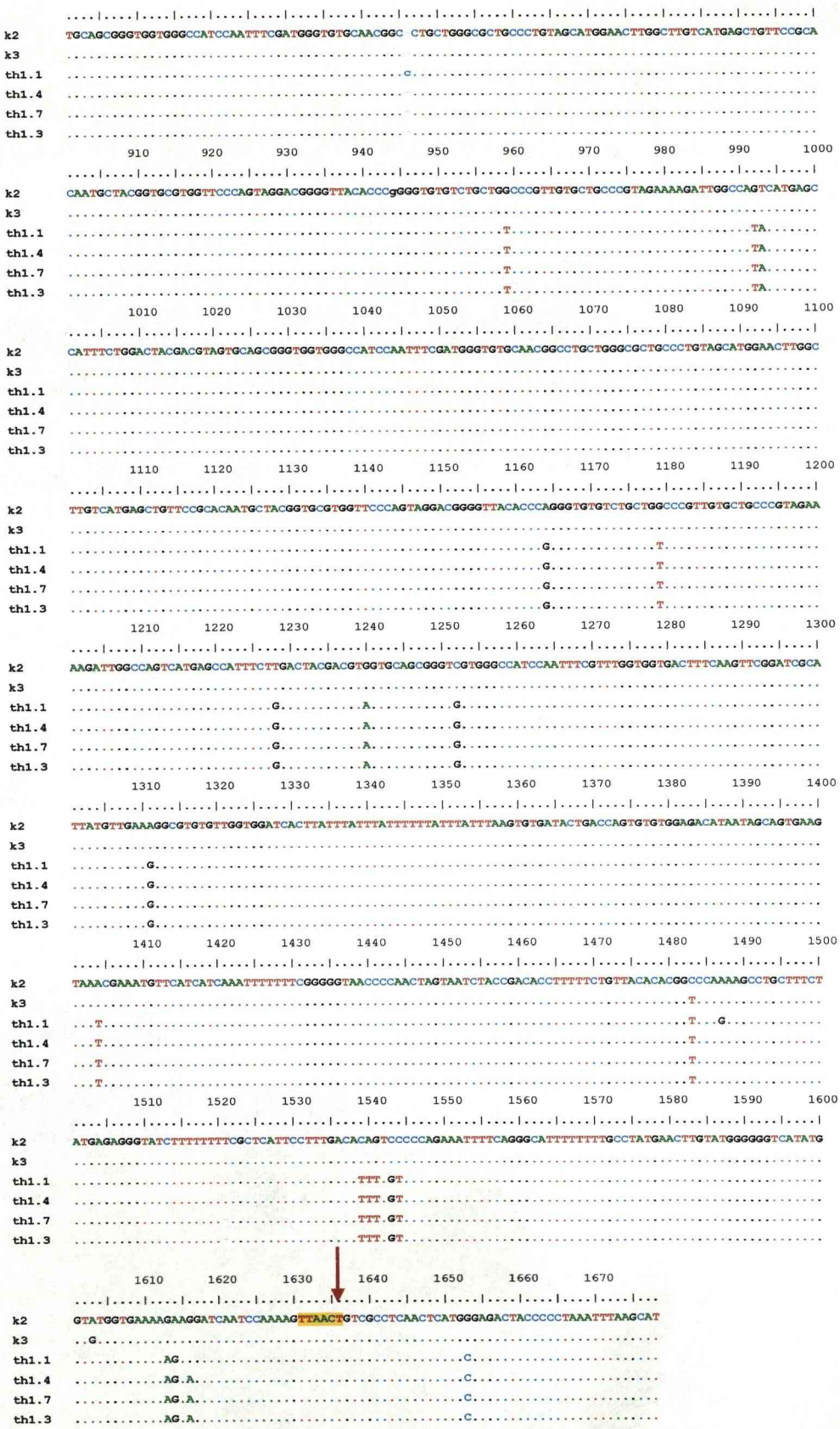
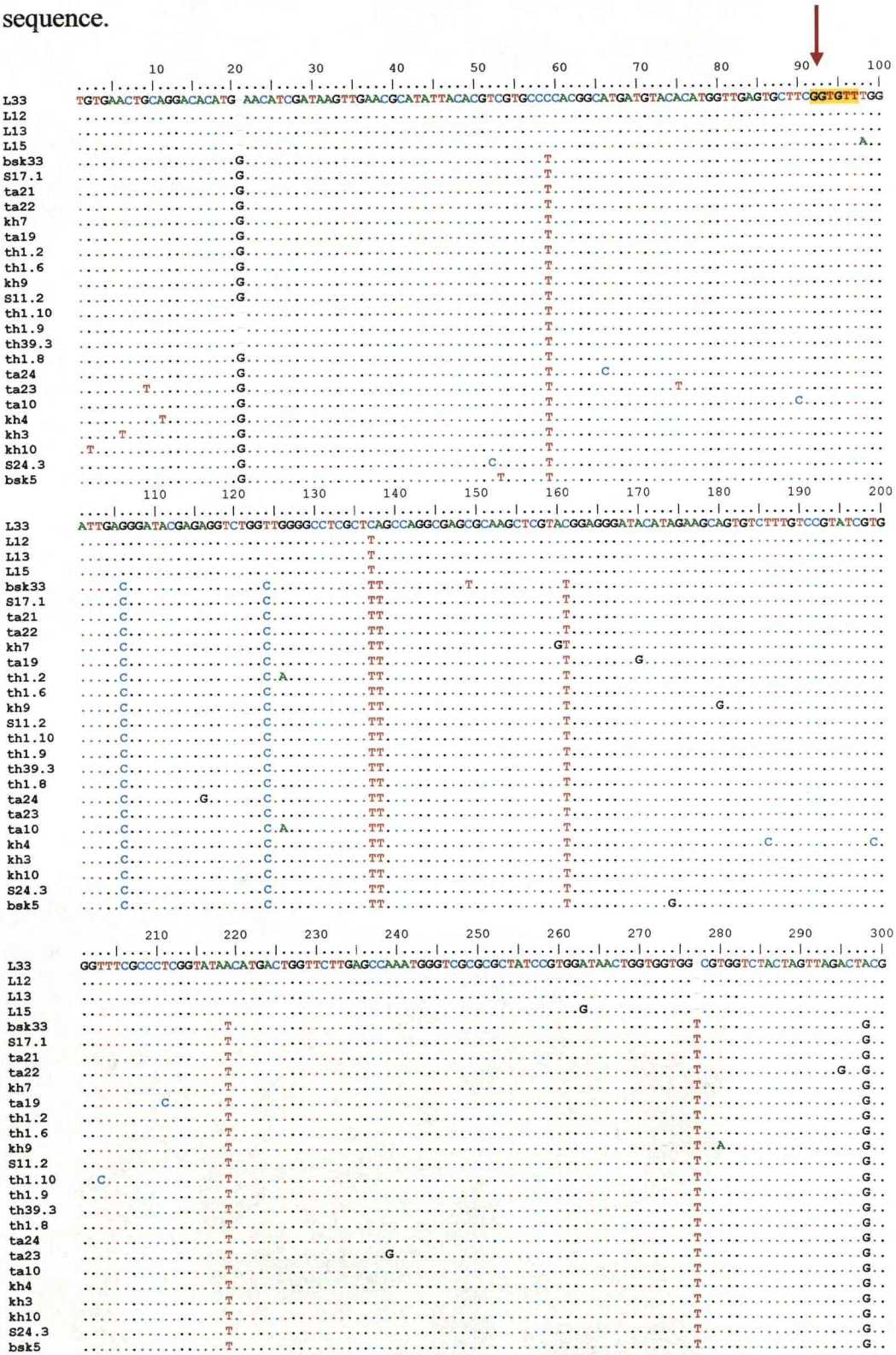
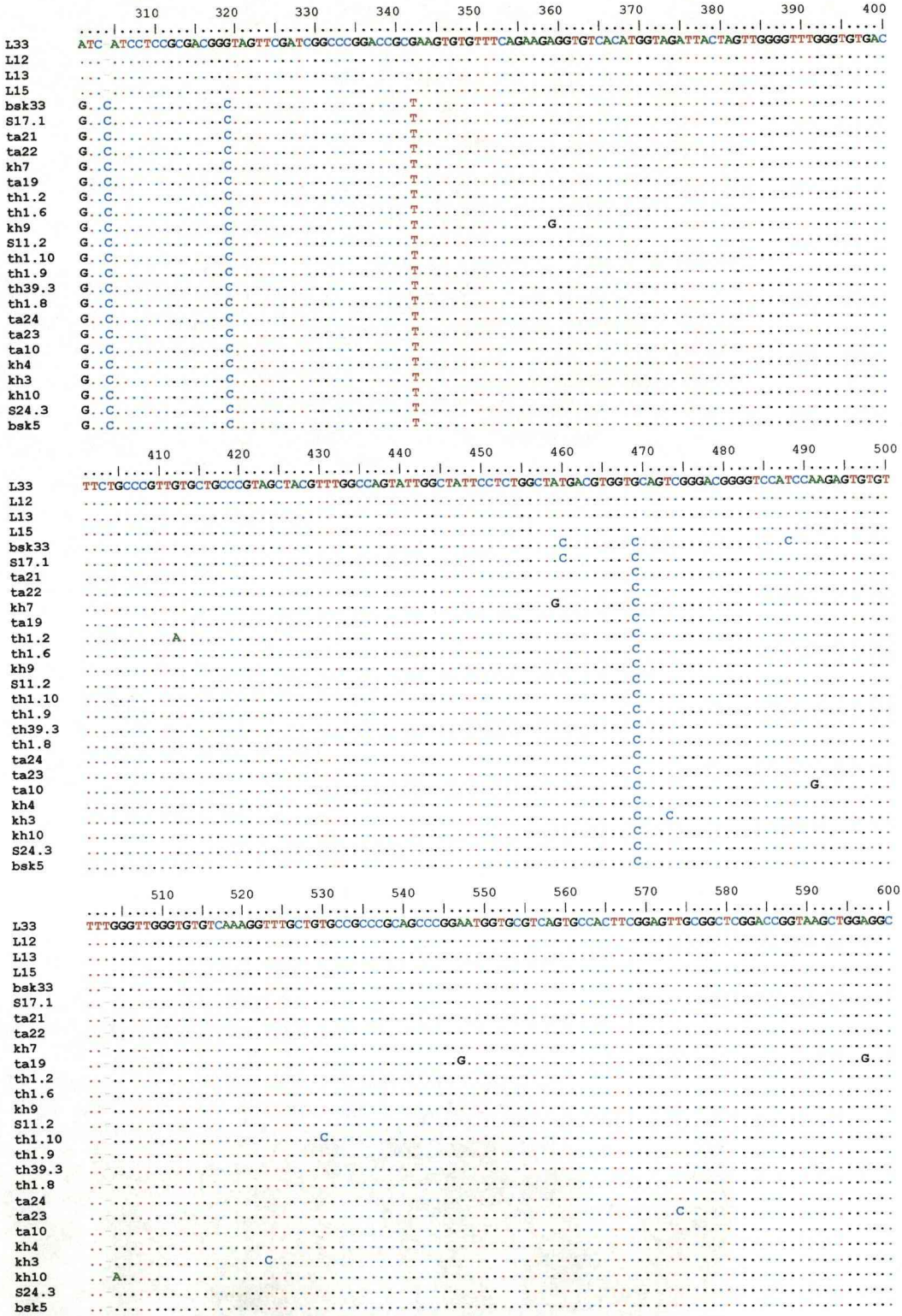
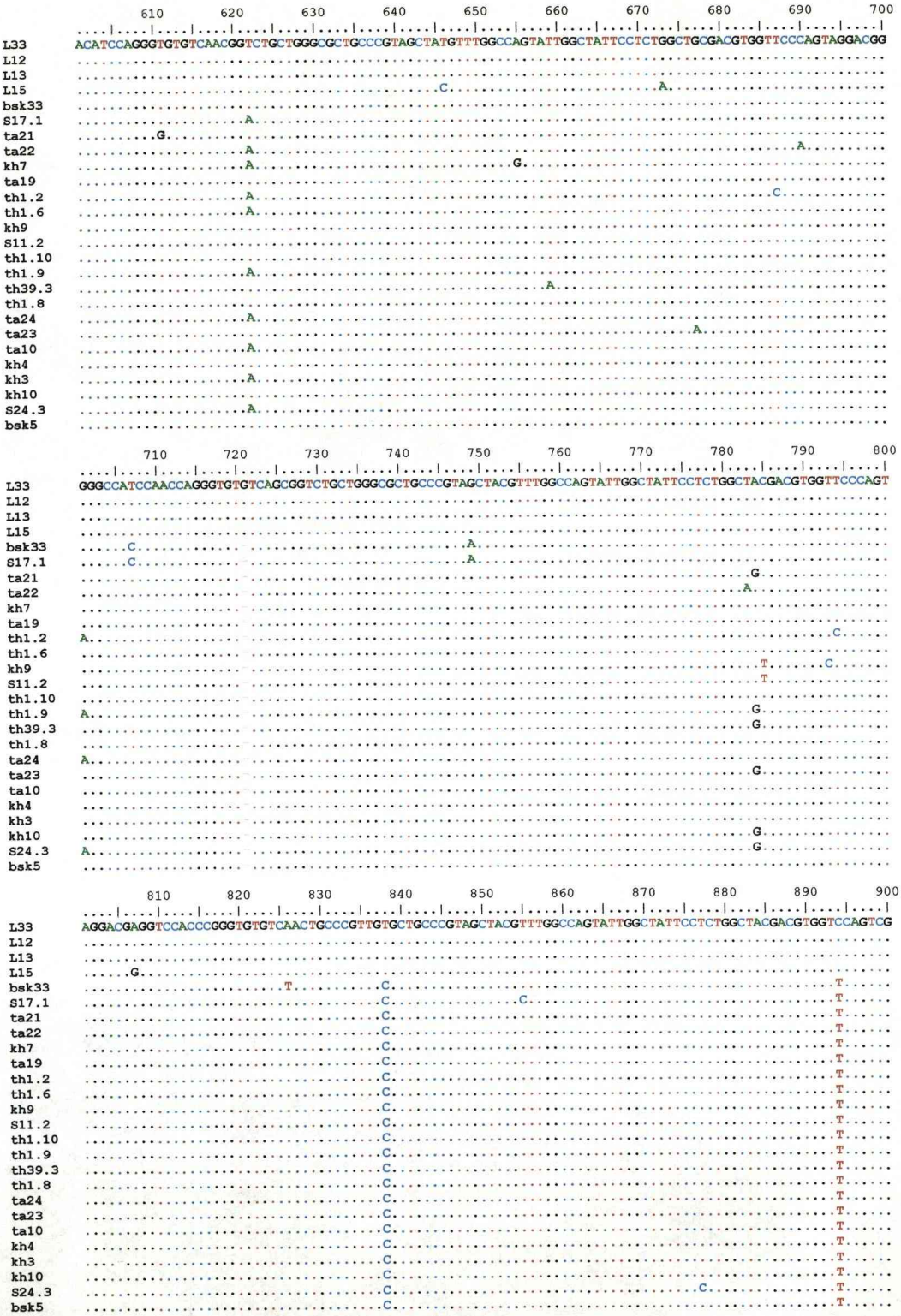
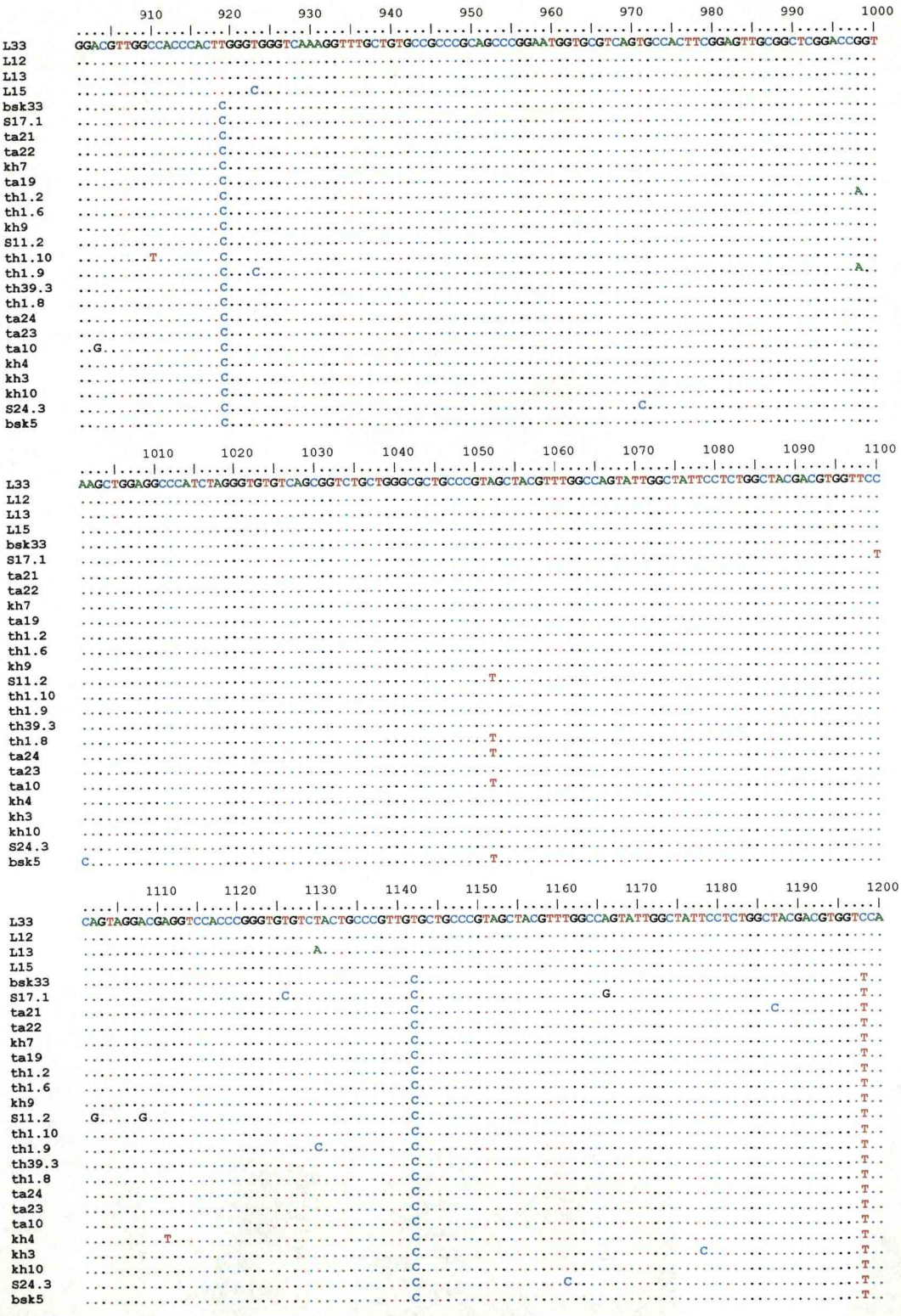

k2
k3
th1.1
th1.4
th1.7
th1.3
910 920 930 940 950 960 970 980 990 1000
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
1610 1620 1630 1640 1650 1660 1670

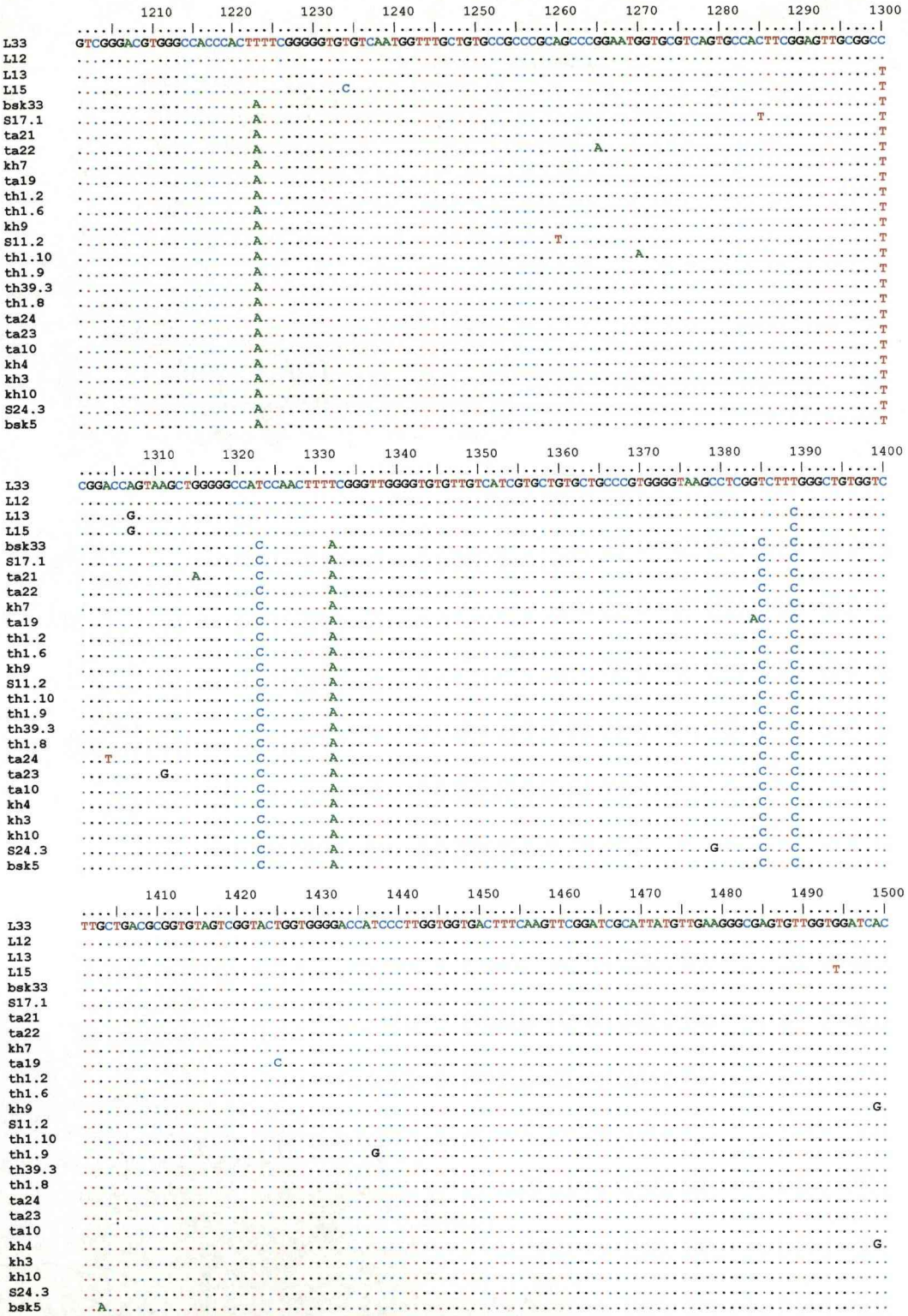
Figure 5.6 Total alignment of the fragment 5.8S (partial) + ITS2 (complete) + 28S (partial) of specimens from clades II (specimens L11- L13, L15) and III (remaining sequences). Arrows indicate start and end of the internal transcribed spacer 2 in the sequence.




310 320 330 340 350 360 370 380 390 400
L33 ATC ATCCTCCGCGACGGGTAGTTCGATCGGCCCGGACCGCGAAGTGTGTTTCAGAAGAGGTGTCACATGGTAGATTACTAGTTGGGGTTTGGGTGTGAC
L12
L13
L15
bsk33 G..C.....C.....T.....
S17.1 G..C.....C.....T.....
ta21 G..C.....C.....T.....
ta22 G..C.....C.....T.....
kh7 G..C.....C.....T.....
ta19 G..C.....C.....T.....
th1.2 G..C.....C.....T.....
th1.6 G..C.....C.....T.....
kh9 G..C.....C.....T.....G.....
S11.2 G..C.....C.....T.....
th1.10 G..C.....C.....T.....
th1.9 G..C.....C.....T.....
th39.3 G..C.....C.....T.....
th1.8 G..C.....C.....T.....
ta24 G..C.....C.....T.....
ta23 G..C.....C.....T.....
ta10 G..C.....C.....T.....
kh4 G..C.....C.....T.....
kh3 G..C.....C.....T.....
kh10 G..C.....C.....T.....
S24.3 G..C.....C.....T.....
bsk5 G..C.....C.....T.....
410 420 430 440 450 460 470 480 490 500
L33 TTCTGCCCGTTGTGCTGCCCGTAGCTACGTTTGGCCAGTATTGGCTATTCCTCTGGCTATGACGTGGTGCAGTCGGGACGGGGTCCATCCAAGAGTGTGT
L12
L13
L15
bsk33 .....C.....C.....C.....
S17.1 .....C.....C.....
ta21 .....C.....
ta22 .....C.....
kh7 .....G.....C.....
ta19 .....C.....
th1.2 ..........A.....C.....
th1.6 .....C.....
kh9 .....C.....
S11.2 .....C.....
th1.10 .....C.....
th1.9 .....C.....
th39.3 .....C.....
th1.8 .....C.....
ta24 .....C.....
ta23 .....C.....
ta10 .....C.....G.....
kh4 .....C.....
kh3 .....C...C.....
kh10 .....C.....
S24.3 .....C.....
bsk5 .....C.....
510 520 530 540 550 560 570 580 590 600
L33 TTTGGGTTGGGTGTGTCAAAGGTTTGCTGTGCCGCCCGCAGCCCGGAATGGTGCGTCAGTGCCACTTCGGAGTTGCGGCTCGGACCGGTAAGCTGGAGGC
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19 .....G.....G...
th1.2
th1.6
kh9
S11.2
th1.10 .....C.....
th1.9
th39.3
th1.8
ta24
ta23 .....C.....
ta10
kh4
kh3 .....C.....
kh10 .. A.....
S24.3
bsk5


610 620 630 640 650 660 670 680 690 700
L33 ACATCCAGGGTGTGTCAACGGTCTGCTGGGCGCTGCCCGTAGCTATGTTTGGCCAGTATTGGCTATTCCTCTGGCTGCGACGTGGTTCCCAGTAGGACGG
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
710 720 730 740 750 760 770 780 790 800
L33 GGGCCATCCAACCAGGGTGTGTCAGCGGTCTGCTGGGCGCTGCCCGTAGCTACGTTTGGCCAGTATTGGCTATTCCTCTGGCTACGACGTGGTTCCCAGT
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
810 820 830 840 850 860 870 880 890 900
L33 AGGACGAGGTCCACCCGGGTGTGTCAACTGCCCGTTGTGCTGCCCGTAGCTACGTTTGGCCAGTATTGGCTATTCCTCTGGCTACGACGTGGTCCAGTCG
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5


910 920 930 940 950 960 970 980 990 1000
L33 GGACGTTGGCCACCCACTTGGGTGGGTCAAAGGTTTGCTGTGCCGCCCGCAGCCCGGAATGGTGCGTCAGTGCCACTTCGGAGTTGCGGCTCGGACCGGT
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
L33 AAGCTGGAGGCCCATCTAGGGTGTGTCAGCGGTCTGCTGGGCGCTGCCCGTAGCTACGTTTGGCCAGTATTGGCTATTCCTCTGGCTACGACGTGGTTCC
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
L33 CAGTAGGACGAGGTCCACCCGGGTGTGTCTACTGCCCGTTGTGCTGCCCGTAGCTACGTTTGGCCAGTATTGGCTATTCCTCTGGCTACGACGTGGTCCA
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
L33 GTCGGGACGTGGGCCACCCACTTTTCGGGGGTGTGTCAATGGTTTGCTGTGCCGCCCGCAGCCCGGAATGGTGCGTCAGTGCCACTTCGGAGTTGCGGCC
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
L33 CGGACCAGTAAGCTGGGGGCCATCCAACTTTTCGGGTTGGGGTGTGTTGTCATCGTGCTGTGCTGCCCGTGGGGTAAGCCTCGGTCTTTGGGCTGTGGTC
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
L33 TTGCTGACGCGGTGTAGTCGGTACTGGTGGGGACCATCCCTTGGTGGTGACTTTCAAGTTCGGATCGCATTATGTTGAAGGGCGAGTGTTGGTGGATCAC
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5

