ORIGINAL ARTICLE

High levels of population structure caused by habitat islands in the malarial vector *Anopheles scanloni*

SM O'Loughlin¹, P Somboon² and C Walton¹

¹Faculty of Life Sciences, University of Manchester, Manchester, UK and ²Department of Parasitology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

The genetic structure of four populations of the malarial vector *Anopheles scanloni* in Thailand was studied using mitochondrial DNA sequences. Four highly divergent lineages were observed, all with signals of population expansion. Since *An. scanloni* is restricted to 'islands' of limestone karst habitat, this suggests there is a metapopulation-type dynamic in this species, with restricted gene flow, extinctions and drift all contributing to lineage divergence. Historical environmental change and marine transgressions may also have contributed to population extinction, expansion

and divergence. Although there is some current gene flow inferred between nearby populations, it is extremely restricted between the northern and southern populations, which also differed by one fixed polymorphism at the ITS2 rDNA locus. Crossing experiments showed no post-mating barriers existing between the north and the south, but the lack of gene flow between these populations could ultimately result in speciation and has implications for malaria control strategies.

Heredity (2007) **99**, 31–40; doi:10.1038/sj.hdy.6800959; published online 11 April 2007

Keywords: *Anopheles dirus*; metapopulation; malaria vectors; limestone karst; divergence

Introduction

Anopheles scanloni (previously known as *An. dirus* C) is a vector of malaria and is a member of the Southeast Asian *An. dirus* complex, which is generally recognized to comprise seven species (Peyton, 1989; Harbach, 2004; Sallum *et al.*, 2005). *An. scanloni* is known only in Thailand from Kanchanaburi at the north of its range to the Malaysian border in the south (Baimai *et al.*, 1988a). It occurs sympatrically with four other members of the species complex, *An. dirus* (previously known as *An. dirus* A), *An. baimaii* (previously known as *An. dirus* D), *An. cracens* (previously known as *An. dirus* B) and *An. nemophilous* (Baimai, 1988, 1988b and personal observation). The remaining two complex members are restricted to peninsular India (*An. elegans*) and Taiwan (*An. takasagoensis*). Like other species of the complex, *An. scanloni* is associated with forest but unlike them it appears to have a non-continuous distribution owing to its dependence on limestone karst habitat (Poopittaya-sataporn and Baimai, 1995), which occurs patchily in this region.

An. scanloni is very closely related to *An. dirus* (Baimai *et al.*, 1987, 1988b; Walton *et al.*, 1999) and they are considered sibling species. However, in a mitochondrial DNA (mtDNA) study of *An. dirus*, *An. scanloni* and

An. baimaii, which included 12 individuals from two populations of *An. scanloni* from northern and southern Thailand, *An. scanloni* showed much higher population differentiation than the other two species over similar geographical distances (Walton *et al.*, 2000). The use of microsatellites in the two populations confirmed that this effect was genome wide (Walton *et al.*, 2001). This very high level of differentiation and lack of shared mitochondrial haplotypes indicated an absence of gene flow between the two populations, which led to the suggestion that the two populations may represent discrete species. If this finding extends to other populations of *An. scanloni*, it would make this a good model system to study the processes that result in divergence during allopatric speciation. In particular, little is known about the order in which different reproductive isolating mechanisms accumulate during allopatric speciation (Coyne and Orr, 1998; Turelli *et al.*, 2001). Determining this requires the use of taxa that have either speciated very recently or are in the process of speciation (to ensure the absence of isolation mechanisms that have been gained post-speciation). Information on the extent and nature of barriers to gene flow between *An. scanloni* populations will improve our understanding of the processes involved in their divergence and in assessing their utility for future studies of speciation.

A further reason for studying gene flow in *An. scanloni* is that a better understanding of gene flow between vector populations and closely related species has important implications for malaria control, particularly with regard to the spread of insecticide resistance or genetic control measures (Catteruccia *et al.*, 2000; Collins *et al.*, 2000). Despite good success in controlling malaria

Correspondence: Dr C Walton, Faculty of Life Sciences, University of Manchester, 3.614 Stopford Building, Oxford road, Manchester M13 9PT, UK.

E-mail: catherine.walton@manchester.ac.uk

Received 27 April 2006; revised 17 October 2006; accepted 26 February 2007; published online 11 April 2007

in much of Thailand, members of the *An. dirus* complex remain a threat as they are highly anthropophilic (Rattanakul et al., 1996) and tend not to rest on interior or exterior insecticide sprayed surfaces (Scanlon and Sandhinand, 1965; Wilkinson et al., 1978). In addition, *An. scanloni* is a particularly early feeder with peak biting occurring between 1800 and 2200 h (Baimai et al., 1988a) when bed nets are not in use.

Here, we determine the genetic structure of *An. scanloni* across the species' known range. To perform this, we have substantially increased the sample number (from 12 to 109), both from the original two populations and from two geographically intermediate sites. We have also increased the amount of mtDNA sequence data from 923 bp of the cytochrome oxidase I (*COI*) gene to 1533 bp (combined *COI* and cytochrome oxidase II (*COII*) genes). Recently, Bazin et al. (2006) have shown, as suggested previously (e.g. Ballard and Whitlock, 2004), that mtDNA does not always evolve neutrally and probably undergoes recurrent selective sweeps. However, providing caution is taken, particularly in the inference of population expansion and effective population size, mtDNA remains one of the most useful markers for studies of population genetic structure and history because of its ease of use, lack of recombination and generally high levels of variability. Here, we have also applied an allele-specific screen to examine ITS2 polymorphism in all populations. Genealogical and coalescent-based methods of analysis have been used to infer historical and contemporary rates of gene flow and to explore whether barriers to gene flow are physical, ecological and/or reproductive in nature. With respect to the latter, laboratory crosses were carried out between two of the most geographically separated populations to determine if there are any post-mating barriers to gene flow.

