Molecular Evidence of Speciation Between Island and Continental Populations of *Anopheles* (*Cellia*) *sundaicus* (Diptera: Culicidae), a Principal Malaria Vector Taxon in Southeast Asia

ISABELLE DUSFOUR,[1] YVONNE-MARIE LINTON,[2] ANNA COHUET,[3] RALPH E. HARBACH,[2] VISUT BAIMAI,[4] HO D. TRUNG,[5] CHANG MOH SENG,[6] ASMAD MATUSOP,[7] AND SYLVIE MANGUIN[1]

J. Med. Entomol. 41(3): 287–295 (2004)

ABSTRACT *Anopheles sundaicus s.l.* is a principal malaria vector taxon on islands and along the coastal areas of Southeast Asia. It has a wide geographical distribution and exhibits a high level of ecological and behavioral variability. Study of this taxon is crucial for understanding its biology and implementing effective vector control measures. We compared populations of *An. sundaicus* from Vietnam, Thailand, and Malaysian Borneo by using two mitochondrial DNA markers: cytochrome oxidase I and cytochrome *b*. Genetic divergence, geographic separation, and cladistic analysis of relationships revealed the presence of two cryptic species: *Anopheles sundaicus s.s.* on Malaysian Borneo and *An. sundaicus* species A in coastal areas of Thailand and Vietnam. A polymerase chain reaction (PCR) assay was developed to easily identify these two species throughout their geographic distributions. The assay was based on sequence characterized amplified region derived from random amplified polymorphic DNA. This PCR identification method needs to be validated and adapted for the recognition of other possible species in the Sundaicus Complex.

KEY WORDS *Anopheles sundaicus*, malaria vector, mitochondrial DNA markers, speciation, identification PCR

The principal malaria vectors in Southeast Asia belong to three species complexes: the *Anopheles dirus* Peyton & Harrison, the *Anopheles minimus* Theobald, and the *Anopheles sundaicus* Rodenwaldt complexes. The systematics of the first two taxa is well studied. The Dirus Complex consists of at least seven species (*An. dirus* [= species A], species B–E, *Anopheles nemophilous* Peyton, and *Anopheles takasagoensis* Morishita) (Peyton 1989) and the Minimus Complex includes at least two species (A and C) (Green et al. 1990). In contrast to these, the systematics of the Sundaicus Complex needs to be investigated. *Anopheles sundaicus s.l.* is a principal malaria vector taxon on islands and along coastal areas of Southeast Asia (Kalra 1978, Brandling-Bennett et al. 1981, Kirnowardoya and Yoga 1987, Nguyen Tang Am et al. 1993, Das et al. 1997). Its distribution reportedly includes coastal areas from northeastern India to southern Vietnam, including Myanmar, Thailand, Cambodia, and Peninsular Malaysia, and the islands of Nicobar, Andaman, Borneo, and Indonesia (Reid 1968, Sukowati and Baimai 1996, Linton et al. 2001b). The wide geographic distribution, especially on numerous islands, and the high level of species variability led to speculation that *An. sundaicus* represents a complex of sibling species. Larvae have been collected in brackish water at many locations (Chow 1970, Brandling-Bennett et al. 1981, Kumari et al. 1993, Nguyen Tang Am et al. 1993, Kumari and Sharma 1994, Das et al. 1997, Chang Moh Seng et al. 2001) and in freshwater near Miri in Sarawak, Malaysia, on Borneo (Linton et al. 2001b), South Tapanuli on Sumatra (Indonesia) (Sukowati and Baimai 1996), and Car Nicobar Island (Das et al. 1997). Chow (1970) found that *An. sundaicus* was mainly anthropophilic, endophilic, and exophagic on the Indonesian island of Java. In Vietnam, *An. sundaicus* was also found to be mainly anthropophilic and exophagic, but also endophagic (Nguyen Tang Am et al. 1993). Kumari et al. (1993) and Kumari and Sharma (1994) reported that *An. sundaicus* on Car Nicobar Island is zoophilic, exophagic, and exophilic. Cytogenetic studies carried out on populations from Thai-

[1] Institute of Research for Development, Center of Biology and Management of Populations, Campus International de Baillarguet CS30016, Montferrier sur Lez 34988, France.

[2] Mosquitoes Programme and Biomedical Sciences Theme, Department of Entomology, The Natural History Museum, Cromwell Rd., London SW7 5BD, United Kingdom.

[3] LIN-IRD, 911 avenue d'Agropolis BP64501 Montpellier 34394 cedex 5, France.

[4] Department of Biology, Mahidol University, Rama IV RD, Bangkok 10400, Thailand.

[5] National Institute of Malariology, Parasitology and Entomology, Luong the Vinh St., BC 10.200 Tu Liem, Hanoi, Vietnam.

[6] WHO representative, 177-179 Street Pasteur, Phnom Penh, Cambodia.

[7] Vector Borne Disease Control, State Health Department, Jalan Tua Abang Hj Openg, Kuching, Sarawak, Malaysia.

0022-2585/04/0287–0295$04.00/0 © 2004 Entomological Society of America

land, Java, and Sumatra resulted in the discovery of three chromosomal forms (A, B, and C) (Sukowati and Baimai 1996). Form A was shown to occur in Sumatra and Thailand, form B was found only in northern Sumatra and central Java, and form C was restricted to Asahan in northern Sumatra. Asahan was the only site where all three cytotypes were found in sympatry (Sukowati and Baimai 1996). Isozyme analysis (Sukowati et al. 1999) provided further evidence that these three forms are separate genetic entities. A fourth cytotype may have been identified from the Nicobar Islands, but evidence to support this remains unpublished (Subbarao et al. 1999).