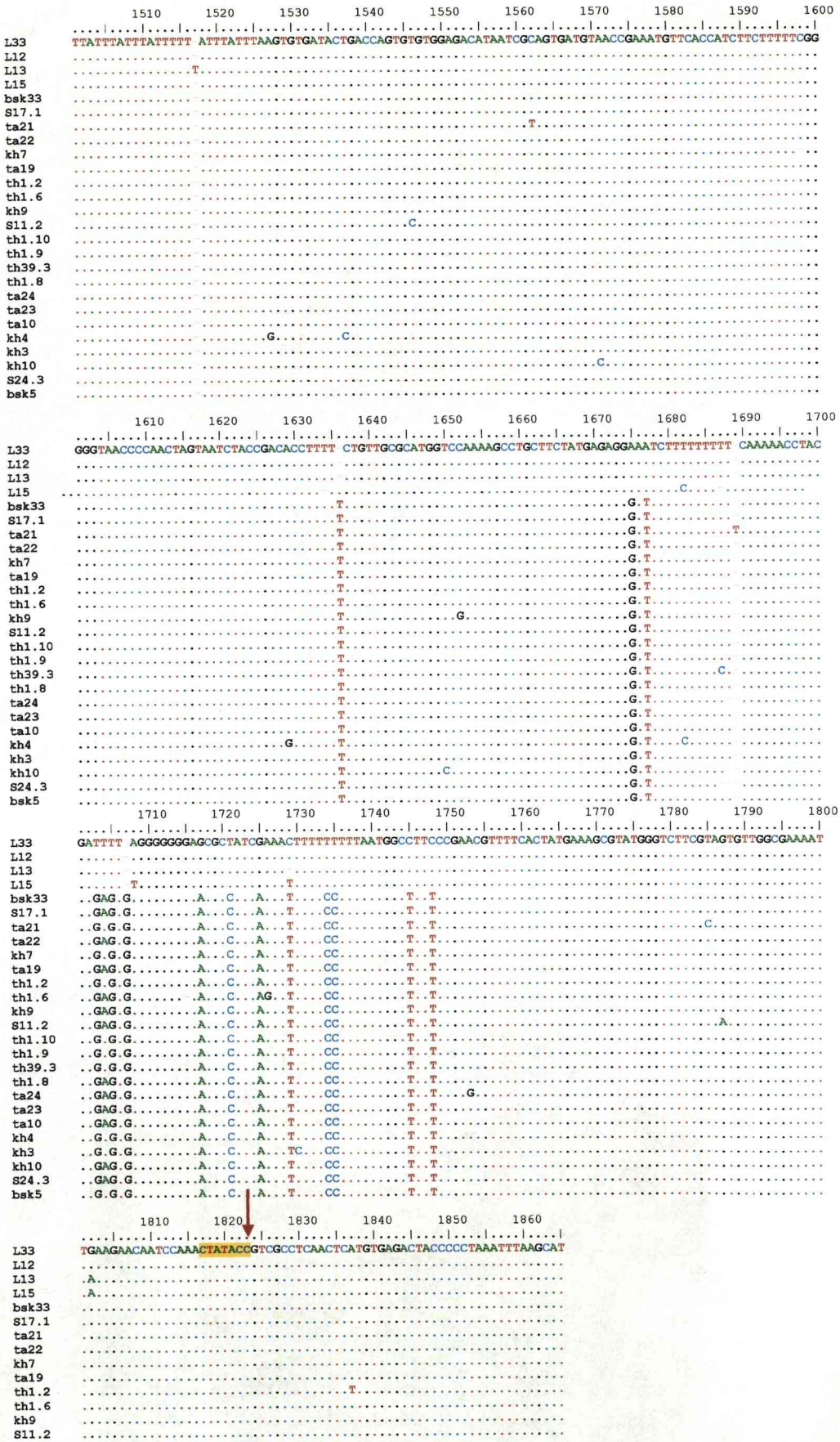
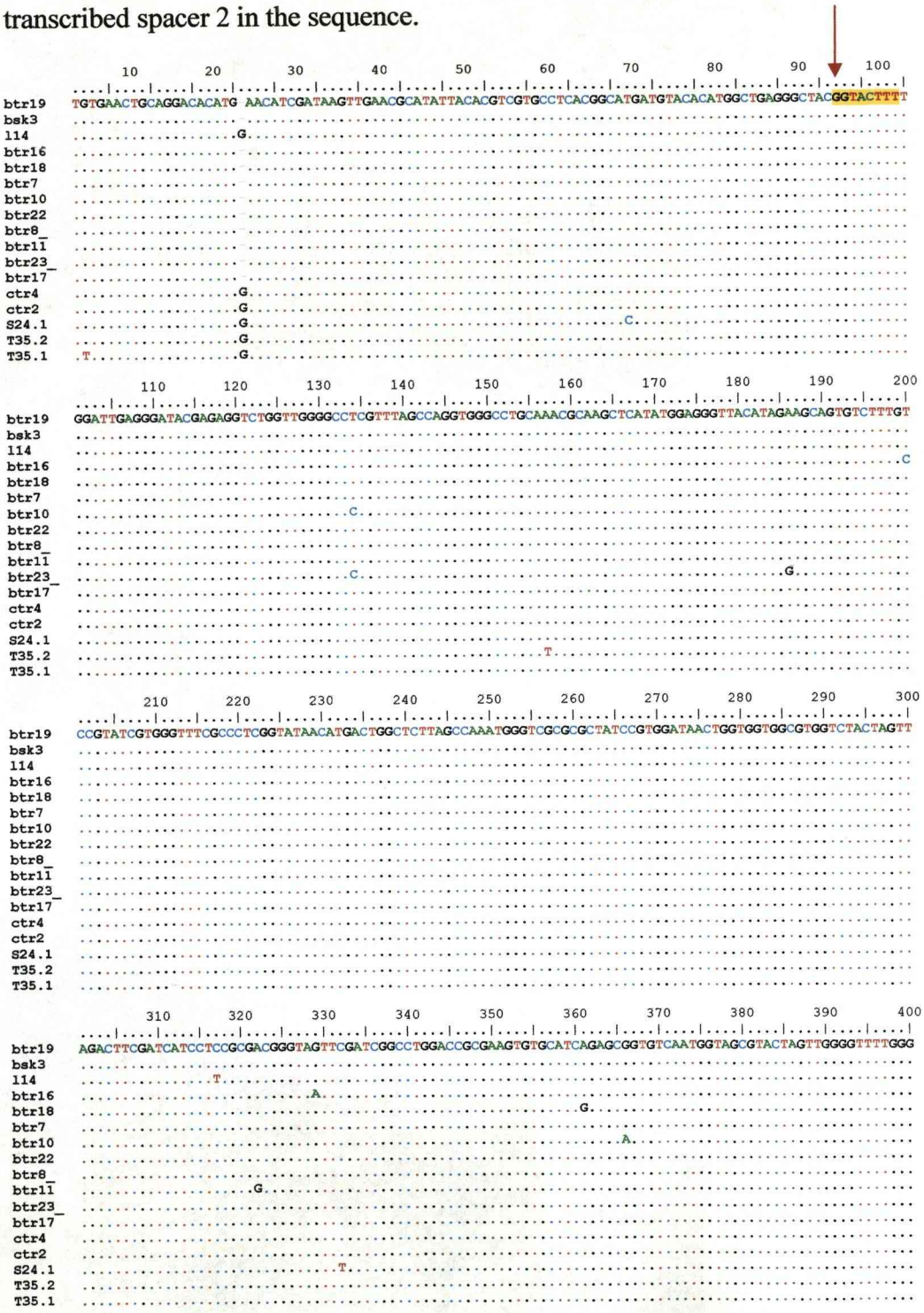
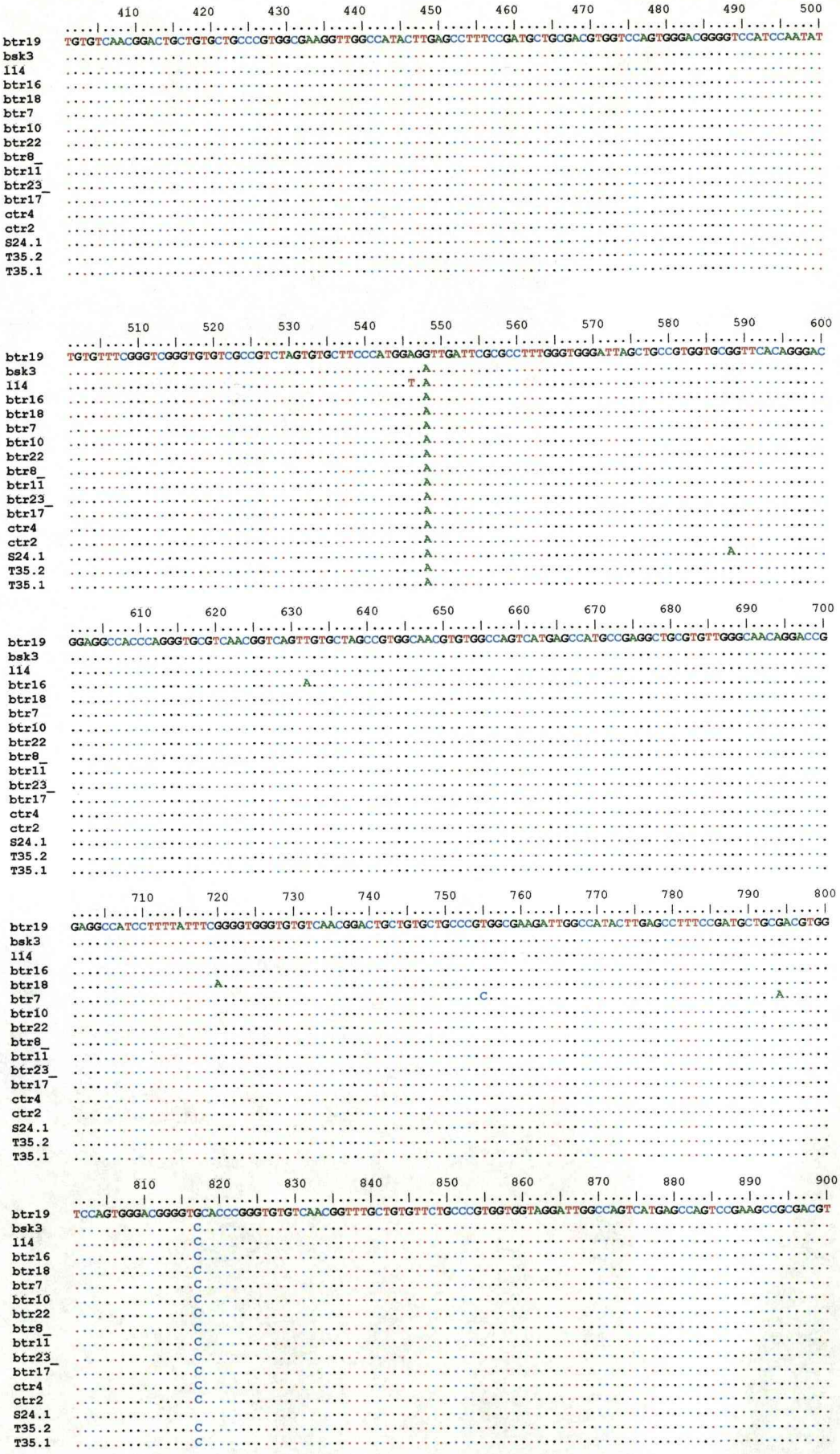
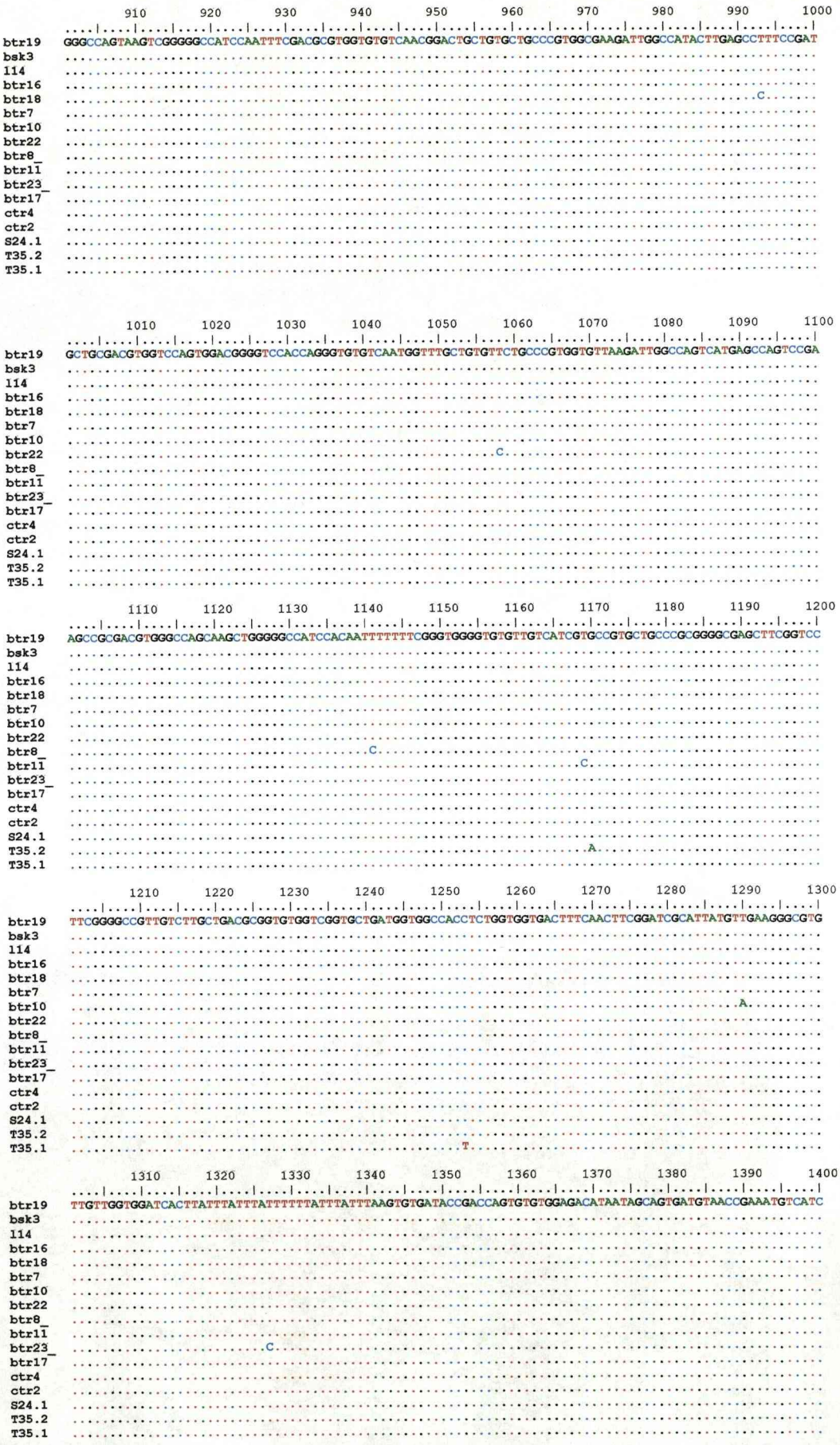

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
L33 TTATTTATTTATTTTT ATTTATTTAAGTGTGATACTGACCAGTGTGTGGAGACATAATCGCAGTGATGTAACCGAAATGTTCACCATCTTCTTTTTCGG
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
L33 GGGTAACCCCAACTAGTAATCTACCGACACCTTTT CTGTTGCGCATGGTCCAAAAGCCTGCTTCTATGAGAGGAAATCTTTTTTTTT CAAAAACCTAC
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
L33 GATTTT AGGGGGGGAGCGCTATCGAAACTTTTTTTTTAATGGCCTTCCCGAACGTTTTCACTATGAAAGCGTATGGGTCTTCGTAGTGTTGGCGAAAAT
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2
th1.10
th1.9
th39.3
th1.8
ta24
ta23
ta10
kh4
kh3
kh10
S24.3
bsk5
1810 1820 1830 1840 1850 1860
L33 TGAAGAACAATCCAAACTATACCGTCGCCTCAACTCATGTGAGACTACCCCCTAAATTTAAGCAT
L12
L13
L15
bsk33
S17.1
ta21
ta22
kh7
ta19
th1.2
th1.6
kh9
S11.2

Figure 5.7 Total alignment of the fragment 5.8S (partial) + ITS2 (complete) + 28S (partial) of specimens from clade IV. Arrows indicate start and end of the internal transcribed spacer 2 in the sequence.




410 420 430 440 450 460 470 480 490 500
btr19 TGTGTCAACGGACTGCTGTGCTGCCCGTGGCGAAGGTTGGCCATACTTGAGCCTTTCCGATGCTGCGACGTGGTCCAGTGGGACGGGGTCCATCCAATAT
bsk3
114
btr16
btr18
btr7
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
510 520 530 540 550 560 570 580 590 600
btr19 TGTGTTTCGGGTCGGGTGTGTCGCCGTCTAGTGTGCTTCCCATGGAGGTTGATTCGCGCCTTTGGGTGGGATTAGCTGCCGTGGTGCGGTTCACAGGGAC
bsk3 A
114 T.A
btr16 A
btr18 A
btr7 A
btr10 A
btr22 A
btr8_ A
btr11 A
btr23_ A
btr17 A
ctr4 A
ctr2 A
S24.1 A A
T35.2 A
T35.1 A
610 620 630 640 650 660 670 680 690 700
btr19 GGAGGCCACCCAGGGTGCGTCAACGGTCAGTTGTGCTAGCCGTGGCAACGTGTGGCCAGTCATGAGCCATGCCGAGGCTGCGTGTTGGGCAACAGGACCG
bsk3
114
btr16 A
btr18
btr7
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
710 720 730 740 750 760 770 780 790 800
btr19 GAGGCCATCCTTTTATTTCGGGGTGGGTGTGTCAACGGACTGCTGTGCTGCCCGTGGCGAAGATTGGCCATACTTGAGCCTTTCCGATGCTGCGACGTGG
bsk3
114
btr16
btr18 A
btr7 C A
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
810 820 830 840 850 860 870 880 890 900
btr19 TCCAGTGGGACGGGGTGCACCCGGGTGTGTCAACGGTTTGCTGTGTTCTGCCCGTGGTGGTAGGATTGGCCAGTCATGAGCCAGTCCGAAGCCGCGACGT
bsk3 C
114 C
btr16 C
btr18 C
btr7 C
btr10 C
btr22 C
btr8_ C
btr11 C
btr23_ C
btr17 C
ctr4 C
ctr2 C
S24.1
T35.2 C
T35.1 C


910 920 930 940 950 960 970 980 990 1000
btr19 GGGCCAGTAAGTCGGGGGCCATCCAATTTCGACGCGTGGTGTGTCAACGGACTGCTGTGCTGCCCGTGGCGAAGATTGGCCATACTTGAGCCTTTCCGAT
bsk3
l14
btr16
btr18 ....C.....
btr7
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
btr19 GCTGCGACGTGGTCCAGTGGACGGGGTCCACCAGGGTGTGTCAATGGTTTGCTGTGTTCTGCCCGTGGTGTTAAGATTGGCCAGTCATGAGCCAGTCCGA
bsk3
l14
btr16
btr18
btr7
btr10
btr22 ....C.....
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
btr19 AGCCGCGACGTGGGCCAGCAAGCTGGGGGCCATCCACAATTTTTTTCGGGTGGGGTGTGTTGTCATCGTGCCGTGCTGCCCGCGGGGCGAGCTTCGGTCC
bsk3
l14
btr16
btr18
btr7
btr10
btr22
btr8_ ....C.....
btr11 ....C.....
btr23_
btr17
ctr4
ctr2
S24.1
T35.2 ....A.....
T35.1
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
btr19 TTCGGGGCCGTTGTCTTGCTGACGCGGTGTGGTCGGTGCTGATGGTGGCCACCTCTGGTGGTGACTTTCAACTTCGGATCGCATTATGTTGAAGGGCGTG
bsk3
l14
btr16
btr18
btr7
btr10 ....A.....
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1 ....T.....
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
btr19 TTGTTGGTGGATCACTTATTTATTTATTTTTTTATTTATTTAAGTGTGATACCGACCAGTGTGTGGAGACATAATAGCAGTGATGTAACCGAAATGTCATC
bsk3
l14
btr16
btr18
btr7
btr10
btr22
btr8_
btr11
btr23_ ....C.....
btr17
ctr4
ctr2
S24.1
T35.2
T35.1

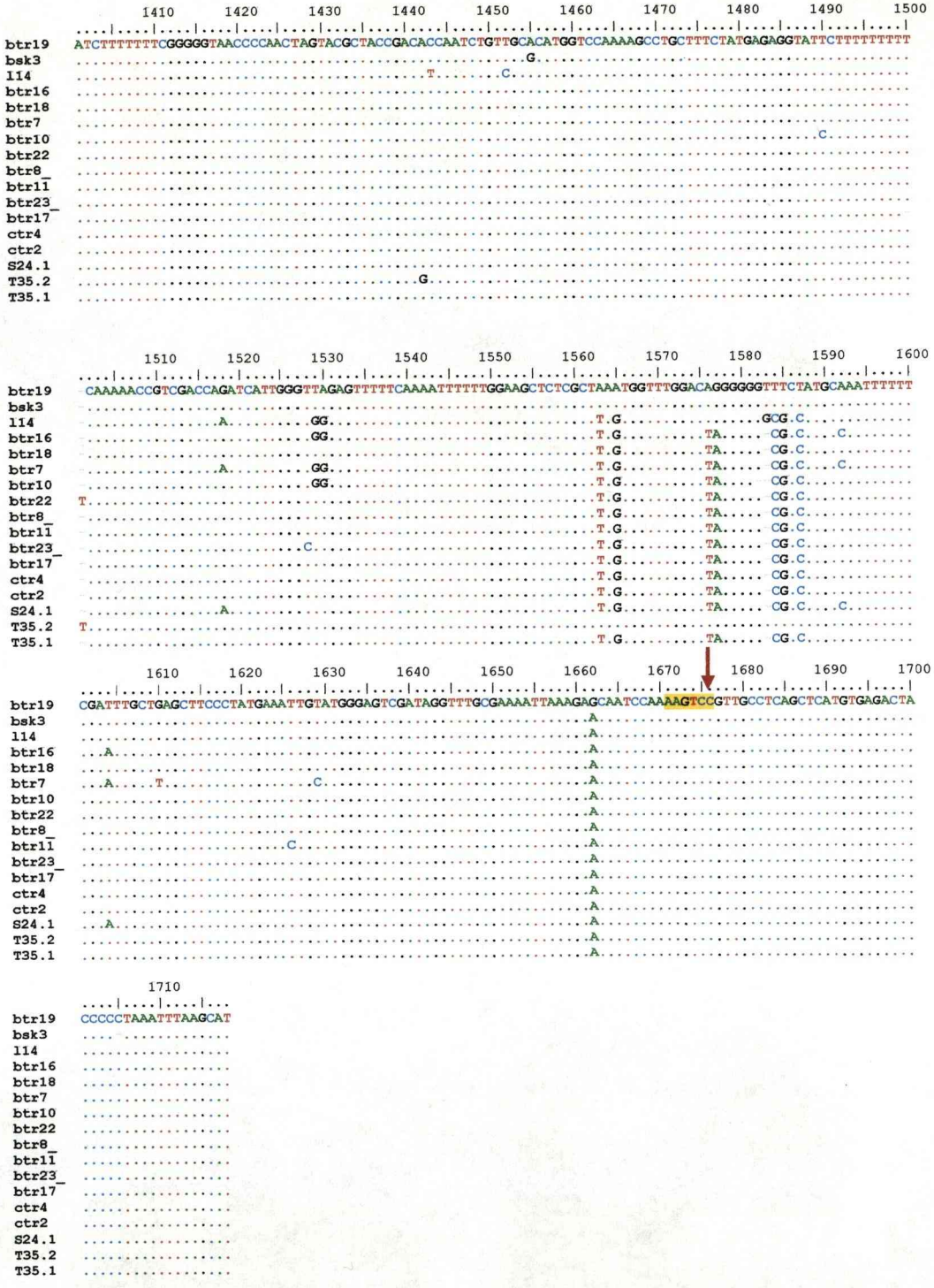
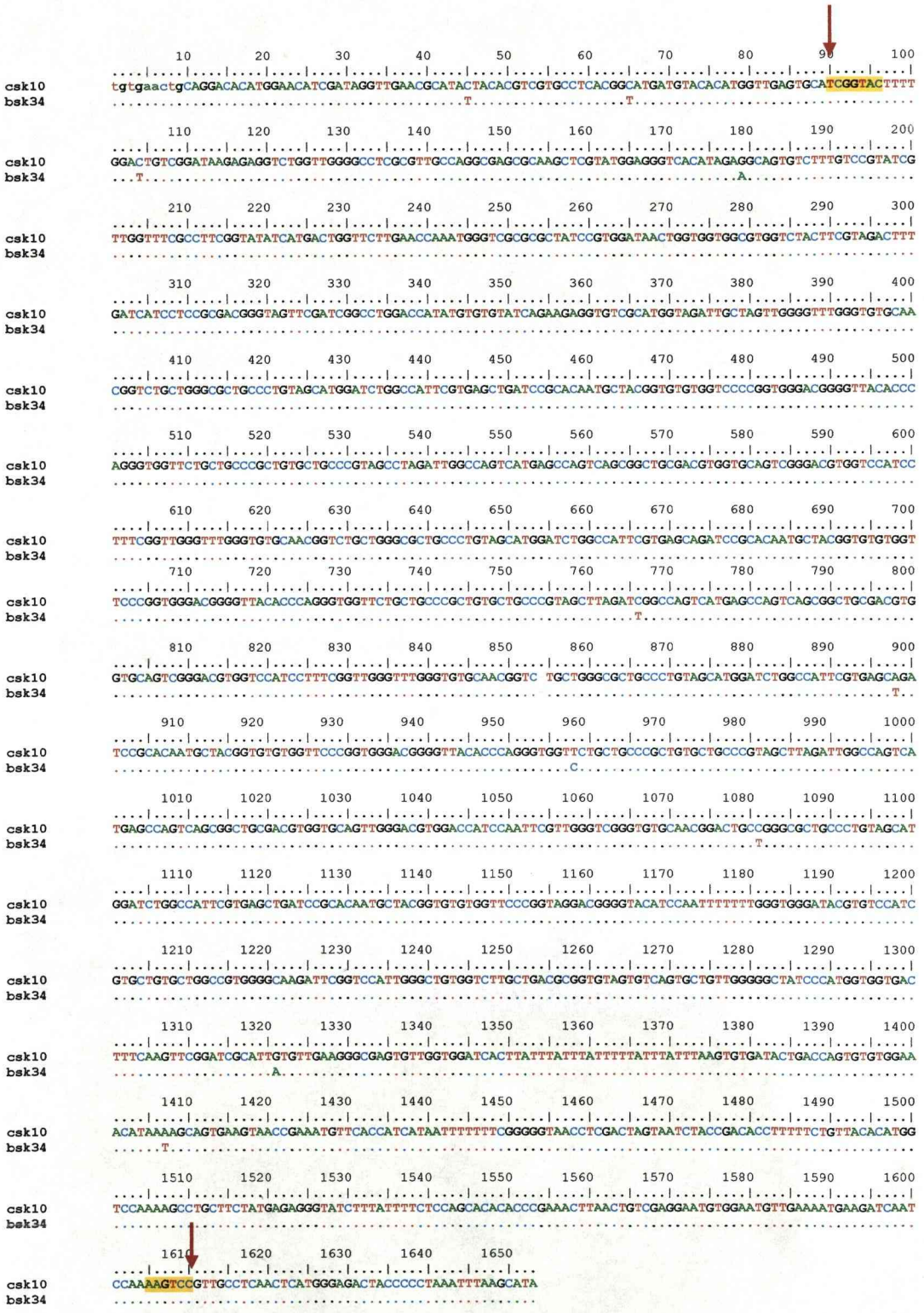

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
btr19 ATCTTTTTTTCGGGGGTAACCCCAACTAGTACGCTACCGACACCAATCTGTTGCACATGGTCCAAAAGCCTGCTTTCTATGAGAGGTATTCTTTTTTTTT
bsk3
l14
btr16
btr18
btr7
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
btr19 CAAAAACCGTCGACCAGATCATTGGGTTAGAGTTTTTCAAAATTTTTTGGAAGCTCTCGCTAAATGGTTTGGACAGGGGGGTTTCTATGCAAATTTTTT
bsk3
l14
btr16
btr18
btr7
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
btr19 CGATTTGCTGAGCTTCCCTATGAAATTGTATGGGAGTCGATAGGTTTGCGAAAATTAAAGAGCAATCCAAAAGTCCGTTGCCTCAGCTCATGTGAGACTA
bsk3
l14
btr16
btr18
btr7
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1
1710
btr19 CCCCCTAAATTTAAGCAT
bsk3
l14
btr16
btr18
btr7
btr10
btr22
btr8_
btr11
btr23_
btr17
ctr4
ctr2
S24.1
T35.2
T35.1

Figure 5.8 Total alignment of the fragment 5.8S (partial) + ITS2 (complete) + 28S (partial) of specimens from clade V. Arrows indicate start and end of the ITS2 in the sequence.



Since the sample sizes are small, no firm conclusion is possible but the differences between the two populations are large, even for allopatric populations within a species (see discussion). In clades II and III, the ITS2 is almost the same size (~1730), but there are 30 fixed substitutions and 5 fixed insertion/deletion events in the ITS2 sequence. In addition, one substitution is present in the conserved 5.8S region (Figure 5.6).

Haplotype diversity in the entire dataset was high, with 47 haplotypes in 51 specimens. The majority of haplotypes comprised a single specimen, with the exception of haplotype 5, in clade I (formed by specimens th1.3, th1.4 and th1.7) and haplotype 11 in clade IV (formed by specimens btr17, ctr2 and ctr4). The number of gaps was directly proportional to the number of specimens found in each clade: Clade I (4 gaps), clade II (2 gaps), cladeIII (7), clade IV (8) and clade V (0). The G+C content of the ITS2 was similar in all clades, 52.3% (clade I), 55.6% (clade II), 55.9% (clade III), 54.3% (clade IV) and 55.1% (clade V). Among the substitutions, the ratio of transitions/transversions was high in clades II and III (2.73), clade IV (2.37) and clade V (8). For clade I the number of transversions was equal.

To sum up: taking into consideration differences in sizes and the presence of fixed substitutions, the entire dataset of specimens from the Barbirostris Subgroup can be divided in five clades: I, II, III, IV and V, which comprise the same specimens observed in the clades obtained from the cytochrome I analysis (Chapter 4).

5.3.3 DNA repeats analysis

The exceptional length of the ITS2 region was due to the presence of repeated elements, size and organization of which varied in the different clades. These repeats were located at the centre of the ITS2, occupying between 55% to 61% of the total length of the spacer. Each repeat was approximately 220 bp in length with evidence of individual duplication (see below). In the following sections there is a broad description of the arrangement of the repeats in each clade, as they varied in number, size and similarity. Analysis is presented on ITS2 size and the arrangement

of repeats, starting with clades I, V and IV (ITS2 = 1519 – 1583 bp) followed by clades II and III (ITS2 = 1727 bp).

5.3.3.1 Repeats in clade I

Detailed examination of ITS2 sequences in the six specimens of this clade, show that they contain 4 consecutive and almost perfect repeats. In Figure 5.4, repeats of specimen k2 are shown as an example. They were all 220 bp in length and showed a similarity of 96.8% (between repeats 1 and 4) and 99.5% (for repeats 1 and 2). Repeat 4 was the most variable, with a minimum of 4 substitution differences from repeats 1, 2 and 3. Seven substitutions were seen between repeats, of which 4 were transversions and 3 were transitions; most of the substitutions observed between the repeats appear in the second half of the sequence (Figure 5.9).

A *GGGTGTG* motif occurs at the beginning of each repeat, again close to its central point (Figure 5.9). When the repeats of the 220 bp fragment are subdivided into two parts of ~110 bp by the motif *GGGTGTG*, the partitions show a 53% similarity (Figure 5.10).

Figure 5.9 Alignment of the repeats found in clade I (specimen k2).

The repeats are consecutive and have been numbered according to their position in the sequence. The GGGTGTG motif is underlined. It appears at the beginning and in the middle of each repeat.

```
                    10        20        30        40        50        60
                    |....|....|....|....|....|....|....|....|....|....|....|
K2 repeats1  GGGTGTGCAACGGACTGCTGGGCGCTGCCCTGTAGCATGGAACTTGGCTTGTCATGAGCT
K2 repeats2  ...........................................................
K2 repeats3  ............C..............................................
K2 repeats4  ............C..............................................

                    70        80        90        100       110       120
                    |....|....|....|....|....|....|....|....|....|....|....|
K2 repeats1  GTTCCGCACAATGCTACGGTGCGTGGTTCCCAGTAGGACGGGGTTACACCCGGGGTGTGT
K2 repeats2  ...........................................................
K2 repeats3  ...........................................................
K2 repeats4  ...........................................................

                    130       140       150       160       170       180
                    |....|....|....|....|....|....|....|....|....|....|....|
K2 repeats1  CTGCTGGCCCGTTGTGCTGCCCGTAGAAAAGATTGGCCAGTCATGAGCTATTTCTGGACT
K2 repeats2  ...........................................................
K2 repeats3  ...........................................................
K2 repeats4  ...........................................................

                    190       200       210       220
                    |....|....|....|....|....|....|....|
K2 repeats1  ACGACGTAGTGCAGCGGGTGGTGGGCCATCCAATTTCGAT
K2 repeats2  ...................................
K2 repeats3  ...................................
K2 repeats4  .......G...........C...............T.
```

Figure 5.10 Alignment of the subrepeats in clade I. The motif *GGGTGTG* was used to divide sequence into 8 subrepeats.

```
                        10        20        30        40        50        60
                        |....|....|....|....|....|....|....|....|....|....|....|
K2 subrepeat1a  GGGTGTGCAACGGACTGCTGGGCGCTGCCCTGTAGCATGGAACTTGGCTTGTCATGAGCT
K2 subrepeat1b  .......TCTGCTGGCC.GTT.T.......  ....A.AA..  .....CA.........
K2 subrepeat2a  ...........................................................
K2 subrepeat2b  .......TCTGCTGGCC.GTT.T.......  ....A.AA..  .....CA.........C
K2 subrepeat3a  ............C..............................................
K2 subrepeat3b  .......TCTGCTGGCC.GTT.T.......  ....A.AA..  .....CA.........C
K2 subrepeat4a  ............C..............................................
K2 subrepeat4b  .......TCTGCTGGCC.GTT.T.......  ....A.AA..  .....CA.........C

                        70        80        90        100       110
                        |....|....|....|....|....|....|....|....|....|....|....
K2 subrepeat1a  GTTCCGCACAATGCTACGGTGCGTGGTTCCCAGTAGGACGG  GGTTACACCCG
K2 subrepeat1b  A..T. TGG.C.A.G...TA.T.CA.CGGGTG..G..C.ATCCAA..T.GAT
K2 subrepeat2a  ...........................................  ..........
K2 subrepeat2b  A..T. TGG.C.A.G...TA.T.CA.CGGGTG..G..C.ATCCAA..T.GAT
K2 subrepeat3a  ...........................................  ..........A
K2 subrepeat3b  A..T. TGG.C.A.G...TA.T.CA.CGGGTG..G..C.ATCCAA..T.GAT
K2 subrepeat4a  ...........................................  ..........A
K2 subrepeat4b  A..T. TTG.C.A.G...TG.T.CA.CGGGTC..G..C.ATCCAA..T.GTT
```

5.3.3.2 Repeats in clade V

In this clade two specimens were sequenced. The repeats show a similar arrangement to clade I, with two marked differences:

1) The first 3 repeats are very similar throughout whereas repeat 4 diverges in sequence after position 112, where the motif *GGGTGGT* (a slight variation of motif *GGGTGTG*) is found (Figure 5.11). The similarity of the repeats varied from 64.7% (between repeats 1 and 4) to 98.6% (between 1 and 2).

2) Repeats show different sizes: 224 bp in repeats 1 and 2, 223 bp in repeat 3 and 222 bp in repeat 4, in contrast with the uniform 220 bp repeats found in clade I (Fig 5.4). Considering only the repeat regions, there were 8 transitions and 5 transversions.

Figure 5.11 Alignment of the repeats found in clade V (specimen bk34 is used as example). Motifs *GGGTGTG* and *GGGTGGT* are underlined. Dots indicate same nucleotide.

```
                        10        20        30        40        50        60
                    ....|....|....|....|....|....|....|....|....|....|....|....|
bsk34 repeat1       GGGTGTGCAACGGTCTGCTGGGCGCTGCCCTGTAGCATGGATCTGGCCATTCGTGAGCTG
bsk34 repeat2       ............................................................A.
bsk34 repeat3       ............................................................
bsk34 repeat4       ............A...........................................

                              70        80        90       100       110       120
                    ....|....|....|....|....|....|....|....|....|....|....|....|
bsk34 repeat1       ATCCGCACAATGCTACGGTGTGTGGTCCCCGGTGGGACGGGGTTACACCCAGGGTGGTTC
bsk34 repeat2       ........................T...................................
bsk34 repeat3       ........................T..............................C.
bsk34 repeat4       ........................T......A.........  ...T...ATT.TT..G

                             130       140       150       160       170       180
                    ....|....|....|....|....|....|....|....|....|....|....|....|
bsk34 repeat1       TGCTGCCCGCTGTGCTGCCCGTAGCCTAGATTGGCCAGTCATGAGCCAGTCAGCGGCTGC
bsk34 repeat2       ........................T...................................
bsk34 repeat3       ........................T...................................
bsk34 repeat4       G.TG.GATA.GTGT.CAT.GTGCTGTGCTGGCC.TGG.G..A..TT.G...CATT.GGCT

                             190       200       210       220
                    ....|....|....|....|....|....|....|....|....
bsk34 repeat1       GACGTGGTGCAGTCGGGACGTGGTCCATCCTTTCGGTTGGGTTT
bsk34 repeat2       ............................................
bsk34 repeat3       ............T.........A......AA.TC.......C
bsk34 repeat4       .TG..CT...T.A..C.GT..A..GTCAGTGC.GTTGG...
```

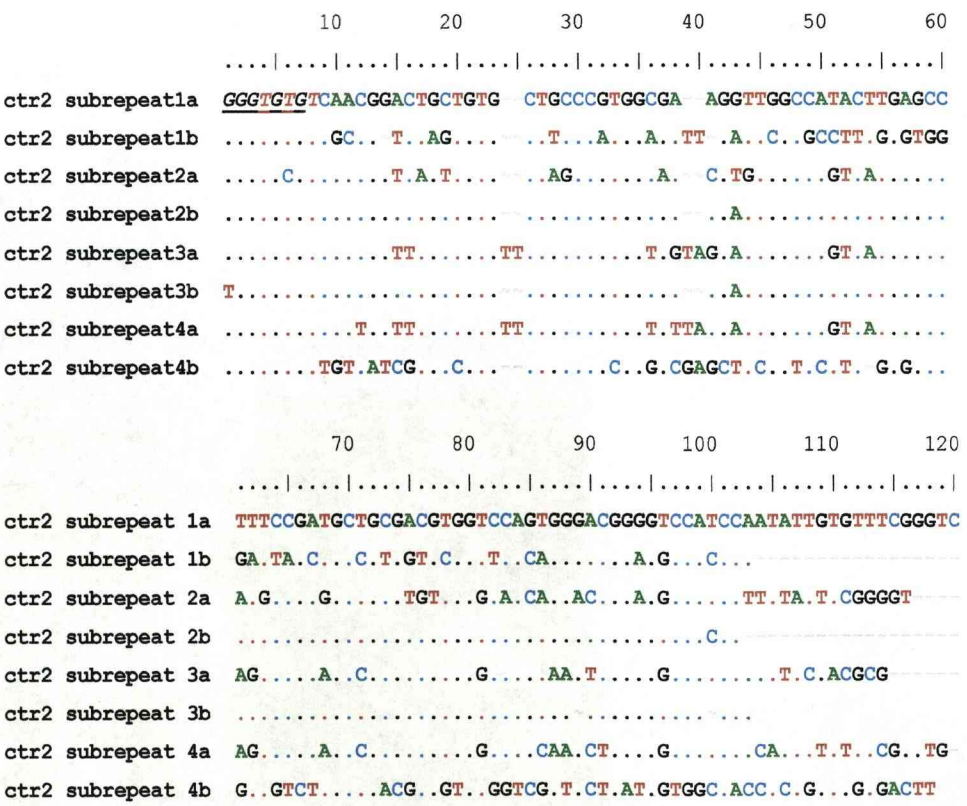
5.3.3.3 Repeats in clade IV

Arrangement of repeats did not follow a clearly organized pattern, as observed in previous clades. To understand their arrangements, insertions were separated where motif *GGGTGTG* (or slight variations) was found; so instead of separating them into 4 DNA repeats, they were split into 8 (Figure 5.12). Motif *GGGTGTG* was also located at the beginning of the 8 repeats, with the exceptions of repeat 3 and repeat 6, in which this motif shows slight variations in the sequence: *GGGTGCG* and *TGGTGTG* respectively 1, 3, 5 and 7.