Materials and methods

Mosquito collection and identification

We collected 109 specimens of *An. scanloni* from four locations in Thailand (Figure 1). The earliest collection

Figure 1 *An. scanloni* collection sites. 1. Kanchanaburi, 2. Huay Lae, 3. Sigma Keed and 4. Thung Song.

was made in Kanchanaburi in 1995 with subsequent collections made in 1996 (Kanchanaburi and Thung Song), 2001 (Kanchanaburi, Huay Lae and Sigma Keed) and 2002 (Kanchanaburi, Huay Lae, Sigma Keed and Thung Song) to make a total of 109 mosquitoes. All collection sites were located in tropical forest or plantations, close to limestone karst and usually near to caves (Kanchanaburi, Huay Lae and Thung Song). Collections were also attempted at other sites in peninsular Thailand (11 locations in forest: two karst and nine non-karst) that yielded no *An. scanloni*. Adult females were collected from all sites and assumed to be unrelated. Mosquitoes were identified to the *An. dirus* complex using morphological characters (Rattanakul and Panthusiri, 1994). Mosquitoes were preserved by desiccation and total DNA extracted from individuals using standard phenol-chloroform extraction (Sambrook, 1989). At these collection sites, the predominant *An. dirus* complex species are *An. scanloni* and *An. baimaii* (unpublished data), which have ITS2 ribosomal DNA sequences of very different lengths (Walton et al., 1999). Therefore, we amplified the ITS2 locus to distinguish individuals of these two species based on product length. Sequencing of ITS2 to confirm species was performed on a random sample of mosquitoes, plus any individuals that showed distinct mtDNA sequences.

Amplification and sequencing of mtDNA

Sequences of the *COI* gene were available from a previous study by Walton et al. (2000) for some mosquitoes. For the remaining majority of mosquitoes, a 1230 bp fragment of the *COI* gene was amplified using *An. dirus*-specific primers from the same study, D1664-86 (5'-GTTCCCTTTAATATTAGGAGCACC-3') and D2893-70 (5'-AATAGATGAAGATAATTGTATAGG-3'). A 777 bp fragment of the *COII* gene was amplified in all mosquitoes using the LEU (5'-TCTAATATGGCAGATTAGTGCA-3') and LYS (5'-ACTTGCTTTCAGTCATCTAATG-3') primers (Sharpe et al., 2000). The polymerase chain reactions (PCRs) were performed in 50 µl volumes on a Perkin Elmer 9700 Geneamp thermocycler (PE Applied Biosystems Incorporated, Foster City, CA, USA). Each reaction mix included DNA equivalent to ~1/800 of a mosquito, *COI* primers at 1 µM or *COII* primers at 0.24 µM, 200 µM dNTP, 1.5 mM MgCl₂, 20 mM (NH₄)₂SO₄, 75 mM Tris-HCl (pH 9.0), 0.01% (w/v) Tween and 2.5 U Thermoprime Plus DNA polymerase (CLP, Northampton, UK). For *COI*, initial denaturation was for 5 min at 94°C, followed by 35 cycles of amplification (20 s at 94°C, 30 s at 45°C and 1 min at 72°C) and a final incubation of 3 min at 72°C. For *COII*, initial denaturation was for 5 min at 94°C, followed by 39 cycles of amplification (1 min at 94°C, 1 min at 51°C and 2 min at 72°C) and a final incubation of 10 min at 72°C. The products were purified on Montage columns (Millipore, Billerica, MA, USA). All products were sequenced in both directions using the amplification primers with the exception of the forward direction of *COI*, which was sequenced with the primer D1908-30 (5'-TTATTACTACTGTAATTAATATA-3'). Sequencing reactions were performed by MWG-Biotech (Ebersberg, Germany) and Macrogen Inc. (Seoul, Korea). DNA sequences were assembled using the Sequencer multiple sequence editor programme (Ann Arbor, MI, USA) and checked manually.

Sequencing of ITS2 and screening for a DNA polymorphism

Amplification and sequencing of ITS2 was conducted using the primers (ITS2A and ITS2B) and amplification protocols in Walton *et al.* (1999). The presence of a G or C at the polymorphic nucleotide site 337 was detected in individual mosquitoes using allele-specific primers. The amplification screen was carried out on each sample in two 25 μ l reaction volumes with three primers per tube; the forward and reverse primers ITS2A and ITS2B were included in both the reactions and one reaction had the reverse allele-specific primer ITS2G (5'-GCACGAACG GCGGTCCCCTTC-3') and the other reaction had the alternative reverse allele-specific primer ITS2C (5'-GCAC GAACGGCGGTCCCCTTG-3') to detect a G or C, respectively, at site 337. Each reaction mix included DNA equivalent to $\sim 1/800$ of a mosquito, ITS2A and ITS2B primers at 1 μ M and either the ITS2G or the ITS2C primer at 0.5 μ M, 200 μ M dNTP, 1.5 mM MgCl₂, 20 mM (NH₄)₂SO₄, 75 mM Tris-HCl (pH 9.0), 0.01% (w/v) Tween and 2.5 U Thermoprime Plus DNA polymerase (CLP, Northampton, UK). The initial denaturation was for 5 min at 94°C, followed by 39 cycles of amplification (20 s at 94°C, 30 s at 58°C and 1 min at 72°C) and a final incubation of 10 min at 72°C. The products were visualized on a 1% agarose gel (Melford Labs Ltd, Ipswich, UK) and scored for the presence or absence of a 377 bp band to determine if a G or C base is present at the polymorphic site.

Genetic diversity analysis and population structure

Estimates of genetic diversity were calculated in ARLEQUIN 2.0. The Tamura and Nei model of evolution was used as this model is closest to the HKY model selected for the data by MODELTEST 3.5 (Posada and Crandall, 2001). The population parameter θ was estimated by θ_π , based on the average pairwise number of differences between sequences (Tajima, 1983) and θ_S based on the expected number of segregating sites in a population (Watterson, 1975).

Pairwise F_{ST} and population divergence between populations were calculated in ARLEQUIN 2.0 and the significance tested by 1000 permutations of the haplotypes among groups of populations. Population divergence was measured as the average number of pairwise differences per site between populations (PiXY). A Mantel test (Mantel, 1967) was carried out in ARLEQUIN with 1000 permutations to determine the relationship between genetic distance (F_{ST}) and geographic distance.

