ORIGINAL ARTICLE

Spatial genetic structure of *Aedes aegypti* mosquitoes in mainland Southeast Asia

Thaung Hlaing,[1,2] Willoughby Tun-Lin,[2] Pradya Somboon,[3] Duong Socheat,[4] To Setha,[4] Sein Min,[2] Sein Thaung,[2] Okorie Anyaele,[5] Babaranda de Silva,[6] Moh Seng Chang,[7] Anil Prakash,[8] Yvonne Linton[9] and Catherine Walton[1]

1 Faculty of Life Sciences, University of Manchester, Manchester, UK
2 Medical Entomology Research Division, Department of Medical Research (Lower Myanmar), Yangon, Myanmar
3 Department of Parasitology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand
4 National Centre for Malaria, Parasitology and Entomology, Phnom Penh, Cambodia
5 Entomology Unit, Department of Zoology, University of Ibadan, Ibadan, Nigeria
6 Department of Zoology, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
7 WHO – Western Pacific Regional Office, Phnom Penh, Cambodia
8 Regional Malaria Research Centre, Indian Council of Medical Research, Dibrugarh, Assam, India
9 Natural History Museum, London, UK

Abstract

Aedes aegypti mosquitoes originated in Africa and are thought to have spread recently to Southeast Asia, where they are the major vector of dengue. Thirteen microsatellite loci were used to determine the genetic population structure of *A. aegypti* at a hierarchy of spatial scales encompassing 36 sites in Myanmar, Cambodia and Thailand, and two sites in Sri Lanka and Nigeria. Low, but significant, genetic structuring was found at all spatial scales (from 5 to >2000 km) and significant F_{IS} values indicated genetic structuring even within 500 m. Spatially dependent genetic-clustering methods revealed that although spatial distance plays a role in shaping larger-scale population structure, it is not the only factor. Genetic heterogeneity in major port cities and genetic similarity of distant locations connected by major roads, suggest that human transportation routes have resulted in passive long-distance migration of *A. aegypti*. The restricted dispersal on a small spatial scale will make localized control efforts and sterile insect technology effective for dengue control. Conversely, preventing the establishment of insecticide resistance genes or spreading refractory genes in a genetic modification strategy would be challenging. These effects on vector control will depend on the relative strength of the opposing effects of passive dispersal.

Introduction

Dengue is a major re-emerging public health problem in the tropics and subtropics outside Africa. An estimated 50 million people are infected annually, which can lead to hospitalization and death in severe cases (WHO 2007). The dengue virus is transmitted to humans by mosquitoes, primarily *Aedes aegypti*. Southeast Asia has experienced severe dengue epidemics since the 1950s, particularly in urban areas (Hammon et al. 1960; Gubler 1998a,b; Muto 1998; Prasittisuk et al. 1998; Thu et al. 2004; Kittayapong 2006). This great burden of disease in Southeast Asia stems from the invasion and spread of *A. aegypti* throughout the region that is thought to have occurred following its introduction from Africa by the shipping trade in the late 19th century (Smith 1956; Tabachnick and Powell 1979). *Aedes aegypti* is a highly domestic species that not only feeds on humans but also breeds in and around human habitation, laying its eggs in water-filled, man-made containers (e.g. water storage jars

and discarded containers). This species, and the disease it transmits, have therefore proliferated in conjunction with human population growth, economic development, increased mobility and uncontrolled urbanization (Gubler and Clark 1995; Gubler 1998b; Guzman and Kouri 2002; WHO 2007).

At present, the reduction of dengue transmission relies on vector control (WHO 2008). For this reason, it is essential to gain a good understanding of the genetic population structure of *A. aegypti*, and the factors underlying this, particularly gene flow (Gooding 1996; Ravel et al. 2001). Genetic population structure usually results from a combination of several contemporary and historical processes such as dispersal ability of the species, mating patterns, environmental barriers to dispersal and demographic history (Balloux and Lugon-Moulin 2002). Disentangling the different roles of these factors in determining population structure is needed if the information is to be useful for vector control. For example, information on the rate of ongoing gene flow would help to predict the spread of insecticide resistance genes, the presence of which can be a major limiting factor to the success of control campaigns (Pasteur et al. 1995; Mousson et al. 2002). An understanding of gene flow and environmental barriers to mosquito dispersal is also essential if plans to genetically modify vector populations are to be realized (Scott et al. 2002; James et al. 2006).