Little is known about the vector capacity of the three forms. Because the design of effective, targeted vector and malaria control measures is dependent on correct species identification, it is essential to characterize all taxa within the Sundaicus Complex. To fix the identity of *An. sundaicus* as a foundation for further investigations of the Sundaicus Complex, a neotype was designated (Linton et al. 2001b) from material collected at Pandan Beach, Lundu District, Sarawak, Malaysia, on Borneo. Further integrated systematics studies are required to determine the composition of the Sundaicus Complex across its range.

The aims of the current study were to conduct a molecular assessment of *An. sundaicus* populations from Thailand, Vietnam, and Malaysian Borneo, and if more than one species was found, to develop a rapid identification method to distinguish the species on a wide scale. Partial regions of two mitochondrial (mtDNA) genes, cytochrome oxidase I (COI) and cytochrome *b* (Cyt-b), were sequenced, and the data were used to assess genetic relationships. These mitochondrial regions have the advantage of being maternally inherited, rapidly evolving, and present in multiple copies per cell, hence they are easy to amplify by PCR. The COI gene is a very useful marker for assessing relationships among closely related insects, with resolution ranging from orders to species (de Brito et al. 2002). The utility of using the COI gene for elucidating relationships of Diptera at the species-complex level is well documented, e.g., the Buzzatii Complex of *Drosophila* (Spicer 1995), the Longipalpis Complex of *Lutzomyia* (Arrivillaga and Feliciangeli 2001), the Imicola Complex of *Culicoides* (Linton et al. 2002), and the Maculipennis Complex of *Anopheles* (Linton et al. 2001a, 2003). Although the Cyt-b gene is widely used in vertebrate systematics, its utility has not been well documented for insects (Simmons and Weller 2001). This gene has been sequenced for few *Anopheles* species, e.g., *An. funestus* and *An. gambiae s.l.* (Besansky et al. 1997, Mukabayire et al. 1999, Krzywinski et al. 2001). Little is known of its utility for distinguishing members of dipteran species complexes (Lachaise et al. 2000); however, this region was informative in studies of sibling species in Coleoptera, Hemiptera, Homoptera, and Hymenoptera (Crozier et al. 1995, Lyman et al. 1999, Morrow et al. 2000, Lin and Wood 2002). The use of Cyt-b jointly with the better known COI (Linton et al. 2001b) in the characterization of *Anopheles* complexes, in particular in the Sundaicus Complex, is novel.

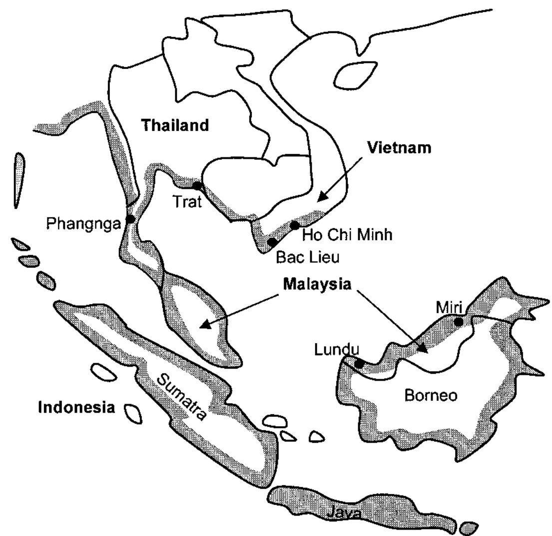

Fig. 1. Reported distribution of *An. sundaicus s.l.* in the studied area (shaded area) and the localities of the six populations studied (black dots).

Materials and Methods

Mosquito Populations. Specimens used in this study were identified as *An. sundaicus s.l.* by using the morphological keys of Reid (1968), and the keys for the *Anopheles* of Vietnam prepared by IMPE (1987). Specimens were collected from six localities, two in southern Vietnam, two in Thailand, and two in Malaysian Borneo (Fig. 1; Table 1). Samples from Lundu Province in Sarawak were collected in association with the neotype series of *An. sundaicus* (Linton et al. 2001b). Sequences of *An. minimus* species A from Hoa Binh Province, Vietnam, and those of *An. gambiae s.s.* (Beard et al. 1993) obtained from GenBank were used to assess genetic relationships between populations of *An. sundaicus s.l.* (see below).

DNA Extraction. DNA was individually extracted following a slightly modified version of the protocol listed in Martinez-Torres et al. (1998). Insects were homogenized in 100 μl of extraction buffer (10 mM Tris-HCl, 60 mM NaCl, 5% sucrose, and 10 mM EDTA, pH 8), to which 125 μl of lysed solution (1.25% SDS, 300 mM Tris-HCl, 5% sucrose, and 10 mM EDTA, pH 8) was added. After a 30-min incubation at 65°C, 38 μl of 3 M potassium acetate (pH 5.2) was added and the homogenates placed at −20°C for 10 min. After centrifugation at 14,000 rpm for 15 min, two volumes of cold 95% ethanol were added and the tubes incubated for at least 10 min at −20°C. DNA was pelleted by centrifugation at 4°C for 15 min at 14,000 rpm. Pellets were then washed with 500 μl of 70% ethanol and dried in speed vacuum. DNA was resuspended in 100 μl of LTE (10 mM Tris, pH 8, 0.1 mM EDTA) and stored at −20°C until required.