Repeats 1a, 2b and 3b show high similarity, although they differ in size (116, 98 and 97 bp respectively). Repeats 3a and 4a show high similarity but differ in size.

Figure 5.12 Alignment of the subrepeats found in clade IV (specimen ctr2).

Each of the four repeats has been divided into subrepeats where the motif *GGGTGTG* or variant occurs.



5.3.3.4 Repeats clades II and III

Clades II and III have the largest ITS2 of all the clades (~1730 bp). They share a similar arrangement of repeats. The size of ITS2 was about 200 bp longer in these clades than in clades I, IV and V, due to the insertion of two additional imperfect DNA repeats. In these clades, the repeats show a complicated arrangement. Repeat numbers correspond to their consecutive arrangement in the sequence.

The motif *GGGTGTG* and its variant *GGGTGGG* are common to all repeats. There is high similarity among repeats 1a, 3a, 4b; 1b, 3b, 5a and 2a, 2b and 4a (Figure 5.13 and Figure 5.14).

Figure 5.13 Alignment of the subrepeats found in clade II (specimen l33).

The motif *GGGTGTG* (underlined) or minor variant was used to divide the sequence into 10 subrepeats. Repeat numbers correspond to their consecutive position within the sequence.

```
                          10        20        30        40        50        60
                  ....|....|....|....|....|....|....|....|....|....|....|....|
133 subrepeat 1a  GGGTGTGACTTCTGCCCGTT  GTGCTGCCCGTAGC TACGTTTGGCCAGTATTGGCTAT
133 subrepeat 1b  .......T.AAAG.TTT.C.  ....C.....C... CCG.AA...TGC..CAGT..C.C
133 subrepeat 2a  .......T.AA.G.T.T.C.GG.C.......... ..T.................
133 subrepeat 2b  .......T.AG.G.T.T.C.GG.C.......... ....................
133 subrepeat 3a  .......T.AA..... ..... ....................
133 subrepeat 3b  .....G.T.AAAG.TTT.C.  ....C.....C... CCG.AA...TGC..CAGT..C.C
133 subrepeat 4a  .......T.AG.G.T.T.C.GG.C.......... ....................
133 subrepeat 4b  .......T..A..... ..... ....................
133 subrepeat 5a  .......T.AATG.TTT.C.  ....C.....C... CCG.AA...TGC..CAGT..C.C
133 subrepeat 5b  .......TTG..AT.GT.C.  ..........G.GG..A.CC.C.GTCT. .G....G.

                            70        80        90       100       110       120
                  ....|....|....|....|....|....|....|....|....|....|....|....|
133 subrepeat 1a  TCCTCTGGCTATGACGTGGT  GCAGTCGGGACGGGGTCCATCCAAGAGTGTGTTTTGGG
L33 subrepeat 1b  . TCGGA.T.GC.G.TC..A  C.G..AA.CTG.A..CA......
L33 subrepeat 2a  ..........GC........TCC....A..ACG....C.ATC.A.CC.
L33 subrepeat 2b  ..........C........TCC....A..ACGA..TC.ACC.
L33 subrepeat 3a  ..........C........ C...........TT.G...C...CTT
L33 subrepeat 3b  . TCGGA.T.GC.G.TC..A  C.G..AA.CTG.A..C.....T.
L33 subrepeat 4a  ..........C........TCC....A..ACGA..TC.ACC.GGG
L33 subrepeat 4b  ..........C........ C...........T..G...C...CTTT.CG.
L33 subrepeat 5a  . TCGGA.T.GC.G.CC..A  C....AA.CTG....C .......CTT.TC.GG..
L33 subrepeat 5b  GG ...T...GACG..GT..AGT.G..ACT.GT....A......CTTG...GT
```

Figure 5.14 Alignment of the repeats found in clade III (specimen ta21). Repeat numbers correspond to their consecutive position within the sequence and are arranged in the same way as those in clade II. The motif *GGGTGTG* (underlined) or its variant was used to divide the sequence into 10 subrepeats.

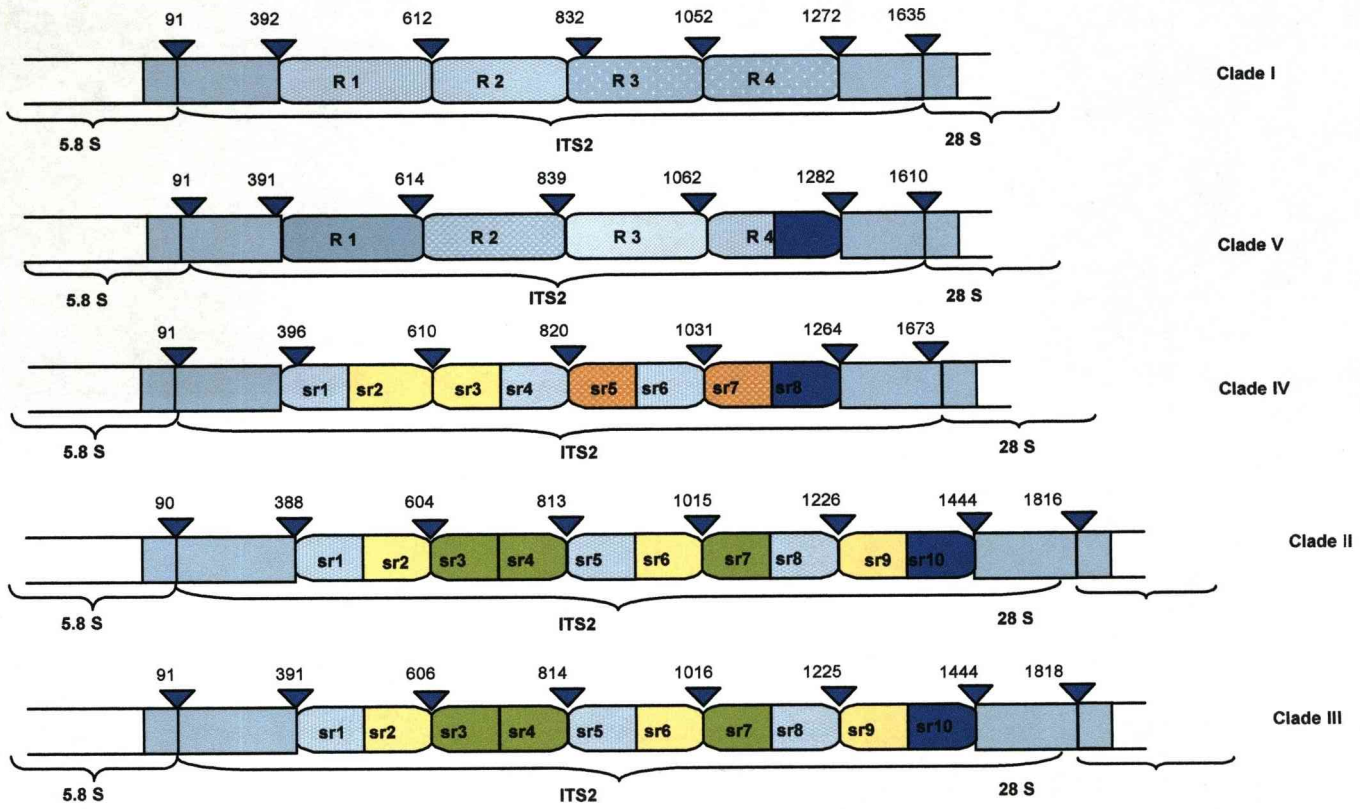
```
                                10        20        30        40        50        60
                      ....|....|....|....|....|....|....|....|....|....|....|....|
ta21 subrepeat 1a     GGGTGTGACTTCTGCCCGTT GTGCTGCCCGTAGC TACGTTTGGCCAGTATTGGCTAT
ta21 subrepeat 1b     .......T.AAAG.TTT.C. ....C.....C... CCG.AA...TGC..CAGT..C.C
ta21 subrepeat 2a     ...... .AA.G.T.T.C.GG.C.............T.....................
ta21 subrepeat 2b     ...... T.AG.G.T.T.C.GG.C.............. ......................
ta21 subrepeat 3a     .......T.AA.......... .C.............. ......................
ta21 subrepeat 3b     .....G.T.AAAG.TTT.C. ....C.....C... CCG.AA...TGC..CAGT..C.C
ta21 subrepeat 4a     .......T.AG.G.T.T.C.GG.C.............. ......................
ta21 subrepeat 4b     .......T..A.......... .C.............. ......................
ta21 subrepeat 5a     .......T.AATG.TTT.C. ....C.....C... CCG.AA...TGC..CAGT..C.C
ta21 subrepeat 5b     .......TTG..AT.GT.C. ..........G.GG..A.CC.C.G.C .TCG....G.

                                70        80        90       100       110       120
                      ....|....|....|....|....|....|....|....|....|....|....|....|
ta21 subrepeat 1a     TCCTCTGGCTATGACGTGGTCCA GTCGGGACGGGGTCCATCCAAGAGTGTGTTTGGG
ta21 subrepeat 1b     TTCGGA.T.GC.G.TC..A..G ..AA.CTG.A..CA......
ta21 subrepeat 2a     ..........GC........T.CCAGTA........G........CCA
ta21 subrepeat 2b     ..........GC........T.CCAGTA.....A......C...T
ta21 subrepeat 3a     ...........C........T.. .........TT.G...C...CT..CA.
ta21 subrepeat 3b     TTCGGA.T.GC.G.TC..A..G ..AA.CTG.A..C.....T..T
ta21 subrepeat 4a     ...........C........T.CCAGTA.....A......C..T
ta21 subrepeat 4b     .........C.C........T.. .........T..G...C...CT.T.CG.AGCAT
ta21 subrepeat 5a     TTCGGA.T.GC.G.TC..A... ..AA.CTA....C ..C....CTT.AC.GG.TCA
ta21 subrepeat 5b     GG..T.GCT.ACGC.GT..A...GGTACT..  T....A......CTTG...GTCA.
```

A diagrammatic representation of the repeats found in the different clades is presented in Figure 5.15.

Figure 5.15 Diagram showing location of repeats in the ITS2 regions for all five clades.

Colours represent similar sequences within a clade. ITS2 in clades II and III is ~1.86 kb, about 200 bp longer than in clades I, IV and V.



5.3.4 Phylogenetic analysis

The Neighbour-joining method was used to infer tree topology with the ITS2 data set, using the Kimura 2-parameter to calculate distance between sequences. A bootstrap analysis was performed to test the reliability of tree topology. Bootstrap values are presented as numbers at internal nodes. The programme used was Mega3, with a bootstrap test of 500 replications.

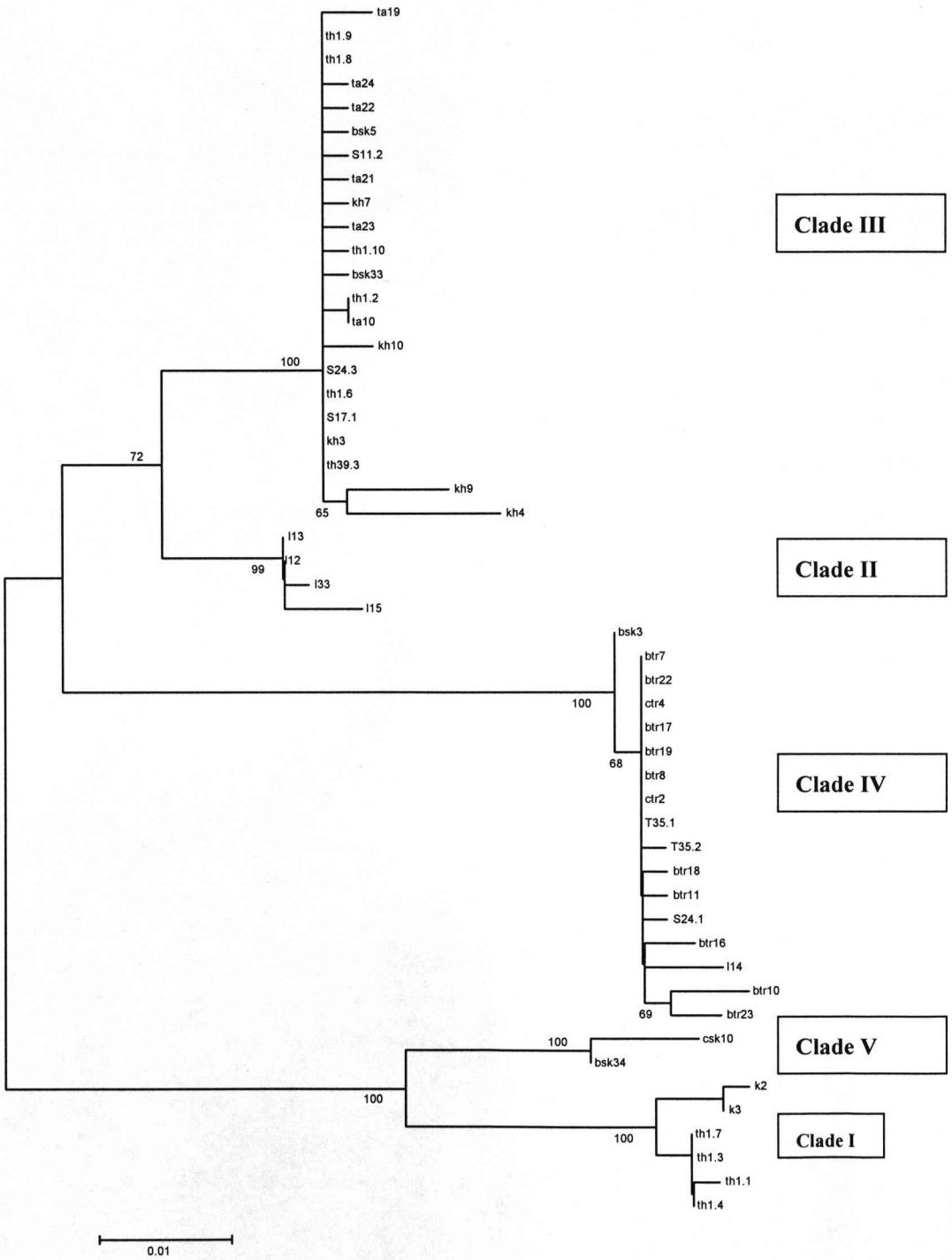
5.3.4.1 Phylogenetic tree obtained using only "alignable regions"

A phylogenetic analysis using the entire data set was extremely problematic due to the difference in size of the sequences and the presence and organization of the repeats. Hence, the internal repeats were omitted and alignment was based on the first 300 bp of the ITS2 at the 5' end previous to repeat 1 and the 250 bp located immediately after the last repeat at the 3' end. These regions were joined and a NJ phylogenetic analysis performed. Although alignment was possible, it demanded significant effort since these regions were of different sizes.

There are five distinct clades in the NJ tree (Figure 5.16). All clades appear to be highly supported. Clades II and III appear to be closely related (72% bootstrap value) and also clades I and V (100%). This reflects the relationships shown in the analysis of the repeats. Clade IV appears as a distinctive clade, separated from the others but closer to clades II and III. Specimens in clade I from Kalimantan and Thailand, appear also as independent clades supported by 100% of the bootstrap analysis. These populations within clade I were well based on the number of fixed substitutions between them.

Phylogenetic analysis was also performed using the entire ITS2 fragments, including repetitive elements and as a result, clades obtained and the relations among them (data not shown) were the same as those obtained in this analysis.

Figure 5.16 Unrooted Neighbour-joining tree (using K 2-parameter). The tree was constructed using only the alignable region of the ITS2: i.e. a ~300 bp fragment preceding the first repeat (5'end) and ~250 bp subsequent to the end of the last repeat (3' end).



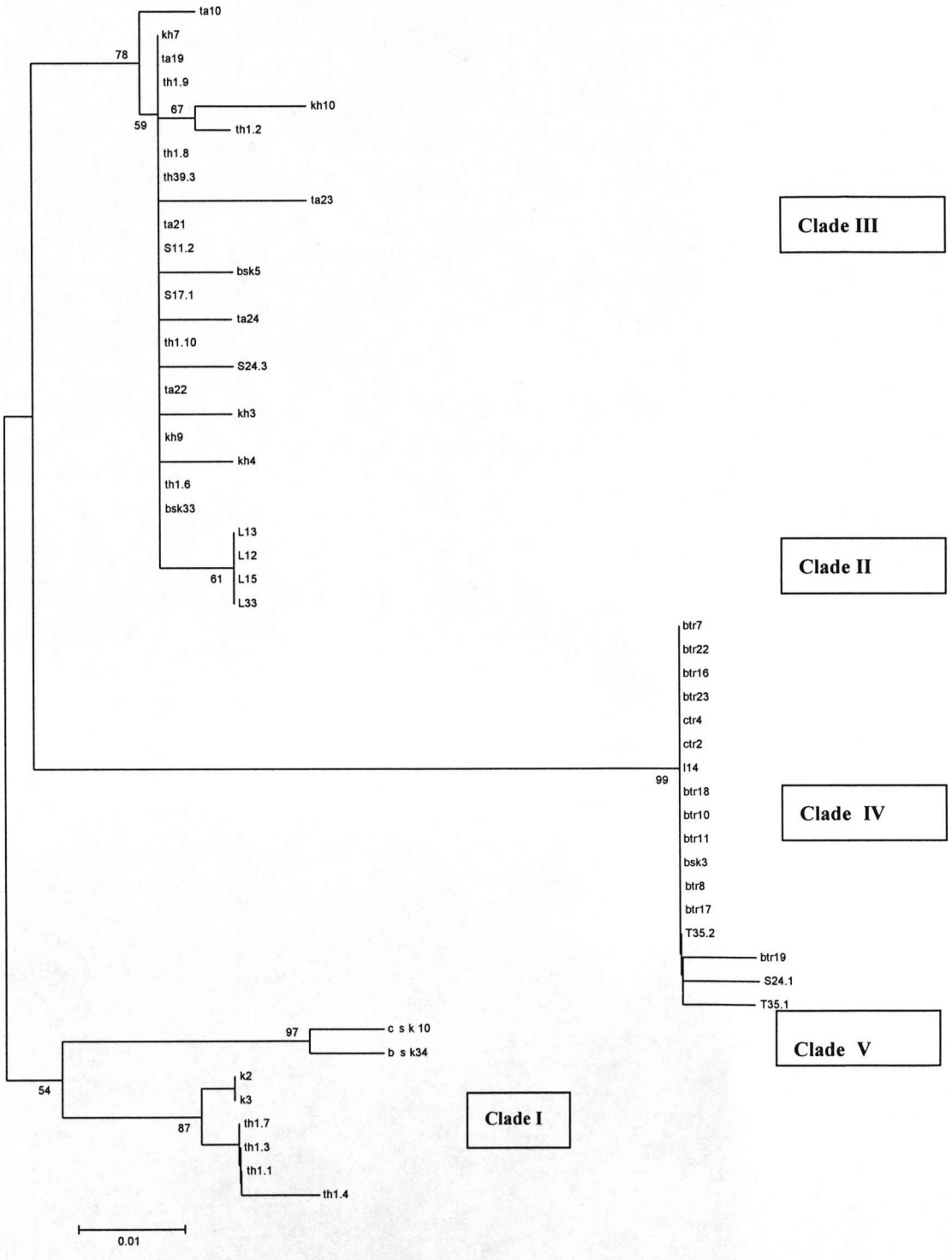
Fifty-one specimens of different localities are included in the analysis. There is no outgroup since no similar sequences were found in GenBank.

5.3.4.2 Tree topology obtained using fragments 5.8S and 28S.

The 5.8S and 28S regions are commonly used to infer higher level phylogenetic relationships, such as between genera or families, due to their lower evolutionary rate in relation to ITS2. However, considering that some substitutions were also observed within these regions, its use for the differentiation of sibling species within the Barbirostris Subgroup was evaluated. The sequence of regions of the 5.8S and 28S were then pasted together, resulting in a ~130 bp fragment used to construct an NJ tree with the K 2-P substitution model. A bootstrap test was performed with 500 replicates. Clades in the Figure 5.17 show lower bootstrap values than those obtained in the ITS2 tree (Figure 5.16), reflecting the small length of sequence used.

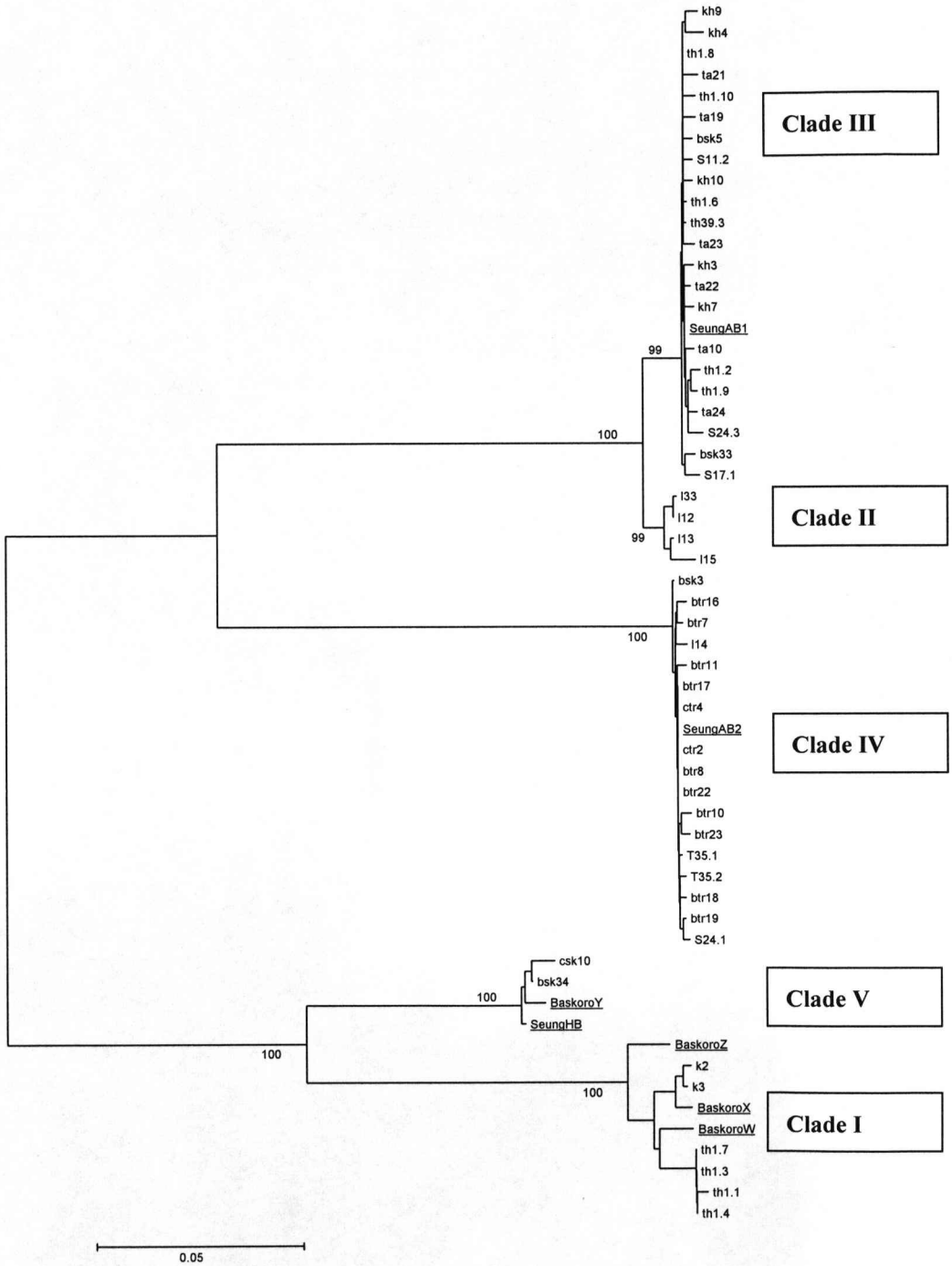
Sequences obtained by Baskoro (2001) of several specimens morphologically identified as *An. barbirostris* were included in the analysis. As can be observed in Figure 5.18, the sequences presented by Baskoro (2001) appear within clade I. These specimens were collected from several islands in Indonesia and identified as adult females in the field, with exception of species Y, which was from Thailand (Baskoro, 2001). Specimens studied by Seung *et al.* (2007) where identified as larvae of *An. campestris* (HB) and *An. barbirostris* (AB1 and AB2). They were all collected in Thailand (see position of these specimens in Figure 5.18)

Figure 5.17 Neighbour-joining (K-2 parameter) phylogenetic tree obtained with the ~ 130 bp fragment based on the 5.8S and 28S regions.



Bootstrap values are positioned next to nodes, these are lower than those obtained with the ITS2 tree.

Figure 5.18 Neighbour-joining tree using 51 sequences of the entire ITS2 fragment in this study, plus those obtained by Baskoro (2001) and Saeung et al. (2007).



Specimens X, W and Z fit well in clade I, whereas Y fits in clade V (Baskoro, 2001). Species identified by Saeung et al. (2007) as *An. barbirostris* AB1 were located in clade III, those identified as *An. barbirostris* AB2 in clade IV, and those identified as *An. campestris* HB, were located in clade V.

5.4 Discussion

5.4.1 Sequence variation

The five clades based on the sequence analysis of the Internal Transcribed Spacer 2 (ITS2) are consistent with results obtained using the Cytochrome Oxidase I (presented in chapter IV). The 51 specimens studied fell into the same clades in both ITS2 + COI analysis. No model of pairwise comparison could be used since sequences were of different sizes; however, it was obvious that the genetic divergence between clades was much higher than within clades. Moreover, the pattern of repeats and subrepeats and the presence of multiple fixed substitutions confirmed the division of the entire dataset into the same clades found in the COI analysis.

The existence of nucleotide substitutions among individuals within clades may indicate that in the Barbirostris Subgroup the process of mutations is faster than homogenization. Assuming that the 5 different clades are distinct species, the process of concerted evolution would not be perfect. Concerted evolution (Dover, 1982) appears to occur in several *Anopheles* species, in which the ITS2 shows little or no variation. For instance, the analysis of the ITS2 in members of the Maculatus Group of *Anopheles* species showed no intraspecific variation within the 5 species studied: *An. maculatus*, *An. dravidicus*, *An. pseudowillmori*, *An. sawadwongporni* and a new unnamed species. This was surprising considering that these specimens were collected from distant localities in Southeast Asia (Walton et al., 2007). In another study, the ITS2 in 5 *Anopheles* species in South America (*An. evansae*, *An. nuneztovari*, *An. rangeli*, *An. strodei* and *An. trinkae*) was studied. Intraspecific sequence variation was only present in two distant populations of *An. rangeli* (Fritz, 1998).

On the other hand, evidence shows that in other species of *Anopheles* there is substantial variation in the ITS2 within species and even within individuals. In the *Anopheles hyrcanus* group, intraspecific variation was reported in *An. junlianensis,*

An. liangshanensis, and *An. pullus* (Ma and Xu, 2005). Beebe et al. (2000) found 13 different ITS2 haplotypes in *Anopheles farauti s.s.* and hypothesised that these intraspecific differences may be related to either the presence of multiple loci for the rDNA gene family, or the slow homogenization of these regions. Moreover, intragenomic variability has been observed in *An. aquasalis,* showing some single nucleotide differences between clones of ribosomal spacers ITS1 and ITS2 obtained from the same individual (Fairley et al., 2005). Intragenomic heterogeneity in the ITS2 of the South American species *An. nuneztovari* has also been reported . Therefore the detection of intraspecific variability of the ITS2, reflected in the high haplotype diversity within clades in specimens studied from the Barbirostris Subgroup, is not surprising.

The G+C content of the ITS2 in the species studied from the Barbirostris Subgroup (52.3-55.9%) is consistent with values obtained for other *Anopheles* species; *An. crucians* A 54%, *An. crucians* B 51%, *An. bradleyi* 54% (Wilkerson et al., 2004), *An. nuneztovari* 55.3- 55.7% (Fritz et al., 1994). Lower G+C content values have been obtained in *An. beklemishevi* 46.6% (Kampen, 2005) and higher values in *An. punctulatus* 61.3-70.9% (Beebe et al., 1999). There is no relation between G+C content and sequence length. In *An. crucians* and *An. beklemishevi*, the ITS2 is larger than in other *Anopheles* mosquitoes, but its G+C content is quite similar to that of other mosquitoes.

5.4.2 Phylogenetic analysis

5.4.2.1 Based on ITS2, fewer internal repeats

The number of fixed substitutions in the ITS2 fragment to differentiate populations of Kalimantan and Mae Hong Son in Thailand was 28 (including insertion/deletion events) and 35 between clades II and III. The number of fixed substitutions in other clades could not be determined since the sequences could not be aligned. Both the COI gene region and the ITS2 analyses agreed on the subdivision of the two populations of clade I, which comprised specimens from Thailand and Kalimantan (Figure 5.16). In Figure 5.18, specimens from Kalimantan appear closely related to

Baskoro's species X, whereas specimens from Mae Hong Son (Thailand) to species W. However, with only 2 specimens from Kalimantan and six from Mae Hong Son, it is difficult to evaluate the species status of these populations.

One of the main disadvantages of the use of the Internal Transcribed Spacer 2 in phylogenetic studies is that it cannot be contrasted with amino acid data, as it is possible to do with mitochondrial markers. On the other hand, it is considered valuable due to its small size in the majority of organisms, which is typically 200-400 bp (Walton et al., 1999b; Young and Coleman, 2004). The ITS2 has proved valuable for identification of members of species complexes, with the exception of the *Anopheles gambiae* complex, where ITS2 is not informative (Scott et al., 1993). Sibling species *An. freeborni* and *An. hermsi* were differentiated by 11 substitutions in ~460 bp of the fragment 5.8S-ITS2-28S, the majority of substitutions having been seen in the ITS2. whereas *An. hermsi* and *An. oxidentalis* differed at 28 sites in a fragment of similar size (Porter and Collins, 1991). On the other hand, in sibling species such as *An. dirus* and *An. scanloni* in Thailand, one fixed substitution is used to differentiate these species (O'Loughlin et al., 2007). If the results obtained in this chapter are judged in comparison to data from other *Anopheles* species, populations of Thailand and Kalimantan in clade I, clade II and III, IV and V would be different species. This hypothesis will be discussed in detail in chapter 8.

There are some discrepancies between the COI and the ITS2 phylogenetic analyses, with respect to the internal branching. The close relation of clades II and III observed in the ITS2 analysis (Figures 5.16-5.18) was not observed in the phylogenetic analysis using the COI (Figure 4.2). In the COI tree they appear as distinctive and non-related clusters, whereas in the ITS2 they appear closely related, as they show the same size of ITS2 and the same pattern of repeats arrangement. Conversely, clades IV and V appear closely related in the COI analysis but as distinctive clades in the ITS2 analysis.

5.4.2.2 Variation in the partial fragments 5.8S and 28S

In contrast to the internal transcribed spacer 2, which is more informative when examining different species within Diptera (Young and Coleman, 2004), the 5.8S and 28S regions have provided valuable information for distinguishing genera and families (Miller et al., 1997). For instance the 5.8S has been used to investigate the origin of the family Culicidae (Miller et al., 1997). In this study, small fragments of the 5.8S (91bp) and 28S (5.8S) were joined together and its value to infer tree topologies was evaluated.

The NJ phylogenetic tree constructed with fragments 5.8S and 28S shows lower resolution in the bootstrap analysis than the ITS2 tree. Surprisingly, these conserved regions contained enough phylogenetic signal to distinguish the same 5 clades, which is consistent with the ITS2 tree and with results obtained from the COI analysis. Nevertheless, clade II appears much closer to clade III in the 5.8S+28S tree than in the ITS2 tree (see Figures. 5.16 and 5.17). This is related to the fact that fragments 5.8S and 28S evolve more slowly than the ITS2 (Hillis and Dixon, 1991) and can be very similar even in distantly related organisms (Nei and Rooney, 2005).