Genealogical relationships

A median joining haplotype network (Bandelt *et al.*, 1999) was constructed in the software NETWORK 4.1.0.7 (Fluxus Technology Ltd, 2004) and also using the statistical parsimony method (Templeton *et al.*, 1992) implemented in TCS (Clement *et al.*, 2000). Nested clade analysis was carried out on the TCS network using the AutoNCPA software (available from Mahesh Panchal, University of Reading), which incorporates TCS and GeoDis (Posada *et al.*, 2000), followed by manual interpretation using the inference key.

Recent demographic history

Two tests of neutrality were carried out in ARLEQUIN 2.0 on each population; Tajima's D (Tajima, 1989) that

compares two estimators of θ , θ_π and θ_S , and Fu's F_S (Fu, 1997), which is negative if there is an excess of alleles from that expected for a given estimate of θ_π . Using ARLEQUIN 3.0, the observed distribution of pairwise nucleotide differences among haplotypes (hereafter called mismatch distribution) was constructed for each lineage and compared with the expectations of the sudden expansion model (Rogers and Harpending, 1992).

Maximum likelihood joint estimates of θ , migration rate and exponential population growth were made in LAMARC 2.0 (Kuhner *et al.*, 2005). Migration is expressed as N_m (number of migrants), and exponential growth rate (g) is estimated from $\theta(t) = \theta_{(present)} \exp(-gt)$, where t is a time before the present. Two search strategies were used; 10 initial chains with 2000 sampled genealogies each and two final chains with 20 000 sampled genealogies (employed four times), and 20 initial chains with 2000 sampled genealogies and four final chains with 20 000 sampled genealogies (employed twice). Both initial and final chains used sampling increments of 20.

Laboratory crosses

Wild-caught, blood-fed *An. scanloni* females were collected from Kanchanaburi (site 1) and Sigma Keed (site 3) for laboratory crosses. The mosquitoes were taken for rearing to the Department of Parasitology, Faculty of Medicine, Chiang Mai University. The virgin progeny from the wild-caught females were used to make reciprocal crosses using the forced mating technique (Ow Yang *et al.*, 1963). Within population, control crosses were also made. Eggs, larvae, pupae and adults were counted for each cross and the testes and ovaries of the adults were dissected to check fertility. Following this, F2 crosses were made and analysed in the same way. Larval hatching rate, pupation rate and overall emergence rate were compared between test crosses and control crosses by one-way analysis of variance (ANOVA) using an arcsine transformation of the data. The comparisons were made between one developmental stage and the next, that is, effects at different stages were not compounded.

Results

ITS2 sequence variation and allelic screening

A total of 11 *An. scanloni* individuals were sequenced for ITS2, which confirmed the presence of a polymorphism at nucleotide site 337 reported in Walton *et al.* (1999) and detected another polymorphism at site 755. Site 755 segregated at A and C bases within individuals. However, site 337 was found to be exclusively G or C within each individual and we designed allele-specific primers to screen all the *An. scanloni* for this polymorphism (see Materials and methods). The ITS2C primer gave a band only in the presence of the C polymorphism, but the ITS2G primer also gave a visible but faint band in the presence of the C polymorphism as well as a strong band for the G polymorphism. Although this may suggest the presence of both bases in a single mosquito individual, sequencing of a random sample of 'G' individuals showed only the G base, and it is likely that the presence of a faint band in the 'C' polymorphism detection PCR is due to some nonspecific primer binding although the presence of a low frequency of 'C' sequences cannot be discounted.

Table 1 Summary haplotype data for each site

Location	N	Haplotypes	S	$\theta_{\pi} \pm s.d.$	$\theta_s \pm s.d.$
1. Kanchanaburi	44	1(5), 2, 3 (2), 4 (5), 5, 6, 7, 8 (3), 9, 10 , 11 (2), 12 (2), 13 (2), 14, 15 (2), 16 (2), 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28	53	0.00516 ± 0.00272	0.00795 ± 0.00247
2. Huay Lae	19	10 (3), 29, 30 (4), 31, 32 (3), 33, 34 (3), 35, 36 (2)	32	0.00608 ± 0.00326	0.00597 ± 0.00226
3. Sigma Keed	12	37 (2), 38, 39 , 40 , 41, 42 (2), 43, 44, 45, 46	26	0.00353 ± 0.00206	0.00562 ± 0.00238
4. Thung Song	34	37 (4), 39 (3), 40 (2), 47(2), 48, 49, 50 (3), 51, 52, 53, 54 (3), 55, 56, 57, 58, 59 (2), 60, 61, 62, 63, 64, 65	35	0.00524 ± 0.00278	0.00558 ± 0.00189
All	109		94	0.00882 ± 0.00442	0.01165 ± 0.00299

Haplotypes in bold occur at more than one location. A number in parentheses indicates the frequency of a haplotype at a particular site. S is the number of segregating sites. θ_{π} and θ_s are estimates of genetic diversity defined in Materials and methods.

mtDNA sequence variation and genetic diversity

A total of 109 individuals were sequenced for *COI* and *COII*; 44 from Kanchanaburi, 19 from Huay Lae, 12 from Sigma Keed and 34 from Thung Song (Table 1). The sequences were concatenated with reading frame maintained to give a total length of 1533 bp for all subsequent analysis. The sequences are deposited in the EMBL database with accession numbers AM180766–AM180874. The sequences are AT-rich (72% A and T bases), which is fairly typical for insect taxa (Simon *et al.*, 1994). Nucleotide substitutions were identified at 94 of the 1533 sites (6.13%), the majority of which were transitions (77.66%). Haplotype diversity was high, with 65 haplotypes found among the 109 individuals. The distribution of haplotypes among the four populations and other summary statistics are shown in Table 1. The levels of genetic diversity are similar between populations, and are similar to those seen in *An. baimaii* but slightly higher than those of *An. dirus* (Walton *et al.*, 2000).