Table 1. Location, number, conservation state, date of collection, and type of larval habitat for specimens of *An. sundaicus s.l.* used in this study

Countries	Localities	Population codes	No. of specimens	Conservation states	Collection dates	Larval habitats	No. of specimens analyzed	
							COI	Cyt-b
Vietnam	Bac Lieu	VBL	7	Frozen adults	03–04/99	Brackish	7	5
	Ho Chi Minh	VHCM	3	Frozen adults	03/99	Brackish	3	2
Thailand	Trat	TTR	6	Dried adults	04/99	Brackish	5	5
	Phangnga	TPG	4	Dried adults	02/00	Brackish	4	4
Malaysia (Sarawak)	Lundu	MAL	7	Larvae in ethanol	11/99	Brackish	4	7
	Miri	MAM	6	Larvae in ethanol	11/99	Freshwater	4	6
						Total	27	29

Numbers of specimens analyzed per mtDNA gene region also are listed.

mtDNA Amplification. A 445-base pair (bp) region of the COI gene was amplified with universal primers C1-J-1718 and C1-N-2191 (Simon et al. 1994), by using the reaction conditions described in Linton et al. (2001b). Initially, an 804-bp fragment of mitochondrial Cyt-b was cloned and sequenced after amplification with primers CB1L (Simon et al. 1994) and CytR (Krzywinski et al. 2001). Specific primers (CBsunA 5′-AATGTTACAAGAATTCA-3′ and CBsunB 5′-TTAGCTATACATTATGC-3′) were subsequently designed to amplify an informative region of 485 bp. For Cyt-b, the reaction conditions were as follows. The 100-μl reaction mix contained 4 μl of 1:10 DNA dilution, 4 pmol of primers, 10 μl of 10× buffer (provided with *Taq* polymerase, 1.5 mM included), 1 mM of $MgCl_2$ added, 0.2 mM of dNTP mix (Eurogentec, Liège, Belgium), and 5 U of *Taq* polymerase (QIAGEN, Valencia, CA). The PCR conditions were 94°C for 5 min, and then 40 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 40 s, with a final extension step of 10 min at 72°C. PCR products were purified (NucleoSpin extract two in 1, Machery-Nagel, Düren, Germany), diluted, and sent for direct sequencing using the original primers (CBsunA). These fragments of *An. sundaicus s.l.* correspond to nucleotides 10580–11062 (COI) and 1649–2093 (Cyt-b) of the mtDNA sequence of *An. gambiae* (NC002084) (Beard et al. 1993). All sequences generated in this study are available in GenBank under the following accession numbers: AY243788-AY243799, AY245283, AY245284, AY256954-AY256957, AY253150-AY253155, AY299094-AY299120, and AY299339-AY299346.

To define the amino acid translation reading frame, sequences were aligned against the complete mtDNA sequence of *An. gambiae* (Beard et al. 1993). The nucleotide sequences were translated into amino acids with MEGA version 2.1 (Kumar et al. 2001), by using the invertebrate mitochondrial code.

Multiplex PCR Conditions. The multiplex PCR for species identification was developed on the base of the sequence characterized amplified regions (SCARs) derived from random amplified polymorphic DNA (RAPD). The protocol used in this approach was described in Kengne et al. (2001) and Manguin et al. (2002).

The multiplex PCR was performed in 25-μl reaction mix containing 3 μl of 1:10 DNA dilution, 4 pmol of each primer (S9A1 5′-GTCCCAAATCCTGCCAATC-3′, S9A2 5′-CGATGGTATTCATTCTGCTTTACTA-3′, S9B1 5′-AAGATCATCGTTTTCATAGATTTCG-3′, and S9B2 5′-CTTTTTGCTGGTTCATTCTATC-TGT-3′), 2.5 μl of 10× buffer (provided with *Taq* polymerase, 1.5 mM included), 0.2 mM of dNTP mix (Eurogentec, Seraing, Belgium), and 1.25 U of *Taq* polymerase (QIAGEN). The conditions of amplification were 94°C for 5 min, then 35 cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s, and a final elongation at 72°C for 10 min.

Sequence Alignment and Phylogenetic Analyses. Multiple sequence alignments were achieved using BioEdit (Hall 1999) with Clustal W 1.4 (Thompson et al. 1994). Transition:transversion ratios (Ti:Tv), sequence diversities, and genetic distances according the Tamura-Nei algorithm (Tamura and Nei 1993) with Gamma correction were calculated in MEGA 2.1 (Kumar et al. 2001). Gamma parameters were estimated within PAUP 4.0b10 (Swofford 2002). The genetic population structure was investigated using pairwise population Fstatistics (Fst) significance tests. Fst was estimated using the method of Slatkin (1991, 1995) and the Weir and Cockerham's Theta (Fst) (Weir and Cockerham 1984) by using ARLEQUIN (Schneider et al. 2000). Significance of Fst estimates were evaluated against the results of 5,000 permutations.

Maximum parsimony (MP) and Neighbor-Joining (NJ) analyses of relationships were performed within PAUP 4.0b10 (Swofford 2002), by using *An. gambiae* (Beard et al. 1993) and *An. minimus* species A as outgroup taxa. *Anopheles sundaicus* and both outgroup species belong to subgenus *Cellia;* however, *An. sundaicus* is more closely related to *An. gambiae* than to *An. minimus* A because they both are members of the Pyretophorus Series (Harbach 1994). MP analysis was performed with unweighted characters and by a heuristic search using 100 random stepwise additions of sequences, with band swapping by tree bisection reconnection (TBR). Congruence of sequence data were tested with the Incongruence Length Difference test (ILD test) (Farris et al. 1994) by using full heu-

Table 2. Percentage of AT richness and relative positions of variable/parsimony informative nucleotides associated with codon positions 1, 2, and 3 for the amino acid translation of the respective partial sequences of COI and Cyt-b generated in this study

	COI				Cyt-b			
	Position 1	Position 2	Position 3	Total	Position 1	Position 2	Position 3	Total
Total sites	149	148	148	445	162	162	161	485
Variable sites	3	0	20	23 (5.1%)	1	0	15	16 (3.3%)
Parsimony informative sites	3	0	15	18 (78.3%)	1	0	13	14 (87.5%)
%AT	56.0	52.7	89.2	66	60.9	62.3	90.7	71.3

Percentage values in parentheses indicate the proportion of variable sites compared to all sites and the proportion of parsimony informative sites compared with all polymorphic sites.

ristic searches based on 1000 replicates. The NJ distance method was performed using Tamura-Nei's model + Gamma + I (Tamura and Nei 1993), which takes account of unequal base frequencies and has six substitution rates that are Gamma-distributed across sites. The distribution of substitution rates at variable sites estimated by the Gamma parameter and the assumed proportion of invariable sites (I) were calculated within PAUP 4.0b10 (Swofford 2002). Bootstrap support values were calculated for MP trees by heuristic search and for NJ trees by NJ search by using 500 stepwise addition sequences for each replicate and TBR branch swapping.