Although it is generally the contemporary processes such as gene flow that are most relevant for vector control, it is also important to determine the demographic history of *A. aegypti*. This could not only help in enabling us to distinguish more clearly the contemporary factors shaping population structure, it could also help to elucidate the genetic basis for the apparent geographic variation in the susceptibility of *A. aegypti* to dengue virus (Gubler et al. 1979; Bosio et al. 1998). There have been several mtDNA-based studies that might have been expected to be informative of demographic history. For example, the apparent presence of two genetically divergent lineages of mtDNA has often been interpreted to indicate that colonization by *A. aegypti* was from multiple, divergent source populations (Gorrochotegui-Escalante et al. 2002; Bosio et al. 2005; Herrera et al. 2006; Scarpassa et al. 2008). However, the presence of nuclear copies of mtDNA (Numts) in *A. aegypti* (Hlaing et al. 2009) means that such inferences may be unreliable. Consequently, as yet we know relatively little about the population history of this species.

Previous large-scale studies of genetic population structure of *A. aegypti* using allozymes (Tabachnick and Powell 1979; Powell et al. 1980; Wallis et al. 1983; Failloux et al. 2002) indicate structuring on a worldwide scale. The overall levels of genetic differentiation are relatively low,

consistent with the recent spread of this species throughout the tropics (Tabachnick and Powell 1979). More recently, microsatellite markers have been used for determining the genetic population structure in *A. aegypti*. Most of these studies have been on small spatial scales: regions within a country, i.e. Mexico (Ravel et al. 2001), Ivory Coast (Ravel et al. 2002), Cameroon (Paupy et al. 2008), or within cities, e.g. Ho Chi Minh City in Vietnam, Phnom Penh in Cambodia and Chiang Mai in Thailand (Huber et al. 2002, 2004; Paupy et al. 2004). Like the studies on a larger spatial scale, these small geographical-scale studies generally conclude that there is a significant level of genetic differentiation. However, these previous microsatellite studies of genetic population structure of *A. aegypti* in Southeast Asia have been very localized and used small numbers of markers (from three to eight loci). Further, there has been no substantial study of the broad-scale patterns of genetic structure of *A. aegypti* in Southeast Asia and the possible demographic factors underlying this.

The aim of this study was therefore to determine the genetic population structure of *A. aegypti* at a hierarchy of spatial scales across the region of mainland Southeast Asia and to determine the factors shaping structure at each scale. In particular, we wish to understand the relative roles of historical and contemporary factors shaping the population structure to enable us to take into account any potentially confounding historical effects in the estimation of contemporary gene flow. Thirteen microsatellite markers were genotyped in mosquitoes sampled from 36 sites in Myanmar, Cambodia and Thailand as well as collections from Sri Lanka and Nigeria. In addition to conventional population-based approaches, we also used landscape genetics methods (Manel et al. 2003; Storfer et al. 2006) to identify the factors shaping genetic structure. Landscape genetic approaches are individual-based rather than population-based, so should not result in misleading conclusions being drawn from the incorrect designation of *a priori* populations. The findings were interpreted in relation to their significance for vector control efforts and for the future utility of using landscape genetics approaches to identify the factors shaping the genetic structure of this species.

Materials and methods

Mosquito sampling

Mosquitoes from mainland Southeast Asia were collected in 2004 and 2005 from a total of 36 sites in Myanmar, Thailand and Cambodia at a hierarchy of spatial scales (Table 1). There were three main collection regions per country each of which comprised four collections sites: three that were ~5 km apart and a fourth that was

Table 1. Sample collection data for *Aedes aegypti* mosquitoes.