Population structure

The allele-specific primers for ITS2 indicated differentiation of the two northern populations from the two southern populations. All the individuals from Sigma Keed and Thung Song were determined to have a C base at site 337. Individuals from Huay Lae and Kanchanaburi all showed a G base at this site with the exception of one individual from Kanchanaburi (*COI-COII* haplotype 8). Two individuals from Huay Lae failed to amplify (*COI-COII* haplotypes 29 and 30) so their ITS2 haplotypes remain undetermined.

Pairwise F_{ST} s estimated using the mitochondrial sequences revealed high levels of differentiation (0.47–0.63) among all populations except Sigma Keed and Thung Song and average pairwise divergence (PiXY) between populations showed the same pattern (Table 2). The Mantel test was nonsignificant ($P = 0.212$) indicating that there is no relationship between geographic distance and genetic distance. However, it should be noted that as there are only four populations, this test will have relatively low power to detect this.

Genealogical relationships between mitochondrial haplotypes

The median joining haplotype network was well resolved with few ambiguities (Figure 2). Four highly divergent lineages were identified (labelled I–IV), separated by 7–18 mutations, and clear geographical structuring was apparent. The Thung Song haplotypes are found only in lineages I and II, and the Sigma Keed haplotypes are almost all confined to lineage I. The Kanchanaburi

Table 2 Pairwise F_{ST} (below diagonal) and PiXY (above diagonal)

Population	Kanchanaburi	Huay Lae	Sigma Keed	Thung Song
Kanchanaburi		0.0112**	0.0106**	0.0097**
Huay Lae	0.515**		0.0132**	0.0120**
Sigma Keed	0.568**	0.626**		0.0049
Thung Song	0.467**	0.538**	0.102*	

* $P < 0.05$, ** $P < 0.0001$ from permutation tests.

and Huay Lae haplotypes were almost entirely restricted to single lineages (IV and III, respectively). One individual from Kanchanaburi (*COI-COII* haplotype 8) was in lineage II (an almost exclusively southern lineage) and also had the southern ITS2 type suggesting that it is a long-distance migrant from the south. The TCS statistical parsimony network was identical to the median joining network. The NCPA nesting algorithm culminated in three 4-step clades; 4-1 containing lineage IV, 4-2 containing lineages II and III and 4-3 containing lineage I (data not shown). NCPA inferences were made on the basis of a non-continuous geographical distribution, which can be justified by the absence of *An. scanloni* at intermediate sampling sites. NCPA found no significant associations between geography and haplotypes below the three-step clade level, except for allopatric fragmentation in lineage IV owing to the clustering of Huay Lae haplotypes at one of the tips. Allopatric fragmentation was inferred in clade 4-2, and in the overall cladogram.

Recent demographic history

The Fu's F_S test of neutrality was negative and significant for all populations except Huay Lae (Table 3); this indicates a high excess of alleles, which is expected after a recent population expansion (but can also be due to a selective sweep or background selection). Tajima's D was significant only for the Sigma Keed population (Table 3) but, since demographic expansion is detected using other methods (see below), this lack of significance is probably due to admixture of mitochondrial lineages in the populations, which will tend to increase θ_{π} relative to θ_s .

The four very divergent lineages do not correspond directly to populations, indicating that each mtDNA lineage has a distinct and separate history. We therefore constructed mismatch distributions for the mtDNA lineages (rather than for the populations) with the aim of inferring whether the lineages had experienced expansions rather than trying to infer demographic expansions in the contemporary populations, which contain

Figure 2 Median joining haplotype network for four populations of *An. scanloni*. Circles represent haplotypes and are proportional to frequency. Branches between haplotypes are one mutational step unless numbered. Southern populations are to the left of the dotted line, northern populations to the right.

Table 3 The neutrality tests Fu's F_S and Tajima's D for each population

Population	F_S	P	Tajima's D	P
Kanchanaburi	-10.95	0.001*	-1.260	0.075
Huay Lae	1.81	0.767	0.011	0.530
Sigma Keed	-3.37	0.046*	-1.676	0.034*
Thung Song	-6.51	0.017*	-0.247	0.475
All	-24.16	0.000**	-0.824	0.235

Significance was tested by permutation. * $P < 0.05$, ** $P < 0.001$.

haplotypes of mixed ancestry. For all lineages, the mismatch distributions were smooth and unimodal, fitting a model of sudden expansion (Figure 3). The output from LAMARC also consistently showed rapid growth for all populations. The results of a representative run using the longer search strategy are shown in Table 4. Present day θ values were high (0.0068–0.0646) for each population, indicative of substantial effective female population sizes. Migration rates were not completely consistent across runs but some general patterns were apparent that are exemplified in Table 4. Generally, migration rates were high between the two southern populations, (particularly from Thung Song into Sigma Keed), and lower between the two northern populations. The highest degree of isolation is between the two northern and the two southern populations, where migration of zero was inferred in 26 out of 48 comparisons. In contrast a zero migration rate was never obtained between the two southern or the two northern populations. In Table 4, the migration that is inferred from Thung Song into Kanchanaburi is most likely a signal of the single Kanchanaburi haplotype in lineage II. However, the apparent migration from the northern populations into Thung Song seems unreasonable as

they do not share any lineages. One assumption of LAMARC is that the migration situation is stable so individuals must theoretically be able to reach all populations, either directly or via other populations, so only one population at the most can have zero immigrants (Kuhner *et al.*, 2005). This low but non-zero migration from Thung Song into the north is probably an artefact of this assumption in LAMARC.

Laboratory crosses

The results of the F1 test crosses between northern and southern individuals and the F1 control crosses are reported in Table 5. The F2 crosses are not reported as numbers were low due to the inherent problems of maintaining lines of this species in the laboratory. All progenies of the F1 and F2 crosses were normal and fertile. For the F1 crosses, the ANOVA detected a high level of experimental variation (i.e. between brood variability), which would make it difficult to detect any subtle effects. The 12 comparisons made of larval hatching, pupation and emergence rates showed no differences between test and control crosses except that the emergence rate of the $K \times S$ crosses was lower than both the $S \times S$ and $K \times K$ control crosses (Table 5). However, these do not retain significance when taking into account the multiple comparisons made. The remaining 10 ANOVA showed no evidence of post-mating barriers to reproduction between the two populations studied with no values approaching significance.

