Genetic diversity and molecular identification of mosquito species in the *Anopheles maculatus* group using the ITS2 region of rDNA

C. Walton a,*, P. Somboon b, S.M. O'Loughlin a, S. Zhang c, R.E. Harbach d, Y.-M. Linton d, B. Chen c, K. Nolan f, S. Duong g, M.-Y. Fong h, I. Vythilingum i, Z.D. Mohammed j, Ho Dinh Trung k, R.K. Butlin l

a Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK
b Department of Parasitology, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand
c School of Biology, University of Leeds, Leeds LS2 9JT, UK
d Department of Entomology, The Natural History Museum, London SW7 5BD, UK
e College of Plant Protection, Southwest Agricultural University, Chongqing 400716, PR China
f Department of Biological Sciences, University of Warwick, Warwick CV4 7AL, UK
g National Center for Malaria, Parasitology and Entomology, Ministry of Health, Phnom Penh, Cambodia
h Department of Parasitology, University of Malaya, Kuala Lumpur, Malaysia
i Institute for Medical Research, Kuala Lumpur, Malaysia
j Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia
k Department of Entomology, National Institute of Malaria, Parasitology and Entomology, Hanoi, Vietnam
l Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Received 6 September 2004; received in revised form 11 May 2006; accepted 12 May 2006
Available online 19 June 2006

Abstract

The species diversity and genetic structure of mosquitoes belonging to the *Anopheles maculatus* group in Southeast Asia were investigated using the internal transcribed spacer 2 (ITS2) of ribosomal DNA (rDNA). A molecular phylogeny indicates the presence of at least one hitherto unrecognised species. Mosquitoes of chromosomal form K from eastern Thailand have a unique ITS2 sequence that is 3.7% divergent from the next most closely related taxon (*An. sawadwongporni*) in the group. In the context of negligible intraspecific variation at ITS2, this suggests that chromosomal form K is most probably a distinct species. Although *An. maculatus* sensu stricto from northern Thailand and southern Thailand/peninsular Malaysia differ from each other in chromosomal banding pattern and vectorial capacity, no intraspecific variation was observed in the ITS2 sequences of this species over this entire geographic area despite an extensive survey. A PCR-based identification method was developed to distinguish five species of the group (*An. maculatus*, *An. dravidicus*, *An. pseudowillmori*, *An. sawadwongporni* and chromosomal form K) to assist field-based studies in northwestern Thailand. Sequences from 187 mosquitoes (mostly *An. maculatus* and *An. sawadwongporni*) revealed no intraspecific variation in specimens from Thailand, Cambodia, mainland China, Malaysia, Taiwan and Vietnam, suggesting that this identification method will be widely applicable in Southeast Asia. The lack of detectable genetic structure also suggests that populations of these species are either connected by gene flow and/or share a recent common history.

© 2006 Elsevier B.V. All rights reserved.

Keywords: *Anopheles maculatus* group; Malaria; Southeast Asia; ITS2; Genetic structure

1. Introduction

Anopheline mosquitoes occur typically as groups of closely related species that cannot always be distinguished reliably using morphological characters. Members of species complexes or groups can differ in biological attributes such as anthropogenicity, exophagy/endophagy, exophily/endophily, longevity and larval habitat preference. These characteristics relate to the vectorial capacity of a species and the means by which effective vector control can be implemented (Subbarao, 1998). Consequently, the development of reliable molecular tools for species identification, and an understanding of
The Anopheles maculatus group is an assemblage of eight recognised species in the Oriental Region (Harbach, 2004). Members of the group occur from the Indian subcontinent through Southeast Asia to Taiwan. Adults are difficult to distinguish morphologically due to overlapping characters. The presence of several species within the group was resolved primarily with the use of cytological methods (Green and Baimai, 1984; Green et al., 1985, 1992; Baimai et al., 1993). The group was revised by Rattanarithikul and Green (1986) and Rattanarithikul and Harbach (1991), who recognised eight morphologically similar species. Table 1 shows how the species designations relate to the cytological forms and their geographic ranges. Chromosomal form K (Baimai, 1989; Baimai et al., 1993) has not yet been formally recognised as a species. Chromosomal forms B and E are currently regarded as cytotypes of An. maculatus. As noted by Green et al. (1985), these chromosomal forms either represent sibling species or reflect geographic variation within An. maculatus. In general, form B is found throughout northern Thailand but is replaced by form E in southern Thailand and peninsular Malaysia (Green et al., 1985). Cross-mating studies found no evidence of post-mating reproductive incompatibility between the two cytotypes (Baimai et al., 1984). The two chromosomal forms can be distinguished using cuticular hydrocarbons (Kittayapong et al., 1990), and using these markers it was inferred that the two chromosomal forms are sympatric at some sites in peninsular Malaysia (Kittayapong et al., 1993). This suggests that the chromosomal forms correspond to separate species, and any barriers to reproduction are likely to be pre-mating.