5.4.3 DNA repetitive elements

The remarkable size of the ITS2 in members of the Barbirostris Subgroup is a result of the presence of repeat sequences that differ in the number of repeats and the extent of their homology. In some clades, the repeats share a high homology, whereas in others, considerable sequence divergence has occurred.

Long ITS2 regions have been found in other organisms, although it does not appear to be a frequent event in nature. The longest ITS2 so far in any species has been reported in the phytopathogen mildew *Plasmopara halstedii*, which has a 2212 bp ITS2. This exceptional size was attributed to the presence of repetitive elements (krThines et al., 2005). *Bremia lactucae*, another phytopathogen fungus, has a 2086 bp ITS2, the length again being due to repetitive elements in the sequence (Choi et al., 2007). In arthropods, long ITS2 regions (>1000 bp) have been found in

rhipicephaline ticks (Barker, 1998; Murrell et al., 2001), probably resulting from the presence DNA repeats.

In *Anopheles* species, repetitive DNA regions of considerable size (>100 bp) have been documented in three species. The first report was for *Anopheles crucians* species B (Wilkerson et al., 2004), yet another species with a long ITS2 (1012bp). Its length was the result of a complex array of repeats in different combinations. Another *Anopheles* species with a long ITS2 is *An. beklemishevi*, a member of the Maculipennis Group which has an ITS2 of 638 bp containing 2 repeats of about 140 bp of similar though not identical sequence (Kampen, 2005). Finally, *An. fluminensis* from Bolivia, has been reported to have three repeats of 125 bp each within an ITS2 of 596 bp (Brelsfoard et al., 2006).

The above three *Anopheles* species and *An. barbirostris*, which share the feature of a relatively long ITS2, are all in the subgenus *Anopheles*, but they are not closely related to each other (Harbach, 2004). They also occur in different geographic regions: *An. fluminensis* in Bolivia, *An. beklemishevi* in northern Eurasia and *An. crucians* B in North America, and *An. barbirostris* in Southeast Asia. Hence, the presence of a large ITS2 region cannot be explained as the consequence of a common evolutionary origin, as postulated by Beebe et al. (1999) to explain the long ITS2 in the *An.dirus* and *An. punctulatus* complexes. Furthermore there are species close to each of the above species with a long ITS2 that have a much shorter ITS2 sequence, for example *An. barbumbrosus* which is in the Barbirostris Group (Baskoro, 2001). An explanation for the longer ITS2 sequences must be sought elsewhere.

In clade I, the 4 repeats in ITS2 are almost perfect (96.8-99.5% similar) and have the same size (220 bp), whereas the repeats found in clades II, III and IV show considerable variability in size and sequence similarity. Imperfect repeats appear to be common in organisms with a large ITS2, but their similarity also varies depending on the species. For instance the 179-194 bp DNA repeats in *B. lactucae*, have a low homology (48-69%), whereas the repeats in *P. halstedii*, are c. 90% similar (Thines, *et al.*, 2005; Choi *et al.* 2007). Repeats in the ITS2 of rhipicephaline

ticks are 104 -109 bp long and imperfect (Murrell et al., 2001); these were explained to be the result of duplication events.

Differences in the similarity levels of repeats may be related to the way in which ITS2 is evolving. There are two possible explanations for the differences found in the sequence and size among repeats in the different clades of the Barbirostris Subgroup. 1) they could be the result of imperfect duplication events of the original fragment or 2) they could be the result of a perfect duplication, which generates fragments of the same size, and these later undergo mutational changes which would alter their size, arrangements and similarity. The latter was the explanation given for the repeats found in *An. beklemishevi* (Kampen, 2005) and it may be a contributing factor in the repeats of the species of the Barbirostris Subgroup. However, an explanation for the generation of the repeats may be found by looking at what is known about the nature of repeats in the rDNA of other organisms.

Variations in the size of spacer regions have been documented in other regions of the ribosomal DNA, including the Intergenic spacer (IGS). Variation in length of this region was seen in parthenogenetic *Daphnia pulex* subjected to divergent selection in the laboratory (Gorokhova et al., 2002). Clonal lines with elevated juvenile growth rates showed an increased preponderance of long IGS variants. The repeats were found to possess a promoter region that was thought to augment the rate of rDNA transcription. The authors concluded that longer IGS regions provide these organisms with the plasticity required to adapt to different environments. There is no evidence that similar promoters occur in the ITS2 region of the mosquitoes.

5.4.3.1 Repetitive regions in the ITS2 of the Babirostris Subgroup and a possible role for transposable elements

Clades do not share the same repeats, but there are some structural features common to the repeats in all clades that suggest they may have a common origin. Thus all show the motif GGGTGTG or similar variants (Figures 5.5-5.8) at the start of the repetitive region. However I was unable to find similar motifs in neither other

studies of insertions within ITS2 (Gorokhova et al., 2002; Hlinka et al., 2002; Kampen, 2005; krThines et al., 2005; Murrell et al., 2001), nor through a BLAST search, using *Anopheles* and *Drosophila* datasets.

In *Drosophila melanogaster,* mobile elements are thought to occupy 10-20% of the entire genome (Engels, 1992). They have been reported in many insect species, including *Anopheles* species. These elements have been classified as 1) DNA transposons, which mobilize from one region to another by a direct cut and paste mechanism, 2) LTR (Long terminal repeats) retrotransposons and 3) Non-LTR (non-long terminal repeats) retrotransposons. LTR and non-LTR retrotransposons are first transcribed into RNA, the reverse transcribed and reintegrated into the genome. Some mobile elements such as Alu (in humans) and L1 lead to deletion and duplication events between the repeats through DNA mispairing and unequal crossing over (Kazazian, 2004).

One feature of particular interest was an AT-rich motif just downstream (3') of the last repetitive element in the ITS2. The motif TATTTATTTATTTTT(T)ATTTATTTAA was found to be highly conserved in sequences of all clades. A BLAST search for this A-T rich motif showed a 96% similarity with a region 26 bp *P*-element like transposon (see below) found in PEST clone AgaP4ref (Quesneville et al., 2006)

Example of Blast result:

```
A-T rich motif              1    TATTTATTTATTT-TTTATTTATTTA 25
                                 ||||||||||||| ||||||||||||
p-like element AgaP4ref  5226    TATTTATTTATTTATTTATTTATTTA 5251
```

This region was also identical to a 16 bp region found in a transposase gene of a *hAT* Herves transposable genetic element in *An. gambiae* (Subramanian et al., 2007). Herves elements are considered important for their potential use in genetically modifying mosquitoes (O'Brochta et al., 2006). It seems likely that the AT-rich region is a relict of TGE, that was active before the divergence of the Barbirostris

Subgroup. This genetic element may have been responsible for the duplication leading to the repeats found in the ITS2 of species of the Barbirostris Subgroup.

CHAPTER 6

6 Morphological characterization of adult female specimens of the Barbirostris Subgroup from Sa Kaeo and Trat (Thailand).

6.1 Introduction

In this chapter, the morphological examination of specimens from the Barbirostris Subgroup and Hyrcanus Group is described. Specimens were identified as *An. barbirostris* and *An. campestris* in the field. However these identifications could not be related to the 5 molecular clades described in chapters 4 and 5. The morphological examination is re-examined in this chapter and results are compared with molecular data. The ultimate objective is to be able to relate species or other taxa based on morphological characters with the entities recognizable from DNA sequence data. Before describing the results of the morphological examination, the literature dealing with the systematics of the Barbirostris Group and related groups is reviewed.

Van der Wulp described *An. barbirostris* in 1884 based on a female collected from Mount Ardjoeno in Eastern Java (Van der Wulp, 1884). Subsequently this species was reported as a non-vector species in several countries in Southeast Asia, until 1935, when *An. barbirostris* was found to be transmitting malaria in Malaysia (Hodgkin and Johnston, 1935). This lead Reid (1942) to link these differences found in vector status with morphological differences in the vector form reported in Malaysia and a non-vector form found in Java (Venhuis, 1939). Reid described these varieties as "dark-winged" and "light-winged *An. barbirostris*, the former being more anthropophilic than the latter (Reid, 1942).

Following several years of research, Reid (1962) described five species of almost identical morphology within the Barbirostris Group as members of the Barbirostris Subgroup. Among them, the previously described light-winged *An. barbirostris* was then linked to the original *An. barbirostris* described by Van der Wulp in 1884 and the dark-winged anthropophilic form, received the name *An. campestris*. In addition,

other forest species were also described: *An. donaldi*, *An. hodgkini*, *An. pollicaris* and *An. franciscoi* (the last species only in the Philippines). Reid's work has been used as a reference in several other keys of *Anopheles* in Southeast Asia (Reid, 1968; Harrison and Scanlon, 1975; Rattanarithikul et al., 2005). Below, the diagnostic characters of *An. barbirostris* (Van der Wulp, 1884) and *An. Campestris* (Reid, 1962) are described.

Reid (1968) summarized the most important diagnostic characters to distinguish *An.* barbirostris from *An.campestris* in the following way:

> "White ventral scales on the abdomen numerous, not confined to the median tufts on each sternite but scattered between these tufts and a row on the lateral sternal margins ; wing darker, usually more than half the scales dark between the basal dark mark on 5 and the apical dark mark on 5.2 ..... *campestris*
>
> White ventral scales fewer, confined, at least in specimens from the west side of Malaya, to the median tufts and a few on the lateral sternal margins, these last usually difficult to see in dried specimens ; wing paler, usually less than half the scales dark between the basal dark mark on 5 and the apical dark mark on 5.2 .....*barbirostris"*

The differentiation of these species is crucial to establish effective malaria and filariasis control programmes. However, since the studies of Reid, several problems have been reported in the distinction of these species. For instance, the wing pattern used to distinguish between *An. barbirostris* and *An. campestris* was considered "not valid" in Thailand by Harrison and Scanlon (1975) and does not appear in more recent illustrated key for *Anopheles* mosquitoes (Rattanarithikul et al., 2005). As a consequence, the distribution of the Barbirostris Subgroup species appear uncertain in several countries in Southeast Asia.

Lien et al. (1968) carried out a brief survey of the mosquitoes of southern Sulawesi, Indonesia. They found that larval and pupa stages fall within the range of *An. barbirostris* as defined by Reid (1968); however the adults showed some characteristics in common with *An. campestris* and *An. barbirostris*:

> Adult female (4 females examined): Wing darker, more than l/2 the scales dark between the

> basal dark mark on 5 and the apical dark mark on 5.2 as in *An. campestris*; however, white ventral scales few in number, confined to the median tufts, and a few on the lateral sternal margins as in *An. barbirostris"* (Lien et al., 1977).

From the characteristics found in the larval and pupal stages, Lien *et al.* (1977) concluded that the specimens were *An. barbirostris*, not *An. campestris* or *An. donaldi.*

An. barbirostris may sometimes resemble *An. donaldi*. Specimens collected from a shaded breeding place at Kuala Krai, Ulu Kelantan (Thailand), where *An. donaldi* was expected, proved to be *An. barbirostris* on examination of the pupal and larval exuviae, though one or two of the adult females had a pale fring spot at 2.1 as in *An. donaldi*; a similar series has been seen from Trat Province, south-west Thailand. (Reid, 1968).

In 2004, an outbreak of malaria in the province of Sa Kaeo in Thailand was attributed to members of the Barbirostris Subgroup (Limrat et al., 2001). However, the authors found it impossible to distinguish between *An. barbirostris* and *An.campestris.* Although the vector was subsequently identified as *An. campestris* (Apiwathnasorn et al., 2002), the authors emphasised the importance of a better understanding of the vector species in the area.

These species can be also confused not only with members within the group but also with species from related groups. Reid (1962) referred to them as very similar in general appearance to those from the *An. bancroftii* group. However the two groups have different distributions, the Bancroftii Group being more oriental and Barbirostris Group more occidental, coinciding only in the Philippines and in Sulawesi (Reid, 1968). Confusion with species from the Umbrosus Group has also been reported, but only male specimens (Harrison and Scanlon, 1975). In this present study, mistakes in differentiating between the Barbirostris and Hyrcanus Groups are reported (see Chapter 4). In theory, specimens from the Hyrcanus Group are simply identified by observing a patch of dark scales on each side of the clypeus, which are absent in Barbirostris Group specimens. The characters used to differentiate the Barbirostris and Hyrcanus Groups are also evaluated.

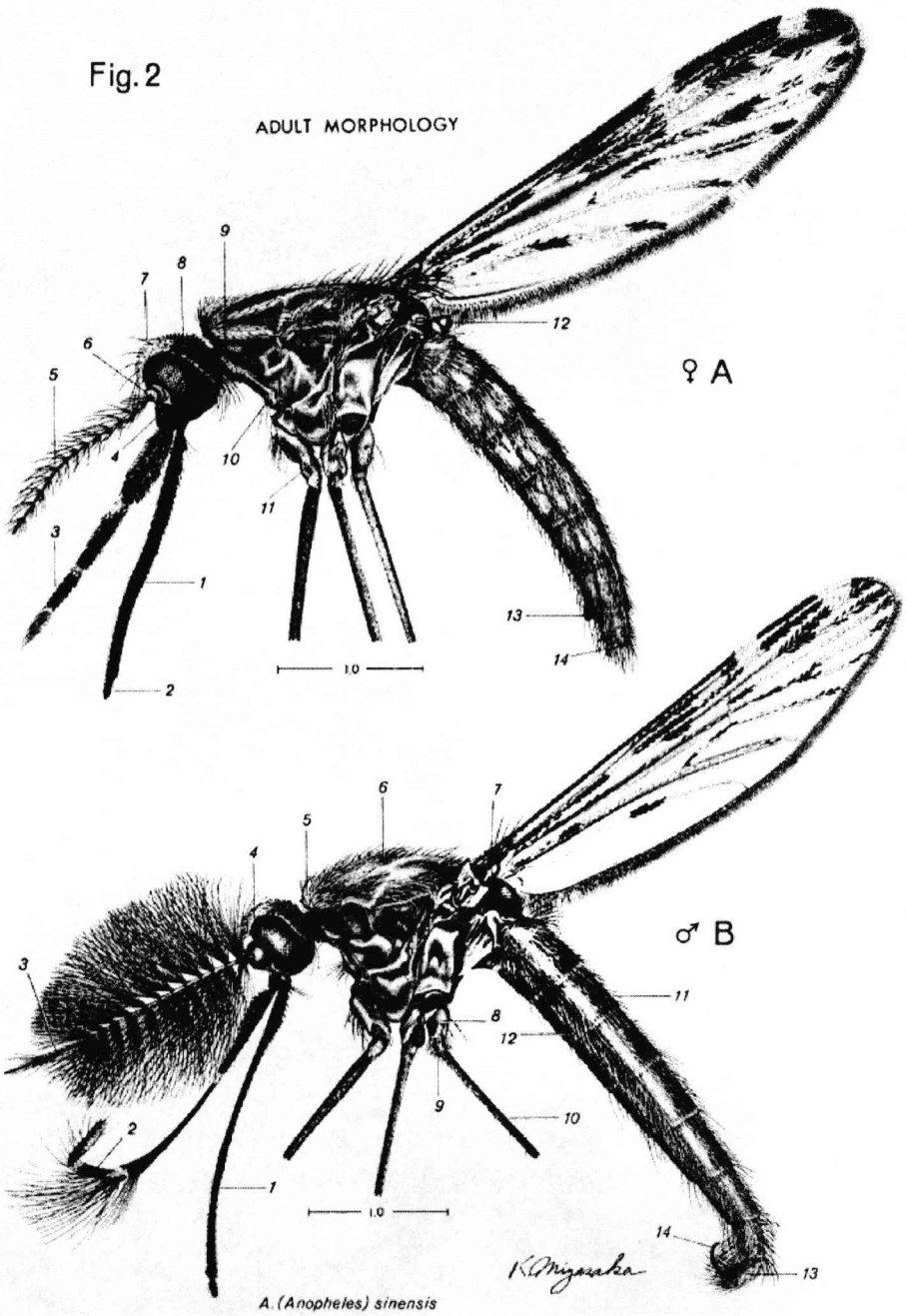
6.2 Material and methods

The morphological analysis was based on adult female specimens collected from the provinces of Sa Kaeo and Trat in Thailand. Specimens were examined under a MEIJI techno binocular stereo microscope. Individuals in bad conditions (with visible fungus or broken legs, abdomens or wings) were separated from the analysis. Specimens were not mounted since they were dry and fragile and furthermore, were required for subsequent PCR. Specimens were observed in Petri dishes, using thin forceps. Photographs of the specimens were taken to keep a record of the most important details.

The morphological analysis was performed as follow: 1) specimens were differentiated as members from the Hyrcanus or Barbirostris Groups, using the morphological key by Harrison and Scanlon summarized in Figure 6.3. Specimens identified as members of the Barbirostris Group were separated and species identified using the keys of Reid (1962) and summarised in Figure 6.5, Harrison & Scanlon (1975) and the illustrated key of *Anopheles* mosquitoes from Thailand (Rattanarithikul et al., 2005). The general morphological features of *Anopheles* species are shown in Figures 6.1 and 6.2.

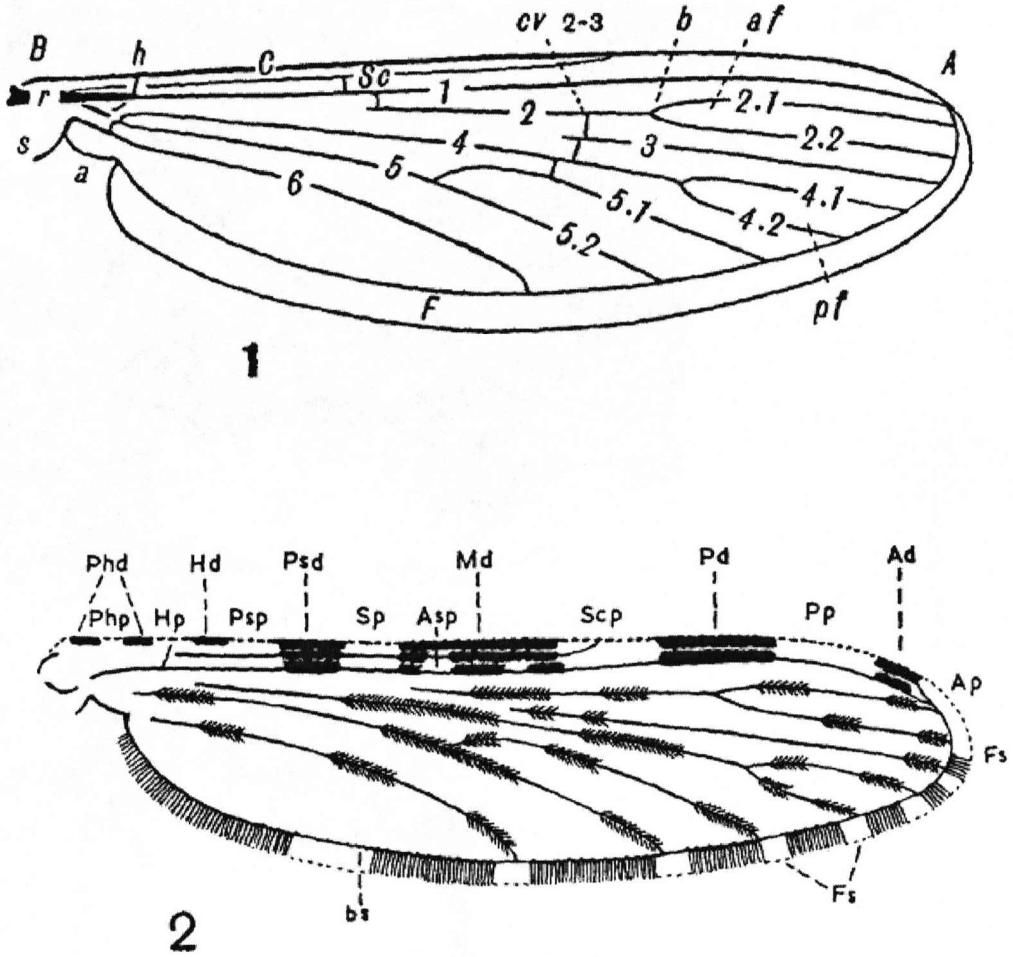
Following their identification, specimens were collected in sterile Eppendorf tubes. Subsequently, each specimen was subjected to DNA extraction, PCR amplification and sequencing of the COI and ITS2 regions to relate the morphological characters with molecular data. Cross contamination was avoided by the surface decontaminant DNA *AWAY* ^{TM} (molecular BioProducts) to remove DNA from forceps and surfaces.

Figure 6.1. Lateral view of *An. sinensis* (Hyrcanus Group) showing morphological characters, taken from Harrison and Scanlon (1975).



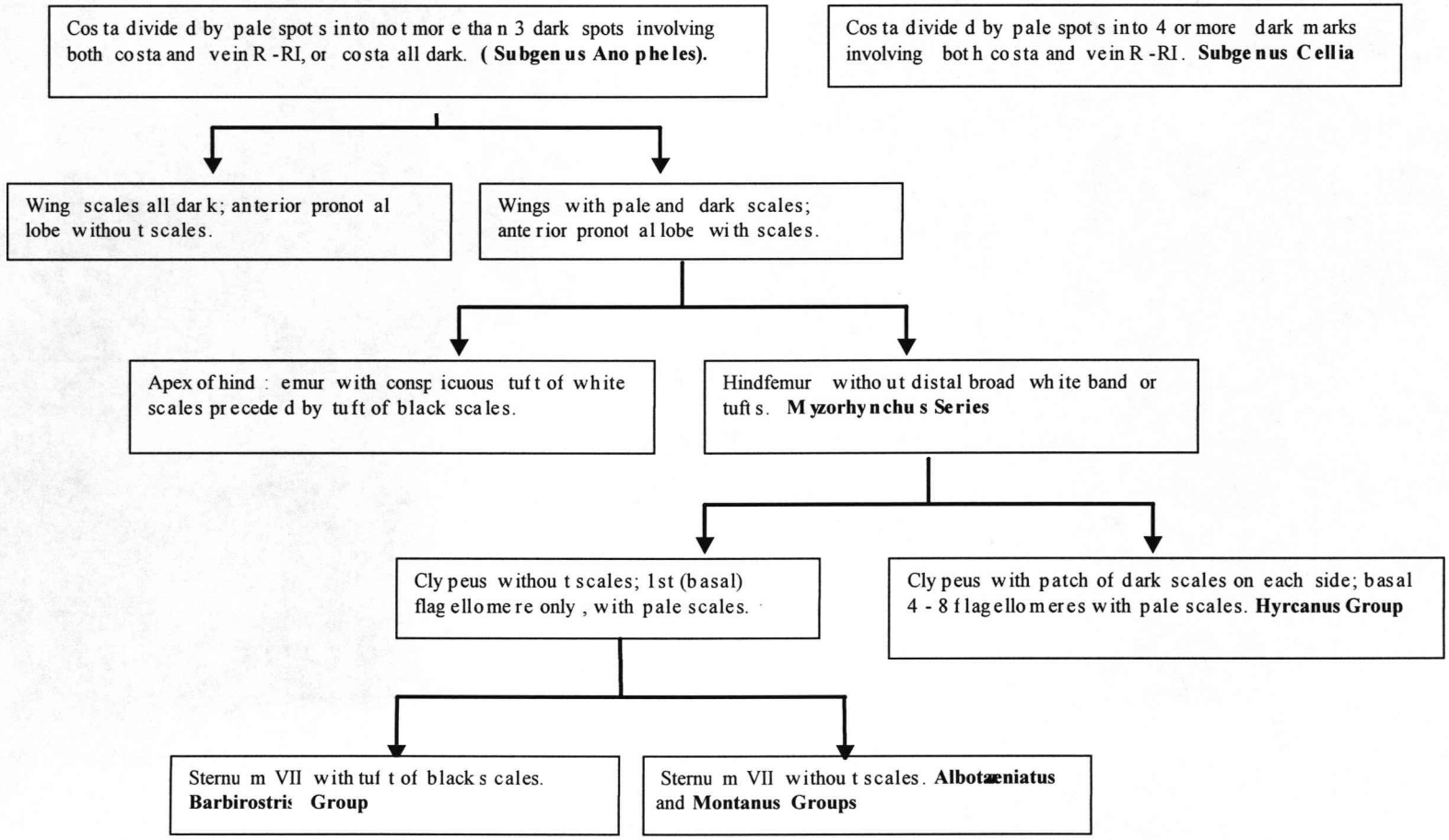
A. Female: 1. labium (proboscis) 2. labellum 3. maxillary palpus 4. clypeus scales 5. flagellomere 6. pedicel7. vertex 8. occiput 9. anterior pronotum 10. propleural setae 11. upper midcoxal setae 12. halter 13. sternum VII scale tuft 14. cercus B. Male: 1. labium (proboscis) 2. maxillary palpus, segment 4 3. flagellomere 12 4. pedicel 5. anterior promontory of scutum 6. scutum 7. scutellum 8. hindcoxa 9. hindtrochanter 10. hindfemur 11. tergum III 12. sternum III 13. basimere 14. distimere

Figure 6.2. Veins and markings of the anopheline wing, taken from Reid (1968).



1, venation ; 2, usual position of light and dark areas in subgenus C&a. 1. A, apex of wing ; a, alula ; af, anterior forked cell ; B, base of wing ; b, bifurcation ; C, costa ; cu 2-3, cross vein betweenveins z and 3 (other cross veins similarly except humeral) ; F, fringe ; lz, humeral cross vein ; $f, posterior forked cell ; Y, remigium (or stem vein) ; SC, subcosta ; s, squama; I, first longitudinal vein (Radius, Rr) ; 2, second longitudinal (Radial sector, Rs) ;2.1, upper or anterior branch of 2nd longitudinal (Rs) ; 2 -2, lower or posterior branchof 2nd longitudinal (Rs) ; 3, third longitudinal (R4+5) ; 4, fourth longitudinal (Media,M) ; 4.1, upper branch of 4th longitudinal (Ml) ; 4.2, lower branch of 4th longitudinal(1212) ; 5, fifth longitudinal (Cubitus, Cu) ; 5.1, upper branch of 5th longitudinal (Cur) ;5.2, lower branch of 5th longitudinal (Cuz) ; 6, sixth longitudinal (Anal, An). 2. Ad, apical dark mark ; Ap, apical pale spot (often merging with apical fringe spot) ; Asp, accessory sector pale spot ; bs, border scales ; Fs, pale fringe spot ; Hd, humeral darkmark ; HP, humeral pale spot ; lbfd, middle dark mark ; Pd, preapical dark mark ; Phd, prehumeral dark mark or marks ; Php, prehumeral pale spot ; Pp, preapical pale spot ; Psd, presector dark mark ; Psp, presector pale spot ; SF, sector pale spot ; Scp, subcostal pale spot.

Figure 6.3 Key to differentiate Hyrcanus and Barbirostris Groups (Harrison and Scanlon, 1975).



6.3 Results

6.4 Morphological differentiation of adult females from the Hyrcanus and Barbirostris Groups

The presence of a tuft of scales on the clypeus (Figure 6.4) is the main character to differentiate Hyrcanus Group from the Barbirostris Group, which lacks this tuft. A second diagnostic feature is the distribution of white scales on the flagellomeres of the antennae (Harrison and Scanlon, 1975). However, as the antennae were lost in many of the specimens in my study, this character could not be observed. Mistakes in the identification carried out by field workers were shown by subsequent DNA analysis.

From the 58 specimens examined in total in this chapter, field workers identified 49 specimens as Barbirostris and 11 as members of the Hyrcanus Group. I re-examined the morphology of these specimens using the morphological key by Harrison and Scanlon (1975) (Figure 6.5) and found 4 mistakes in the identification of the groups (Table 6.1). Specimens identified by field staff as *An. campestris,* (2) *An. barbirostris* (1) and *An. barbumbrosus* (1) showed scales on the clypeus,

Table 6.1. Specimens from Hyrcanus Group morphologically identified.

Specimen	Locality	Field worker identification	Identification in this study
HT8.1	Trat	Hyrcanus Group	Hyrcanus Group
HT8.2	Trat	Hyrcanus Group	Hyrcanus Group
HT8.3	Trat	Hyrcanus Group	Hyrcanus Group
HT8.4	Trat	Hyrcanus Group	Hyrcanus Group
HT8.5	Trat	Hyrcanus Group	Hyrcanus Group
HS43.1	Sakaeo	Hyrcanus Group	Hyrcanus Group
HS36.1	Sakaeo	Barbumbrosus G.	Hyrcanus Group
HS36.2	Sakaeo	Hyrcanus Group	Hyrcanus Group
HS40.2	Sakaeo	*An. campestris*	Hyrcanus Group
HS40.3	Sakaeo	*An. campestris*	Hyrcanus Group
HT25.3	Trat	*An. barbirostris*	Hyrcanus Group

Figure 6.4. Photos showing the absence or presence of clypeus scales in the Barbirostris (A) and Hyrcanus (B) Groups respectively.

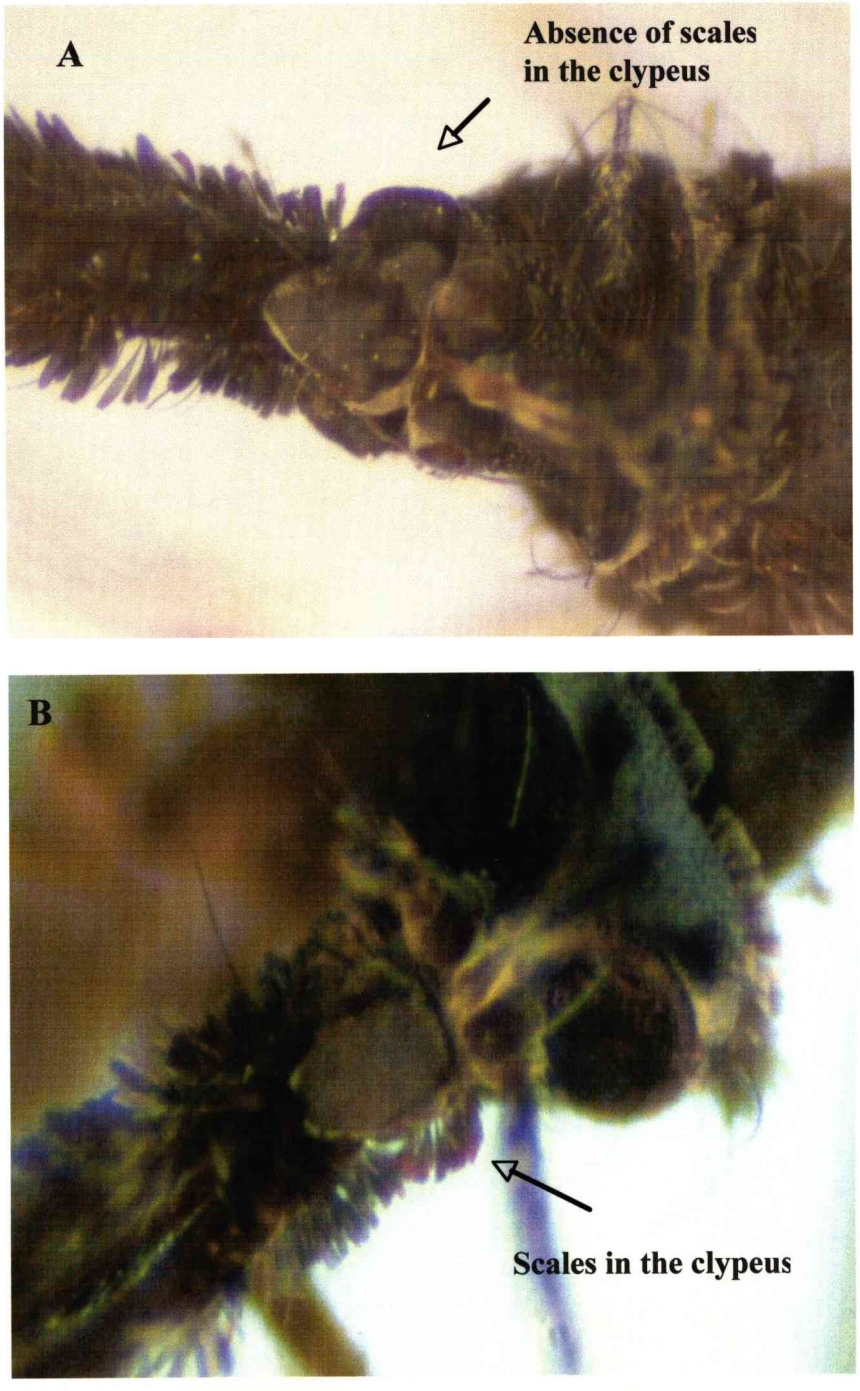
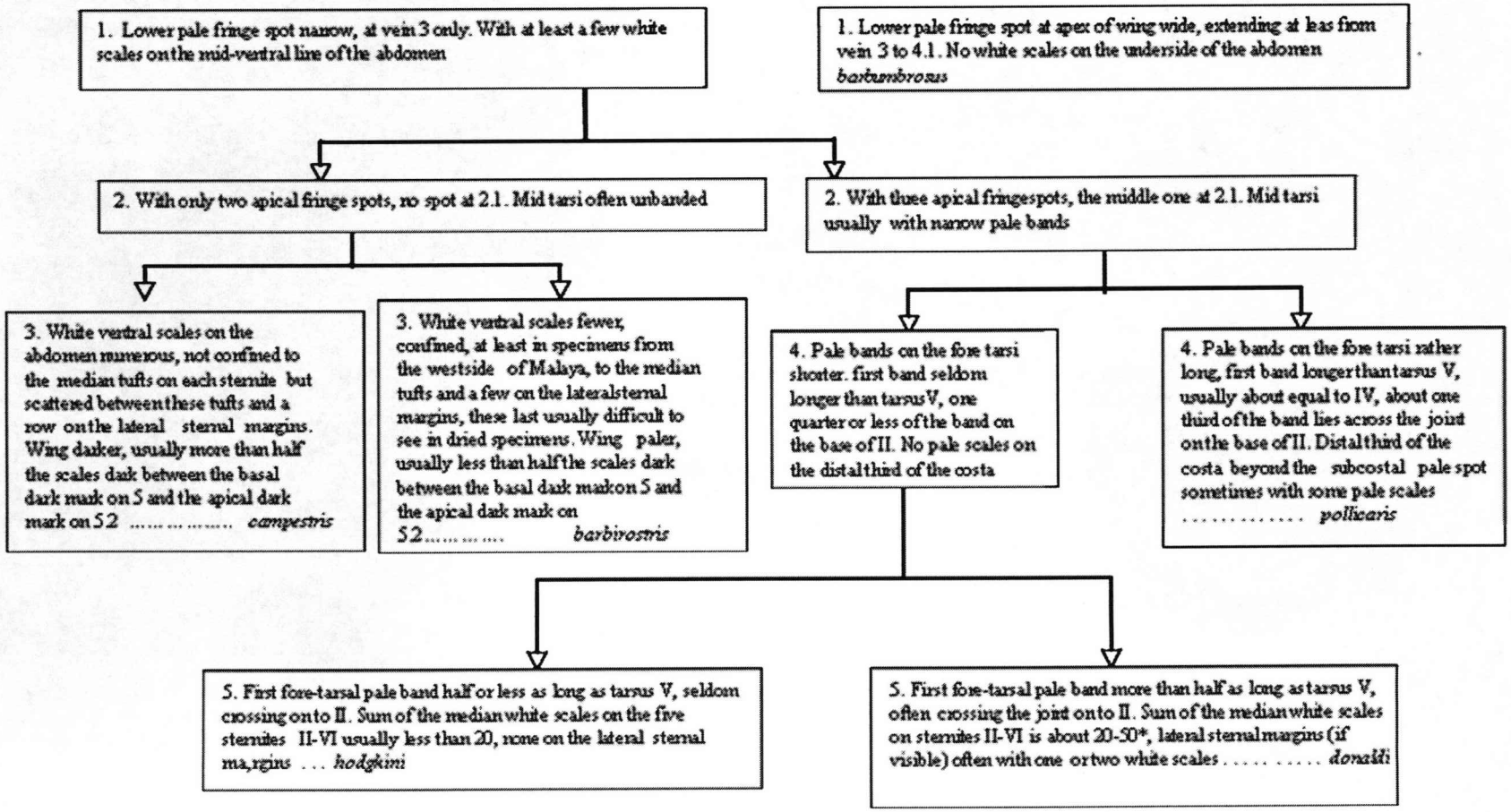


Figure 6.5. Morphological key to differentiate species of the Barbirostris Group (Reid, 1962).



typical of Hyrcanus Group species. The molecular analysis corroborated that field workers made mistakes in the identification of specimens mentioned and that these were in fact from the Hyrcanus Group.