Results

Mitochondrial DNA Analysis. Twenty-seven and 29 specimens (2–7 specimens/population) were sequenced for the COI and Cyt-b genes, respectively (Table 1), with 23 specimens being sequenced for both regions. The 930-bp fragment, generated in two separate PCR reactions, comprised 445 and 485 bp of the COI and Cyt-b genes, respectively. Due to the protein coding nature of the two mitochondrial genes, indels and gaps were absent and the resulting alignments were unambiguous.

The mean AT richness of both gene regions reached 68.3% (COI, 66%; Cyt-b, 71.3%), which is congruent with published values for mtDNA of other *Anopheles* (Beard et al. 1993, Mitchell et al. 1993). This AT bias is strongly associated with the third codon position, where it accounts for an average of 90% (COI, 89.2%; Cyt-b, 90.7%) of all nucleotide substitutions at this position (Table 2).

Using the expected reading frame based on the whole mitochondrial sequence of *An. gambiae* (Beard et al. 1993), both partial nucleotide gene sequences were translated into amino acids (AA). Translated protein sequences were 148 AA for COI and 161 AA for Cyt-b. Neither sequence contained stop codons, indicating that the data were free from pseudogenes, nonfunctional nuclear copies of mitochondrial genes. With the exception of nucleotide position 387 (AA129) in the COI sequences of both specimens TTR15 and MAL12 from Trat, Thailand and Lundu, Malaysia, which induced an isoleucine (ATT) to methionine (ATA) change, all other amino acid substitutions were silent. The amino acid sequences generated for the Cyt-b fragments were identical.

The 445-bp fragment of the COI gene amplified for 27 specimens showed a polymorphism of 5.1%. From 23 variable sites, 18, representing 78.3% of all polymorphic sites, were parsimony informative (Table 2). Comparatively lower levels of variability were detected in the Cyt-b sequences of the 29 specimens examined. In the 485 bp of Cyt-b, 16 polymorphic sites (3.3%) were detected, 14 of which (87.5%) were parsimony informative (Table 2). Both markers were highly conserved at the first and second codon positions; the polymorphism was mainly located at the third codon position with 20 variable sites on 23 (86.9%) for COI and 15 on 16 (93.8%) for Cyt-b (Table 2). Transition was the major substitution type for both markers, with an overall Ti:Tv equal to 3.0 for COI and 42.6 for Cyt-b. Transitions seemed to have a linear relationship with Tamura-Nei genetic distances over the range of divergence within the ingroup. However, when substitutions between ingroup and outgroup were considered, the level of transition increased and saturation effect was detected for Cyt-b when sequence divergence reached ≈8–9%. In mitochondrial genes, saturation of transitions often occurs when the total divergence reaches 8–10% (Hackett 1996, Griffiths 1997, Bloomer and Crowe 1998, Martin and Bermingham 1998). Apart from the distances associated with the outgroup taxa, the largest distance for Cyt-b data among haplotypes was 2.6%. Thus, the distances within the ingroup were all less than the limit where saturation occurs.

Twenty-two and 10 nucleotide haplotypes were determined for COI and Cyt-b, with a total sequence diversity of 1.6 and 1.5%, respectively.

Intrapopulation sequence diversities ranged from 0.3 to 1.1% per site for COI and up to 0.5% per site for Cyt-b, indicating a low level of divergence (Table 3). Genetic distances for Cyt-b were small between Miri and Lundu populations (0.3%) and among Thai and Vietnamese populations, ranging up to 0.7% (Table 3). Malaysian Borneo populations (shaded area in Table 3) were genetically more distant to the Thai and Vietnamese populations, with values of 2.2 to 2.8% (Table 3). Genetic distances based on COI revealed the same pattern of genetic difference, with 0.4% divergence between both populations from Vietnam (VBL and VHCM) and 1.1% between those from Thailand (TTR and TPG), and also values ranging from 1.8 to 2.6% between the Thai or Vietnamese populations and those from Malaysian Borneo (shaded area in Table

Table 3. Values of Tamura-Nei genetic distances between populations

		Vietnam		Thailand		Malaysian Borneo	
		VBL	VHCM	TTR	TPG	MAL	MAM
VBL			0	0.006	0.007	0.026	0.028
VHCM		0.004		0.006	0.007	0.026	0.028
TTR		0.018	0.019		0.006	0.022	0.025
TPG		0.017	0.017	0.011		0.023	0.025
MAL		0.021	0.022	0.026	0.023		0.003
MAM		0.018	0.018	0.026	0.023	0.009	
Sequence diversities	COI	0.003	0.006	0.005	0.011	0.010	0.005
	Cyt-b	0	0	0.004	0.005	0.001	0.002

Genetic distances for COI (below diagonal line) and Cyt-b (above diagonal line), as well as sequence diversities for COI and Cyt-b.