Discussion

Allopatric fragmentation and habitat islands

Four highly divergent mtDNA lineages have been found in the four *An. scanloni* populations (Figure 2), including

Figure 3 Observed mismatch distributions among haplotypes in *An. scanloni* for the four divergent lineages. Bars are observed distribution, the line shows the distribution simulated under a sudden expansion model. The *P*-values are from the sum of squared deviations goodness of fit test for the sudden expansion model (Rogers and Harpending, 1992).

a lineage from Huay Lae and Kanchanaburi that is separated from the others by at least 12 mutations that was not sampled previously. This high level of intraspecies divergence over small distances is not typical of *Anopheles* species. For example, one of the most well-known physical barriers to gene flow in *An. gambiae*, the Rift Valley, was identified from a maximum F_{ST} in mtDNA of 0.123 between populations 700 km apart (Lehmann *et al.*, 2000), which is considerably lower than all pairwise F_{ST} comparisons in *An. scanloni* except

between the Sigma Keed and Thung Song populations. *An. scanloni*'s sibling species, *An. dirus*, shows no genetic structure in mtDNA or microsatellites across a much greater geographical distance than those sampled here (Walton *et al.*, 2001; unpublished data). In the absence of geographic barriers, most *Anopheles* species exhibit inferred panmixia to about 200 km (Conn *et al.*, 1999; De Merida *et al.*, 1999; Donnelly and Townson, 2000; Fairley *et al.*, 2000). Isolation by distance is then usually seen, but this does not appear to be the case in *An. scanloni*.

Allopatric fragmentation is indicated by the NCPA as the likely explanation for the highly divergent lineages in *An. scanloni*. It has previously been noted (Poopittayasatporn and Baimai, 1995) and our collection records concur (see Materials and methods), that *An. scanloni* is associated with limestone karst, which is patchily distributed in the Thai peninsular (Figure 4). If *An. scanloni* is strictly dependent on these 'islands' of habitat there will be restrictions both on population size within islands and gene flow between them. F_{ST} is significant between all populations, but in the two southern populations (Sigma Keed and Thung Song) LAMARC infers historical non-negligible gene flow between them in both directions (Table 4). The distance between these populations is small (~110 km) and it can be seen in Figure 4 that there are small patches of limestone between the populations. *An. scanloni* has been sampled previously from at least one of these intermediate locations (Baimai *et al.*, 1988a) and it is likely that at present these facilitate a low level of dispersal. Between the northern populations (Huay Lae and Kanchanaburi) gene flow is low, and appears to be recent and probably ongoing, as many of the haplotypes involved are at the tips of lineages and therefore more recently derived. Limestone patches do exist between these populations but there is a large limestone-free area to the north of Huay Lae, which may act as a barrier to dispersal (Figure 4).

This restriction to karst habitat will result in the typical metapopulation model of population extinction and recolonization (Hastings and Harrison, 1994). It might be expected that gene flow would be more likely between geographically closer habitat islands, but the Mantel test showed no correlation between geographic and genetic distance in *An. scanloni*. This implies that genetic drift plays a greater role than gene flow in generating divergence between populations (Hutchison and Templeton, 1999). Wade and McCauley (1988) explored the genetic consequences of two models of colonization; 'propagule-pool' where the founding individuals arise from the same population, and 'migrant-pool' where founders arise from random populations and may be genetically diverse. The genetic bottlenecks (and consequential increased drift) associated with the first model of colonization can contribute to high levels of divergence as is seen in *An. scanloni*, whereas colonization from more than one source (as in the second model) can lead to lineage mixing, and lowering of between population differentiation as observed in the Thung Song and Sigma Keed populations. High levels of intraspecies divergence have been observed in other taxa, which are dependent on patchy habitat, for instance Colorado springsnails and Iowa Pleistocene snails (Ross, 1999; Hurt, 2004).

Table 4 Estimates of migration (shown as N_m , the number of migrants moving from one population into another population), θ , and exponential growth rate (g) from LAMARC

Population		Into				θ	g
		Kanchanaburi	Huay Lae	Sigma Keed	Thung Song		
From	Kanchanaburi		2.57	0.00	0.87	0.0514	3318
	Huay Lae	2.30		0.00	0.87	0.0068	3996
	Sigma Keed	0.00	0.00		7.95	0.0646	7244
	Thung Song	2.30	0.00	46.39		0.0436	2208

Results shown are from a representative run employing the longer search strategy. Shaded cells denote migration estimates between the two northern and two southern populations.

Table 5 Results from laboratory crosses between Kanchanaburi (K) and Sigma Keed (S) individuals

Crosses ^a		Eggs				Emergence rate ^b (%)		
		No. of broods	Total	Mean no./brood	Hatching rate ^b (%)	Pupation rate ^b (%)	F	M
K × S	2	124	62.0	77.4	74.2	29.8	37.9	67.7 ^c
S × K	7	709	101.3	46.7	43.4	19.3	22.8	42.2
<i>Control crosses</i>								
S × S	3	293	97.7	41.0	38.2	16.4	21.5	37.9
K × K	2	194	97.0	62.9	60.3	31.4	25.8	57.2

^aAll crosses are female × male.

^bEach rate was calculated by (total number of individuals that reached each developmental stage)/(total number of eggs) × 100.

^cThe emergence rate (relative to the pupation rate) of the K × S cross (i.e. a drop from 74.2 to 67.7%) was lower than that of both control crosses, K × K ($P = 0.04$) and S × S ($P = 0.05$) using one-way ANOVA.