6.4.1 Morphological differentiation of specimens within the Barbirostris Subgroup

6.4.1.1 Subgroup identification (Figure 6.5)

The first step was to distinguish between the two subgroups comprising the Barbirostris Group: Barbirostris and Vanus Subgroups. The fringe spots on the wing serve to differentiate these subgroups (see Figure 6.5); However scales on the apical part of the wing were easily lost in dry specimens and this character could not be seen in any specimen. The presence of scales on the abdomen was used instead. Barbirostris Subgroup specimens show scales on the abdominal sterna, whereas *An. barbumbrosus* (Vanus Subgroup) does not. As a result, 47 specimens were identified as Barbirostris Subgroup and no specimens from the Vanus Subgroup were seen.

6.4.1.2 Identification of species within the Barbirostris Subgroup based on morphology

Forty-seven specimens of the Barbirostris Subgroup from Thailand were identified to species level (Table 6.2). There were nineteen differences between the identifications by field workers and those of the author. Seventeen specimens identified as *An. barbirostris* by field workers were based on the diagnostic morphological characters identified as *An. campestris.* Two specimens identified as *An. pollicaris /donaldi /hodgkini* by the author had been identified by field workers as *An. campestris*.

To differentiate *An. campestris* and *An. barbirostris* from *An. donaldi, An. hodgkini* and *An. pollicaris,* the character: presence/absence of pale bands in midtarsomeres (tarsomeres of midlegs) was used. This character is not the principal one for the differentiation of these species (Figure 6.5), but it had to be used because the apex of

the wings had been lost. Of the 47 specimens examined, only 2 showed pale bands on the midtarsomeres. Although bands were evident under the microscope, they appear faint to the naked eye and may explain why field staff considered these to be absent. Specimens with these bands were identified as possible *An. pollicaris /donaldi /hodgkini* by the author. The identification of the remaining 45 specimens was carried out following step 3 (Figure 6.5).

As described in the morphological key (Figure 6.5), *An. barbirostris* is identified by white ventral scales confined to the median tufts and few on the lateral sternal margins (Figure 6.6). *An. campestris* has numerous white scales on the abdomen, not confined to median tufts but scattered between these tufts and a row on the lateral external margins (Harrison and Scanlon, 1975; Reid, 1962). The main disadvantage with this character is that some *An. campestris* specimens can lose their abdominal scales and then resemble *An. barbirostris*.

The number of scales on the abdominal sternites is considered by Harrison and Scanlon (1975) and by Rattanarithikul et al. (2005) as the only character to distinguish adults of *An. campestris* and *An. barbirostris*. The authors identification of these species was based on this character. As a result 18 specimens were identified as *An. barbirostris* and 27 were identified as *An. campestris*.

Table 6.2 List of specimens of the Barbirostris Subgroup morphologically identified in this study. Squares indicate differences in identification.

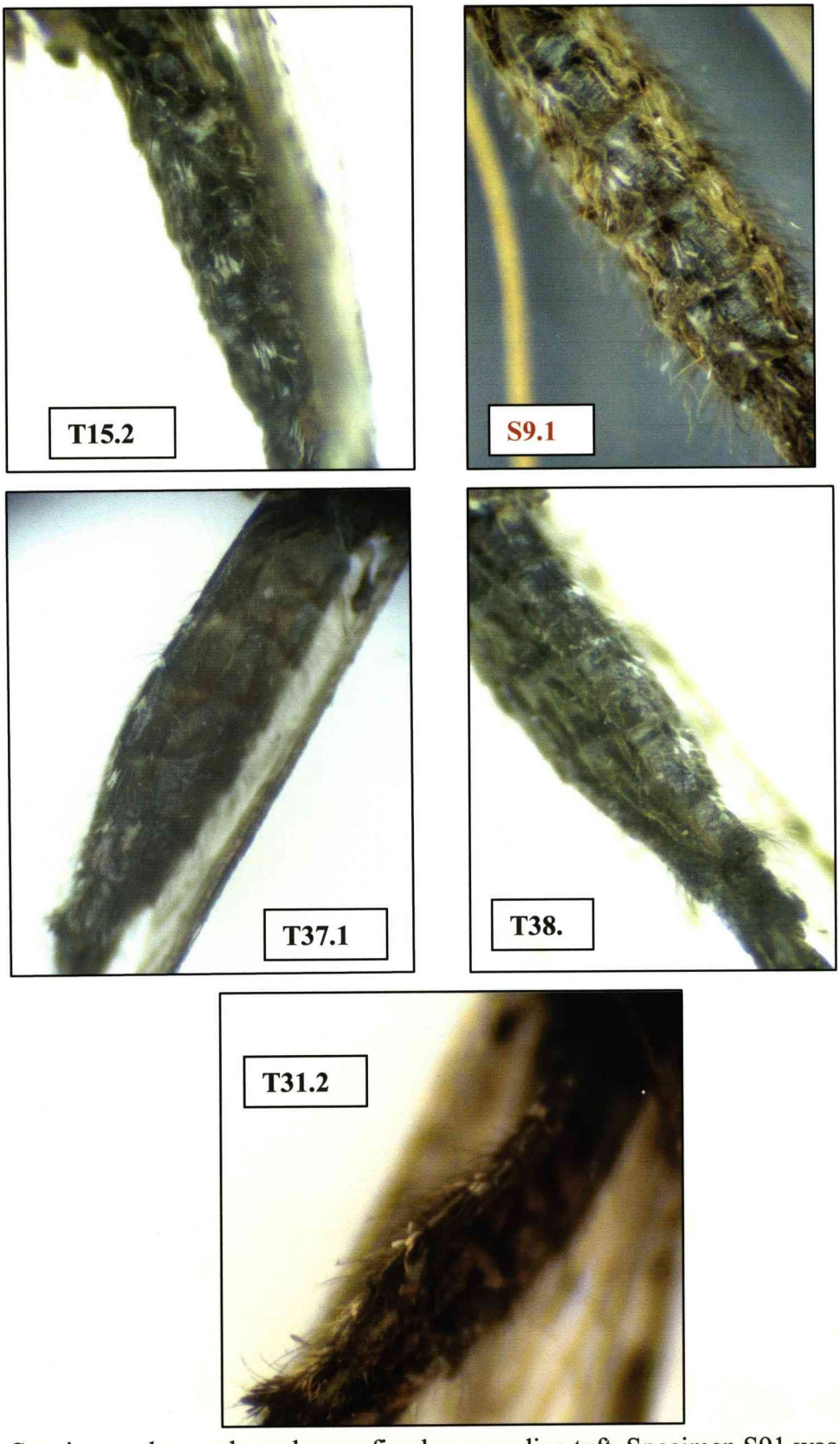
Number	Specimen	Locality	Field worker Identification	Identification in this study	Molecular identity
1	S11.2	Sakaeo	*An. barbirostris*	*An..barbirostris*	Clade III
2	S13.1	Sakaeo	*An. barbirostris*	*An. barbirostris*	Clade III
3	S13.2	Sakaeo	*An. barbirostris*	*An. barbirostris*	Clade III
4	S15.2	Sakaeo	*An. barbirostris*	*An. barbirostris*	Clade III
5	S15.4	Sakaeo	*An. barbirostris*	*An. barbirostris*	Clade III
6	S17.1	Sakaeo	*An. barbirostris*	*An. barbirostris*	Clade III
7	T6.1	Trat	*An. barbirostris*	*An. barbirostris*	Clade III
8	S9.1	Sakaeo	*An. barbirostris*	*An. barbirostris*	Clade III
9	T10.2	Trat	*An. barbirostris*	*An. barbirostris*	Clade III
10	T10.3	Trat	*An. barbirostris*	*An. barbirostris*	Clade III
11	T29.3	Trat	*An. barbirostris*	*An. barbirostris*	Clade III
12	T31.2	Trat	*An. barbirostris*	*An. barbirostris*	Clade III
13	T37.1	Trat	*An. barbirostris*	*An. barbirostris*	Clade III
14	T38.2	Trat	*An. barbirostris*	*An. barbirostris*	Clade III
15	S27.5	Sakaeo	*An. campestris*	*An. campestris*	Clade IV
16	S19.1	Sakaeo	*An. barbirostris*	*An. barbirostris*	Clade IV
17	T31.1	Trat	*An. barbirostris*	*An. barbirostris*	Clade IV
18	T33.1	Trat	*An. barbirostris*	*An. barbirostris*	Clade IV
19	T35.2	Trat	*An. barbirostris*	*An. barbirostris*	Clade IV
20	S1	Sakaeo	*An. barbirostris*	*An. campestris*	Clade IV
21	S10.1	Sakaeo	*An. barbirostris*	*An. campestris*	Clade IV
22	S11.1	Sakaeo	*An. barbirostris*	*An. campestris*	Clade IV
23	S15.1	Sakaeo	*An. barbirostris*	*An. campestris*	Clade IV
24	S24.1	Sakaeo	*An. campestris*	*An. campestris*	Clade IV
25	S24.4	Sakaeo	*An. campestris*	*An. campestris*	Clade IV
26	S24.5	Sakaeo	*An. campestris*	*An. campestris*	Clade IV
27	S27.1	Sakaeo	*An. campestris*	*An. campestris*	Clade IV
28	S27.3	Sakaeo	*An. campestris*	*An. campestris*	Clade IV
29	S30.1	Sakaeo	*An. campestris*	*An. campestris*	Clade IV
30	T12.3	Trat	*An. barbirostris*	*An. campestris*	Clade IV
31	T13.1	Trat	*An. barbirostris*	*An. campestris*	Clade IV
32	T22.1	Trat	*An. barbirostris*	*An. campestris*	Clade IV
33	T29.2	Trat	*An. barbirostris*	*An. campestris*	Clade IV
34	T30.1	Trat	*An. barbirostris*	*An. campestris*	Clade IV
35	T32.2	Trat	*An. barbirostris*	*An. campestris*	Clade IV
36	T35.1	Trat	*An. barbirostris*	*An. campestris*	Clade IV
37	T35.3	Trat	*An. barbirostris*	*An. campestris*	Clade IV
38	T36.1	Trat	*An. barbirostris*	*An. campestris*	Clade IV
39	T36.2	Trat	*An. barbirostris*	*An. campestris*	Clade IV
40	T37.2	Trat	*An. barbirostris*	*An. campestris*	Clade IV
41	T38.1	Trat	*An. barbirostris*	*An. campestris*	Clade IV
42	T7.1	Trat	*An. barbirostris*	*An. campestris*	Clade IV
43	S23.2	Sakaeo	*An. campestris*	*An. donaldi/pollicaris*	Clade IV
44	S40.1	Sakaeo	*An. campestris*	*An. donaldi/pollicaris*	Clade IV
45	S24.2	Sakaeo	*An. campestris*	*An. campestris*	Clade V
46	S25.1	Sakaeo	*An. campestris*	*An. campestris*	Clade V
47	S27.4	Sakaeo	*An. campestris*	*An. campestris*	Clade V

Table 6.3 List of specimens whose wings were examined.

Those with more than 50% of scales pale between the apical dark marks of veins 5-5.2 were called "pale wings" and those with less than 50% "dark wings"

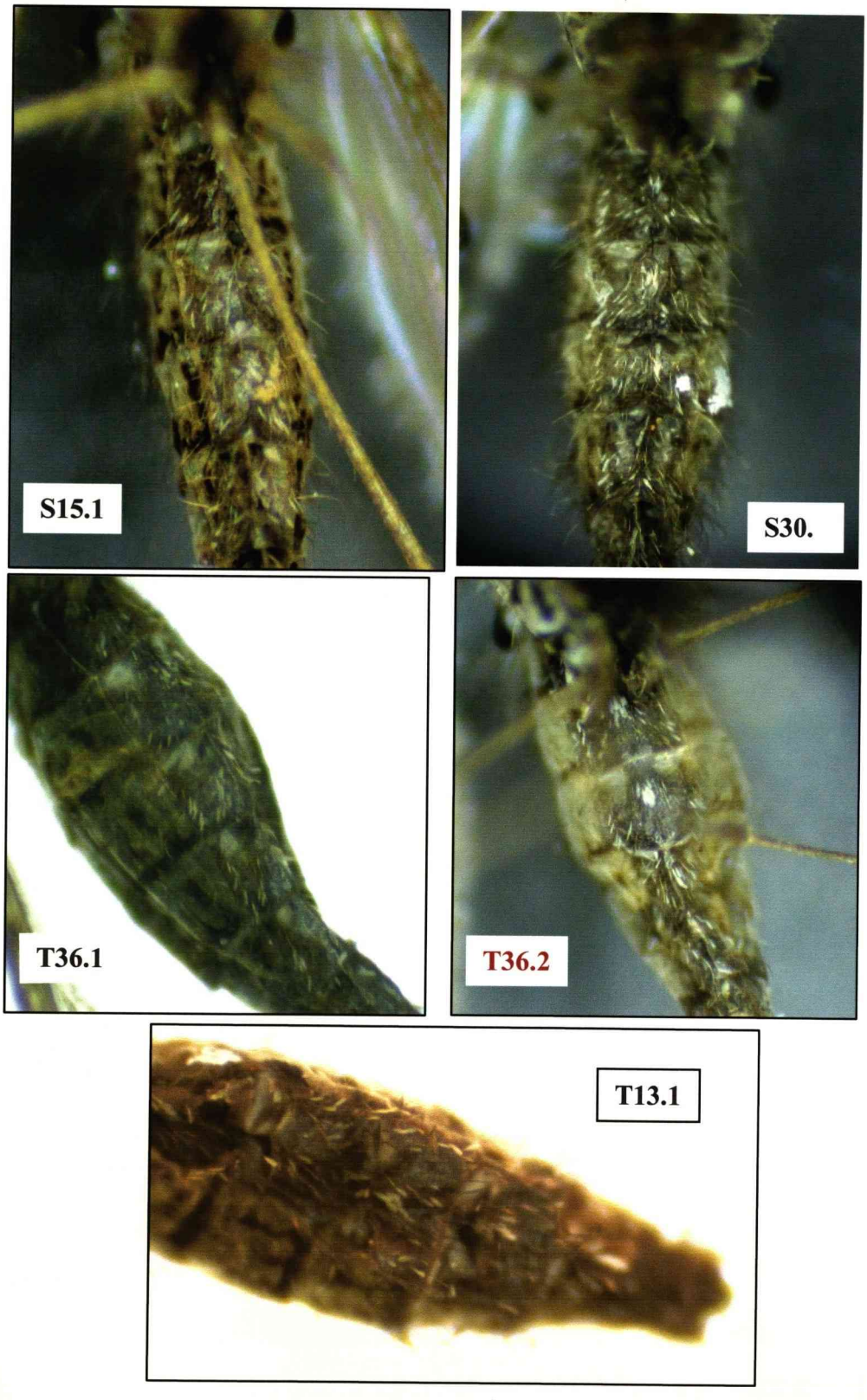
Number	Specimen	Locality	Morphological identification	Molecular identity	Wing pattern
1	T10.2	Trat	*An. barbirostris*	Clade III	pale
2	T29.3	Trat	*An. barbirostris*	Clade III	pale
3	T37.1	Trat	*An. barbirostris*	Clade III	pale
4	T38.2	Trat	*An. barbirostris*	Clade III	pale
5	T6.1	Trat	*An. barbirostris*	Clade III	pale
6	T31.1	Trat	*An. barbirostris*	Clade IV	pale
7	T33.1	Trat	*An. barbirostris*	Clade IV	pale
8	S30.1	Sakaeo	*An. campestris*	Clade IV	pale
9	T12.3	Trat	*An. campestris*	Clade IV	pale
10	T13.1	Trat	*An. campestris*	Clade IV	pale
11	T22.1	Trat	*An. campestris*	Clade IV	dark
12	T29.2	Trat	*An. campestris*	Clade IV	pale
13	T36.1	Trat	*An. campestris*	Clade IV	pale
14	T36.2	Trat	*An. campestris*	Clade IV	pale
15	T37.2	Trat	*An. campestris*	Clade IV	pale
16	T38.1	Trat	*An. campestris*	Clade IV	pale
17	T7.1	Trat	*An. campestris*	Clade IV	pale
18	S25.1	Sakaeo	*An. campestris*	Clade V	dark
19	S27.4	Sakaeo	*An. campestris*	Clade V	dark

Figure 6.6. Photographs of abdominal scales in specimens from Clade III.



Specimens show pale scales confined as a median tuft. Specimen S91 was originally identified as *campestris*.

Figure 6.7. Photographs of abdominal scales in specimens from Clade IV.



Specimen in red was originally identified as *An. barbirostris* Van der Wulp.

Figure 6.8. Photographs of abdominal scales in specimens from Clade IV.

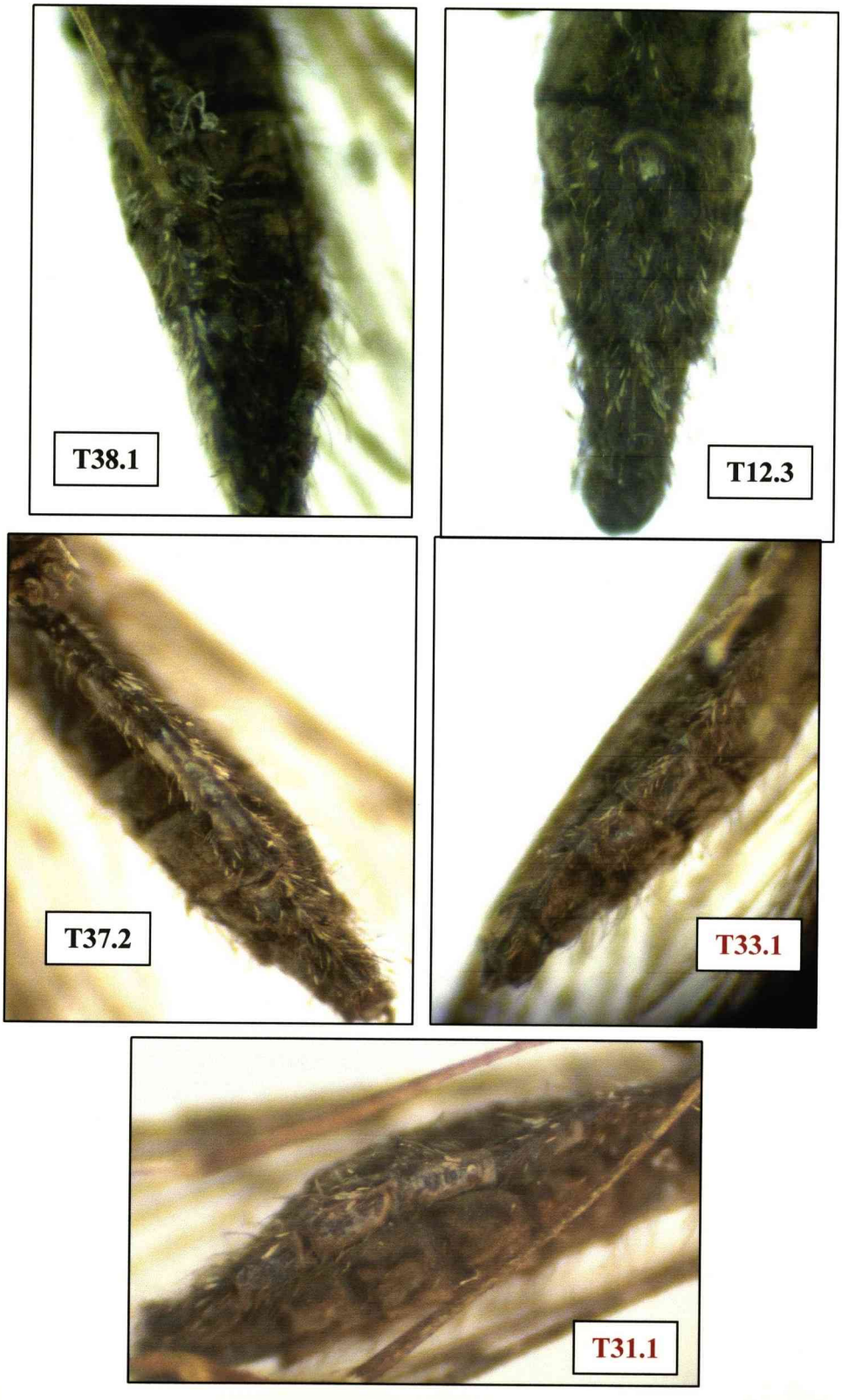


Figure 6.9. Photographs of abdominal scales in specimens from Clade V.

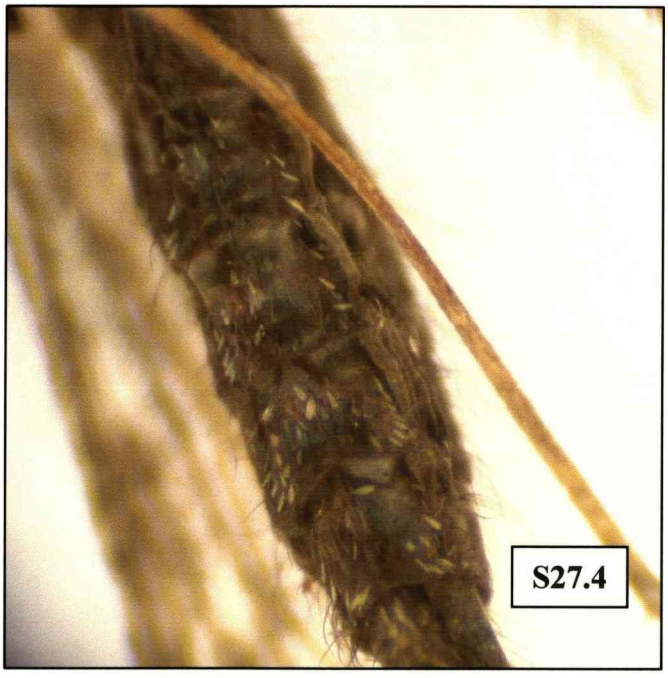


Figure 6.10. Wing pattern of specimens of clade III.

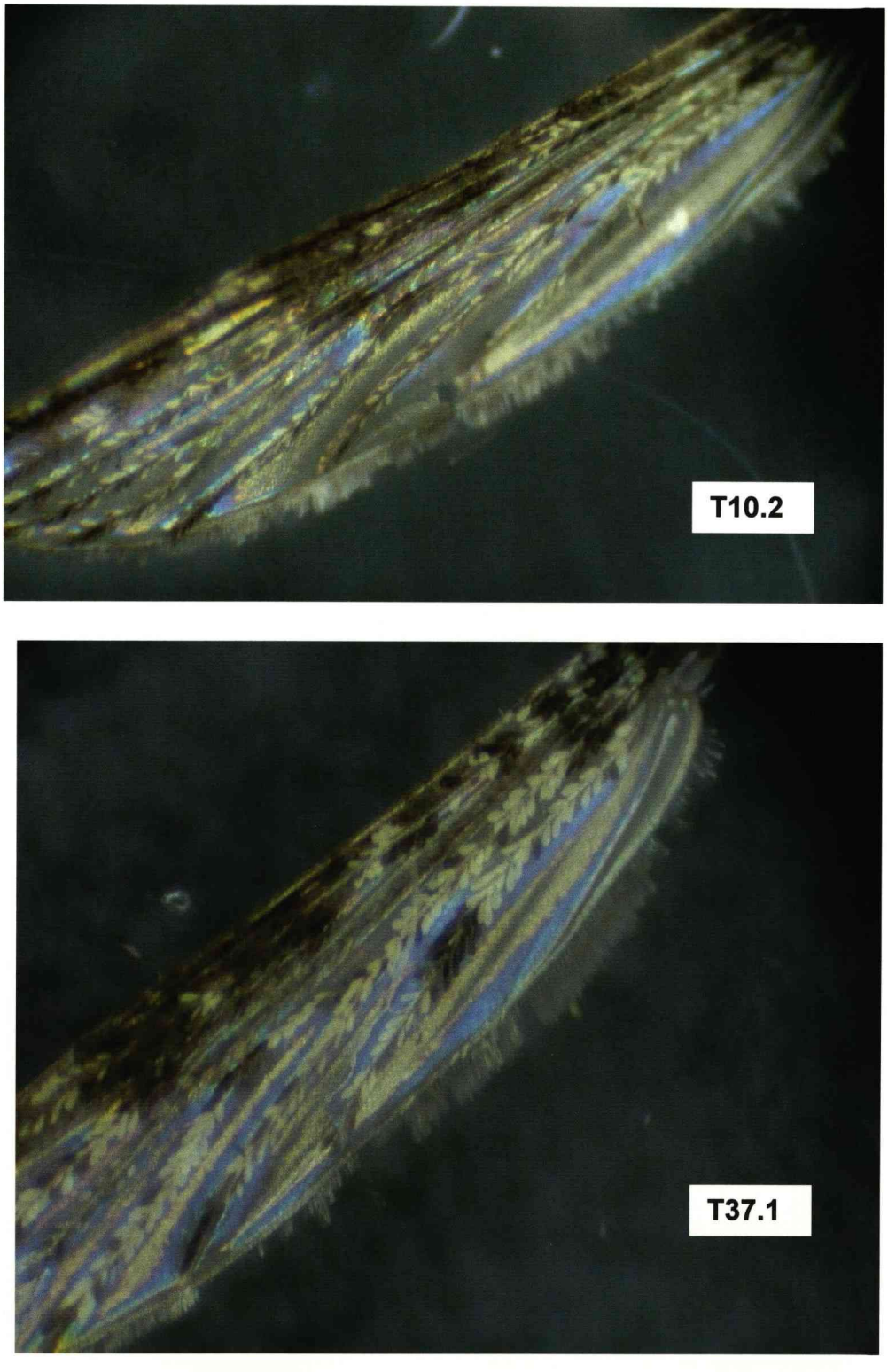


Figure 6.11. Wing pattern of specimens of clade IV.

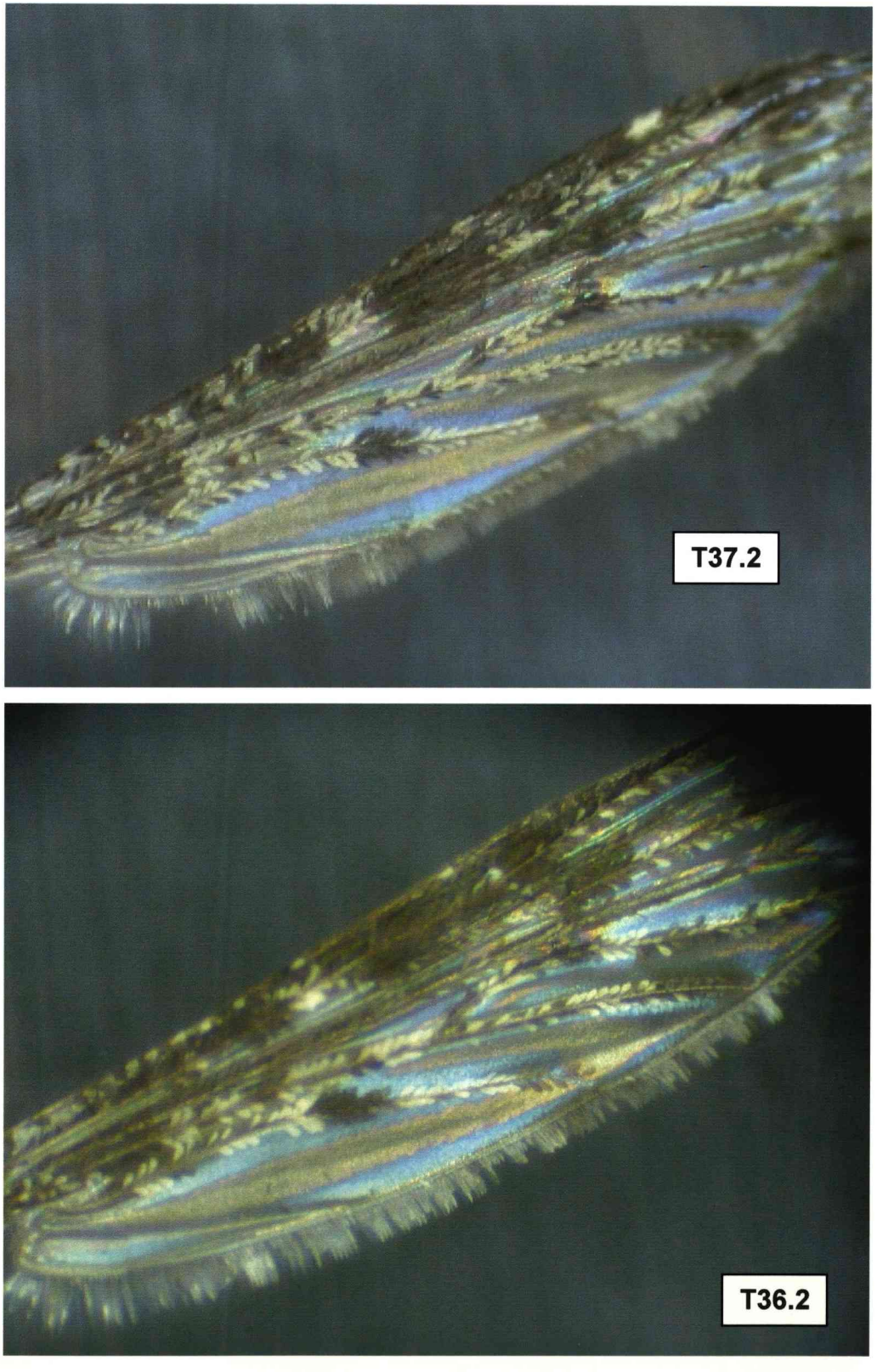


Figure 6.12. Wing pattern of specimens of clade V.

6.5 Discussion

6.5.1 Distinction between Barbirostris and Hyrcanus Groups

The presence or absence of scales on the clypeus is a reliable character to distinguish between the Hyrcanus and Barbirostris Groups. Several of the mistakes made by field workers could be related to the fact that the scales on the clypeus appear sometimes not to be attached to the clypeus surface (Harbach, personal communication). The other simplest explanation is the absence of trained staff for the identification of mosquitoes in this region.

6.5.2 Morphological identification of species and their relation to the clades within the Barbirostris Subgroup

All specimens from Clade III showed characters like *An. barbirostris* (Reid, 1962): few pale scales confined to median tufts on the sternites and wings paler (Figure 6.6 and Figure 6.10). Only one specimen, S9.1, was originally identified as *An. campestris* based on the pattern of pale scales on the abdomen. However when re-examining its morphology in the photographs, I observed that this specimen was in fact *An. barbirostris.* The mistake in the identification was due to some few scales are located in a lateral row on the abdominal sternites (Figure 6.5). Based on this evidence, it would appear that specimens from clade III are *An. barbirostris*.