Map reference	Country	Region	Sample site	Sample code	N	Longitude	Latitude	Collection date
1	Myanmar	Yangon	North Okkalapa*	M1	48	96.17E	16.45N	Dec-2004
			South Okkalapa†	M2	39	96.18E	16.85N	Dec-2004
			Thaketa†	M3	22	96.20E	16.78N	Dec-2004
			Hlegu†	M4	32	96.19E	17.09N	Dec-2004
2		Meiktila	Aung San*	M5	28	95.83E	20.85N	Jul-2005
			Yadana Man Aung†	M6	18	95.86E	20.88N	Jul-2005
			Tawma‡	M7	12	95.91E	20.87N	Jul-2005
			Yemethin†	M8	16	96.14E	20.42N	Jul-2005
3		Myitkyina	Yangyi Aung†	M9	24	97.39E	25.36N	Aug-2005
			Shwe Nyaungbin†	M10	11	97.38E	25.38N	Aug-2005
			Pamma Tee‡	M11	20	97.31E	25.37N	Aug-2005
			Moe Kaung†	M12	34	96.93E	25.30N	Aug-2005
4	Thailand	Chiang Mai	Chiang Mai moat*	T1	43	98.98E	18.78N	Oct-2004
			Mae Hia†	T2	14	98.95E	18.74N	Oct-2004
			Ban Pong Noi†	T3	16	98.94E	18.76N	Oct-2004
			Lamphun*	T4	45	99.01E	18.58N	Oct-2004
5		Ubon Ratchathani	Ban Khamyai Moo†	T5	41	104.86E	15.29N	Nov-2004
			Wat Hat Tai*	T6	28	104.87E	15.22N	Nov-2004
			Wat San Sumran*	T7	24	104.85E	15.19N	Nov-2004
			Ban Kudkrasean†	T8	46	104.55E	15.33N	Nov-2004
6		Songkla	Tee Main†	T9	21	100.59E	7.20N	Jun-2005
			Kao Seng†	T10	28	100.62E	7.17N	Jun-2005
			Ban Bang Dan†	T11	42	100.59E	7.14N	Jun-2005
			Ban Bo Tru†	T12	33	100.40E	7.65N	Jun-2005
7	Cambodia	Battambang	Cham Kasamrong†	C1	21	103.19E	13.11N	Sep-2005
			Preak Preahsdech†	C2	38	103.21E	13.09N	Sep-2005
			Takdol†	C3	43	103.20E	13.14N	Sep-2005
			Reusey Krok‡	C4	39	103.02E	13.54N	Oct-2005
8		Phnom Penh	Chrang Chamreh Pir*	C5	27	104.89E	11.63N	Sep-2005
			Svay Pak*	C6	35	104.86E	11.66N	Sep-2005
			Bakeng†	C7	13	104.91E	11.67N	Oct-2005
			Sala Leak Pram†	C8	45	104.71E	11.93N	Oct-2005
9		Kratie	Thma Kre†	C9	40	106.00E	12.55N	Nov-2005
			Rokakanda†	C10	40	106.02E	12.45N	Nov-2005
			Oresey villa‡	C11	41	106.03E	12.49N	Nov-2005
			Kbal Snoul‡	C12	44	106.42E	12.07N	Nov-2005
10	Northeast India	Assam	Dibrugarh*	IND	6	96.27E	26.76N	Dec-2005
11	Sri Lanka	Columbo	Mattakkuliya†	SRI	17	79.89E	6.96N	Mar-2006
	Kenya	Rabai	Rabai†	KEN	4	39.61E	3.97S	Feb-1999
	Nigeria	Ibadan	Akobo*	NIG	46	3.91E	7.39N	Feb-2006

The fourth sites in each cluster (i.e. M4, M8, M12, T4, T8, T12, C4, C8 and C12) are ~50 km distant from the others within each corresponding cluster.