Members of the Maculatus Group are known to be involved in malaria transmission, but the vectorial capacity of individual species remains unclear from previous studies due to the difficulty in species identification using morphology alone. This uncertainty is aggravated by the observation that the ability of a species to transmit malaria can vary depending upon local factors such as environmental conditions and population size. For example, although An. willmori has been recorded as a major vector in Nepal (Pradhan et al., 1970), it has never been implicated as such in Thailand where it is apparently rare. An. maculatus is widespread but is only considered to be a major vector in southern Thailand and peninsular Malaysia (Hodgkin, 1956; Rahman et al., 1993). Since the presence of malaria transmitted by An. maculatus correlates with the presence of form E, it is possible that only this chromosomal form is able to transmit malaria to any significant extent (Rongnoparut et al., 1996; Kittayapong et al., 1993; Upatham et al., 1988). As noted by Green et al. (1985), this chromosomal form is considered to be a principal vector in Java (Barcus et al., 2002), but its cytotype and specific identity remain undetermined. Anopheles pseudovollmori (Green et al., 1991) and An. maculatus and An. sawadwongporni (Rattanarithikul et al., 1996; Somboon et al., 1998), have been found infected with malaria parasites in Thailand.

As part of a large-scale study carried out in northwestern Thailand to understand the ecology and biting behaviour of all potential malaria vector species in relation to land cover and land use change, it was necessary to be able to distinguish species of the Maculatus Group. This excludes An. dispar and An. greenii which are confined to the Philippines (Rattanarithikul and Harbach, 1991), and which can readily be identified using the ITS2-RFLP assay of Torres et al. (2000). The other six species of the An. maculatus group potentially occur in northwestern Thailand, although only An. maculatus and An. sawadwongporni are considered to be widespread (Green et al., 1991). Two molecular methods have been developed to distinguish some members of the An. maculatus group in China (Ma et al., 2002; Li et al., 2003), but these were unavailable at the start of this study and to our knowledge they have not been tested for use in Thailand. The number of specimens to be screened in such studies is often large. Cytological methods of identification are not suitable as they are stage-specific, time-consuming and laborious to perform. PCR-based methods of identification are preferable as they are relatively quick, straightforward and reliable. Regions of the ribosomal DNA (rDNA) are often the markers of choice in Anopheles for this purpose as there are often fixed differences even between closely related species (Collins and Paskewitz, 1996; Walton et al., 1999a).

The aim of this work was two-fold: (1) to explore the genetic diversity of the An. maculatus group and (2) to develop a PCR-
based identification method to reliably distinguish species of the group in northwestern Thailand, which could be used in a large-scale epidemiological and ecological study. The marker used for these purposes was the second internal transcribed spacer (ITS2) that separates the 5.8S and 28S rDNA subunits. The diversity of the An. maculatus group was investigated, not only within the study foci in northwestern Thailand, but also in the rest of Thailand and other Southeast Asian countries. This enabled us to evaluate the geographic extent over which the identification method is potentially applicable since this is dependent upon the distribution of genetic diversity of ITS2 within and between species.

2. Materials and methods

2.1. Mosquito collection and morphological and chromosomal identification

Adult mosquitoes were collected using animal and human baits at sites 1 and 2 in northwestern Thailand (Fig. 1, Table 2), and progeny broods were raised from some females. Larvae were collected from the edges of running streams and reared to adulthood. Collections of larvae and adults were made at the mountain site 3 (Doi Inthanon) in an attempt to collect the higher altitude species An. willmori, but only An. maculatus and An. pseudowillmori were found. Specimens were identified to species based on adult and egg morphology (Rattanarithikul and Green, 1986) (keys are unavailable for immature stages of the An. maculatus group) and/or metaphase karyotypes (Baimai et al., 1993). Mosquitoes that could be reliably identified to species (some of which were from progeny broods and some of which were field-caught mosquitoes) were designated as reference specimens. Some siblings of the progeny broods were retained as vouchers in The Natural History Museum, London. In collections from sites 4 to 8, made during routine collections in our ecological and epidemiological study, the mosquitoes were identified to the group level only and subsequently sequenced or tested using the identification assay developed herein. Mosquitoes from other localities in Thailand, Cambodia, mainland China, Malaysia, Taiwan and Vietnam were also sequenced to examine geographical diversity.