Pleistocene climatic effects

Another factor known to have a significant effect on the genetic structure of many populations is Pleistocene climatic change (Hewitt, 2000). In tropical regions such as Thailand the climate was cooler and drier during glacial periods, which may have resulted in the replacement of most of the tropical forest with pine forest and savannah (Penny, 2001; Hope *et al.*, 2004). The effect on peninsular Thailand is less clear; pollen core evidence from western Java points to the survival of forests, which has been extrapolated to peninsular Malaysia (Hope *et al.*, 2004), but archaeology of the Thai peninsular reports a drop in arboreal markers and the presence of grazing animals, suggesting a mosaic of forest and grassland (White *et al.*, 2004). Despite this, a recent reanalysis of various geographical evidence from the last glacial period supports the presence of a 'savannah corridor', which at its most extreme may have spanned the Thai-Malay peninsular (Bird *et al.*, 2005). Indirect support for a lack of forest habitat comes from studies of two forest-related species in Thailand, *An. baimaii* (unpublished data) and *Simulium tani* (Pramual *et al.*, 2005). These species have reduced genetic diversity in the south of Thailand relative to the north with evidence of recent expansion indicating that the peninsular region was recently re-colonised from the north, and that prior to this forest habitat was scarce or absent. However, another study of forest and non-forest dependent bats found inconclusive relationships between ecotype and population history (Campbell *et al.*, 2006). In the case of *An. scanloni*, a pattern of lower genetic diversity in the south than the north is not seen which suggests that it has been less affected by climatic change than *An. baimaii* and *S. tani*. It is possible that the limestone karst, which

Figure 4 Outline map showing areas of limestone (in grey) on the Thai peninsular. Collection sites from this study are marked by circles. Adapted from a geological map from the website of the Department of Mineral Resources, Thailand (<http://www.dmr.go.th>) (in Thai).

has unique ecosystems often associated with aquifers (Wong *et al.*, 2001), may have been able to provide a moist microclimate, which allowed populations of *An. scanloni* to persist. Despite this, overall reduced forest cover and greater aridity during glacial periods might be expected to exacerbate the extinction component of the metapopulation dynamic and to restrict gene flow between habitat islands. The signal of population expansion seen from LAMARC and NCPA may therefore be related to changes in the overall extent of forest cover and level of aridity associated with climate change. Alternatively (or in addition to), the genetic signals of expansion could be attributed to the expansions of mitochondrial lineages following metapopulation founder events. Finally, since only a single marker is used, we cannot exclude the possibility that the expansion signal does not reflect population history but is actually a locus-specific effect. Background selection cannot be excluded as a possible explanation, although selective sweeps, which may occur commonly in mtDNA (see Introduction), are unlikely here since they would have to have occurred independently in several populations/lineages.

The north–south divide: dispersal or vicariance?

The haplotype network and LAMARC results for *An. scanloni* show that there is very little gene flow between the two northern and the two southern populations. The northern and southern populations being fixed for different ITS2 haplotypes further suggests greater connectivity within, rather than between, these groupings (although as there are only four sampled populations, the possibility that this pattern of haplotype fixation could just be due to chance fixation of ancestral polymorphism in each of the four populations cannot be excluded). A possible barrier to gene flow between the north and south is the paucity of karst habitat between the northern and southern populations (Figure 4). In addition, at the narrowing of the Thai peninsular known as the Isthmus of Kra (Figure 4), there are two other factors that could affect gene flow. First, the mountains north of this point, which almost span the peninsular, may be a physical barrier to dispersal. Second, there is a transition between the Indochinese and Sundaic biotas across the Isthmus corresponding to a shift from mixed deciduous forests in the north to tropical evergreen rain forest in the south, as well as differences in seasonal rainfall (Hughes *et al.*, 2003; Woodruff, 2003). The different ecological conditions between the north and south may lead to differential adaptation as has been inferred to have occurred in the two distinct chromosomal forms of *S. tani* in the north and south of Thailand (Pramual *et al.*, 2005). If the northern and southern populations of *An. scanloni* are ecologically differentiated, this could be an important barrier to gene flow between the regions as long-distance migrants or their hybrid offspring may have lower fitness than native individuals. In this context, it is interesting to note that rare long-distance migration between north and south is possible (based on the Kanchanaburi individual with a southern ITS2 haplotype, that falls into southern lineage II). However, this may not necessarily result in successful gene flow as this would be evidenced, for example, by a mixture of northern and southern mitochondrial or nuclear markers in the same individual.

There is evidence that rising sea levels may have created seaways at the Isthmus of Kra at various times during the Miocene (24–13 Myr BP) and Pliocene (5.5–4.5 Myr BP) with anecdotal evidence that these marine transgressions may have occurred as recently as the first century AD (Woodruff, 2003). Vicariance in the form of a seaway at the Isthmus of Kra may therefore have contributed to north–south differentiation in *An. scanloni*. It is highly likely that seaways at the Isthmus have played a role in Pliocene speciation or differentiation; for example, the giant freshwater prawn (*Macrobrachium rosenbergii*) has two divergent mitochondrial lineages on either side of the Isthmus of Kra that have been attributed to the seaways (de Bruyn *et al.*, 2005). However, if the Isthmus of Kra has played some role in population divergence in *An. scanloni*, it cannot easily explain the origin of at least four highly divergent lineages, indicating that a metapopulation dynamic associated with karst habitat islands is the most important factor in defining the population structure in this species.

Are the high levels of divergence indicative of speciation?

Based on the very high levels of differentiation at microsatellite and mtDNA loci between the Kanchanaburi and Thung Song populations, it was previously suggested that what is recognized as *An. scanloni* may comprise more than one species (Walton *et al.*, 2001). It has also been observed that *An. scanloni* from Kanchanaburi displays a morphological difference from other populations on the hind tibia (Sallum *et al.*, 2005), but it was not reported whether other northern populations were examined. Here, we have inferred extremely little or no gene flow between the north and the south that may be due to physical barriers and/or the ecological transition near the Isthmus of Kra. Although this may indicate a northern and southern species, no post-mating barriers to reproduction were detected from cross-mating experiments, but since these experiments are carried out using a forced-mating technique we know nothing about the possible existence of any premating barriers to reproduction. Defining species in *Anopheles* can be difficult as even well-recognized species can have substantial gene flow and/or share ancestral polymorphism at some loci. The ITS2 locus is often used for species identification in *Anopheles* as it tends to vary even between very closely related species. Even in ITS2, sometimes few differences are seen between well-defined species (e.g. *An. scanloni* and *An. dirus* have only one fixed difference and vary in the repeat number of a microsatellite, and between *An. gambiae* and *An. arabiensis* there are seven fixed differences (Gentile *et al.*, 2002)). However, generally small variations in ITS2 can be attributed to lack of gene flow between populations (Onyabe and Conn, 1999; Beebe *et al.*, 2000; Di Luca *et al.*, 2004). The latter appears to be the most likely explanation in *An. scanloni*.