The majority of specimens in Clade IV were morphologically identified as *An. campestris* based on the pattern of pale scales on the abdomen (Figure 6.7 and Figure 6.8) but surprisingly these also showed "pale wings" as in *An. barbirostris* (Figure 6.11 and Table 6.3). Wing pattern was used by Reid (1962) to differentiate these species, but this character was considered "not valid" and only 80-85% reliable in Thailand by Harrison and Scanlon (1975). These authors affirmed that "many Thai *An. campestris* had pale wings". Moreover, a more recent illustrated key by Rattanarithikul et al. (2005) did not include this wing character to differentiate *campestris* and *barbirostris*. The observations on the pale wings in *An. campestris* reported by Harrison and Scanlon (1975) agree with observations in this present

study in specimens from clade IV. This shows intermediate characters between *An. barbirostris* and *An. campestris*. Clade IV also comprised 4 specimens morphologically identified as *An. barbirostris* by the author. This was based on the low number of pale scales on the abdomen. As these scales are easily lost in dry specimens, my confusion in the identification of some individuals was not surprising.

Following Reid's (1962) key, two specimens in clade IV were identified as *An. pollicaris/donaldi/franciscoi* based on the presence of pale bands on the midtarsi. The presence of bands on the midtarsomeres is a character that is considered "usually" absent in *An. barbirostris* and *An. campestris* (Harrison and Scanlon, 1975; Reid, 1962), and as a result "not reliable". The fact that variation of this character did not relate to any molecular differences in the specimens indicates that this morphological character should not be used in the identification of these species.

Specimens from clade V were morphologically identified as *An. campestris*, based on both pale scales on the abdomen (Figure 6.9) and "dark-wings." (Figure 6.12), as described by Reid (1962). However, it is important to consider that only three specimens examined corresponded to this clade (wings were observed only in two specimens) and this may be insufficient to arrive at further conclusions.

To sum up, differentiating species within the Barbirostris Subgroup based on adult morphology is complicated as indicated by Reid (1962) and Harrison and Scanlon (1975). This chapter presents evidence for morphological differentiation of specimens of clades III, IV and V.

CHAPTER 7

7 Molecular analysis of ITS2 and COI in specimens of the Hyrcanus Group

7.1 Introduction

This chapter presents an analysis of the internal transcribed spacer 2 (ITS2) and cytochrome oxidase I (COI) in several specimens from the Hyrcanus Group. Specimens from this group were originally identified as members of the Barbirostris Subgroup by field workers, but errors in the identification were detected first by sequence analysis (see chapters 4 and 5) and subsequently by morphological re-examination of some specimens (see chapter 6).

The *Anopheles hyrcanus* species group (Hyrcanus Group) is the most species-rich group within the Myzorhynchus Series (Harbach, 2004). *Anopheles hyrcanus* was first described by Pallas in 1771 as *Culex hyrcanus* (Harrison and Scanlon, 1975) and subsequently assigned to the *An. hyrcanus* group (Reid, 1953). Twenty-seven species were included in the Hyrcanus Group in the latest revision of the classification of the genus *Anopheles* (Harbach, 2004) and subsequently 2 more species have been reported (Rueda, 2005).

Species of the Hyrcanus Group are widely distributed in the Palaearctic and Oriental regions, from Spain and North Africa in the west to Japan in the east and through India, the Malay Archipelago, the Philippines, Sulawesi and Moluccas in the south (Reid, 1968). Some of the species have been incriminated as vectors of malaria and filariasis (Manh et al., 2000; Sen, 1956; Simpson, 1951).

The Hyrcanus Group includes species with highly variable morphology and the identification of related species has been considered taxonomically problematic (Ma and Xu, 2005). Molecular markers are an important addition to traditional methods based on morphology and, in particular, ITS2 appears to be valuable for the identification of species in the group (Hwang, 2007; Ma and Xu, 2005; Min et al., 2002). Based on analysis of ITS2 in the species *An. pullus* and *An. yatsushiroensis*,

the synonymy of these species was demonstrated (Hwang et al., 2004). These authors supported that uncertainties in the extent of dark scales on the wings may be seasonal, perhaps linked to temperature. This molecular marker ITS2 was also used to detect two new species within the Hyrcanus Group, which were "unknown species 1," and "unknown species 2 (Li et al., 2005). Both species were morphological similar to *An. sinensis*, but their ITS2 was 9.1% and 10.7 % different from *An. sinensis,* respectively (Li et al., 2005). This led Rueda (2005) to examine the morphology of these species and name them *Anopheles belenrae* (unknown species 1) and *Anopheles kleini* (unknown species 2).

7.2 Material and methods

The cytochrome oxidase I region, in the mitochondrial genome and ITS2, in the nuclear ribosomal DNA, were amplified following the procedures described in chapter 3. ITS2 sequences were cloned prior to sequencing, whereas COI fragments were sequenced directly.

In total, 28 specimens from the Hyrcanus Group were studied in this present study. Twenty-two were originally identified as members of the Barbirostris Group:14 identified as *An. barbirostris*, 7 as *An. barbumbrosus* and 1 as *An. campestris* and six specimens were identified in the field as species from the Hyrcanus Group. These specimens were obtained from Kalimantan in Indonesia and the localities of Sa Kaeo, Tak and Trat in Thailand (Table 7.1). For comparison, 13 ITS2 and 2 COI sequences available in Gen Bank were also included (Table 7.2).

Multiple sequence alignments were carried out using Clustal W and phylogenetic trees were obtained using the Neighbour-joining method, with a Kimura two-parameter distance model. A bootstrap test of 500 replicates was used to assess the reliability of branches. The COI tree used the corresponding *An. gambiae* sequence as an outgroup.

Table 7.1. A list of specimens from the Hyrcanus Group analysed in this study. Many of these were erroneously identified by field staff as members of the Barbirostris Group. Specimens HT8.1, HT8.2 and HT8.3 from Trat in Thailand were examined by the author prior to sequencing and identified as Hyrcanus Group from their morphology.

Species	Locality	Identification in the field	Sequence amplified
a11	Kalimantan	*An. barbirostris*	ITS2
a13	Kalimantan	*An. barbirostris*	ITS2, COI
a14	Kalimantan	*An. barbirostris*	COI
a22	Kalimantan	*An. barbirostris*	ITS2, COI
a21	Kalimantan	*An. barbirostris*	ITS2, COI
a23	Kalimantan	*An. barbirostris*	ITS2, COI
a24	Kalimantan	*An. barbirostris*	ITS2, COI
a25	Kalimantan	*An. barbirostris*	ITS2, COI
a26	Kalimantan	*An. barbirostris*	ITS2, COI
bbs1	Sa Kaeo	*An. barbumbrosus*	COI
bbs3	Sa Kaeo	*An. barbumbrosus*	COI
bbs4	Sa Kaeo	*An. barbumbrosus*	COI
bbs5	Sa Kaeo	*An. barbumbrosus*	COI
bbs7	Sa Kaeo	*An. barbumbrosus*	COI
bbs9	Sa Kaeo	*An. barbumbrosus*	COI
btr20	Trat	*An. barbirostris*	ITS2, COI
ta1	Tak	*An. barbirostris*	ITS2, COI
ta2	Tak	*An. barbirostris*	ITS2, COI
sa1	Sa Kaeo	*An. barbirostris*	ITS2, COI
sa2	Sa Kaeo	*An. barbirostris*	ITS2, COI
HS36.1	Sa Kaeo	*An. barbumbrosus*	COI
HS36.2	Sa Kaeo	Hyrcanus Group	COI
HS40.3	Sa Kaeo	*An. campestris*	COI
HS43.1	Sa Kaeo	Hyrcanus Group	COI
HT8.1	Trat	Hyrcanus Group	ITS2
HT8.2	Trat	Hyrcanus Group	ITS2
HT8.3	Trat	Hyrcanus Group	ITS2
HT8.4	Trat	Hyrcanus Group	COI
HT8.5	Trat	Hyrcanus Group	COI

Table 7.2. List of sequences of species obtained from GenBank.

Species	Region	Accession number	Author
An. sinensis	COI	AY339281	(Hwang et al., 2004)
An. gambiae	COI	AF417706	(Sallum et al., 2002)
An. pullus	COI	AY339279	(Hwang et al., 2004)
An. lesteri	ITS2	AY576906	(Gao et al., 2004)
An. belenrae	ITS2	DQ137142	(Park et al. 2003)
An. kweiyangensis	ITS2	AF261150	(Ma and Xu, 2005)
An. hyrcanus	ITS2	EF613304	(Djadid et al., 2007)
An. pullus	ITS2	AY339274	(Hwang et al., 2004)
An. sinensis	ITS2	AY339278	(Hwang et al., 2004)
An. engarensis	ITS2	AB159604	(Sawabe et al. 2004)
An. sineroides	ITS2	AB159605	(Sawabe et al. 2004)
An. kleini	ITS2	DQ177501	(Park et al. 2003)
An. liangshanensis	ITS2	AF146750	(Ma and Xu, 2005)
An. junlianensis	ITS2	AY316155	(Ma and Xu, 2005)
An. crawfordi	ITS2	AF261949	(Ma and Xu, 2005)
An. peditaeniatus	ITS2	AF543862	(Gao et al., 2004)

7.3 Results

7.3.1 Analysis of the Internal Transcribed Spacer 2 (ITS2)

The phylogenetic relationships among members of the Hyrcanus group were studied with a Neighbour-joining tree, using ITS2 sequences. The fragment amplified included 5.8S (partial fragment), ITS2 (complete) and 28S (partial fragment). Only ITS2 was used in the phylogenetic analysis. Boundaries of the ITS2 were defined by comparison with those in *Anopheles gambiae* (Paskewitz et al., 1993).The trees constructed were unrooted, since ITS2 sequences of these closest groups in GenBank, *An. koreicus* (Barbirostris Group) and *An. bancroftii* (Bancroftii Group) were unalignable.

Specimens sa1, sa2 collected from Sa Kaeo, Thailand and originally identified as *An. barbirostris* were located in the same phylogenetic clade as *An. sinensis* (Figure 7.1). These differed by only two nucleotides from the *sinensis* sequence deposited in GenBank: AY339281 (Hwang et al., 2004).

Specimens ta1 and ta2 were also received as *An. barbirostris*. Molecular analysis of the ITS2 proved that these specimens were related to *An. peditaeniatus*. This was based on sequence comparisons and phylogenetic analysis. ITS2 sequences for *An. peditaeniatus* deposited in GenBank belonged to a specimen collected from central China (Gao et al., 2004).

Specimens HT8.2 and HT8.3 and btr20 were morphologically examined by the author and identified as Hyrcanus Group species. The ITS2 Neighbour-joining tree showed that these formed a distinctive clade, which could not be related to any ITS2 sequence deposited in GenBank and that appear distant from the other 12 species compared (Figure 7.1). These appear closely related to the clade formed by specimens from Kalimantan (a11, a21, a23, a25, a13, a26, a22, a24), whose identity could also not be determined.

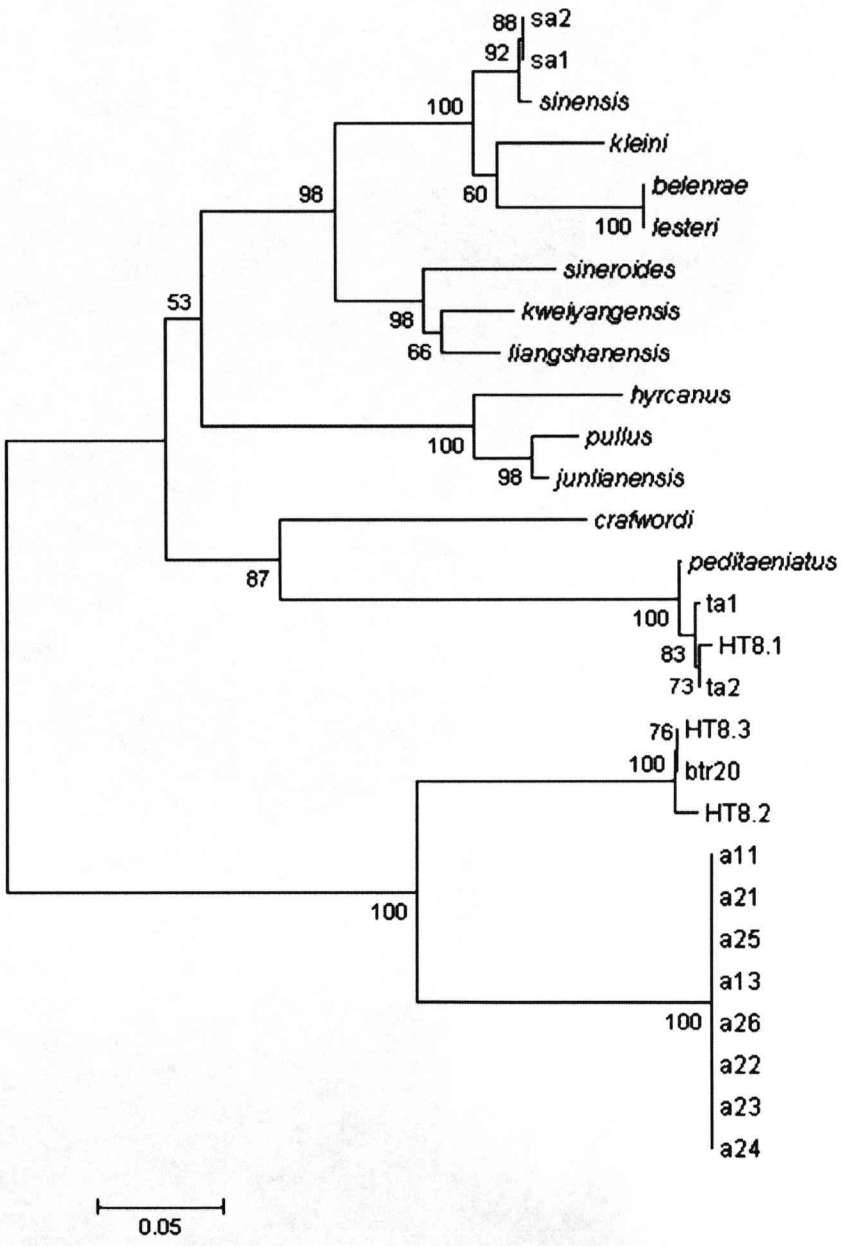
The length of the ITS2 in sequences studied varied, being 448 bp in specimens identified as *An. peditaeniatus,* 465 bp in *An. sinensis*, 476 bp in specimens H8T3, btr20, HT8.2 and 503 bp in specimens from Kalimantan. All ITS2 sequences were AT-rich (53.4- 56%).

7.4 Analysis of cytochrome oxidase I (COI)

The COI sequences were 756 bp long. Sequences were AT rich, ranging from 69.31% (a25) to 71.43% (HT8.5). The main disadvantage of using this marker in the Hyrcanus group is the limited number of sequences available in GenBank. Only sequences of species *An. sinensis and An. pullus* were found. Most studies have used the ITS2 gene region for identification purposes.

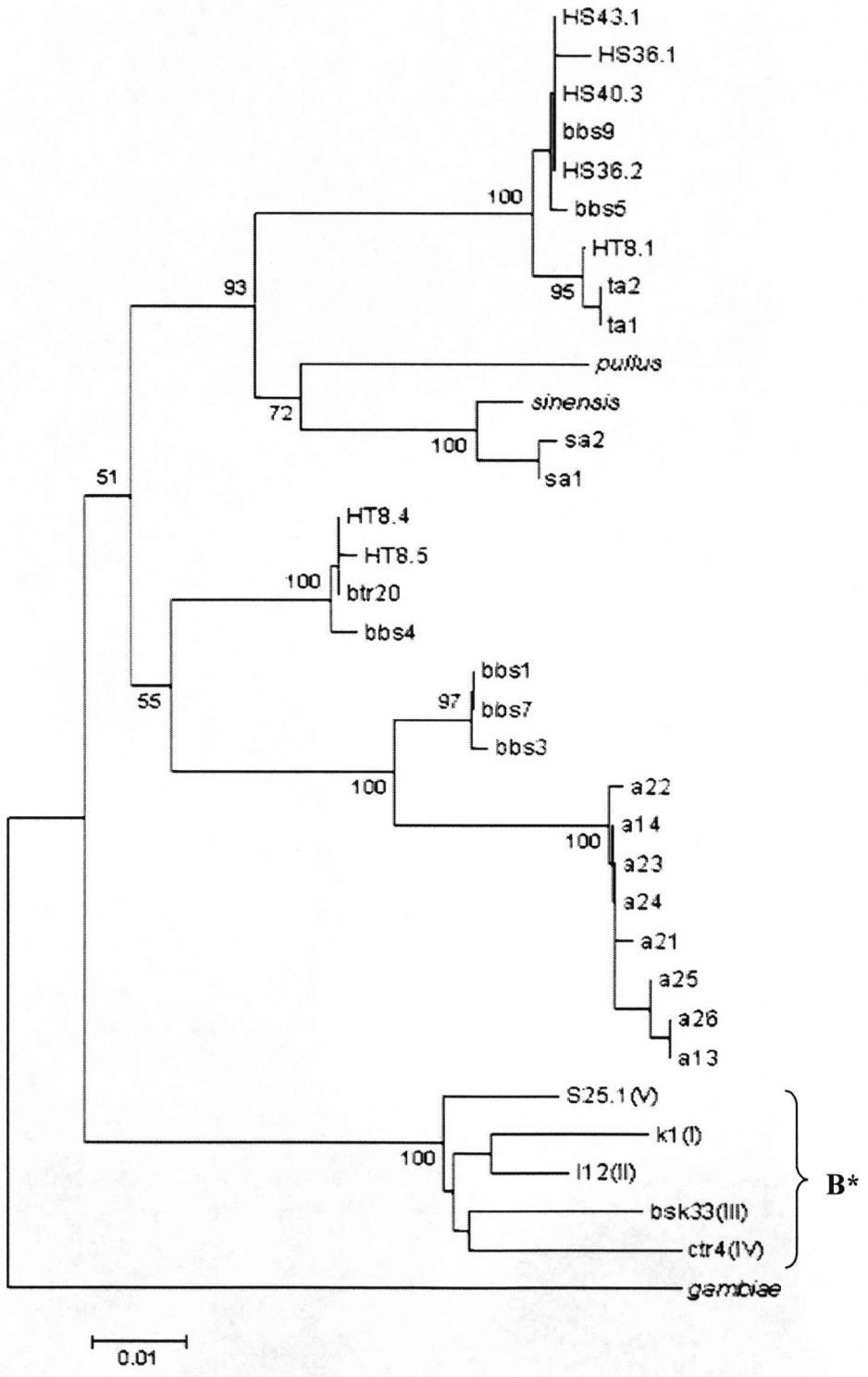
The method used to infer tree topology was Neighbour-joining, with a Kimura 2-parameter substitution model. The tree was rooted and the COI sequences of *An. gambiae* and of the 5 clades of the Barbirostris Subgroup were used as outgroups (Figure 7.2). Six clades were obtained. Specimens sa1 and sa2 were located in the same clade as *An. sinensis* AY339278 from Korea and only differed from it by 6 nucleotide substitutions in the 525 bp region compared. The remaining specimens could not be related to sequences deposited in GenBank.

Figure 7.1. Neighbour-joining tree using ITS2 sequences of specimens from the Hyrcanus Group.



Sequences for named species were obtained from GenBank. Those for the numbered specimens were sequenced as part of the study. Specimens starting with a capital letter were morphologically identified Hyrcanus Group species.

Figure 7.2. Neighbour-joining tree using COI sequences.



Sequences for named species were obtained from GenBank. Those for the numbered specimens were sequenced as part of the study. Specimens starting with a capital letter were morphologically identified Hyrcanus Group species. B*=clades I-V of the Barbirostris Subgroup.

7.5 Discussion

The identity of some specimens was deciphered by means of phylogenetic analysis. Specimens sa1 and sa2 (from Sa Kaeo) were *An. sinensis,* as demonstrated with the analysis of the COI and ITS2 regions. Originally these had been identified as *An. barbirostris* by field staff. The reference sequence from GenBank belonged to a specimen of *An. sinensis* collected in Korea (Hwang et al., 2004). It is interesting to observe that, considering the large distance between localities in Thailand and Korea (about 3000 km), the Thai specimens only differed in 2 nucleotide substitutions in the ITS2 sequences and in 6 nucleotides in 525 bp of the COI fragment. *An. sinensis* is widely distributed in the orient: it is found in Assam, Myanmar, Thailand, Peninsular Malaysia, Sumatra (Indonesia), Cambodia, Vietnam and China (WHO 1984).

Based on the comparison of ITS2 sequences, specimens ta1 and ta2 collected from Tak were identified as *An. peditaeniatus.* Interestingly these specimens only differed in 2 nucleotide substitutions with relation to the GenBank sequence, which belong to a specimen collected from China. Both trees presented agreed that specimen HT8.1 is in the same clade as these specimens and as a result, likely to be *An. peditaeniatus.*

Both *An. sinensis* and *An. peditaeniatus* appear widely distributed in Thailand, and have been reported in every province of the country (Rattanarithikul et al., 2005). Subsequent to their molecular identification of specimens ta1, ta2 and sa1, sa2, the morphology of other specimens kept by the providers was re-examined and their identity confirmed as *An. peditaeniatus* and *An. sinensis* respectively. This demonstrates that these species can be reliably identified with molecular markers ITS2 and COI.

The identity of the remaining specimens is still unknown, as they did not resemble any sequences from GenBank. Those collected in Kalimantan (Indonesia) are likely to be *An. nigerrimus* (Ralph Harbach, personal communication), but considering the poor condition of specimens when morphologically identified, it is difficult to reach

conclusions. Species *An. nitidus* and *An. argyropus* have also been recorded in Indonesia (O'Connor, 1980) and should be also considered as the possible species of specimens from Kalimantan. The ITS2 tree shows that these "unknown" species are distant to all Hyrcanus Group species sequences recorded to date (Figure 7.1). It is possible that these specimens belong to *An. nigerrimus* or *An. nitidus* from the Nigerrimus Subgroup, although no sequences from this group have been recorded in GenBank. It is also possible that these specimens may belong to a species not yet recorded. Similarly, the "unknown" species from Thailand may belong to species whose ITS2 and COI sequences have not been deposited in GenBank, for example *An. argyropus, An. nigerrimus, An. nitidus, An. paraliae* and *An. pursati.*

Much additional taxonomic work is required on the Hyrcanus Group if we are to determine the distribution and phylogenetic relationships of almost 30 species. Such studies are also essential if we are to determine their potential role in the transmission of vector borne diseases (Manh et al., 2000; Reid, 1968).

CHAPTER 8

8 General discussion

A previous study of *Anopheles barbirostris* Van der Wulp by Baskoro (2001) uncovered evidence for cryptic species within this taxon in Indonesia and Thailand. Based on sequences of the mtDNA COI region, Baskoro identified four major clades, some of which occurred in sympatry and others which were allopatric. The same clades also showed major differences in sequences in the nuclear rDNA ITS-2 spacer region. Baskoro regarded these clades as species to which he applied the letters W, X, Y and Z. Species W and X occurred in sympatry in Java and Flores, Indonesia, whereas Z was only found in Sulawesi. A sample from one site in Thailand yielded species Y. Two specimens from China appeared to be distinct from these four species.

The original objective of the current thesis was to extend the work of Baskoro (2001) to a broader area of South-East Asia, with the emphasis mainly on *Anopheles barbirostris*. However, during the course of molecular studies of field-collected material, it became apparent that another species of the Barbirostris Subgroup, *An. campestris*, was being confused with *An. barbirostris*, and that this confusion extended not only to other species in the larger Barbirostris Group, but possibly to other anopheline mosquitoes in Southeast Asia. As a result the focus of the work became broader to include the species with which *An. barbirostris* was being confused. It also became clear that taxonomic entities inferred from molecular data needed to be related to the diagnostic morphological characters in standard taxonomic keys.

This chapter reviews the results from the analysis of cytochrome oxidase I (COI) and Internal transcribed spacer 2 (ITS2) and discusses these in relation to the morphology of members of these clades.

During the final stages of the writing up of this thesis, a study was published describing cytogenetic and molecular systematic studies of species of the Barbirostris Subgroup in Thailand (Saeung et al., 2007). The authors inferred that there were two species in what they described as 'the *An. barbirostris* complex'. Hence, in this chapter it is important to compare my results with those of Saeung et al. as well as those of (Baskoro, 2001).

At most sample sites there were insufficient specimens to draw conclusions about the population structure within the clades. However, because this has been a subject of intensive study in other anopheline vectors, the data on population structure obtained are briefly considered.

The findings from this molecular systematics study have possible implications for the epidemiology and control of malaria and lymphatic filariasis, and this is also briefly discussed.

8.1 Species identification, distribution and feeding preferences

A comprehensive analysis of regions COI and ITS2 showed that the species within the Barbirostris Subgroup identified as *An. barbirostris* and *An. campestris* belong to 5 distinct clades. Based on the evidence presented in previous chapters, each of these 5 clades corresponds to a different species, i.e. five species in total. The subsequent morphological examination of specimens from Sa Kaeo and Trat showed that clades III, IV and V have distinctive morphology. As a result, three species have been found to correspond to the current taxon *An. barbirostris*, one species corresponds to *An. campestris* and one appears to be a previously unrecognised species with some features of both *campestris* and *barbirostris.* Below is a summary of these findings.

8.1.1 *Anopheles barbirostris* (Van der Wulp, 1884)

Adult female specimens from clade III showed the morphology of *An. barbirostris*, as described by Reid (1962): a tuft of pale scales confined only to the middle of the

abdominal sternites, and pale scales on wing veins 5-5.2. Specimens from this clade were originally identified as both larvae and adult specimens in the Natural History Museum; hence an assumption that they were *An. barbirostris* was justified. The comparison of the ITS2 sequence data of clade III with sequences published by Seung et al. (2007) showed that this clade is the same as the karyotypic form B of *An. barbirostris* identified by these authors. Based on this evidence we can affirm that clade III is one of the species within the *An. barbirostris* complex. The two specimens from China examined by Baskoro (2001) seem close to clade III. The region of COI sequenced by Baskoro was shorter (600 bp) than that used in the present study, resulting in a 756-bp region of only partial overlap.

The large number of specimens of clade III that were captured using animal bait, both in this study and that of Saeung *et al.,* indicates that clade III is a zoophilic species as described by Reid (1962). Therefore, its role in malaria and filariasis transmission may be limited, at least in the regions studied. *Anopheles barbirostris* (clade III) appears widely distributed in Thailand, in the provinces of Mae Hong Son, Tak, Kanchanaburi, Trat Sa Kaeo (collections sites of this present study), Chiang Mai, Ubon Ratchathani and Petchaburi (specimens collected by Saeung *et al.* 2007). It also appears to be present in the province of Yunan, in south China based on Baskoro's COI data used in the present analysis (full data not shown).

Specimens from clades I and II showed adult female morphology similar to *An. barbirostris*. Clade I specimens were collected from Mae Hong Son (Thailand) and Kalimantan (Indonesia) and those from clade II were found in the islands of Sumatra and Java (Indonesia). Baskoro (2001) carried out a study of *An. barbirostris* in Indonesia. The forms found by Baskoro, named species W, X and Z are closely related to those from clade I in the present study. Species W and X were founf living in sympatry in Java and Flores, whereas species Z was found in Sulawesi. On the other hand, species Y (from Thailand), was found to be in clade V. In this study, specimens from Kalimantan appear close to Baskoro's species X, whereas those from Mae Hong Son to species W. Baskoro's species Z appears in one separate cluster (Figure 5.18).

Figure 8.1 Geographic distribution of the 5 clades (species) of the Barbirostris Subgroup in Thailand and Indonesia in this study.

- Clades I (*Anopheles barbirostris*)
- Clade II (*Anopheles barbirostris*)
- Clade III (*Anopheles barbirostris*)
- Clade IV (Unknown species, intermediate form)
- Clade V (*Anopheles campestris*)
- Specimens studied by Baskoro (2001), clade I (*Anopheles barbirostris*)

Based on the morphological examination of adult specimens, clade I appears to be a second form of *An. barbirostris*, widely distributed in Indonesia and also present in Thailand. Clade I is also found in the province of Binh Thuan, Vietnam; detected by COI sequence comparison of a partial fragment of 500 bp. The clade II specimens have been identified as *An. barbirostris* by morphological examination of seven adult specimens. Clade II would be a third form of *An. barbirostris*. It is likely that clade I or II are conspecific with the *An. barbirostris* described by Van der Wulp.

The evidence presented in this thesis strongly suggests that *An. barbirostris* is a species complex. This would help explain the differences reported in the vector status of this taxon. Zoophilic and anthropophilic forms of *An. barbirostris* have been reported in the past (Lien et al., 1977). However these variations in vector status could not be associated with distinctive morphological characters (Reid, 1979). Subsequent studies demonstrated that *An. barbirostris* comprises four distinct chromosomal forms, three of these in Thailand, A, B, C (Baimai et al., 1995; Saeung et al., 2007) and a fourth in Java, D (Baimai et al., 1995). The authors maintained that it was difficult to know if forms in Thailand were the result of intraspecific or interspecific differences but form D from Java was thought to be a distinct species. By sequence comparison with the ITS2 sequence presented by Saeung et al. (2007), it is concluded that forms B and C are the same species, i.e. clade III (*Anopheles barbirostris* Van der Wulp), whereas form A is in clade IV.

To summarise, at least three species are found within *Anopheles barbirostris* Van der Wulp, clades I, II and III. Clade I is widely distributed in the Indonesian archipelago, Thailand and Vietnam; clade II was found on the islands of Sumatra and Java in Indonesia and clade III from Thailand and China. Some of these species were found in sympatry in Mae Hong Son, Thailand (clades I and III) and in Java (clade I and II) (Figure 8.1). There is limited information regarding the host preferences of these species, except for clade III which appears to be primarily zoophilic. This would suggest that the species represented by clade III do not have a major role in the transmission of malaria and filariasis in Southeast Asia.

8.1.2 *Anopheles campestris* Reid (1962)

Although only three specimens were morphologically examined in Clade V, they appear to be *An. campestris* as described by Reid (1962). These showed numerous white ventral scales on the abdomen and their wings were darker than *An. barbirostris*, with more than 50% of dark scales on wing veins 5-5.2. Sequence comparison demonstrated that clade V is the "*An. campestris*-like" group of Saeung et al. (2007), identified at the pupal stage. These specimens were reared in the lab as progeny broods from the adult females collected in the field using human-baits, suggesting that their behaviour was primarily anthropophilic (Saeung et al., 2007). In this present study specimens from the province of Sa Kaeo were collected using animal bait and only 10% were *An. campestris,* reflecting the anthropophilic nature of this species. Apiwathnasorn et al. (2002) found that in the provinces of Sa Kaeo, the proportion of *An. campestris* in animal landing catches was only 7.1% of the total number of all females captured, but 78.6% when using human landing catches.

An. campestris has been reported in Thailand (Harrison and Scanlon, 1975), Cambodia (Harrison and Klein, 1975) and Malaysia (Reid, 1968). In this present study, specimens from *An. campestris* (clade V) were collected only from Sa Kaeo, Thailand; it did not appear in the samples from other collection sites, possibly because most collections were from animal bait. *An. campestris* inhabits places of low elevation in Thailand (Harrison and Scanlon, 1975), from 1 to 200 m elevation. Altitude may be an important limiting factor in its distribution and could explain why there are no records of this species from the mountainous northern provinces Thailand.

Even though *An. campestris* were characterized at the molecular level by Saeung et al. (2007), the authors failed to place this species in the correct taxonomic position. This so called "*An. campestris*-like form" was placed within the *An. barbirostris* complex. However, all the available evidence suggests that *An. campestris* should be placed not within the *An. barbirostris* complex, but as a member of the Barbirostris Subgroup, as originally suggested by Reid (1962). This present study confirms the existence of *An. barbirostris* complex, but this would comprise three sibling species

of very similar, if not identical morphology, clade I, clade II and clade III, in addition to those previously identified in Indonesia (Baskoro, 2001).

8.1.3 Unknown species in clade IV

Interestingly, clade IV included specimens with "abdominal sterna with many white scales scattered between median patch and lateral rows" (Rattanarithikul et al., 2005), as described for *An. campestris*, but also showed pale wings as described for *An. barbirostris* (Reid, 1962). This clade corresponds to the karyotypic form A of *An. barbirostris* described by Saeung et al. (2007), which were identified morphologically as pupae. Saeung et al. (2007) showed that this species is zoophilic, which appears consistent with the high number of specimens obtained from animal-bait traps used in this present study. This does not correspond to the highly anthropophilic behaviour described for *An. campestris*.

It is indisputable that clade IV represents a different species. However as it has some morphological characters from both *An. campestris* and *An. barbirostris*, its relation to either of these species is confusing. Considering that larvae and pupae are the best stages for identification of *An. campestris* (Harrison and Scanlon, 1975; Rattanarithikul et al., 2005), clade IV would be identified as a new species within the *An. barbirostris* complex. Nonetheless, taking into consideration that ITS2 size in clade IV specimens (1583 bp) is different from that reported for *An. barbirostris* (clade III~1730) and for *An.* campestris (clade V~ 1519) and that morphological examinations in the field are mainly based on the examination of adult specimens, I suggest that clade IV should be identified as a new species within the Barbirostris Subgroup. It differs from those described to date, as it appears not only distinctive on molecular criteria but also at the morphological level (pale-winged *campestris*). A study with more comprehensive morphological examination and a larger number of specimens is required to finally settle the status of this species.

8.1.4 Comments on morphological characters

Current morphological keys for identification of *Anopheles barbirostris and An. campestris* in Thailand as adult stages (Harrison and Scanlon, 1975; Rattanarithikul et al., 2005) cannot distinguish the zoophilic clade IV from the anthropophilic clade V, since they are only based on the examination of ventral pale scales on the abdomen. A revision of these keys is necessary.