2.2. DNA extraction, amplification and sequencing of ITS2

DNA was extracted from whole individual mosquitoes using a salting-out protocol (Sunnucks and Hales, 1996). One microlitre of DNA (equivalent to 1/800 of a mosquito) was used in each 50 μl PCR reaction. The rDNA ITS2 was amplified using primers 5.8F (5'-TGTGAACTGCAGGACACATG-3') and 28R (5'-ATGCTTAAATTTAGGGGTAA-3') (Collins and Paskewitz, 1996). The concentrations of the reactants were: 0.2 μM of each primer, 200 μM dNTP, 2.5 mM MgCl2, 20 mM (NH4)2SO4, 75 mM Tris–HCl (pH 8.8) and 0.01% (w/v) 'Tween'. One unit of Thermoprime Plus DNA Polymerase (ABgene, Epsom, UK) was used per reaction. The samples were heated at 94 °C for 5 min before 35 cycles of amplification at 94 °C for 1 min, 61 °C for 30 s and 72 °C for 30 s followed by a final extension step of 5 min. The amplification products were purified on columns and sequenced using the PCR primers and fluorescent chemistry (Applied Biosystems, Warrington, UK). Sequences were aligned and checked manually in SeqEdit (version 1.0.3) (Applied Biosystems, Warrington, UK). Most of the sequencing was done in both directions, including that for all reference specimens, all specimens of An. pseudowillmori, chromosomal form K and Anopheles dravidicus from Thailand, and at least two specimens from each sampling site. The remainder were confirmed to be one of the established sequences by comparison of the ITS2 sequence generated by sequencing in a single direction, but if there was any ambiguity they were then sequenced in both directions.

![Fig. 1. Outline map of Thailand and part of Southeast Asia showing the 33 mosquito collection sites listed in Table 2. The box indicates the approximate coverage of the epidemiological and entomological study area in northwestern Thailand.](image-url)
2.3 Sequence alignment and phylogenetic analysis

DNA sequences were aligned using Clustal W version 1.7 (Thompson et al., 1994). Phylogenetic relationships were inferred using maximum-likelihood (ML), maximum parsimony and neighbour-joining methods in PHYLIP version 3.5c (Felsenstein, 1989). Kimura two-parameter distances with a transition/transversion ratio of two were used for tree construction with the neighbour-joining method. The model for the ML method used one category of substitution rates, empirical base-frequencies and an expected transition/transversion ratio of 2. The global search option was also used in ML.