*Citwntown or urban areas

†Suburban or peri-urban settlements.

‡Rural areas.

~50 km away from the trio. For example, sampling within a region could include three clustered sites in suburban and/or periurban areas of a town or city and a village or town ~50 km away. Although choice of sample sites was governed largely by spatial position, we also attempted to collect from a range of city, periurban and rural sites in order to determine if connectivity (presumably highest in cities and lowest in rural sites) influenced genetic structure within and among sites. Each collection site covered an area ~500 m in diameter based on the ability of *A. aegypti* to fly up to several hundred metres (Christophers 1960; Reiter et al. 1995; Honório et al. 2003). Third- and fourth-stage larvae and pupae were collected from ~50 different water storage containers (such as outside water storage jars, indoor cisterns and discarded cans, coconut shells, tyres) in and around

people's houses, sampling as evenly as possible throughout the collection site. Larvae and the adults (which hatched from the collected pupae) were examined under a light microscope and morphologically identified to species using taxonomic keys (Rattanarithikul and Panthusiri 1994). Identified larvae were preserved in 95% ethanol and adults were preserved by desiccation using silica gel. The locations of the sampling sites were recorded using a global positioning system. For the microsatellite genotyping, a single individual was selected at random from each container to avoid incidental sampling of close relatives.

DNA extraction and microsatellite genotyping

DNA was extracted from individual mosquitoes using a standard phenol/chloroform method (Sambrook and Russell 2001) or ammonium acetate precipitation method (Nicholls et al. 2000). Thirteen dinucleotide microsatellite loci have been characterized previously in *A. aegypti* that have suitable levels of variation (Slotman et al. 2007). These microsatellites were amplified in two sets of multiplex PCR in 2 μL reaction volumes. Each reaction comprised: DNA template (1 μL of a 1:400 dilution); 1 μL of primer mix (containing each primer at 0.2 mM with the forward primer of each pair fluorescently labelled with HEX, FAM or NED); and 1 μL of Qiagen Master Mix (QIAGEN, Crawley, UK). The amplification conditions used were an initial activation and denaturation step at 95°C for 15 min; 35 cycles at 94°C for 30 s, 55°C for 90 s and 72°C for 90 s; and a final extension step of 10 min at 72°C. The amplified products were run on an ABI 3730 capillary sequencer (Applied Biosystems, Warrington, UK) at the Molecular Genetics Facility, University of Sheffield (SMGF). A subset of amplified fragments was included in all runs to ensure consistency in size estimation using GeneScan ROX500-bp internal size standards (Applied Biosystems). In addition, for genotype scores of low amplitude the DNA was re-amplified and re-genotyped to confirm the genotyping. The amplified microsatellite markers were genotyped using the GeneMapper software 3.7 (Applied Biosystems).

Genetic analyses

Allelic richness (R_S), observed and expected heterozygosity (H_O and H_E), the inbreeding coefficient (F_{IS}) and deviation from Hardy–Weinberg equilibrium (HWE) were estimated for each locus in each population using ARLEQUIN 3.01 (Excoffier et al. 2006). ARLEQUIN was also used to test for the presence of linkage disequilibrium (LD) between all possible pairs of loci in each population for a total of 38 populations. To determine if significantly positive F_{IS} values were due to isolation by distance

within a 500-m diameter collection site, we determined if the genetic and geographic distances between individuals within a site were correlated using a Mantel test (ARLEQUIN 3.01). The genetic distance measure between individuals was the proportion of shared alleles (Bowcock et al. 1994), which was calculated in Microsatellite analyser 4.05 with 10 000 permutations (Dieringer and Schlötterer 2003).

Genetic differentiation between pairs of populations was estimated in ARLEQUIN 3.01 (Excoffier et al. 2006) using both F_{ST} (Weir and Cockerham 1984) and R_{ST} (Slatkin 1995) as the distance metric. Significance was estimated at the 5% level by 1000 permutations of the genotypes among populations. A Mantel test was implemented in ARLEQUIN 3.01 to test for isolation by distance using $F_{ST}/(1 - F_{ST})$ as a linearized estimate of pairwise genetic distance between populations and the logarithm of geographic distance (Slatkin 1987).