Altogether, despite the clear restrictions to gene flow between the northern populations and the southern populations, we have insufficient evidence to justify them as distinct species. However, the long-term lack of gene flow between these populations does make this a potentially interesting study system for allopatric speciation as it is very rare to see this level of intraspecies

divergence, which may represent the very early stages of speciation. Furthermore, the isolation and distinctiveness of the northern and southern populations means that they should be considered separately in epidemiological studies, which aim to determine characteristics related to malaria transmission (e.g. vector competence, anthropogenicity, resting behaviour), and for malaria control (e.g. genetic modification using malaria refractory genes).

Conclusions

We have observed four highly divergent lineages in *An. scanloni* that are most likely due to a dependence on limestone habitat islands that results in a metapopulation dynamic. Historical environmental change and marine transgressions may also contribute to population extinction, expansion and divergence. Particularly, high restriction to gene flow was observed between the northern and the southern populations but there is no evidence that these represent separate species. The low levels of gene flow between populations show that even in species such as *Anopheles*, which have a good potential for dispersal, habitat fragmentation can cause severe restrictions to gene flow.

Acknowledgements

We are indebted to the staff of the Office of Disease Prevention and Control in Chiang Mai, Thailand (in particular Ruksakul Kanthawong) and to Kay Nolan, for valuable assistance in the collection of mosquitoes for this study. We also thank Damrongpan Thongwat for his assistance in the insectary. This work was funded by a NERC studentship to SM O'Loughlin and the Japan Society for Promotion of Sciences, Grant No. 16300086.

References

- Baimai V (1988). Population cytogenetics of the malaria vector *Anopheles leucosphyrus* group. *Southeast Asian J Trop Med Public Health* **19**: 667–680.
- Baimai V, Andre RG, Harrison BA, Kijchalao U, Panthusiri L (1987). Crossing and chromosomal evidence for 2 additional sibling species within the taxon *Anopheles dirus* Peyton and Harrison (Diptera, Culicidae) in Thailand. *Proc Entomol Soc Washington* **89**: 157–166.
- Baimai V, Kijchalao U, Sawadwongporn P, Green CA (1988a). Geographic distribution and biting behaviour of four species of the *Anopheles dirus* complex (Diptera: Culicidae) in Thailand. *Southeast Asian J Trop Med Public Health* **19**: 151–161.
- Baimai V, Poopittayasatoporn A, Kijchalao U (1988b). Cytological differences and chromosomal rearrangements in 4 members of the *Anopheles dirus* complex (Diptera, Culicidae). *Genome* **30**: 372–379.
- Ballard JWO, Whitlock MC (2004). The incomplete natural history of mitochondria. *Mol Ecol* **13**: 729–744.
- Bandelt HJ, Forster P, Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**: 37–48.
- Bazin E, Glemin S, Galtier N (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**: 570–572.
- Beebe NW, Cooper RD, Foley DH, Ellis JT (2000). Populations of the south-west Pacific malaria vector *Anopheles farauti* s.s. revealed by ribosomal DNA transcribed spacer polymorphisms. *Heredity* **84**: 244–253.
- Bird MI, Taylor D, Hunt C (2005). Environments of insular Southeast Asia during the last glacial period: a savanna corridor in Sundaland? *Quatern Sci Rev* **24**: 2228–2242.
- Campbell P, Schneider CJ, Adnan AM, Zubaid A, Kunz TH (2006). Comparative population structure of *Cynopterus* fruit bats in peninsular Malaysia and southern Thailand. *Mol Ecol* **15**: 29–47.
- Catteruccia F, Nolan T, Loukeris TG, Blass C, Savakis C, Kafatos FC et al. (2000). Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* **405**: 959–962.
- Clement M, Posada D, Crandall KA (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1659.
- Collins FH, Kamau L, Ranson HA, Vulule JM (2000). Molecular entomology and prospects for malaria control. *Bull World Health Organ* **78**: 1412–1423.
- Conn JE, Rosa-Freitas MG, Luz SL, Momen H (1999). Molecular population genetics of the primary neotropical malaria vector *Anopheles darlingi* using mtDNA. *J Am Mosq Control Assoc* **15**: 468–474.
- Coyne JA, Orr HA (1998). The evolutionary genetics of speciation. *Philos Trans R Soc London B Biol Sci* **353**: 287–305.
- de Bruyn M, Nugroho E, Hossain MM, Wilson JC, Mather PB (2005). Phylogeographic evidence for the existence of an ancient biogeographic barrier: the Isthmus of Kra Seaway. *Heredity* **94**: 370–378.
- De Merida AM, Palmieri M, Yurrita M, Molina A, Molina E, Black WCt (1999). Mitochondrial DNA variation among *Anopheles albimanus* populations. *Am J Trop Med Hyg* **61**: 230–239.
- Di Luca M, Boccolini D, Marinucci M, Romi R (2004). Intrapopulation polymorphism in *Anopheles messeae* (*An. maculipennis* complex) inferred by molecular analysis. *J Med Entomol* **41**: 582–586.
- Donnelly MJ, Townson H (2000). Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in Eastern Africa. *Insect Mol Biol* **9**: 357–367.
- Fairley TL, Renaud TM, Conn JE (2000). Effects of local geographic barriers and latitude on population structure in *Anopheles punctipennis* (Diptera: Culicidae). *J Med Entomol* **37**: 754–760.
- Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- Gentile G, della Torre A, Maegga B, Powell JR, Caccone A (2002). Genetic differentiation in the African malaria vector, *Anopheles gambiae* s.s., and the problem of taxonomic status. *Genetics* **161**: 1561–1578.
- Harbach RE (2004). The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bull Entomol Res* **94**: 537–553.
- Hastings A, Harrison S (1994). Metapopulation dynamics and genetics. *Annu Rev Ecol Syst* **25**: 167–188.
- Hewitt G (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.
- Hope G, Kershaw AP, Kaars Svd, Xiangjun S, Liew P-M, Heusser LE et al. (2004). History of vegetation and habitat change in the Austrial-Asian region. *Quatern Int* **118–119**: 103–126.
- Hughes JB, Round PD, Woodruff DS (2003). The Indochinese-Sundaic faunal transition at the Isthmus of Kra: an analysis of resident forest bird species distributions. *J Biogeogr* **30**: 569–580.
- Hurt CR (2004). Genetic divergence, population structure and historical demography of rare springsnails (*Pyrgulopsis*) in the lower Colorado River basin. *Mol Ecol* **13**: 1173–1187.
- Hutchison DW, Templeton AR (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**: 1898–1914.