### Table 2

Number of specimens of each species sequenced for ITS2 from each site

<table>
<thead>
<tr>
<th>Site no.</th>
<th>Date (month/year)</th>
<th>Collection sites (country, province, village/town)</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MAC</td>
</tr>
<tr>
<td>1</td>
<td>November 2000</td>
<td>Thailand, Mae Hong Son, Ban Mae Top Nua</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>November 2000</td>
<td>Thailand, Lampang, Ban Den Udom</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>April 2001</td>
<td>Thailand, Chiang Mai, Doi Inthanon</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>June 2001–December 2002</td>
<td>Thailand, Mae Hong Son, Ban Nong Kha Klang</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>June 2001–December 2002</td>
<td>Thailand, Mae Hong Son, Ban Huai Pong Khan Nai</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>June 2001–December 2002</td>
<td>Thailand, Ma Hong Son, Huai Chang Kham</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>June 2001–December 2002</td>
<td>Thailand, Chiang Mai, Ban Huai Ngui</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>June 2001–December 2002</td>
<td>Thailand, Lamphun, Ban Pang</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>October 1996</td>
<td>Thailand, Loei, Ban Pa Kow Lam</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>October 1996</td>
<td>Thailand, Sukhon Nakhon, Ban Kok Klang</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>July 2001</td>
<td>Thailand, Kanchanaburi, Ban Phu Toei</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>July 2001</td>
<td>Thailand, Prachub Khiri Khan, Hue Rae</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>July 2001</td>
<td>Thailand, Prachub Khiri Khan, Palau-U</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>July 2001</td>
<td>Thailand, Chumphon, Ban Noi Chok Kwa</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>July 2001</td>
<td>Thailand, Ranong, Ban Hing Chang</td>
<td>6</td>
</tr>
<tr>
<td>17</td>
<td>July 2001</td>
<td>Thailand, Song Khla, Pedang Besar</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>August 2003</td>
<td>Malaysia, Terengganu, Kampungs Jenagor, Basong,</td>
<td>11</td>
</tr>
<tr>
<td>19</td>
<td>August 2003</td>
<td>Payah Kayu and Dura</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>August 2003</td>
<td>Malaysia, Johor, Kota Tinggi</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>August 2001</td>
<td>China, Guangxi, Pubei, Chenggu</td>
<td>13</td>
</tr>
<tr>
<td>22</td>
<td>August 2001</td>
<td>China, Guangdong, Huidong, Daling</td>
<td>14</td>
</tr>
<tr>
<td>23</td>
<td>September 2000</td>
<td>China, Taiwan, Taung</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>June 2001</td>
<td>China, Hainan, Changjiang, Shilu</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>June 2003</td>
<td>Vietnam, Lang Son, Trang Dinh, Chi Minh</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>July 2000</td>
<td>Vietnam, Ninh Binh, Cucphuong National Forest</td>
<td>5</td>
</tr>
<tr>
<td>27</td>
<td>May 2004</td>
<td>Vietnam, Nghe An, Thanh Chuong, Thanh Lam</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>2003</td>
<td>Vietnam, Quang Binh, Le Thuy, Ngn Thuy</td>
<td>3</td>
</tr>
<tr>
<td>29</td>
<td>June 2004</td>
<td>Vietnam, Quang Ninh, Ba Che, Thanh Lam</td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td>October 2003</td>
<td>Cambodia, Preah Vihear, Ror Vieng, Romeny</td>
<td>3</td>
</tr>
<tr>
<td>31</td>
<td>June 2003</td>
<td>Cambodia, Ratanakiri, Ochum, Chaongchan</td>
<td>6</td>
</tr>
<tr>
<td>32</td>
<td>October 2003</td>
<td>Cambodia, Pailin, Pang Rolim</td>
<td>1</td>
</tr>
<tr>
<td>33</td>
<td>June 2003</td>
<td>Cambodia, Pursat, Dey Kra Hom</td>
<td>3</td>
</tr>
</tbody>
</table>

Total per species 122 48 4 4 9

The morphologically identified specimens came from sites 1 and 2 and cytologically identified specimens of form K were collected at site 11. Other specimens were identified from field collections as belonging to the An. maculatus group and identified to species based on ITS2 sequences. Collection sites were at or near the localities indicated and their approximate locations are shown in Fig. 1. MAC, An. maculatus; SAW, An. sawadwongporni; PSEU, An. pseudowillmori; DRAV, An. dravidicus; K, form K (see text).

### Table 3

Primers used in the multiplex PCR and expected sizes of the fragments amplified from each species

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Length (bp)</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8F</td>
<td>5'-ATCACAGGCTCGCTGGGATC-3'</td>
<td>20</td>
<td>Universal forward</td>
</tr>
<tr>
<td>MAC</td>
<td>5'-GGCGGTCAAGTCTGGGTAAAGT-3'</td>
<td>20</td>
<td>An. maculatus 180</td>
</tr>
<tr>
<td>PSEU</td>
<td>5'-GCCGCCGGTGGTCAAAACAG-3'</td>
<td>19</td>
<td>An. pseudowillmori 203</td>
</tr>
<tr>
<td>SAW</td>
<td>5'-ACCGGCTCGATCGGTGTCG-3'</td>
<td>19</td>
<td>An. sawadwongporni 242</td>
</tr>
<tr>
<td>K</td>
<td>5'-TCATCGCTCGCCTTACCA-3'</td>
<td>20</td>
<td>Form K 301</td>
</tr>
<tr>
<td>DRAV</td>
<td>5'-GCTACTTTGACGCGGAC-3'</td>
<td>20</td>
<td>An. dravidicus 477</td>
</tr>
</tbody>
</table>
tree construction to further explore alternative tree topologies. Bootstrap supports were based on 1000 re-sampled datasets using SEQBOOT in PHYLIP. Trees were visualised using TREEVIEW (Page, 2001).