Two Bayesian clustering methods, TESS 2.0 (Francois et al. 2006; Chen et al. 2007) and GENELAND 3.1.4 (Guillot et al. 2005), were used to detect genetic clusters. Individuals are assigned probabilistically to genetic clusters based on their multilocus genotypes to maximize HWE and minimize LD. As these methods do not assume predefined populations, they are useful for identifying spatial discontinuities between samples. Both methods used here are based on the reasonable assumption that spatially proximate individuals are more likely to be genetically related than those that are not and they therefore use the spatial location of individuals as a prior. Although ideally individuals should be sampled in a spatially continuous manner, these approaches have been shown to be effective even when this is not the case (Chen et al. 2007). Explicitly taking into account spatial information is particularly useful for the optimal assignment of individuals into *k* clusters when the overall level of population structure is low (Chen et al. 2007), as is the case here.

In GENELAND, we first estimated the number of clusters (*k*) from five separate runs allowing *k* to vary from 1 to 38. Each run comprised 10 independent Markov Chain Monte Carlo (MCMC) chains of 600 000 iterations with a thinning interval of 50 and a short burn-in of 200. All runs used the correlated frequency model and spatial model. Each of the five runs generated a distribution of the posterior probability density for different values of *k* from which the modal value of *k* was taken to indicate the number of genetic clusters. Four longer runs of GENELAND were then performed using 1 000 000 iterations for a fixed value of *k* estimated from the previous runs to determine the consistency with which individuals were assigned to clusters.

Unlike GENELAND, *k* is a fixed parameter in the model used by TESS. The optimal value of *k* was inferred by

running the software for sequentially higher values of the maximal number of clusters, k_{max} (from 2 to 38) until the estimated number of clusters, k, was less than k_{max}. For each value of k, we ran 10 independent MCMC runs each with 50 000 sweeps and a burn-in period of 10 000 sweeps. As recommended by the authors, we used the no admixture and no F model. The spatial interaction parameter (ψ), which determines the extent of spatial dependence of the algorithm, was fixed at 0.6 according to Chen et al. (2007). Similar trial runs with ψ set at 0.4, 0.8 or 1.0 generated consistent results. For some values of k, we performed five longer runs of TESS each with 200 000 sweeps and a burn-in of 20 000 sweeps. For each value of k used for the long runs, the overall assignment of individuals to clusters was determined using the Greedy algorithm in CLUMPP 1.1.1 (Jakobsson and Rosenberg 2007).

To identify the environmental factors that determine the genetic population structure of *A. aegypti* in Southeast Asia, we applied the software GESTE 2.0 (Foll and Gaggiotti 2006) to the 36 populations from Myanmar, Thailand and Cambodia. Northeast India was excluded due to its small sample size. The analysis was restricted to mainland Southeast Asia as it is more likely that the same factors underlie the population structure across this region whereas additional factors are likely to determine the wider-scale genetic structure including Sri Lanka and Africa. The GESTE modelling process uses a hierarchical Bayesian method to test the effect of different environmental factors on population structure. F_{ST} values are estimated for each local population (population specific F_{ST} values) and provide information on how genetically distinct a population is relative to other populations in the sample. For example, under a model of diffusive dispersal following a single colonization event, populations furthest from the origin would have the highest F_{ST} values due to the cumulative effects of drift from repeated founder events. Population-specific F_{ST} values are related to environmental factors using a generalized linear model. Posterior probabilities are estimated for alternative models with differing environmental variables and the model with the highest posterior probability best explains the data. Deviation from the regression, i.e. how well the data fit the model, is measured by σ^2 and the extent of uncertainty of model parameter values is estimated by the 95% highest probability density interval (HPDI), the smallest interval containing 95% of the values.