The presence of pale bands on the midtarsomeres was said by Reid (1962) to separate other members of the Barbirostris Subgroup from *An. campestris* and *An. barbirostris.* However, in this study specimens with the molecular features of *An. campestris* were found to have pale bands on the midtarsomeres, suggesting that this is not a reliable character. The validity of other morphological characters to distinguish species has been thrown into doubt by molecular data. The fringe spots on the wings were considered the key character to differentiate *An. pullus* and *An. yatsushuroensis* (Miyasaky, 1951 *In* Hwang *et al.* 2004). However, the analysis of ITS2 demonstrated the synonymy of these species (Hwang et al., 2004). This character appears to vary with season, possibly influenced by temperature. Wing fringe morphology has also been used to distinguish *An. bancroftii* from *An. pseudobarbirostris,* but a restriction length polymorphism and a heteroduplex analysis of the ITS2, showed no correlation between genotypes and fringe patterns (Beebe et al., 2001). The fringe spots are also key characters used to distinguish members of the Barbirostris Subgroup (see Figure 6.5). The validity of this character should be tested using molecular sequence data. This was impossible in the present study due to the poor state of many of the specimens received. However, some morphological characters evaluated in the present study, like the presence/absence of scales in the clypeus have proved efficient to separate the Barbirostris and Hyrcanus Groups.

The presence of a sunken patch of 7–10 sensilla borne in an unusual location in the antenna has been reported in *An. barbirostris* by examination with electron microscope (Kaur, 2005). Such sensilla have not been reported in other species of *Anopheles*, *Culex* or *Aedes*. Whether this character occurs in more than one species

within the *An. barbirostris* species complex or is present in *An. campestris* has yet to be determined.

Both molecular markers used in this present study agreed in the identification of 5 distinctive clades (species) in the Barbirostris Subgroup (species identified morphologically as *An. barbirostris* and *An.campestris*). The main disadvantage in the use of the ITS2 was the difficulty of amplification and subsequent sequencing, due to its large size and the presence of internal repeats. There was high interspecific and low intraspecific variability, confirming that ITS2 is a valuable tool to identify species within the Barbirostris Subgroup.

The differences in sizes of ITS2: *Anopheles barbirostris* (clade III=1730 bp), *An. campestris* (clade V= 1519 bp) and "pale winged" *An. campestris* (Clade IV= 1583 bp) could be used to distinguish these species, as difference in sizes could be visualised with a simple PCR, followed by examination on 1% agarose gel. However, its use in the identification of these species would depend on the absence of clades I and II in these regions, since these showed very similar sizes for ITS2: 1545 and 1727 bp respectively. PCR-based techniques have proven valuable to distinguish sibling species of *Anopheles* species (Cornel et al., 1996; Goswami et al., 2005; Manguin et al., 2002; Phuc et al., 2003; Porter and Collins, 1991).

The use of restriction enzymes in a restriction length polymorphism PCR (RFLP) may allow a rapid identification of the 5 species, as they may present a different restriction profile. Nevertheless, the success of this technique depends on the quality of the PCR product, being affected by the presence of spurious products. This study showed that a touchdown PCR enhance the quality of the amplification and reduced the presence of spurious products and is highly recommended in the design of a diagnostic test. Time constraints in this research prevented the evaluation of RFLP as a species diagnostic tool.

The COI has proven valuable in resolving the phylogenetic relationships in *Anopheles* species. However, it does not appear useful in all *Anopheles* species. One of the principal disadvantages with mtDNA is the presence of shared haplotypes

between sibling species. In the *Anopheles dirus* complex, species *An. dirus* and *An. baimaii* (previously known as *An. dirus* A and D respectively) share several haplotypes. Similarly, members of the *Anopheles gambiae* complex: *An. gambiae, An arabiensis* and *An. bwambae* show a large number of shared haplotypes (Besansky et al., 1997; Donnelly et al., 2001). No shared haplotypes were found among species of the Barbirostris Subgroup.

It has been suggested that Cytochrome Oxidase Subunit I be considered the standard region to identify animals by the Barcoding of Life Consortium. However, in addition to the disadvantages cited above, it has been found that the evolutionary rates in the structural regions of the COI can vary between species (Roe and Sperling, 2007). Hence, the value of COI to delimit species will vary. The present study also showed evidence of the variation in the evolutionary rates of these regions in different *Anopheles* species. Roe and Sperling (2007) maintained that Barcoding of life is likely to be more efficient in the identification of species that diverged earlier but appears less appropriate to distinguish sibling species. They recommended that care be taken to choose an appropriate region to analyse, to obtain reliable inferences. In this present study, COI appears useful to distinguish between different species within the Barbirostris Subgroup.

8.2 Population structure of species of the Barbirostris Subgroup

Based on the analysis of the COI gene region, this study provided evidence for population expansion of *Anopheles barbirostris* (clade III) in Thailand and of its deviation from mutation drift equilibrium (MDE). Due to the limited number of specimens and collection sites for the other clades, it was not possible to draw conclusions about the population structure of other clades.

In *Anopheles* mosquitoes, deviation from MDE and population expansion have been associated with their relationship to human populations. Donelly *et al.*(2002) maintained that primary vectors, including *Anopheles* species, share some attributes in common. These are 1) wide geographical distribution, 2) high local abundance, 3) dispersal and colonization ability, 4) adaptation to invade human-made

environments, 5) high anthropophily and 6) high susceptibility to parasites (Donnelly et al., 2002). The authors maintained that the agricultural revolution, between 5000 – 10 000 years ago, may have played an important role in the dispersal and specialization of *Anopheles* mosquitoes as vectors of malaria. In Africa, agricultural practices are though to contribute to the phenomenon of recent expansion in *Anopheles gambiae* in Africa (Coluzzi, 1982; Donnelly et al., 2001). One exception is *Anopheles darlingi*, a primary malaria vector in South America. The distribution of this vector is more likely to have been determined by climatic changes than human-mediated environmental change, since this species is less dependant in humans for blood meals or larval habitats (Mirabello and Conn, 2006b).

Donnelly et al. (2002) suggested that the major vectors of malaria have a shallow genetic structure (reflected by low F_{ST} values), with a limited effect of distance on differentiation. This review was based on empirical estimations of F_{ST} values in the principal vectors in Africa, South America and Asia, and culicine mosquitoes such as *Culex pipiens* and *Aedes aegypti* (Donnelly et al., 2002). The authors highlighted the need for more studies in secondary vectors to arrive at conclusions. In the present study, low F_{ST} values obtained within clade III also indicate a shallow genetic structure and show clear evidence of population expansion in Thailand. This is not consistent with the assumption that this species is not a primary vector of malaria or filariasis. However, this species/clade may be present in other countries in Southeast Asia as its identification at the molecular level has not been reported previously; therefore general assumptions about its genetic structure would require more extensive sampling.

The time since population expansion of *Anopheles barbirostris* (clade III) has been estimated from the data in this thesis to occur approximately 12 000 years ago, based on the assumptions of Powell (1986) (see section 4.3.7), later than the expansion of human populations, which is postulated to have occurred between 60 000 and 262 000 years ago (Rogers and Harpending, 1992; Slatkin and Hudson, 1991). Thus, it is likely that agricultural activities played a role in the evident population expansion found in *Anopheles barbirostris* (clade III) in Thailand.

Nonetheless, considering that this species utilizes a variety of oviposition sites (Harrison and Scanlon, 1975) and that females do not depend on human blood, agricultural practices may have played a secondary role in the distribution of this clade.

The limited number of specimens obtained in clade I (four from Kalimantan and six from Thailand) make it impossible to arrive at firm conclusions about the population structure of this clade. Specimens studied by Baskoro (2001) from Flores, Java and Sulawesi appear to be closely related to clade I, and this suggests a wide distribution of this clade in the Indonesian islands and adjacent continental regions. Relatively recent geological events may be responsible for the current distribution of clade I. During the glaciation periods, large quantities of the water were bound up in ice, the sea level fell between 100 and 200 meters and the Sunda shelf and the islands of Sumatra, Java and Bali became an extension of the mainland (Mayr, 1944 *In* O' Connor, 1980). Under these circumstances, the migration of continental species to the islands was facilitated and this may explain why clade I populations are similar in populations from Mae Hong Son, Thailand and Kalimantan, which are separated by ~ 3000 km.

8.3 Factors determining size and sequence of the ITS2 in species from the Barbirostris Subgroup

A large ITS2 (>1.5 kb) was found in all species within the Barbirostris Subgroup in this study. *Anopheles barbirostris*, with clades II and III, showed the largest ITS2 of ~1730 bp. This is the largest ITS2 reported from arthropods to date. The presence of repeats within the ITS2 is the reason for its outstanding size. The closely related *Anopheles bancroftii* group and *Anopheles hyrcanus* group have a much smaller ITS2 (< 600 bp) and hence can easily be differentiated through a simple PCR, followed by an electrophoresis in 1% agarose gel.

In chapter 5, a study of the repeat structure of ITS2 is presented. The central region of ITS2 is made up of four repeat sequences (c. 220 bp) of varying sequence similarity in different clades. Each of these repeats is a partial duplication of a 110

bp region. On the 3' end of the last of these four repeats in all Barbirostris Subgroup clades and specimens examined, was a sequence which is almost identical to a 27 bp region of the transposase found in transposable elements (TGEs) of other mosquitoes (Quesneville et al., 2006)

This is not a complete transposase, let alone a complete transposable element, but it may be assumed that this sequence is the relict of a once active TGE. It is possible that this may have played a role in duplication of the repeats within the ITS2 region, producing the increase on its size. Mobile elements are numerous in many animal species and are thought to play an important role in speciation (Kazazian, 2004), but whether they played any role in the evolution of the structure of the ITS2 region within the Barbirostris Subgroup is not known.

8.4 Implications of this study for vector control

Anopheles campestris has been considered as the third most anthropophilic *Anopheles* species in Southeast Asia (Reid, 1968). It has been incriminated as an important vector of filariasis and malaria in Malaysia. However, in Thailand, *An. campestris* was not thought to be a vector of human disease until very recently. Based on its abundance, anthropophilic behaviour and capacity to complete the sporogonic cycle of *Plasmodium vivax,* Apiwathnasorn (2002) concluded that *An. campestris* was an important vector in Sa Kaeo, Thailand. In areas where *An. campestris* is a malaria vector, *Plasmodium vivax* tends to be the most abundant malaria species (Reid, 1962).

Programmes to control *Anopheles campestris* have been successful because of its highly endophagic behaviour; it is readily controlled by house spraying. This species is apparently "one of the few vectors in the world that has behavioural traits that permit rapid elimination by chemical control methods" (Harrison and Scanlon, 1975). Confusion in the identification of this species could lead to misdirected control if there is confusion with other closely related species. The present study has found two species that would be identified in the field as *An. campestris,* the

anthropophilic clade V and the zoophilic clade IV. This would justify the development of simple PCR-based diagnostic tests to help distinguish these species.

An. barbirostris is considered an important vector of malaria and Brugian filariasis in Sulawesi, Flores and Timor (Atmosoedjono et al., 1977b; Lien et al., 1977; Reid, 1979; WHO, 2000), whereas it appears to be a non-vector species in other regions (Reid, 1962). A recent survey of the *Anopheles* fauna in northern Sumatra reported the presence of *An. barbirostris* and it was incriminated as a potential mosquito vector (Syafruddin et al., 2007). Variations in the vector status of this species could not be related to morphological characters in the past (Reid, 1979). With the development of molecular tools, it should eventually be possible to clarify the vector status of the species in the *An. barbirostris* complex.

Since clades III and IV appear to be predominantly zoophilic, their importance in the transmission of vector-borne diseases in humans may be limited. There was limited information on methods of collection for specimens of clade I and clade II, and no information on their feeding preferences. Clade I seems to be related to species studied by Baskoro (2001) in Flores, Java and Sulawesi in Indonesia. Baskoro showed that specimens from these localities were attracted to humans, particularly those from Flores and Sulawesi (see table 4.2 in Baskoro, 2001), however their zoophilic behaviour was not evaluated. Coincidently "vector forms" of *An. barbirostris* have been previously reported in these islands (Lien et al., 1977; Reid, 1979). Considering that the analysis of the ITS2 showed that clade I is more closely related to the anthropophilic clade V (*An. campestris*) than to other clades, it is possible that the "vector forms" of *An. barbirostris* belong in clade I. Nonetheless, to confirm this hypothesis, a more comprehensive study would be required.

8.5 Conclusions

Anopheles barbirostris Van der Wulp is a species complex. It comprises at least 3 different species identified as clade I, clade II and clade III. Clade III appears to be zoophilic and is widely distributed in Thailand. Clade III is also present in southern

China. There is limited information regarding the feeding preferences of clades I and II.

Clade I is the most widely distributed species. It is found in Thailand and in Kalimantan, Java, Flores and Sulawesi, in Indonesia. This clade may comprise 3 species X, W and Z, as named by Baskoro (2001). More exhaustive work is needed to determine the species status of this clade and populations therein.

Anopheles (clade IV) is a zoophilic species, with morphology that does not completely correspond to any species within the Barbirostris Group. This species is also distinct, at the molecular level, from both *An. barbirostris* and *An. campestris*.

Anopheles campestris has been characterized at the molecular level as clade V.

There is evidence of the population expansion of *Anopheles barbirostris* clade III in Thailand

8.6 Further studies

A weakness of this study was the limited information on the ecology and behaviour of the specimens collected, reflecting a more general problem when using specimens collected by others. Considering that the Barbirostris Subgroup comprises some important vectors of malaria and filariasis, future studies should incorporate comprehensive field work, to determine host preferences and biting behaviour. It will be important to identify more fully the vector and non-vector forms of *An. barbirostris*, particularly in Indonesia, where vector forms have been reported.

Further studies of the Barbirostris Subgroup will require specimens from a wider range of countries, particularly those where the Barbirostris Subgroup appears to be associated with malaria and filariasis transmission. Such studies should include other members of the Barbirostris Subgroup not examined in this study, *Anopheles donaldi, An. franciscoi, An. pollicaris* and *An. hodgkini*.

Crossing experiments have been carried out between *An. campestris* and *Anopheles barbirostris*, demonstrating the reproductive isolation of these species (Saeung et al., 2007). Crosses between the different members that comprise the *An. barbirostris* complex (clade I, II and III) with *Anopheles campestris* (clade V) and clade IV could help confirm their identity as separate species.

The true identity of *Anopheles barbirostris* Van der Wulp (1884) has to be established. In this study, two clades (I and II) are likely to be conspecific with this species, since they were found in Java. A comprehensive study of the specimens found in Mount Ardjoeno, the type locality, in Eastern Java is required.

REFERENCES

Abbott, C. L., M. C. Double, J. W. H. Trueman, A. Robinson, and A. Cockburn. 2005. An unusual source of apparent mitochondrial heteroplasmy: Duplicate mitochondrial control regions in thalassarche albatrosses. Molecular Ecology 14:3605-3613.

Abu Hassan, A., W. A. Rahman, M. Z. A. Rashid, M. R. Shahrem, and C. R. Adanan. 2001. Composition and biting activity of *anopheles* (diptera : Culicidae) attracted to human bait in a malaria endemic village in peninsular malaysia near the thailand border. Journal of Vector Ecology 26:70-75.

Agapow, P. M., O. R. P. Bininda-Emonds, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall, and A. Purvis. 2004. The impact of species concept on biodiversity studies. Quarterly Review of Biology 79:161-179.

Alvarez, I., and J. F. Wendel. 2003. Ribosomal its sequences and plant phylogenetic inference. Molecular Phylogenetics And Evolution 29:417-434.

Amstutz, H., P. Munz, W. D. Heyer, U. Leupold, and J. Kohli. 1985. Concerted evolution of transfer-rna genes - intergenic conversion among 3 unlinked serine transfer-rna genes in s-pombe. Cell 40:879-886.

Andersson, L. 1990. The driving force - species concepts and ecology. Taxon 39:375-382.

Apiwathnasorn, C., S. Prommongkol, Y. Samung, D. Limrat, and B. Rojruthai. 2002. Potential for *anopheles campestris* (diptera : Culicidae) to transmit malaria parasites in pa rai subdistrict (aranyaprathet, sa kaeo province), thailand. Journal Of Medical Entomology 39:583-586.

Atmosoedjono, S., F. Partono, D. T. Dennis, and Purnomo. 1977a. Anopheles-barbirostris (diptera-culicidae) as a vector of timor filaria on flores-island - preliminary-observations. Journal Of Medical Entomology 13:611-613.

Atmosoedjono, S., F. Partono, D. T. Dennis, and Purnomo. 1977b. *Anopheles barbirostris* (diptera-culicidae) as a vector of timor filaria on flores island preliminary observations. Journal of Medical Entomology 13:611-613.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography - the mitochondrial-DNA bridge between population-genetics and systematics. Annual Review of Ecology and Systematics 18:489-522.

Ayala, F. J., and J. R. Powell. 1972. Allozymes as diagnostic characters of sibling species of drosophila. Proceedings Of The National Academy Of Sciences Of The United States Of America 69:1094-&.

Baimai, V., R. G. Andre, B. A. Harrison, U. Kijchalao, and L. Panthusiri. 1987. Crossing and chromosomal evidence for two additional sibling species within the taxon *anopheles dirus* peyton and harrison (diptera: Culicidae) in thailand. Proceedings of the Entomological Society of Washington 89:157-166

Baimai, V., R. Rattanarithikul, and U. Kijchalao. 1995. Metaphase karyotypes of *anopheles* of thailand and southeast-asia .4. The barbirostris and umbrosus species groups, subgenus *anopheles* (diptera, culicidae). Journal Of The American Mosquito Control Association 11:323-328.

Ballard, J. W. O., G. J. Olsen, D. P. Faith, W. A. Odgers, D. M. Rowell, and P. W. Atkinson. 1992. Evidence from 12s ribosomal-rna sequences that onychophorans are modified arthropods. Science 258:1345-1348.

Ballard, J. W. O., and D. M. Rand. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Annual Review of Ecology Evolution and Systematics 36:621-642.

Ballinger-Crabtree, M. E., W. C. Black, and B. R. Miller. 1992. Use of genetic polymorphisms detected by the random-amplified polymorphic DNA-polymerase chain-reaction (rapd-pcr) for differentiation and identification of *aedes aegypti* subspecies and populations. American Journal of Tropical Medicine And Hygiene 47:893-901.

Barker, S. C. 1998. Distinguishing species and populations of rhipicephaline ticks with its 2 ribosomal rna. Journal of Parasitology 84:887-892.

Baskoro, T. 2001. Cryptic species within the *anopheles barbirostris* van der wulp, 1884 inferred from nuclear and mitochondrial gene sequence variation. Pages 243 *in* Liverpool School of Tropical Medicine University of Liverpool Liverpool.

Bauzer, L., N. A. Souza, R. D. C. Maingon, and A. A. Peixoto. 2007. *Lutzomyia longipalpis* in brazil: A complex or a single species? A mini-review. Memorias Do Instituto Oswaldo Cruz 102:1-12.

Bazin, E., S. Glemin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312:570-572.

Beebe, N. W., J. T. Ellis, R. D. Cooper, and A. Saul. 1999. DNA sequence analysis of the ribosomal DNA its2 region for the *anopheles punctulatus* group of mosquitoes. Insect Molecular Biology 8:381-390.

Beebe, N. W., J. Maung, A. F. van den Hurk, J. T. Ellis, and R. D. Cooper. 2001. Ribosomal DNA spacer genotypes of the *anopheles bancroftii* group (diptera : Culicidae) from australia and papua new guinea. Insect Molecular Biology 10:407-413.

Bellon, H., R. C. Maury, Sutanto, R. Soeria-Atmadja, J. Cotten, and M. Polve. 2004. 65 my-long magmatic activity in sumatra (indonesia), from paleocene to present. Bulletin De La Societe Geologique De France 175:61-72.

Benesh, D. P., T. Hasu, L. R. Suomalainen, E. T. Valtonen, and M. Tiirola. 2006. Reliability of mitochondrial DNA in an acanthocephalan: The problem of pseudogenes. International Journal For Parasitology 36:247-254.

Bensasson, D., D. X. Zhang, and G. M. Hewitt. 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Molecular Biology And Evolution 17:406-415.

Bermingham, E., T. Lamb, and J. C. Avise. 1986. Size polymorphism and heteroplasmy in the mitochondrial-DNA of lower-vertebrates. Journal Of Heredity 77:249-252.

Besansky, N. J., T. Lehmann, G. T. Fahey, D. Fontenille, L. E. O. Braack, W. A. Hawley, and F. H. Collins. 1997. Patterns of mitochondrial variation within and between african malaria vectors, *anopheles gambiae* and *an. Arabiensis*, suggest extensive gene flow. Genetics 147:1817-1828.

Birky, C. W. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. Annual Review Of Genetics 35:125-148.

Birungi, J., and L. E. Munstermann. 2002. Genetic structure of aedes albopictus (diptera : Culicidae) populations based on mitochondrial nd5 sequences:

Evidence for an independent invasion into brazil and united states. Annals of the Entomological Society of America 95:125-132.
Black, W. C., and J. Piesman. 1994. Phylogeny of hard-tick and soft-tick taxa (acari, ixodida) based on mitochondrial 16s rdna sequences. Proceedings of the National Academy of Sciences of the United States of America 91:10034-10038.
Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Research 27:1767-1780.
Brelsfoard, C. L., G. N. Fritz, and R. Rodriguez. 2006. Sequence analysis of the rdna internal transcribed spacer 2 and polymerase chain reaction identification of anopheles fluminensis (diptera : Culicidae : Anopheles) in bolivia. Journal of Medical Entomology 43:460-466.
Brown, D. D., P. C. Wensink, and E. Jordan. 1972. Comparison of ribosomal dnas of xenopus-laevis and xenopus-mulleri - evolution of tandem genes. Journal Of Molecular Biology 63:57-&.
Brown, T. A. 1999. Genomes. BIOS Scientific Publishers Ltd, Oxford.
Caccone, A., B. A. Garcia, and J. R. Powell. 1996. Evolution of the mitochondrial DNA control region in the anopheles gambiae complex. Insect Molecular Biology 5:51-59.
Cameron, S. L., C. L. Lambkin, S. C. Barker, and M. F. Whiting. 2007. A mitochondrial genome phylogeny of diptera: Whole genome sequence data accurately resolve relationships over broad timescales with high precision. Systematic Entomology 32:40-59.
Caterino, M. S., S. Cho, and F. A. H. Sperling. 2000. The current state of insect molecular systematics: A thriving tower of babel. Annual Review Of Entomology 45:1-54.
Chadee, D. D., P. Kittayapong, A. C. Morrison, and W. J. Tabachnick. 2007. A breakthrough for global public health. Science 316:1703-1704.
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. Systematic Entomology 28:57-69.
Chen, B., R. E. Harbach, and R. K. Butlin. 2004. Genetic variation and population structure of the mosquito anopheles jeyporiensis in southern china. Molecular Ecology 13:3051-3056.
Chinnery, P. F., D. C. Samuels, J. Elson, and D. M. Turnbull. 2002. Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: Is there a common mechanism? Lancet 360:1323-1325.
Choi, Y. J., S. B. Hong, and H. D. Shin. 2007. Extreme size and sequence variation in the its rdna of bremia lactucae. Mycopathologia 163:91-95.
Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial-DNA molecule of drosophila-yakuba - nucleotide-sequence, gene organization, and genetic-code. Journal of Molecular Evolution 22:252-271.
Clement, M., D. Posada, and K. A. Crandall. 2000. Tcs: A computer program to estimate gene genealogies. Molecular Ecology 9:1657-1659.
Cockburn, A. F., S. E. Mitchell, and J. A. Seawright. 1990. Cloning of the mitochondrial genome of anopheles-quadrimaculatus. Archives of Insect Biochemistry and Physiology 14:31-36.
Collet, T., K. M. Ferreira, M. C. Arias, A. E. E. Soares, and M. A. Del Lama. 2006. Genetic structure of africanized honeybee populations (apis mellifera l.)

from brazil and uruguay viewed through mitochondrial DNA coi-coii patterns. Heredity 97:329-335.

Collins, F. H., M. A. Mendez, M. O. Rasmussen, P. C. Mehaffey, N. J. Besansky, and V. Finnerty. 1987. A ribosomal-rna gene probe differentiates member species of the anopheles-gambiae complex. American Journal of Tropical Medicine and Hygiene 37:37-41.

Collins, F. H., and S. M. Paskewitz. 1996. A review of the use of ribosomal DNA (rdna) to differentiate among cryptic anopheles species. Insect Molecular Biology 5:1-9.

Coluzzi, M. 1982. Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. Pages 143-153 *in* Bargozzi c. Mechanism of speciation (A. R. Liss, ed.), New York.

Coluzzi, M., A. Sabatini, A. della Torre, M. A. Di Deco, and V. Petrarca. 2002. A polytene chromosome analysis of the anopheles gambiae species complex. Science 298:1415-1418.

Coluzzi, M., A. Sabatini, V. Petrarca, and M. A. Dideco. 1979. Chromosomal differentiation and adaptation to human environments in the anopheles-gambiae complex. Transactions of the Royal Society of Tropical Medicine and Hygiene 73:483-497.

Cornel, A. J., C. H. Porter, and F. H. Collins. 1996. Polymerase chain reaction species diagnostic assay for anopheles quadrimaculatus cryptic species (diptera: Culicidae) based on ribosomal DNA its2 sequences. Journal of Medical Entomology 33:109-116.

Cracraft, J. 1989. Speciation and it ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. Pages 28-59 *in* Speciation and its consequences (E. D. Otte and J.A. Endler, ed.) Sinauer, Sunderland.

Crandall, K. A., and A. R. Templeton. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics 134:959-969.

Crozier, R. H., Y. C. Crozier, and A. G. Mackinlay. 1989. The co-i and co-ii region of honeybee mitochondrial-DNA - evidence for variation in insect mitochondrial evolutionary rates. Molecular Biology and Evolution 6:399-411.

Davidson, G., H. E. Paterson, M. Coluzzi, G. F. Mason, and D. W. Micks. 1967. Genetics of insect vectors of disease*in* Genetics of insect vectors of disease (J. W. Wright, and R. Pal, eds.). World Health Organization, Amsterdam.

De Oliveira, M. T., A. M. L. De Azeredo-Espin, and A. C. Lessinger. 2005. Evolutionary and structural analysis of the cytochrome c oxidase subunit i (coi) gene from haematobia irritans, stomoxys calcitrans and musca domestica (diptera : Muscidae) mitochondrial DNA. DNA Sequence 16:156-160.

de Queiroz, K. 2005. Ernst mayr and the modern concept of species. Proceedings Of The National Academy Of Sciences Of The United States Of America 102:6600-6607.

Desalle, R., J. Gatesy, W. Wheeler, and D. Grimaldi. 1992. DNA-sequences from a fossil termite in oligomiocene amber and their phylogenetic implications. Science 257:1933-1936.

Djadid, N. D., S. Gholizadeh, E. Tafsiri, R. Romi, M. Gordeev, and S. Zakeri. 2007. Molecular identification of palearctic members of anopheles maculipennis in northern iran. Malaria Journal 6.

Dondorp, A. M., P. N. Newton, M. Mayxay, W. Van Damme, F. M. Smithuis, S. Yeung, A. Petit, A. J. Lynam, A. Johnson, T. T. Hien, R. McGready, J. J. Farrar, S. Looareesuwan, N. P. J. Day, M. D. Green, and N. J. White. 2004. Fake antimalarials in southeast asia are a major impediment to malaria control: Multinational cross-sectional survey on the prevalence of fake antimalarials. Tropical Medicine & International Health 9:1241-1246.

Donnelly, M. J., M. C. Licht, and T. Lehmann. 2001. Evidence for recent population expansion in the evolutionary history of the malaria vectors anopheles arabiensis and anopheles gambiae. Molecular Biology and Evolution 18:1353-1364.

Donnelly, M. J., F. Simard, and T. Lehmann. 2002. Evolutionary studies of malaria vectors. Trends in Parasitology 18:75-80.

doValle, L. S., and A. M. L. deAzeredoEspin. 1995. Mitochondrial DNA variation in two brazilian populations of cochliomyia macellaria (diptera: Calliphoridae). Brazilian Journal Of Genetics 18:521-526.

Dover, G. 1982. Molecular drive - a cohesive mode of species evolution. Nature 299:111-117.

Dusfour, I., Y. M. Linton, A. Cohuet, R. E. Harbach, V. Baimai, H. D. Trung, C. M. Seng, A. Matusop, and S. Manguin. 2004. Molecular evidence of speciation between island and continental populations of anopheles (cellia) sundaicus (diptera : Culicidae), a principal malaria vector taxon in southeast asia. Journal of Medical Entomology 41:287-295.

Ekrem, T., and E. Willassen. 2004. Exploring tanytarsini relationships (diptera : Chironomidae) using mitochondrial coii gene sequences. Insect Systematics & Evolution 35:263-276.

Engels, W. R. 1992. The origin of p elements in drosophila-melanogaster. Bioessays 14:681-686.

Engstrom, T. N., H. B. Shaffer, and W. P. McCord. 2004. Multiple data sets, high homoplasy, and the phylogeny of softshell turtles (testudines : Trionychidae). Systematic Biology 53:693-710.

Erpenbeck, D., J. N. A. Hooper, and G. Worheide. 2006. Co1 phylogenies in diploblasts and the 'barcoding of life' - are we sequencing a suboptimal partition? Molecular Ecology Notes 6:550-553.

Fairley, T. L., C. W. Kilpatrick, and J. E. Conn. 2005. Intragenomic heterogeneity of internal transcribed spacer rdna in neotropical malaria vector anopheles aquasalis (diptera : Culicidae). Journal of Medical Entomology 42:795-800.

Fantini, B. 1994. Anophelism without malaria: An ecological and epidemiological puzzle. Parassitologia 36:83-106.

Felsenstein, J. 1985. Confidence-limits on phylogenies - an approach using the bootstrap. Evolution 39:783-791.

Foley, D. H., R. C. Wilkerson, R. D. Cooper, M. E. Volovsek, and J. H. Bryan. 2007. A molecular phylogeny of anopheles annulipes (diptera : Culicidae) sensu lato: The most species-rich anopheline complex. Molecular Phylogenetics and Evolution 43:283-297.

Fritz, G. N. 1998. Sequence analysis of the rdna internal transcribed spacer 2 of five species of south american human malaria mosquitoes. DNA Sequence 8:215-221.

Fritz, G. N., J. Conn, A. Cockburn, and J. Seawright. 1994. Sequence-analysis of the ribosomal DNA internal transcribed spacer-2 from populations of anopheles nuneztovari (diptera, culicidae). Molecular Biology and Evolution 11:406-416.

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

Ganley, A. R. D., and T. Kobayashi. 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rdna repeat variation revealed by whole-genome shotgun sequence data. Genome Research 17:184-191.

Gao, Q., N. W. Beebe, and R. D. Cooper. 2004. Molecular identification of the malaria vectors anopheles anthropophagus and anopheles sinensis (diptera : Culicidae) in central china using polymerase chain reaction and appraisal of their position within the hyrcanus group. Journal of Medical Entomology 41:5-11.

Gerbi, S. A. 1985. Evolution of ribosomal DNA. Pages 419-517 *in* Molecular evolutionary genetics (R.J.MacIntyre, ed.) Plenum, New York.

Gorokhova, E., T. E. Dowling, L. J. Weider, T. J. Crease, and J. J. Elser. 2002. Functional and ecological significance of rdna intergenic spacer variation in a clonal organism under divergent selection for production rate. Proceedings Of The Royal Society Of London Series B-Biological Sciences 269:2373-2379.

Goswami, G., K. Raghavendra, N. Nanda, S. K. Gakhar, and S. K. Subbarao. 2005. Pcr-rflp of mitochondrial cytochrome oxidase subunit ii and its2 of ribosomal DNA: Markers for the identification of members of the anopheles culicifacies complex (diptera : Culicidae). Acta Tropica 95:92-99.

Hackett, L. W., and A. Misiroli. 1935. The varieties of anopheles maculipennis and their relation to the distribution of malaria in europe. Rivista di Malariologia 14:45-109.

Han, H. Y., and K. E. Ro. 2005. Molecular phylogeny of the superfamily tephritoidea (insecta : Diptera): New evidence from the mitochondrial 12s 16s, and coii genes. Molecular Phylogenetics and Evolution 34:416-430.

Harbach, R. E. 2003. Mosquito systematics: From organisms to molecules - a tribute to kenneth l. Knight. Journal of the American Mosquito Control Association 19:452-460.

Harbach, R. E. 2004. The classification of genus anopheles (diptera : Culicidae): A working hypothesis of phylogenetic relationships. Bulletin Of Entomological Research 94:537-553.

Harbach, R. E., E. Parkin, B. Chen, and R. K. Butlin. 2006. Anopheles (cellia) minimus theobald (diptera : Culicidae): Neotype designation, characterization, and systematics. Proceedings of the Entomological Society of Washington 108:198-209.

Harpending, H. C., S. T. Sherry, A. R. Rogers, and M. Stoneking. 1993. The genetic-structure of ancient human-populations. Current Anthropology 34:483-496.

Harrison, B., and J. M. Klein. 1975. The revisited list of the *anopheles* of cambodia. Mosquito Systematics 7:9-12.

Harrison, B., and J. Scanlon. 1975. The subgenus anopheles in thailand (diptera: Culicidae), 1 edition. Ann Arbor.

Harrison, R. G. 1989. Animal mitochondrial-dna as a genetic-marker in population and evolutionary biology. Trends In Ecology & Evolution 4:6-11.

Hebert, P. D. N., S. Ratnasingham, and J. R. deWaard. 2003. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London Series B-Biological Sciences 270:S96-S99.

Hemingway, J. 2004. Taking aim at mosquitoes. Nature 430:936-936.