2.4. PCR conditions for the identification method

Reactants: 0.2 μM of primers 5.8F, MAC, DRAV, K and 0.1 μM of primers SAW and PSEU (primers defined in Table 3); 2.5 mM MgCl₂; 200 μM dNTP; 20 mM (NH₄)₂SO₄, 75 mM Tris–HCl (pH 8.8) and 0.01% (w/v) ‘Tween’; 0.5 units of Thermoprime Plus DNA Polymerase (ABgene, Epsom, UK); and 1 μl of genomic DNA sample in a total reaction volume of 25 μl. The samples were heated at 94 °C for 5 min before 35 cycles of amplification at 94 °C for 1 min, 61 °C for 30 s and 72 °C for 30 s followed by a final extension step of 5 min.

3. Results

Adult and larval mosquitoes from sites 1 to 3 in northwestern Thailand (Fig. 1, Table 2) were identified morphologically to obtain a set of reference specimens for the study area. Four species of the An. maculatus group were identified: An. maculatus, An. sawadwongporni, An. dravidicus and An. pseudowillmori. Despite numerous entomological surveys, An. notanandai has not been recorded from this region of Thailand, so it is possible that it does not occur there. Although An. willmori has been recorded in northwestern Thailand, it is associated with higher altitudes and is therefore unlikely to occur in the majority of entomological and epidemiological surveys.

The reference specimens from sites 1 to 3 (16 An. maculatus, 14 An. sawadwongporni, 3 An. pseudowillmori and 3 An. dravidicus (Table 2, Fig. 1)) were sequenced for ITS2. The sequences for the four species are quite distinct from each other and no intraspecific variation was found (Fig. 2). To further establish that only these species were present at the field sites in northwestern Thailand, the ITS2 sequences of 27 additional specimens from sites 4 to 8 (Table 2, Fig. 1) were also obtained. Each was found to have a sequence identical to one of the four reference sequences.

To assess the level of geographical variation, another 146 mosquitoes, collected from other areas within Thailand and from other countries in Southeast Asia, were sequenced for ITS2 (Table 2). Despite the wide geographical range covered, no intraspecific variation was found. In the case of An. maculatus, this involved specimens from northern and southern Thailand, Cambodia, China (Guangxi, Guangdong and Taiwan), Malaysia and Vietnam, and for An. sawadwongporni specimens were from mainland China, Cambodia, Vietnam and northern and southern Thailand. However, three specimens that were identified as chromosomal form K from eastern Thailand (site 11, Table 2 and Fig. 1) had a quite distinct ITS2 sequence (sequence K in Fig. 2) differing by 14 base substitutions and the length of an indel from the next most closely related sequence, of An. sawadwongporni (Fig. 2). This sequence was also obtained from three specimens from Cambodia (site 30) and three from Vietnam (site 28).

Comparisons were made between the species-specific ITS2 sequence found in this study and ITS2 sequences of the same species from mainland China and Malaysia that were available from GenBank. Several bases (~20) from the beginning and end of the database sequences were excluded in these comparisons as the multiple differences observed in these regions, in comparisons with our sequences, are most likely due to sequencing errors in the database sequences (as the region near to the primer can be difficult to read), or possibly due to the inclusion of primer sequences with the submitted sequence. On this basis, ITS2 sequences from An. maculatus (AF261950), An. sawadwongporni (AF512551) and An. dravidicus (AF261951) from mainland China were identical to those found in our study. The ITS2 sequence of An. pseudowillmori (AF512550), however, differed at seven sites from the sequence of Thai specimens. A second (partial) sequence from An. pseudowillmori (AF261952) was more similar to the sequence of Thai specimens although it still differed by at least one base. (This sequence was not included in the phylogenetic analysis because it was incomplete.) The two database sequences of An. maculatus from Malaysia (AF500072 and AF500073) both differed at three bases (32, 150 and 240, Fig. 2) from the sequence of this species that we obtained from Thai, Chinese, Malaysian and Cambodian mosquitoes. AF500072 (MAC Malay1) differed at another two bases and AF500073 (MAC Malay2) differed at another six bases from our sequence of An. maculatus.