3). Genetic divergences between the Thai and Vietnamese populations were higher for COI (1.7–1.9%) than Cyt-b (0.6–0.7%). Based on genetic divergence, two genetically distant groups are apparent, one consisting of populations from Malaysian Borneo (MAL and MAM) and the other comprising the continental populations of Thailand and Vietnam (VBL, VHCM, TTR, and TPG). Mean genetic distance values between these two groups reached 2.2% for COI and 2.5% for Cyt-b. Moreover, no haplotype was common between these two groups, whereas one haplotype was shared by Vietnamese populations for COI, two Cyt-b haplotypes present in the populations from Bac Lieu (Vietnam), Ho Chi Minh (Vietnam), and Trat (Thailand), and two in the populations from Miri and Lundu. Intragroup sequence diversities were close to intrapopulation values, with 0.8% in specimens from Malaysian Borneo and 1.2% in the continental ones for COI, and 0.2 and 0.5%, for Cyt-b, respectively, indicating homogeneity among haplotypes in each group. Genetic structure, as described above, is supported by the coefficients of differentiation: Fst (Slatkin 1995) between these two groups reached 0.5 ($p \ll 0.01$) for COI and 0.82 ($p \ll 0.01$) for Cyt-b. Also, Theta values (Weir and Cockerham 1984) were higher than Fst, with 0.7 ($p \ll 0.01$) for COI and 0.90 ($p \ll 0.01$) for Cyt-b. Highly significant Fst values for Cyt-b indicate negligible gene flow between groups. Although Fst for COI was lower, it still described limited gene flow. These values are indicative of two distinct species. Fixed nucleotide substitutions characterized each species. Three substitutions (positions 216, 294, and 378) among 19 polymorphic sites in COI and seven (positions 12, 24, 138, 267, 297, 469, and 474) among 16 in Cyt-b were identified. Position 216 of the COI fragment showed three nucleotide substitutions, which were country-specific.

Evolutionary Relationships. MP and Tamura-Nei-based NJ analyses were rooted with the outgroup taxa, *An. gambiae* and *An. minimus* species A. The congruence of the mitochondrial markers (ILD test $P = 0.08 > 0.05$) allowed the joined analysis of the COI and Cyt-b sequences in the MP approach. Trees constructed in both analyses showed congruent basal topologies with *An. minimus* A most basal, followed by *An. gambiae* and a monophyletic *An. sundaicus* clade (Fig. 2). Both analyses revealed two distinct clades for *An. sundaicus* populations, one (clade I) corresponding to *An. sundaicus s.s.* from both sites in Malaysian Borneo, and the second (clade II) representing the continental populations from Thailand and Vietnam (Fig. 2, NJ trees not shown). The joined data MP analysis strongly supported the separation of these clades with bootstrap values of 97 for clade I and 86 for clade II (Fig. 2). The analyses confirmed the occurrence of two genetic species in the Sundaicus Complex: *An. sundaicus s.s.* in Sarawak, Malaysia on the island of Borneo (Linton et al. 2001b), and *An sundaicus* species A in coastal areas of Vietnam and Thailand.

Differentiation of Continental and Island Populations by Using Multiplex PCR. Of the 62 RAPD primers screened (list available upon request from S.M.), only five (B7, F6, S5, S8, and S9) showed polymorphism with reproducible banding patterns discriminating continental populations (Thailand and Vietnam) from those of Malaysian Borneo. Of these five primers, only the S9 produced SCAR fragments with the expected RAPD polymorphism showing two

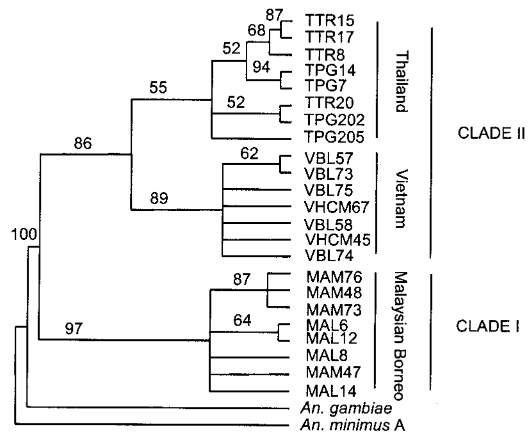

Fig. 2. Maximum parsimony tree derived from the analysis of 930 bp of combined COI and Cyt-b nucleotide sequences. The analysis was performed with unweighted characters and by a heuristic search by using 100 random stepwise additions of sequences, with band swapping by tree bisection-reconnection. Bootstrap support values were calculated by a heuristic search by using 500 stepwise addition sequences for each replicate and TBR branch swapping. See Table 1 for abbreviations for populations.

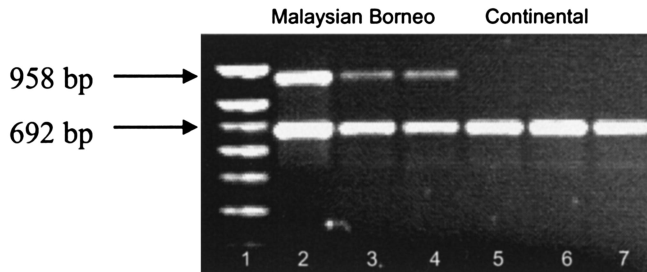

Fig. 3. Multiplex PCR for the identification of *An. sundaicus s.s.* (lanes 2–4; Malaysian Borneo) and *An. sundaicus* species A from coastal areas of continental Southeast Asia (lane 5, Thailand; lanes 6–7, Vietnam). Lane 1, 100-bp ladder.

bands, one at 870 bp, specific to the continental species, and one at 970 bp for *An. sundaicus s.s.* in Malaysian Borneo (data not shown). After cloning and sequencing these two SCAR fragments, four species-specific primers were designed for an optimization of the multiplex PCR. Amplified products obtained with these species-specific SCAR primers presented bands at 692 and 958 bp for *An. sundaicus* from Malaysian Borneo, and only one band at 692 bp for the continental populations (Fig. 3). The band at 692 bp served as the common band to identify *An. sundaicus s.l.* and the one at 958 bp was specific to *An. sundaicus s.s.*