- Kuhner MK, Yamato J, Beerli P, Smith L, Rynes E, Walkup E et al. (2005) LAMARC v 2.0. <http://evolution.gs.washington.edu/lamarc.html>. University of Washington.
- Lehmann T, Blackston CR, Besansky NJ, Escalante AA, Collins FH, Hawley WA (2000). The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya: the mtDNA perspective. *J Hered* **91**: 165–168.
- Mantel N (1967). Detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209.
- Onyabe DY, Conn JE (1999). Intragenomic heterogeneity of a ribosomal DNA spacer (ITS2) varies regionally in the neotropical malaria vector *Anopheles nuneztovari* (Diptera: Culicidae). *Insect Mol Biol* **8**: 435–442.
- Ow Yang CK, Sta Maria FL, Wharton RW (1963). Maintenance of a laboratory colony of *Anopheles maculatus* Theobald by artificial mating. *Mosq News* **23**: 34–35.
- Penny D (2001). A 40,000 year palynological record from north-east Thailand; implications for biogeography and palaeo-environmental reconstruction. *Palaeogeogr Palaeoclimatol Palaeoecol* **171**: 97–128.
- Peyton E (1989). A new classification for the Leucosphyrus group of *Anopheles* (Cellia). *Mosq Systematics* **21**: 197–205.
- Poopittayasataporn A, Baimai V (1995). Polytene chromosome relationships of 5 species of the *Anopheles dirus* complex in Thailand. *Genome* **38**: 426–434.
- Posada D, Crandall KA (2001). Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Mol Biol Evol* **18**: 897–906.
- Posada D, Crandall KA, Templeton AR (2000). GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol Ecol* **9**: 487–488.
- Pramual P, Kuvangkadilok C, Baimai V, Walton C (2005). Phylogeography of the black fly *Simulium tani* (Diptera: Simuliidae) from Thailand as inferred from mtDNA sequences. *Mol Ecol* **14**: 3989–4001.
- Rattananarithikul R, Konishi E, Linthicum KJ (1996). Observations on nocturnal biting activity and host preference of anophelines collected in southern Thailand. *J Am Mosq Control Assoc* **12**: 52–57.
- Rattananarithikul R, Panthusiri P (1994). Illustrated keys to the medically important mosquitoes of Thailand. *Southeast Asian J Trop Med Public Health* **25**: 1–66.
- Rogers AR, Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* **9**: 552–569.
- Ross TK (1999). Phylogeography and conservation genetics of the Iowa Pleistocene snail. *Mol Ecol* **8**: 1363–1373.
- Sallum MAM, Peyton EL, Wilkerson RC (2005). Six new species of the *Anopheles leucosphyrus* group, reinterpretation of *An. elegans* and vector implications. *Med Vet Entomol* **19**: 158–199.
- Sambrook J (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory: Cold Spring Harbor, NY.
- Scanlon JE, Sandhinand U (1965). The distribution and biology of *Anopheles balabacensis* in Thailand. *J Med Entomol* **2**: 61–69.
- Sharpe RG, Harbach RE, Butlin RK (2000). Molecular variation and phylogeny of members of the Minimus Group of *Anopheles* subgenus Cellia (Diptera: Culicidae). *System Entomol* **25**: 263–272.
- Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (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.
- Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Templeton AR, Crandall KA, Sing CF (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.
- Turelli M, Barton NH, Coyne JA (2001). Theory and speciation. *Trends Ecol Evol* **16**: 330–343.
- Wade MJ, McCauley DE (1988). Extinction and recolonization – their effects on the genetic differentiation of local populations. *Evolution* **42**: 995.
- Walton C, Handley JM, Collins FH, Baimai V, Harbach RE, Deesin V et al. (2001). Genetic population structure and introgression in *Anopheles dirus* mosquitoes in south-east Asia. *Mol Ecol* **10**: 569–580.
- Walton C, Handley JM, Kuvangkadilok C, Collins FH, Harbach RE, Baimai V et al. (1999). Identification of five species of the *Anopheles dirus* complex from Thailand, using allele-specific polymerase chain reaction. *Med Vet Entomol* **13**: 24–32.
- Walton C, Handley JM, Tun-Lin W, Collins FH, Harbach RE, Baimai V et al. (2000). Population structure and population history of *Anopheles dirus* mosquitoes in southeast Asia. *Mol Biol Evol* **17**: 962–974.
- Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**: 256–276.
- White JC, Penny D, Kealhofer L, Maloney B (2004). Vegetation changes from the late Pleistocene through the Holocene from three areas of archaeological significance in Thailand. *Quatern Int* **113**: 111–132.
- Wilkinson RN, Gould DJ, Boonyakanist P, Segal HE (1978). Observations on *Anopheles balabacensis* (Diptera: Culicidae) in Thailand. *J Med Entomol* **14**: 666–671.
- Wong T, Hamilton-Smith E, Chape S, Friederich H (2001). *Proceedings of the Asia-Pacific Forum on Karst Ecosystems and World Heritage*. IUCN: Gland, Switzerland.
- Woodruff DS (2003). Neogene marine transgressions, palaeogeography and biogeographic transitions on the Thai-Malay Peninsula. *J Biogeogr* **30**: 551–567.