A molecular phylogeny (Fig. 3) was constructed using all the unique ITS2 sequences of members of the An. maculatus group available from this study and from GenBank, using the alignment in Fig. 2. The gene tree in Fig. 3 was constructed using maximum-likelihood and is unrooted because an outgroup with easily aligned ITS2 is not available. ML and maximum parsimony methods of tree construction resulted in the same tree topology. However, the relative positions of An. willmori, An. dravidicus, An. dispar and An. greeni were altered when the neighbour-joining method was used, although the latter two species still clustered with An. maculatus. There is low bootstrap support (≤56%) for the deeper branching events. Nevertheless, Fig. 3 illustrates clearly that the specimens of chromosomal form K are most closely related to An. sawadwongporni, yet show a level of sequence divergence comparable to that between other species of the group. These sequences have been deposited in GenBank (accession numbers DQ518615–DQ518629).

The identification method is based on the principle of allele-specific amplification in which Thermus aquaticus (Taq) DNA polymerase is unable to extend primers that are mismatched to their template DNA (Ugozzoli and Wallace, 1991; Scott et al., 1993). The alignment of the sequences from each species (Fig. 2) was used to design species-specific amplification primers with a reverse orientation (Table 3, Fig. 2). A single universal primer (5.8F) binds to the 5.8S gene in all species in the forward orientation (Collins and Paskewitz, 1996). The reaction conditions were optimised with respect to annealing temperature, magnesium concentration, primer concentration and polymerase concentration to ensure that each species-
Fig. 2. Alignment of ITS2 sequences of members of the *An. maculatus* group collected and sequenced in this study: *An. maculatus* (MAC), *An. sawadwongporni* (SAW), *An. dravidicus* (DRAV) and chromosomal form K (K); together with differing sequences obtained from GenBank: *An. pseudowillmori* (PSEU) China; accession number: AF512550) and *An. willmori* from China (WILL; AF512552), *An. dispar* (DISPAR; AF234778) and *An. greeni* (GREENI; AF234779) from the Philippines; and two sequences of *An. maculatus* (MAC Malay1; AF500072) and (MAC Malay2; AF500073). Dots indicate identity with the reference sequence from *An. maculatus* and a dash denotes a deletion with respect to the other sequences. Boxes indicate the binding sites of the species-specific primers (Table 2).
specific primer (in combination with the universal primer) only amplifies DNA from the corresponding species and there is no cross-amplification with any of the other primers. The optimised reaction conditions for the identification method are given in Section 2. Inclusion of all the primers in a single PCR reaction with DNA from one of the five species generates a PCR product of a diagnostic length (Table 3) that can be detected by agarose gel electrophoresis (Fig. 4).

Since large number of mosquitoes need to be identified for meaningful epidemiological and ecological field studies, we sought to develop a quick, reliable method of processing mosquitoes. The use of a crude extraction method (see Section 2), rather than a complex DNA extraction method that yields DNA of high quality, was therefore investigated. Ten-fold serial dilutions of crude DNA samples (prepared from heads) were used in the multiplex PCR. Amplification was successful across four orders of magnitude of sample concentration, indicating that this simple method of DNA extraction is extremely robust. Extracted samples have been used successfully for up to 2 weeks after preparation when stored frozen. The DNA is unlikely to be suitable for long-term storage but the method does enable the identification procedure to be repeated if necessary. Legs or heads can be used for DNA extraction enabling the remainder of the mosquito to be used for other assays, preserved for long-term storage for other studies, or retained as a voucher. Routinely, whole mosquitoes are boiled for 15 min in 200 μl of water and 1 μl used in the above PCR identification method.