Discussion

The existence of a species complex within *An. sundaicus* was first suggested from ecological and behavioral variations (Chow 1970, Brandling-Bennett et al. 1981, Kumari et al. 1993, Nguyen Tang Am et al. 1993, Kumari and Sharma 1994, Das et al. 1997, Chang Moh Seng et al. 2001). Subsequently, Sukowati and Baimai (1996) and Sukowati et al. (1999) defined cytogenetic and isoenzymatic forms among *An. sundaicus* populations in Thailand and Indonesia. This study is the first to present molecular evidence for recognizing two separate genetic species within the complex. Based on COI and Cyt-b partial mtDNA gene sequences, two well supported clades, each representing a separate species, were revealed by MP and NJ analyses of relationships among *An. sundaicus s.l.* populations. In addition to this differentiation, a genetic homogeneity within each entity, unlikely gene flow between entities, and a genetic distance >2% between the two defined taxa (Foley et al. 1998, Beebe et al. 2000), further support the recognition of the two species.

The neotype of *An. sundaicus* from Lundu Province in Sarawak on the island of Borneo designated by Linton et al. (2001a) is geographically isolated and genetically distinct from the species A that occurs in coastal areas of Vietnam (Ho Chi Minh and Bac Lieu Provinces) and Thailand (Trat and Phangnga Provinces). Populations of *An. sundaicus s.s.* in Malaysian Borneo breed in both brackish and freshwater habitats (Chang Moh Seng et al. 2001, Linton et al. 2001b); thus, differences in larval ecology in this area do not support the hypothesis of a Sundaicus Complex. Tolerance to different levels of salinity is not an exception for species of subgenus *Cellia.* Larvae of *An. farauti s.s.* of the Australasian Punctulatus Group (Sweeney 1987), and *An. bwambae*, *An. gambiae*, *An. merus*, and *An. melas* of the Afrotropical Gambiae Complex (Thelwell et al. 2000, Krzywinski and Besansky 2003) commonly occur in both brackish and freshwater habitats.

The two species reported here are completely allopatric due to their wide geographic separation by the South China Sea, and introgression is not likely to occur. Consequently, due to this allopatric situation and also to the haploid property of the mtDNA (Page and Holmes 1998), heterozygotes are absent and gene flow cannot be fully studied. According to the biological species concept, extended allopatry is considered as a mechanism for speciation by enforced reproductive isolation of populations and subsequent genetic divergence (Turelli et al. 2001). Based on calculations of the mitochondrial molecular clock with 2.3% divergence per million years (myr) (Brower 1994), the genetic divergence between these two species was estimated at 0.9 myr for both markers, with approximated values of 0.91 myr for COI and 0.94 myr for Cyt-b. This estimated time of divergence seems to be correlated with the historical geography of Southeast Asia. Alternating glaciation and warm periods occurred during the Pleistocene between 1.8 million and 11,000 yr ago. During ice ages, sea levels decreased from 50 to 200 m and land links formed between the islands and mainland of Southeast Asia. The creation of a land mass (Sundaland) that included Borneo, the continent, and Indonesia was possible only when sea levels decreased by 200 m (Tougard 2001). This situation occurred 0.8 myr before present (BP). After 0.8 myr BP, sea levels fluctuated from only −50 to −135 m, thus providing a water barrier between continental and island populations that facilitated species divergence and speciation.

Further investigation with molecular markers, in particular by increasing the number of collection sites within the geographical distribution of *An. sundaicus s.l.*, is necessary to more fully understand the Sundaicus Complex. Moreover, the results of previous cytogenetic and isoenzymatic studies need to be confirmed by molecular studies. Chromosomal inversions and molecular markers are not always in congruence for species discrimination in *Anopheles* species com-

plexes. Mukabayire et al. (1999) examined the mtDNA (Cyt-b) and rDNA (ITS2) of karyotyped *An. funestus* and found that neither was correlated with different chromosomal arrangements. Similarly, the M and S molecular forms of *An. gambiae* do not always coincide with the Mopti and Savana + Bamako chromosomal forms (Wondji et al. 2002, Krzywinski and Besansky 2003). Clearly molecular study of Indonesian populations of *An. sundaicus s.l.* that are characterized by different polytene chromosome banding patterns (Sukowati and Baimai 1996) and enzyme profiles (Sukowati et al. 1999) is essential to better understand the Sundaicus Complex. Because Sukowati and Baimai (1996) found only chromosomal form A in Trat Province of Thailand, it is most likely the continental species recognized in this article as *An. sundaicus* species A. Sukowati et al. (1999) also found that form A is widespread in Indonesia; hence, the first step to test for congruence between the molecular, cytogenetic, and enzymatic markers will be to use the identification PCR method presented here to confirm the presence of the continental species in Indonesia. In addition, the PCR technique will need to be redesigned and validated if the existence of other species of the Sundaicus Complex is confirmed with molecular markers. The development of a cheap, rapid, reproducible method of identification is important for studying the ecology and behavior of vector species, a prerequisite for improving the effectiveness and selectivity of vector control methods.

Acknowledgments

We thank the staff in Malaysia, Thailand, and Vietnam for precious help in field mosquito collections. We are especially grateful to Prof. M. Coosemans (ITM, Antwerp, Belgium), the coordinator of the EC grant, INCO-DC research project no. ERBIC 18CT970211, which financially supported this work.