Herrera, F., L. Urdaneta, J. Rivero, N. Zoghbi, J. Ruiz, G. Carrasquel, J. A. Martinez, M. Pernalete, P. Villegas, A. Montoya, Y. Rubio-Palis, and E. Rojas. 2006. Population genetic structure of the dengue mosquito aedes aegypti in venezuela. Memorias Do Instituto Oswaldo Cruz 101:625-633.

Hey, J. 2001. The mind of the species problem. Trends In Ecology & Evolution 16:326-329.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA - molecular evolution and phylogenetic inference. Quarterly Review of Biology 66:410-453.

Hlinka, O., A. Murrell, and S. C. Barker. 2002. Evolution of the secondary structure of the rrna internal transcribed spacer 2 (its2) in hard ticks (ixodidae, arthropoda). Heredity 88:275-279.

Hodgkin, E. P., and R. S. Johnston. 1935. Malaria at batu gajah, perak: Transmitted by *anopheles barbirostris* van der wulp. Bulletin of the Institute of Medical Research 1:19.

Hudson, R. R., D. D. Boos, and N. L. Kaplan. 1992. A statistical test for detecting geographic subdivision. Molecular Biology and Evolution 9:138-151.

Hung, L. Q., P. J. de Vries, P. T. Giao, N. V. Nam, T. Q. Binh, M. T. Chong, N. Quoc, T. N. Thanh, L. N. Hung, and P. A. Kager. 2002. Control of malaria: A successful experience from viet nam. Bulletin Of The World Health Organization 80:660-666.

Hwang, U. W. 2007. Revisited its2 phylogeny of anopheles (anopheles) hyrcanus group mosquitoes: Reexamination of unidentified and misidentified its2 sequences. Parasitology Research 101:885-894.

Hwang, U. W., T. S. Yong, and H. I. Ree. 2004. Molecular evidence for synonymy of anopheles yatsushiroensis and an. Pullus. Journal of the American Mosquito Control Association 20:99-104.

Imbernon, J. 1999. Changes in agricultural practice and landscape over a 60-year period in north lampung, sumatra. Agriculture Ecosystems & Environment 76:61-66.

Isaac, N. J. B., J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: Its influence on macroecology and conservation. Trends In Ecology & Evolution 19:464-469.

Jongwutiwes, S., C. Putaporntip, T. Iwasaki, T. Sata, and H. Kanbara. 2004. Naturally acquired plasmodium knowlesi malaria in human, thailand. Emerging Infectious Diseases 10:2211-2213.

Junkum, A., N. Komalamisra, A. Jitpakdi, N. Jariyapan, G. S. Min, M. H. Park, K. H. Cho, P. Somboon, P. A. Bates, and W. Choochote. 2005. Evidence to support two conspecific cytological races of anopheles aconitus in thailand. Journal of Vector Ecology 30:213-224.

Kampen, H. 2005. The its2 ribosomal DNA of anopheles beklemishevi and further remarks on the phylogenetic relationships within the anopheles maculipennis group of species (diptera : Culicidae). Parasitology Research 97:118-128.

Kann, L. M., E. B. Rosenblum, and D. M. Rand. 1998. Aging, mating, and the evolution of mtdna heteroplasmy in drosophila melanogaster. Proceedings

Of The National Academy Of Sciences Of The United States Of America 95:2372-2377.
Kaur, S. D. 2005. Unusual sensilla on the antennae of female anopheles barbirostris. Journal of the American Mosquito Control Association 21:464-465.
Kaya, N., and K. Isik. 2006. Segregation and linkage relationships of allozymes in pinus brutia ten. Silvae Genetica 55:155-160.
Kazazian, H. H. 2004. Mobile elements: Drivers of genome evolution. Science 303:1626-1632.
Keller, I., I. C. Chintauan-Marquier, P. Veltsos, and R. A. Nichols. 2006. Ribosomal DNA in the grasshopper podisma pedestris: Escape from concerted evolution. Genetics 174:863-874.
Kijewski, T. K., M. Zbawicka, R. Vainola, and R. Wenne. 2006. Introgression and mitochondrial DNA heteroplasmy in the baltic populations of mussels mytilus trossulus and mytilus edulis. Marine Biology 149:1371-1385.
Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. Journal of Molecular Evolution 16:111-120.
Kmiec, B., M. Woloszynska, and H. Janska. 2006. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Current Genetics 50:149-159.
Kolaczinski, J., and J. Webster. 2003. Malaria control in complex emergencies: The example of east timor. Tropical Medicine & International Health 8:48-55.
Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. Annual Review of Genomics and Human Genetics 1:539-559.
krThines, M., H. Komjati, and O. Spring. 2005. Exceptional length of its in plasmopara halstedii is due to multiple repetitions in the its-2 region. European Journal of Plant Pathology 112:395-398.
Krzywinski, J., and N. J. Besansky. 2003. Molecular systematics of anopheles: From subgenera to subpopulations. Annual Review Of Entomology 48:111-139.
Krzywinski, J., R. C. Wilkerson, and N. J. Besansky. 2001. Evolution of mitochondrial and ribosomal gene sequences in anophelinae (diptera : Culicidae): Implications for phylogeny reconstruction. Molecular Phylogenetics and Evolution 18:479-487.
Kunz, W. 2002. When is a parasite species a species? Trends in Parasitology 18:121-124.
Lehmann, T., W. A. Hawley, H. Grebert, M. Danga, F. Atieli, and F. H. Collins. 1999. The rift valley complex as a barrier to gene flow for anopheles gambiae in kenya. Journal of Heredity 90:613-621.
Lehr, M. A., C. W. Kilpatrick, R. C. Wilkerson, and J. E. Conn. 2005. Cryptic species in the anopheles (nyssorhynchus) albitarsis (diptera : Culicidae) complex: Incongruence between random amplified polymorphic DNA-polymerase chain reaction identification and analysis of mitochondrial DNA coi gene sequences. Annals of the Entomological Society of America 98:908-917.
Li, C., J. S. Lee, J. L. Groebner, H. C. Kim, T. A. Klein, M. L. O'Guinn, and R. C. Wilkerson. 2005. A newly recognized species in the anopheles hyrcanus group and molecular identification of related species from the republic of south korea (diptera : Culicidae). Zootaxa:1-8.
Lien, J. C., B. A. Kawengian, F. Partono, B. Lami, and J. H. Cross. 1977. Brief survey of mosquitos of south sulawesi, indonesia, with special reference to

identity of anopheles-barbirostris (diptera-culicidae) from margolembo area. Journal Of Medical Entomology 13:719-727.

Limrat, D., B. Rojruthai, C. Apiwathanasorn, Y. Samung, and S. Prommongkol. 2001. Anopheles barbirostris/campestris as a probable vector of malaria in aranyaprathet, sa kaeo province. SOUTHEAST ASIAN JOURNAL OF TROPICAL MEDICINE AND PUBLIC HEALTH 32:739-744.

Linton, Y. M., L. Smith, G. Koliopoulos, A. Samanidou-Voyadjoglou, A. K. Zounos, and R. E. Harbach. 2003. Morphological and molecular characterization of anopheles (anopheles) maculipennis meigen, type species of the genus and nominotypical member of the maculipennis complex. Systematic Entomology 28:39-55.

Looareesuwan, S., J. D. Chulay, C. J. Canfield, and D. B. A. Hutchinson. 1999. Malarone (r) (atovaquone and proguanil hydrochloride): A review of its clinical development for treatment of malaria. American Journal Of Tropical Medicine And Hygiene 60:533-541.

Looareesuwan, S., T. Harinasuta, and Chongsuphajaisiddhi. 1992. Drug resistant malaria, with special reference to thailand. Southeast Asian J Trop Med Public Health. 23:621-34.

Lopez, J. V., N. Yuhki, R. Masuda, W. Modi, and S. J. Obrien. 1994. Numt, a recent transfer and tandem amplification of mitochondrial-dna to the nuclear genome of the domestic cat. Journal Of Molecular Evolution 39:174-190.

Loxdale, H. D., and G. Lushai. 1998. Molecular markers in entomology. Bulletin Of Entomological Research 88:577-600.

Lunt, D. H., D. X. Zhang, J. M. Szymura, and G. M. Hewitt. 1996. The insect cytochrome oxidase i gene: Evolutionary patterns and conserved primers for phylogenetic studies. Insect Molecular Biology 5:153-165.

Ma, Y. J., S. Z. Li, and J. N. Xu. 2006. Molecular identification and phylogeny of the maculatus group of anopheles mosquitoes (diptera : Culicidae) based on nuclear and mitochondrial DNA sequences. Acta Tropica 99:272-280.

Ma, Y. J., and J. N. Xu. 2005. The hyrcanus group of anopheles (anopheles) in china (diptera : Culicidae): Species discrimination and phylogenetic relationships inferred by ribosomal DNA internal transcribed spacer 2 sequences. Journal of Medical Entomology 42:610-619.

Macdonald, C., and H. D. Loxdale. 2004. Molecular markers to study population structure and dynamics in beneficial insects (predators and parasitoids). International Journal Of Pest Management 50:215-224.

Magnacca, K. N., and B. N. Danforth. 2006. Evolution and biogeography of native hawaiian hylaeus bees (hymenoptera : Colletidae). Cladistics 22:393-411.

Mahon, R. J., C. A. Green, and R. H. Hunt. 1976. Diagnostic allozymes for routine identification of adults of anopheles-gambiae complex (diptera, culicidae). Bulletin Of Entomological Research 66:25-&.

Mallet, J. 2004. Species problem solved 100 years ago. Nature 430:503-503.

Mallet, J. 2006. Species concepts. Pages 367-373 *in* Evolutionary genetics: Concepts and case studies (C. W. F. a. J. B. Wolf, ed.) Oxford University Press, Oxford.

Manguin, S., P. Kengne, L. Sonnier, R. E. Harbach, V. Baimai, H. D. Trung, and M. Coosemans. 2002. Scar markers and multiplex pcr-based identification of isomorphic species in the anopheles dirus complex in southeast asia. Medical and Veterinary Entomology 16:46-54.

Manh, N. D., T. D. Hinh, R. E. Harbach, J. Elphick, and Y. M. Linton. 2000. A new species of the hyrcanus group of anopheles, subgenus anopheles, a secondary vector of malaria in coastal areas of southern vietnam. Journal of the American Mosquito Control Association 16:189-198.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning. A laboratory manual. Cold Spring Harbor, New York.

Manonmani, A., H. Townson, T. Adeniran, P. Jambulingam, S. Sahu, and T. Vijayakumar. 2001. Rdna-its2 polymerase chain reaction assay for the sibling species of anopheles fluviatilis. Acta Tropica 78:3-9.

Marinucci, M., R. Romi, P. Mancini, M. Di Luca, and C. Severini. 1999. Phylogenetic relationships of seven palearctic members of the maculipennis complex inferred from its2 sequence analysis. Insect Molecular Biology 8:469-480.

Mayr, E. 1942. Systematics and the origin of the species, New York.

Mayr, E. 1991. The principles of systematic zoology 2nd edition. McGraw-Hill, New York.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the adh locus in drosophila. Nature 351:652-654.

Meyer, A. 1994. Shortcomings of the cytochrome-b gene as a molecular marker. Trends In Ecology & Evolution 9:278-280.

Miller, B. R., M. B. Crabtree, and H. M. Savage. 1997. Phylogenetic relationships of the culicomorpha inferred from 18s and 5.8s ribosomal DNA sequences (diptera: Nematocera). Insect Molecular Biology 6:105-114.

Min, G. S., W. Choochote, A. Jitpakdi, S. J. Kim, W. Kim, J. Jung, and A. Junkum. 2002. Intraspecific hybridization of anopheles sinensis (diptera : Culicidae) strains from thailand and korea. Molecules and Cells 14:198-204.

Mirabello, L., and J. E. Conn. 2006a. Molecular population genetics of the malaria vector anopheles darlingi in central and south america. Heredity 96:311-321.

Mirabello, L., and J. E. Conn. 2006b. Molecular population genetics of the malaria vector anopheles darlingi in central and south america (vol 97, pg 438, 2006). Heredity 97:438-438.

Miya, M., K. Saitoh, R. Wood, M. Nishida, and R. L. Mayden. 2006. New primers for amplifying and sequencing the mitochondrial nd4/nd5 gene region of the cypriniformes (actinopterygii : Ostariophysi). Ichthyological Research 53:75-81.

Molina-Cruz, A., A. M. P. De Merida, K. Mills, F. Rodriguez, C. Schoua, M. M. Yurrita, E. Molina, M. Palmieri, and W. C. Black. 2004. Gene flow among anopheles albimanus populations in central america, south america, and the caribbean assessed by microsatellites and mitochondrial DNA. American Journal of Tropical Medicine and Hygiene 71:350-359.

Moorhouse, D. E. a. W., R.H. . 1965. Studies on malayan vectors of malaria; methods of trapping, and obesrvations on biting cycles. 1:359-370.

Morel, F., M. Renoux, and S. Alziari. 2006. Mitochondrial biochemical activities and heteroplasmy evolution in established d-subobscura cell line. In Vitro Cellular & Developmental Biology-Animal 42:201-207.

Mousset, S., N. Derome, and M. Veuille. 2004. A test of neutrality and constant population size based on the mismatch distribution. Molecular Biology and Evolution 21:724-731.

Murrell, A., N. J. H. Campbell, and S. C. Barker. 2001. Recurrent gains and losses of large (84-109 bp) repeats in the rdna internal transcribed spacer 2 (its2) of rhipicephaline ticks. Insect Molecular Biology 10:587-596.

Narang, S. K., T. A. Klein, O. P. Perera, J. B. Lima, and A. T. Tang. 1993. Genetic-evidence for the existence of cryptic species in the anopheles-albitarsis complex in brazil - allozymes and mitochondrial-dna restriction-fragment-length-polymorphisms. Biochemical Genetics 31:97-112.

Nardi, F., A. Carapelli, P. P. Fanfciulli, R. Dallai, and F. Frati. 2001. The complete mitochondrial DNA sequence of the basal hexapod tetrodontophora bielanensis: Evidence for heteroplasmy and trna translocations (vol 18, pg 1293, 2001). Molecular Biology And Evolution 18:2139-2139.

Naylor, G. J. P., and W. M. Brown. 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Systematic Biology 47:61-76.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America 70:3321-3323.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press Oxford.

Nei, M., and W. H. Li. 1979. Mathematical-model for studying genetic-variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America 76:5269-5273.

Nei, M., and A. P. Rooney. 2005. Concerted and birth-and-death evolution of multigene families. Annual Review Of Genetics 39:121-152.

NOAA. 2007. Weather conditions indonesia/new guinea(http://www.cpc.noaa.gov/products/assessments/assess_97/indo.html, ed.).

Noor, M. A. F. 2002. Is the biological species concept showing its age? Trends In Ecology & Evolution 17:153-154.

O'Brochta, D. A., R. A. Subramanian, J. Orsetti, E. Peckham, N. Nolan, P. Arensburger, P. W. Atkinson, and D. J. Charlwood. 2006. Hat element population genetics in anopheles gambiae s.L. In mozambique. Genetica 127:185-198.

O'Connor, C. T. 1980. The *anopheles hyrcanus* group in indonesia. Mosquito Systematics 12:293-305.

O'Loughlin, S. M., P. Somboom, and C. Walton. 2007. High levels of population structure caused by habitat islands in the malarial vector anopheles scanloni. Heredity 99:31-40.

Oshaghi, M. A., F. Yaaghoobi, and M. R. Abaie. 2006. Pattern of mitochondrial DNA variation between and within anopheles stephensi (diptera : Culicidae) biological forms suggests extensive gene flow. Acta Tropica 99:226-233.

Page, R., and E. Holmes. 1998. Molecular evolution. A phylogenetic approach. Blackwell Science, Oxford.

Park, S. J., W. Choochote, A. Jitpakdi, A. Junkum, S. J. Kim, N. Jariyapan, J. W. Park, and G. S. Min. 2003. Evidence for a conspecific relationship between two morphologically and cytologically different forms of korean anopheles pullus mosquito. Molecules and Cells 16:354-360.

Parr, R. L., J. Maki, B. Reguly, G. D. Dakubo, A. Aguirre, R. Wittock, K. Robinson, J. P. Jakupciak, and R. E. Thayer. 2006. The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. Bmc Genomics 7.

Paskewitz, S. M., and F. H. Collins. 1990. Use of the polymerase chain-reaction to identify mosquito species of the anopheles-gambiae complex. Medical and Veterinary Entomology 4:367-373.

Paskewitz, S. M., D. M. Wesson, and F. H. Collins. 1993. The internal transcribed spacers of ribosomal DNA in five members of the anopheles gambiae species complex. Insect Molecular Biology 4:247-257.

Peyton, E. L. 1989. A new classification for the leucosphyrus group of. Anopheles (cellia). Mosquito Systematics 21:197 205.

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. Medical and Veterinary Entomology 17:423-428.

Porter, C. H., and F. H. Collins. 1991. Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species anopheles-freeborni and anopheles-hermsi (diptera, culicidae). American Journal of Tropical Medicine and Hygiene 45:271-279.

Powell, J. R., A. Caccone, G. D. Amato, and C. Yoon. 1986. Rates of nucleotide substitution in drosophila mitochondrial-DNA and nuclear-DNA are similar. Proceedings of the National Academy of Sciences of the United States of America 83:9090-9093.

Prokopowich, C. D., T. R. Gregory, and T. J. Crease. 2003. The correlation between rdna copy number and genome size in eukaryotes. Genome 46:48-50.

Quesneville, H., D. Nouaud, and D. Anxolabehere. 2006. P elements and mite relatives in the whole genome sequence of anopheles gambiae. Bmc Genomics 7.

Rao, R. 1981. The anophelines of india. council of Medical Research, New Dehli.

Rathod, P. K., T. McErlean, and P. C. Lee. 1997. Variations in frequencies of drug resistance in plasmodium falciparum. Proceedings Of The National Academy Of Sciences Of The United States Of America 94:9389-9393.

Rattanarithikul, R., B. Harrison, and P. Panthusiri. 2005. Ilustrated keys to the mosquitoes of thailand i. Background; geographic distribution; list of genera, subgenera, and species; and a key to the genera. The Southeast Asian Journal of Tropical Medicine and Public Health 36:7.

Reid, J. A. 1942. A preliminary note in malayan forms of *an. Barbirostris* Journal of the Malaya Branch of the British Medical Association 5:71-72.

Reid, J. A. 1953. The anopheles hyrcanus group in south-east asia (diptera: Culicidae). Bulletin of Entomological Research 44.

Reid, J. A. 1962. The *anopheles barbirostris* group (diptera, culicidae). Bulletin of Entomological Research 53:1-57.

Reid, J. A. 1968. Anophelines of malaya and borneo. Studies from the institute for medical research malaysia. Staples Printers Limited, London.

Reid, J. A., Harrison, B.A. Atmosoedjono S. 1979. Variation and vector status in *anopheles barbirostris*. Mosquito Systematics 11:235-250.

Reiff, D. M., A. Kaneko, G. Taleo, M. Amos, and J. K. Lum. 2007. Population structure and gene flow of anopheles farauti s.S. (diptera : Culicidae) among ten sites on five islands of vanuatu: Implications for malaria control. Journal of Medical Entomology 44:601-607.

Richly, E., and D. Leister. 2004. Numts in sequenced eukaryotic genomes. Molecular Biology And Evolution 21:1081-1084.

Rodriguez-Perez, M. A., C. A. Nunez-Gonzalez, C. Lizarazo-Ortega, A. Sanchez-Varela, M. C. Wooten, and T. R. Unnasch. 2006. Analysis of genetic variation in ribosomal DNA internal transcribed spacer and the nadh dehydrogenase subunit 4 mitochondrial genes of the onchocerciasis vector simulium ochraceum. Journal of Medical Entomology 43:701-706.

Roe, A. D., and F. A. H. Sperling. 2007. Patterns of evolution of mitochondrial cytochrome c oxidase i and ii DNA and implications for DNA barcoding. Molecular Phylogenetics and Evolution 44:325-345.

Roehrdanz, R. L. 1993. An improved primer for pcr amplification of mitochondrial DNA in a variety of insect species. Insect Molecular Biology 2:89-91.

Rogers, A. R., and H. Harpending. 1992. Population-growth makes waves in the distribution of pairwise genetic-differences. Molecular Biology and Evolution 9:552-569.

Rokas, A., and S. B. Carroll. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Molecular Biology and Evolution 22:1337-1344.

Rowland, M., N. Durrani, M. Kenward, N. Mohammed, H. Urahman, and S. Hewitt. 2001. Control of malaria in pakistan by applying deltamethrin insecticide to cattle: A community-randomised trial. Lancet 357:1837-1841.

Rozas, J., and R. Rozas. 1999. Dnasp version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174-175.

Rueda, L. M. 2005. Two new species of anopheles (anopheles) hyrcanus group (diptera : Culicidae) from the republic of south korea. Zootaxa:1-26.

Saccone, C., C. De Giorgi, C. Gissi, G. Pesole, and A. Reyes. 1999. Evolutionary genomics in metazoa: The mitochondrial DNA as a model system. Gene 238:195-209.

Saeung, A., Y. Otsuka, V. Baimai, P. Somboom, Pitasawat., B. Taetun, A. Junkum, H. Takaoka, and W. Choochote. 2007. Cytogenetics and molecular evidence for two species in the *anopheles barbirostris* complex (diptera:Culicidae) in thailand. Parasitology Research 101:1337-1344.

Saito, M., S. Kojima, and K. Endo. 2000. Mitochondrial coi sequences of brachiopods: Genetic code shared with protostomes and limits of utility for phylogenetic reconstruction. Molecular Phylogenetics and Evolution 15:331-344.

Saitou, N., and M. Nei. 1987. The neighbor-joining method - a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Salemi, M., and A. Vandamme. 2003. The phylogenetic handbook. A practical approach to DNA and protein phylogeny. Cambridge University Press, Cambridge.

Sallum, M. A. M., P. G. Foster, C. Li, R. Sithiprasasna, and R. C. Wilkerson. 2007. Phylogeny of the leucosphyrus group of anopheles (cellia) (diptera : Culicidae) based on mitochondrial gene sequences. Annals of the Entomological Society of America 100:27-35.

Sallum, M. A. M., E. L. Peyton, and R. C. Wilkerson. 2005. Six new species of the anopheles leucosphyrus group, reinterpretation of an. Elegans and vector implications. Medical and Veterinary Entomology 19:158-199.

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. Systematic Entomology 27:361-382.
Saraste, M. 1990. Structural features of cytochrome-oxidase. Quarterly Reviews Of Biophysics 23:331-366.
Scarpassa, V. M., and J. Conn. 2006. Molecular differentiation in natural populations of *anopheles oswaldoi* sensu lato (diptera: Culicidae) from the brazilian amazon, using cytochrome oxidase i gene mitochondrial DNA. Genetics and Molecular Research In press.
Scott, J. A., W. G. Brogdon, and F. H. Collins. 1993. Identification of single specimens of the anopheles-gambiae complex by the polymerase chain-reaction. American Journal of Tropical Medicine and Hygiene 49:520-529.
Sen, P. 1956. Anopheles-hyrcanus as a vector of filariasis in west-bengal. Nature 178:749-750.
Service, M. W., and H. Townson. 2002. The *anopheles* vector. Pages 59-84 *in* Essential malariology (D. A. Warell, and H. M. Gilles, eds.). Arnold, London.
Shigenobu, Y., K. Saitoh, K. I. Hayashizaki, and H. Ida. 2005. Nonsynonymous site heteroplasmy in fish mitochondrial DNA. Genes & Genetic Systems 80:297-301.
Silva-Brandao, K. L., A. V. L. Freitas, A. V. Z. Brower, and V. N. Solferini. 2005. Phylogenetic relationships of the new world troidini swallowtails (lepidoptera : Papilionidae) based on coi, coii, and ef-1 alpha genes. Molecular Phylogenetics and Evolution 36:468-483.
Simon, C., T. R. Buckley, F. Frati, J. B. Stewart, and A. T. Beckenbach. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annual Review Of Ecology Evolution And Systematics 37:545-579.
Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Annals Of The Entomological Society Of America 87:651-701.
Simpson, T. W. 1951. A note on filariasis among the natives of okinawa, with particular reference to possible transmission of wuchereria-bancrofti by anopheles-hyrcanus-sinensis. American Journal of Tropical Medicine 31:614-616.
Singh, B., L. K. Sung, A. Matusop, A. Radhakrishnan, S. S. G. Shamsul, J. Cox-Singh, A. Thomas, and D. J. Conway. 2004. A large focus of naturally acquired plasmodium knowlesi infections in human beings. Lancet 363:1017-1024.
Singh, O. P., D. Chandra, N. Nanda, S. K. Sharma, P. T. Htun, T. Adak, S. K. Subbarao, and A. P. Dash. 2006. On the conspecificity of anopheles fluviatilis species s with anopheles minimus species c. Journal of Biosciences 31:671-677.
Singhasivanon, P. 2003. Mekong malaria 2003. The Ssoutheast Asian Journal of Tropical Medicine and Public Health 34.
Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. Genetics 129:555-562.

Soltis, P. S., and D. E. Soltis. 2003. Applying the bootstrap in phylogeny reconstruction. Statistical Science 18:256-267.

Steel, D. J., S. A. Trewick, and G. P. Wallis. 2000. Heteroplasmy of mitochondrial DNA in the ophiuroid astrobrachion constrictum. Journal Of Heredity 91:146-149.

Subramanian, R. A., P. Arensburger, P. W. Atkinson, and D. A. O'Brochta. 2007. Transposable element dynamics of the hat element herves in the human malaria vector anopheles gambiae s.S. Genetics 176:2477-2487.

Sugino, R. P., and H. Innan. 2006. Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. Trends In Genetics 22:642-644.

Surendran, S. N., M. S. Ramasamy, B. De Silva, and R. Ramasamy. 2006. Anopheles culicifacies sibling species b and e in sri lanka differ in longevity and in their susceptibility to malaria parasite infection and common insecticides. Medical and Veterinary Entomology 20:153-156.

Suroso, T., S. S. Margono, T. Wandra, and A. Ito. 2006. Challenges for control of taeniasis/cysticercosis in indonesia. Parasitology International 55:S161-S165.

Syafruddin, D., P. Asih, I. Wahid, R. M. Dewi, S. Tuti, I. Laowo, W. Hulu, P. Zendrato, F. Laihad, and A. H. Shankar. 2007. Malaria prevalence in nias district, north sumatra province, indonesia. Malaria Journal 6:116.

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

Takada, H., T. Ono, H. Torikura, and T. Enokiya. 2006. Geographic variation in esterase allozymes of aulacorthum solani (homoptera : Aphididae) in japan, in relation to its outbreaks on soybean. Applied Entomology And Zoology 41:595-605.

Tasanor, O., R. Ruengweerayut, J. Sirichaisinthop, K. Congpuong, W. H. Wernsdorfer, and K. Na-Bangchang. 2006. Clinical-parasitological response and in-vitro sensitivity of plasmodium vivax to chloroquine and quinine on the western border of thailand. Transactions of the Royal Society of Tropical Medicine and Hygiene 100:410-418.

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. Molecular Ecology 7:381-397.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data .3. Cladogram estimation. Genetics 132:619-633.

Thimasarn, K., S. Jatapadma, S. Vijaykadga, J. Sirichaisinthop, and C. and Wongsrichanalai. 1995. Epidemiology of malaria in thailand. Journal of travel medicine 2:59-65.

Tibayrenc, M. 2006. The species concept in parasites and other pathogens: A pragmatic approach? Trends in Parasitology 22:66-70.

Tirados, I., C. Costantini, G. Gibson, and S. J. Torr. 2006. Blood-feeding behaviour of the malarial mosquito anopheles arabiensis: Implications for vector control. Medical and Veterinary Entomology 20:425-437.

Torres, E. P., D. H. Foley, and J. H. Bryan. 2006. Molecular systematics of the philippine malaria vector anopheles flavirostris. Medical and Veterinary Entomology 20:44-52.

Townson, H., R. E. Harbach, and T. A. Callan. 1999. DNA identification of museum specimens of the anopheles gambiae complex: An evaluation of pcr as a tool for resolving the formal taxonomy of sibling species complexes. Systematic Entomology 24:95-100.

Trung, H. D., W. Van Bortel, T. Sochantha, K. Keokenchanh, O. J. T. Briet, and M. Coosemans. 2005. Behavioural heterogeneity of anopheles species in ecologically different localities in southeast asia: A challenge for vector control. Tropical Medicine & International Health 10:251-262.

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. Tropical Medicine & International Health 9:230-237.

Uyeda, J. C., and S. R. Kephart. 2006. Detecting species boundaries and hybridization in camassia quamash and c-leichtlinii (agavaceae) using allozymes. Systematic Botany 31:642-655.

Van Bortel, W., R. E. Harbach, H. D. Trung, P. Roelants, T. Backeljau, and M. Coosemans. 2001. Confirmation of anopheles varuna in vietnam, previously misidentified and mistargeted as the malaria vector anopheles minimus. American Journal Of Tropical Medicine And Hygiene 65:729-732.

Van der Wulp, F. M. 1884. Note 38. On exotic diptera. Notes of the Leiden Museum 6:248-256.

Vanvalen, L. 1976. Ecological species, multispecies, and oaks. Taxon 25:233-239.

Vargas, M. E. I., and A. Espin. 1995. Genetic-variability in mitochondrial-dna of the screwworm, cochliomyia-hominivorax (diptera, calliphoridae), from brazil. Biochemical Genetics 33:237-256.

Venhuis, W. G. 1939. Voorloopige entomologische mededeelingen omtrent *an.Barbirostris* van celebes. Ned Tijdschr Geneeskd 79:2515-2519.

Verle, P., N. T. Ruyen, N. T. Huong, N. T. Be, A. Kongs, P. Van Der Stuyft, and M. Coosemans. 1998. A simple field test for detecting pyrethroids on impregnated nets. Tropical Medicine & International Health 3:833-836.

Walton, C., R. K. Butlin, and K. A. Monk. 1997. A phylogeny for grasshoppers of the genus chitaura (orthoptera : Acrididae) from sulawesi, indonesia, based on mitochondrial DNA sequence data. Biological Journal Of The Linnean Society 62:365-382.

Walton, C., J. M. Handley, C. Kuvangkadilok, F. H. Collins, R. E. Harbach, V. Baimai, and R. K. Butlin. 1999a. Identification of five species of the anopheles dirus complex from thailand, using allele-specific polymerase chain reaction. Medical And Veterinary Entomology 13:24-32.

Walton, C., J. M. Handley, W. Tun-Lin, F. H. Collins, R. E. Harbach, V. Baimai, and R. K. Butlin. 2000. Population structure and population history of anopheles dirus mosquitoes in southeast asia. Molecular Biology And Evolution 17:962-974.

Walton, C., R. G. Sharpe, S. J. Pritchard, N. J. Thelwell, and R. K. Butlin. 1999b. Molecular identification of mosquito species. Biological Journal Of The Linnean Society 68:241-256.

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, H. D. 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. Infection Genetics and Evolution 7:93-102.
Warrior, R., and J. Gall. 1985. The mitochondrial-dna of hydra-attenuata and hydra-littoralis consists of 2 linear-molecules. Archives Des Sciences 38:439-445.
Weil, G. J., and R. M. R. Ramzy. 2007. Diagnostic tools for filariasis elimination programs. Trends in Parasitology 23:78-82.
Weisrock, D. W., and A. Larson. 2006. Testing hypotheses of speciation in the plethodon jordani species complex with allozymes and mitochondrial DNA sequences. Biological Journal Of The Linnean Society 89:25-51.
Wharton, R. H., A. B. G. Laing, and W. H. Cheong. 1963. Studies on distribution and transmission of malaria and filariasis among aborigines in malaya. Annals Of Tropical Medicine And Parasitology 57:235-&.
White, G. B. 1977a. The place of morphological studies in the investigation of *anopheles* species complexes. Mosquito Systematics 9:1-24.
White, J. B. 1977b. The place of morphological* studies in the investigation of *anopheles* species complexes. Mosquito Systematics VBC/SC/76.14.
.
White, N. J. 2004. Antimalarial drug resistance. Journal Of Clinical Investigation 113:1084-1092.
WHO. 2000. Rbm complex emergency malaria database east timor.
WHO. 2006. Revise strategy for malaria control in the south-east asia region. 24th meeting of ministers of healthRegional Office for South-East Asia, Dhaka.
WHO. 2007. Malaria. Disease burden in southeast asia(R. O. f. S. Asia, ed.).
Wilkerson, R. C., J. F. Reinert, and L. Cong. 2004. Ribosomal DNA its2 sequences differentiate six species in the anopheles crucians complex (diptera : Culicidae). Journal of Medical Entomology 41:392-401.
Williams, S. M., R. W. Debry, and J. L. Feder. 1988. A commentary on the use of ribosomal dna in systematic studies. Systematic Zoology 37:60-62.
Wright, S. 1965. The interpretation of population-structure by f-statistics with special regard to systems of mating. Evolution 19:395-420.
Wu, C. I. 2001. The genic view of the process of speciation. Journal Of Evolutionary Biology 14:851-865.
Young, I., and A. W. Coleman. 2004. The advantages of the its2 region of the nuclear rdna cistron for analysis of phylogenetic relationships of insects: A drosophila example. Molecular Phylogenetics and Evolution 30:236-242.
Zhang, D. X., and G. M. Hewitt. 1996. Nuclear integrations: Challenges for mitochondrial DNA markers. Trends in Ecology & Evolution 11:247-251.
Zhang, D. X., and G. M. Hewitt. 1997. Assessment of the universality and utility of a set of conserved mitochondrial coi primers in insects. Insect Molecular Biology 6:143-150.
Zharkikh, A., and W. H. Li. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide-sequences .2. 4 taxa without a molecular clock. Journal of Molecular Evolution 35:356-366.