4. Discussion

4.1. Species diversity

The sequencing survey and phylogenetic analysis (Fig. 3) indicate greater species diversity in the An. maculatus group than has been recognised formally up to now. In the general context of low intraspecific variation, the high degree of differentiation of the specimens of chromosomal form K from the other species (3.7% divergence from the most closely related species, An. sawadwongporni; Fig. 3) is strong evidence that these belong to another species. Divergence among members of mosquito species complexes varies but can be substantial (e.g. ~10–18% at ITS2 among members of the An. annularis group, unpublished data). However, divergence at ITS2 can be much lower than this even between well-recognised species; for example, it is only 0.6% between An. dirus and An. baimaii (Walton et al., 1999b), 1.0% between An. maculipennis and An. dacie (Nicolescu et al., 2004), and 0.4–1.6% among members of the An. gambiae complex (Paskewitz et al., 1993). The species status of this chromosomal and genetic form needs to be confirmed by observing sympatry without heterozygotes using chromosomal or other markers. Although it is possible that form K corresponds to An. notanandai, it is unlikely since An. notanandai corresponds to chromosomal form G and has previously only been reported from eastern Thailand (Table 1, Rattanarithikul and Green, 1986). According to R. Rattanarithikul (personal communication to REH), form K is morphologically similar to, but distinguishable from, An. notanadai, providing additional evidence that form K represents another species of the An. maculatus group.

There are two sequences for An. pseudowillmori from Yunnan Province, China in the database. One sequence (AF261952) appears to be the same as the Thai and Vietnamese sequences generated in this study (see below). The other sequence (AF512550) exhibits seven differences from the sequences of Thai An. pseudowillmori even after the ends are trimmed. It is therefore possible that these sequences correspond to two different, yet very closely related species. However, since it is unknown whether the Chinese specimens
were collected from the same or different places within Yunnan, they may merely represent geographically isolated and differentiated populations of *An. pseudowillmori*.

The taxonomic status of chromosomal forms B and E of *An. maculatus* remains unclear, although the apparent difference in vectorial capacity between them makes this distinction particularly worthwhile to clarify. Previous collection records with associated chromosomal data indicate the northernmost collection of chromosomal form E is from Phato, Rangong (at latitudes of 09°46′N and less) (Green et al., 1985), but according to Rongnoparut et al. (1999) this form occurs as far north as ~12°N. The majority of our specimens are therefore expected to be form B, but those from sites 13, 14, 15, 16 and 17 in peninsular Thailand, and especially those from sites 18, 19 and 20 in Malaysia, are most likely to be chromosomal form E. Despite the likelihood that our specimens include both chromosomal forms, only a single ITS2 sequence was detected.

The two GenBank sequences of Malaysian *An. maculatus* differ by several bases from the sequences that we obtained from specimens of *An. maculatus* collected in Thailand, China, Cambodia, Vietnam, Malaysia and Vietnam. This difference could correspond to the two chromosomal forms if the Malaysian mosquitoes from the database are all form E and our mosquitoes are all chromosomal form B, although, as argued above, this seems unlikely. Rather than indicating any form of intraspecific variation or the presence of cryptic species, the differences in the Malaysian sequences may be due to sequencing errors. This is supported by the fact that the sequencing we report here was carried out independently in two laboratories without any conflict. Furthermore, errors in GenBank sequences have been noted previously (Linton et al., 2002). It is clearly important to gather more sequence data (from more variable loci), ideally coupled with chromosomal data and epidemiological data, from *An. maculatus* in Malaysia.

### 4.2. Phylogenetic relationships

Fig. 3 shows clearly that the putative species corresponding to chromosomal form K is the sister of *An. sawadwongporni*. Ongoing studies have shown that the adults of both taxa have overlapping morphological characters, but their eggs are clearly distinct. Moreover, crossing experiments revealed post-zygotic isolation between chromosomal form K and both *An. maculatus* (form B) and *An. sawadwongporni* (P. Somboon, unpublished data). It can also be seen that the two genetic forms of *An. pseudowillmori* are substantially divergent from other members of the *An. maculatus* group. However, the low bootstrap support (≤56%) for the deeper branching events in Fig. 3, illustrates that the rapidly evolving ITS2 locus is unable to resolve the deeper relationships. A full understanding of the phylogenetic relationships within the group will require the analysis of other loci, such as more conserved regions of the rDNA or mtDNA genes.