References Cited

Arrivillaga, J. C., and M. D. Feliciangeli. 2001. *Lutzomyia pseudolongipalpis:* the first new species within the longipalpis (Diptera: Psychodidae: Phlebotominae) complex from La Rinconada, Curarigua, Lara State, Venezuela. J. Med. Entomol. 38: 783–790.

Beard, C. B., D. M. Hamm, and F. H. Collins. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae:* DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. 2: 103–124.

Beebe, N. W., R. D. Cooper, D. A. Morrison, and J. T. Ellis. 2000. A phylogenetic study of the *Anopheles punctulatus* group of malaria vectors comparing rDNA sequence alignments derived from the mitochondrial and nuclear small ribosomal subunits. Mol. Phylogenet. Evol. 17: 430–436.

Besansky, N. J., T. Lehmann, G. T. Fahey, D. Fontenille, L. E. 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.

Bloomer, P., and T. M. Crowe. 1998. Francolin phylogenetics: molecular, morphobehavioral, and combined evidence. Mol. Phylogenet. Evol. 9: 236–254.

Brandling-Bennett, A. D., E. B. Doberstyn, and S. Pinichpongse. 1981. Current epidemiology of malaria in Southeast Asia. Southeast Asian J. Trop. Med. Public Health 12: 289–297.

Brower, A. V. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. U.S.A. 91: 6491–6495.

Chang Moh Seng, Y. M. Linton, and R. E. Harbach. 2001. The *Anopheles sundaicus* species complex - a preliminary study of larval ecology from two sites in Sarawak, Malaysia. Mekong Malaria Forum 8: 91–95.

Chow, C. Y. 1970. Bionomics of malaria vectors in the western pacific region. Southeast Asian J. Trop. Med. Public Health 1: 40–57.

Crozier, R. H., N. Dobric, H. T. Imai, D. Graur, J. M. Cornuet, and R. W. Taylor. 1995. Mitochondrial-DNA sequence evidence on the phylogeny of Australian jack-jumper ants of the *Myrmecia pilosula* complex. Mol. Phylogenet. Evol. 4: 20–30.

Das, M. K., T. Adak, and V. P. Sharma. 1997. Genetic analysis of a larval color mutant, yellow larva, in *Anopheles sundaicus.* J. Am. Mosq. Control Assoc. 13: 203–204.

de Brito, R. A., M. H. Manfrin, and F. M. Sene. 2002. Nested cladistic analysis of Brazilian populations of *Drosophila serido.* Mol. Phylogenet. Evol. 22: 131–143.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing significance of congruence. Cladistics 10: 315–319.

Foley, D. H., J. H. Bryan, D. Yeates, and A. Saul. 1998. Evolution and systematics of *Anopheles:* insights from a molecular phylogeny of Australasian mosquitoes. Mol. Phylogenet. Evol. 9: 262–275.

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

Griffiths, C. S. 1997. Correlation of functional domains and rates of nucleotide substitution in cytochrome b. Mol. Phylogenet. Evol. 7: 352–365.

Hackett, S. J. 1996. Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). Mol. Phylogenet. Evol. 5: 368–382.

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

Harbach, R. E. 1994. Review of the internal classification of the genus *Anopheles* (Diptera: Culicidae): the foundation for comparative systematics and phylogenetic research. Bull. Entomol. Res. 84: 331–342.

[IMPE] Institute of Malariology, Parasitology and Entomology. 1987. Keys to *Anopheles* in Vietnam (adults-pupae-larvae). Department of Entomology, Institute of Malariology, Parasitology and Entomology, Hanoi, Vietnam.

Kalra, N. L. 1978. National Malaria Eradication Programme, India - its problems, management and research needs. J. Commun. Dis. 10: 1–20.

Kengne, P., H. D. Trung, V. Baimai, M. Coosemans, and S. Manguin. 2001. A multiplex PCR-based method derived from random amplified polymorphic DNA (RAPD) markers for the identification of species of the *Anopheles minimus* group in Southeast Asia. Insect Mol. Biol 10: 427–435.

Kirnowardoya, S., and G. P. Yoga. 1987. Entomological investigations of an outbreak of malaria in Chilacap on

south coast of central Java, Indonesia during 1985. J. Commun. Dis. 19: 121–127.

Krzywinski, J., and N. J. Besansky. 2003. Molecular systematics of *Anopheles:* from subgenera to subpopulations. Annu. Rev. Entomol. 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. Mol. Phylogenet. Evol. 18: 479–487.

Kumar, S., K. Tamura, I. B. Jacobsen, and N. Masatoshi. 2001. MEGA Molecular Evolutionary Genetics Analysis version 2.1 Software, Arizona State University, Tempe, AZ.

Kumari, R., and V. P. Sharma. 1994. Resting and biting habits of *Anopheles sundaicus* in Car Nicobar Island. Indian J. Malariol. 31: 103–114.

Kumari, R., H. Joshi, A. Giri, and V. P. Sharma. 1993. Feeding preferences of *Anopheles sundaicus* in Car Nicobar Island. Indian J. Malariol. 30: 201–206.

Lachaise, D., M. Harry, M. Solignac, F. Lemeunier, V. Benassi, and M. L. Cariou. 2000. Evolutionary novelties in islands: *Drosophila santomea,* a new melanogaster sister species from Sao Tome. Proc. R. Soc. Lond. B Biol. Sci. 267: 1487–1495.

Lin, C. P., and T. K. Wood. 2002. Molecular phylogeny of the North American *Enchenopa binotata* (Homoptera: Membracidae) species complex. Systematics 95: 162–171.

Linton, Y. M., A. Samanidou-Voyadjoglou, L. Smith, and R. E. Harbach. 2001a. New occurrence records for *Anopheles maculipennis* and *Anopheles messae* in the northern Greece based on DNA sequence data. Eur. Mosq. Bull. 11.

Linton, Y. M., R. E. Harbach, M. S. Chang, T. G. Anthony, and A. Matusop. 2001b. Morphological and molecular identity of *Anopheles* (*Cellia*) *sundaicus* (Diptera: Culicidae), the nominotypical member of a malaria vector species complex in Southeast Asia. Syst. Entomol. 26: 357–366.

Linton, Y. M., A. J. Mordue, R. H. Cruickshank, R. Meiswinkel, P. S. Mellor, and J. F. Dallas. 2002. Phylogenetic analysis of the mitochondrial cytochrome oxidase subunit I gene of five species of the *Culicoides imicola* species complex. Med. Vet. Entomol. 16: 139–146.

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. Syst. Entomol. 28: 39–55.

Lyman, D. F., F. A. Monteiro, A. A. Escalante, C. Cordon-Rosales, D. M. Wesson, J. P. Dujardin, and C. B. Beard. 1999. Mitochondrial DNA sequence variation among triatomine vectors of Chagas' disease. Am. J. Trop. Med. Hyg. 60: 377–386.

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. Med. Vet. Entomol. 16: 46–54.

Martin, A. P., and E. Bermingham. 1998. Systematics and evolution of lower Central American cichlids inferred from analysis of cytochrome b gene sequences. Mol. Phylogenet. Evol. 9: 192–203.

Martinez-Torres, D., F. Chandre, M. S. Williamson, F. Darriet, J. B. Berge, A. L. Devonshire, P. Guillet, N. Pasteur, and D. Pauron. 1998. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae s. s.* Insect Mol. Biol. 7: 179–184.

Mitchell, S. E., A. F. Cockburn, and J. A. Seawright. 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome 36: 1058–1073.

Morrow, J., L. Scott, B. Congdon, D. Yeates, M. Frommer, and J. Sved. 2000. Close genetic similarity between two sympatric species of tephritid fruit fly reproductively isolated by mating time. Evolution 54: 899–910.

Mukabayire, O., D. Boccolini, L. Lochouarn, D. Fontenille, and N. J. Besansky. 1999. Mitochondrial and ribosomal internal transcribed spacer (ITS2) diversity of the African malaria vector *Anopheles funestus.* Mol. Ecol. 8: 289–297.

Nguyen Tang Am, R. Le Quy, H. Vu Thi, and L. Nguyen Bieh. 1993. Entomo-epidemiological studies of malaria in the coastal zone of Ho Chi Minh City, 1990–1992. Santé 3: 464–473.

Page R.D.M., and E. C. Holmes. 1998. Molecular evolution: a phylogenetic approach. Blackwell, Oxford, United Kingdom.

Peyton, E. L. 1989. A new classification for the *Leucosphyrus* group of *Anopheles* (*Cellia*). Mosq. Syst. 21: 197–205.

Reid, J. A. 1968. Anopheline mosquitoes of Malaya and Borneo. Studies from the Institute of Medical Research Malaysia 31.

Schneider, S., D. Roessli, and L. Excoffier. 2000. ARLEQUIN version 2.000 A software for population genetics data analysis. (http://anthro.unige.ch/arlequin).

Simmons, R. B., and S. J. Weller. 2001. Utility and evolution of cytochrome b in insects. Mol. Phylogenet. Evol. 20: 196–210.

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. Ann. Entomol. Soc. Am. 87: 651–701.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. Genet. Res. 58: 167–175.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457–462.

Spicer, G. S. 1995. Phylogenetic utility of the mitochondrial cytochrome oxidase gene: molecular evolution of the *Drosophila buzzatii* species complex. J. Mol. Evol. 41: 749–759.

Subbarao, S. K., N. Nanda, and K. Raghavandra. 1999. Malariogenic stratification in India using *Anopheles culicifacies* sibling species prevalence. ICMR Bull. 29.

Sukowati, S., and V. Baimai. 1996. A standard cytogenetic map for *Anopheles sundaicus* (Diptera: Culicidae) and evidence for chromosomal differentiation in populations from Thailand and Indonesia. Genome 39: 165–173.

Sukowati, S., V. Baimai, S. Harun, Y. Dasuki, H. Andris, and M. Efriwati. 1999. Isozyme evidence for three sibling species in the *Anopheles sundaicus* complex from Indonesia. Med. Vet. Entomol. 13: 408–414.

Sweeney, A. W. 1987. Larval salinity tolerances of the sibling species of *Anopheles farauti.* J. Am. Mosq. Control. Assoc. 3: 589–592.

Swofford, D. L. 2002. PAUP Phylogenetic analysis using parsimony (and others methods) version 4.0b10. Sinauer, Sunderland, MA.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512–526.

Thelwell, N. J., R. A. Huisman, R. E. Harbach, and R. K. Butlin. 2000. Evidence for mitochondrial introgression between *Anopheles bwambae* and *Anopheles gambiae.* Insect Mol. Biol. 9: 203–210.

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

Tougard, C. 2001. Biogeography and migration routes of large mammal faunas in South-East Asia during the Late Middle Pleistocene: focus on the fossil and extant faunes from Thailand. Palaeogeogr. Palaeoclimatol. Palaeoecol. 168: 337–358.

Turelli, M., N. H. Barton, and J. A. Coyne. 2001. Theory and speciation. Trends Ecol. Evol. 16: 330–343.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

Wondji, C., F. Simard, and D. Fontenille. 2002. Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. Insect Mol. Biol. 11: 11–19.

Received for publication 28 May 2003; accepted 30 September 2003.

Dusfour Isabelle, Linton Y.M., Cohuet Anna, Harbach R.E., Baimai V., Trung H.D., Seng C.M., Matusop A., Manguin Sylvie
(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 (3), 287-295

ISSN 0022-2585