### 4.3. Intraspecific diversity

Despite considerable geographic sampling, no intraspecific variation was found in ITS2 in the 187 specimens that we sequenced. (This assumes that we have correctly interpreted the level of divergence observed between putative species K and other members of the group as representing interspecific divergence rather than intraspecific variation.) The only indication of intraspecific variation came from comparisons made with sequences obtained from sequence databases. The difference of a one base insertion in *An. pseudowillmori* from China (AF261952) relative to the sequences of Thai specimens could be due to sequencing error since the other three species in Thailand and China (*An. sawadwongporni, An. maculatus* and *An. dravidicus*) were identical to each other (once starting and ending sequences are edited from the Chinese sequences). Given that we found no variation in *An. maculatus* from Thailand, China, Cambodia, Vietnam and Malaysia, the variation observed between the sequences of the two Malaysian specimens is unexpectedly high. It would be helpful if this locus could be sequenced from more individuals from countries neighbouring Thailand, particularly China and Malaysia, to establish the extent of any intraspecific variation within members of the group.

The lack of intraspecific variation that is generally observed over a large geographic area could imply sufficient gene flow to allow homogenisation of sequences of the rDNA genes by concerted evolution (Elder and Turner, 1995). Alternatively, it could reflect a demographic history of these populations in which they have been derived sufficiently recently to have prevented diversification from the ancestral sequence. The very low level of population structure detected in *An. maculatus* in Thailand using microsatellites (Rongnoparut et al., 1999) is consistent with both of these hypotheses.

### 4.4. Applicability of the identification method

When the molecular identification method was applied to 240 adult mosquitoes of the *An. maculatus* group from our field study in northwestern Thailand, it was able to amplify >94% of the specimens, with the unidentified specimens being attributed to degraded DNA due to poor preservation. *An. maculatus* and *An. sawadwongporni* are the most abundant species (51.7% and 42.1% of the total, respectively), correlating with their widespread distribution indicated in Table 2. *An. dravidicus* and *An. pseudowillmori* were rarely encountered (2.1% and 4.2% of the total, respectively), but this could, in part, be due to differences in feeding preferences as all specimens were captured on human bait. Chromosomal form K was never encountered, which together with its incidence at sites 11, 28 and 30, suggests that it has an eastern distribution in mainland Southeast Asia.

The general lack of intraspecific variation in ITS2 sequence makes the identification method likely to be very useful over a large geographic area—apparently at least in Malaysia, most of Thailand, parts of China, and most probably Cambodia, Vietnam and Taiwan. However, the method will not distinguish the two sequences of ITS2 obtained from *An. pseudowillmori* in China. The inclusion of chromosomal form K in the identification method will help to extend its usefulness to eastern Thailand, and to Vietnam and Cambodia where this form also occurs. In some areas it will be necessary to adapt the
method to include the identification of additional species, for example, *An. willmori* in high altitude areas and *An. notanandai* in west-central Thailand. The sequence alignment (Fig. 2) illustrates that there appears to be ample variability between species, and the region is sufficiently lengthy to enable primers to be designed for additional species. Before the method is deployed in new areas, it is clearly advisable to assess which species are present and the extent of intraspecific variation by sequencing the ITS2 region of specimens from the area.

In conclusion, the identification method presented here is likely to work over a large geographic area with scope to modify it to include additional species. Furthermore, it is very robust to the use of a simple and rapid DNA extraction method and to the concentration of DNA used. For this reason, and the fact that the method requires only a single step, a PCR reaction, before running the samples on an agarose gel, the method is an eminently practical tool for large-scale field-based studies where reliable species identification is important.

Acknowledgements

We thank the many people who contributed to the mosquito collections for this study, including: Dr. Samsak Prajakwong, former director of the Office of Vector-Borne Disease Control No. 2 in Chiang Mai, Thailand, and his staff, particularly Mr. RaksaKul Kantawong, for their assistance in collection and identification of mosquito specimens in Thailand; Conor Cahill and Mark Isenstadt for assistance with collections in north-western Thailand; Dr. Hwa-Jen Teng, Division of Vector-borne Infectious Diseases, Taiwan, and Prof. Masahiro Takagi and Dr. Yoshio Tsuda, Institute of Tropical Medicine, Nagasaki University, Japan, for organizing the collection of specimens in Taiwan; Dr. Nguyen Duc Manh and staff of the National Institute of Malariology, Parasitology and Entomology, Hanoi, Vietnam, for collecting and providing specimens from Vietnam; Dr. Tho Sochanta and other staff of the National Center for Malaria, Parasitology and Entomology, Cambodia, for their assistance in the collection and identification of mosquitoes in Cambodia. This work was part of the RISKMODEL and MALVECASIA projects funded by the European Union, grant numbers QLK2-CT-2000-01787 and ERBIC 18CT970211, respectively.

References