We considered three different environmental scenarios each of which includes two environmental factors and explored the effect of the factors individually and their interaction in determining population structure. The first scenario was connectivity which included spatial separation and town size (population size of the town) as the factors. Spatial separation on its own is effectively an isolation-by-distance model. The second factor, town size, was expected to increase the connectivity between *A. aegypti* populations due to increased movement of people. The second scenario was a land-based range expansion, assuming a major point of entry of mosquitoes into Southeast Asia from Africa followed by diffusive dispersal, making latitude and longitude the two environmental factors. The third scenario investigates the effect of human transportation routes using distance from a port and town size as the factors. All these scenarios are of course more simplistic than the reality is likely to be, perhaps particularly the idea of a single colonization event and simple spread of *A. aegypti* across Southeast Asia. However, they serve as a means to attempt to distinguish the most important environmental factors shaping population structure. To check for consistency, each scenario was tested using three short and one long run. Each short run had a total of 250 000 iterations with a thinning interval of 20 including a burn in of 50 000. Each long run had a total of 2 050 000 iterations with a thinning interval of 20 including a burn in of 50 000.

Tests of population expansion

The populations were tested for deviation from mutation–drift equilibrium using the intra-locus k-test and inter-locus g-test (Reich et al. 1999). Both tests were conducted using the Excel Macro KGTESTS (Bilgin 2007). The k statistic tests whether the frequency distribution of allele lengths is more peaked than would be expected for a population of constant size. A constant-sized population is expected to have a ragged, multimodal distribution due to ancient bifurcations caused by stochastic lineage loss whereas an expanding population tends to have a unimodal distribution due to a lack of deep bifurcations in the genealogy. A model of constant population size can be rejected if there are a lower than expected number of positive k statistics for the number of loci tested. The g statistic tests for there being lower variance among loci in the variance of allele frequency sizes than expected for a constant sized population. In populations of a constant size, the dates of deep bifurcations will vary among loci whereas, in an expanding population, bifurcations for all loci will tend to date to the time of expansion. A significantly low value for the g-test was determined empirically by the fifth-percentile cut-off (Reich et al. 1999).

Results

Summary statistics

We analysed variation in 13 microsatellite loci for a total of 1111 individuals encompassing 36 sites in mainland

Southeast Asia. We also included 17 individuals from Sri Lanka and 46 individuals from West Africa (Table 1). The summary statistics, including, allelic richness (R_S), observed heterozygosity (H_O) and deviation from HWE are shown for each locus in each population in the Appendix. All loci were polymorphic in all populations. Allelic richness (R_S) was the highest for the Nigerian population, which had an average R_S of 8 over all loci compared with an average of 5.6 and a maximum of 7 for all other populations. For the 494 tests of HWE, 181 tests showed significant deviation (at $P \leq 0.05$), all due to a deficiency of observed heterozygotes, but this was not associated with particular populations or particular loci.

Genetic differentiation within sites

Out of the 2964 pairwise tests of LD among loci, 462 were significant compared with 148 expected at $P \leq 0.05$. This is unlikely to be due to tight physical linkage or selection since all pairs of loci showed LD in at least one population and no particular pairs of loci consistently showed LD. The LD together with the deviation from HWE could therefore be indicative of some form of inbreeding within collection sites. An alternative explanation for the deficiency of heterozygotes is the presence of null alleles. At the start of the study, we investigated this possibility for five loci (*AG7*, *AC4*, *CT2*, *AG3* and *AC7*) by amplifying the same 259 individuals using both the primers designed by Slotman et al. (2007) and more exterior primer pairs that were designed from the available genome sequence (Table 2). The amplification of every individual for all loci consistently generated a heterozygote or homozygote with both primer pairs, even though overall there were slightly fewer heterozygotes than expected (157 rather than 168; Table 2). This indicates that the large number of significant F_{IS} values in the final dataset is unlikely to be due to null alleles and instead indicates genetic structuring within a site. Although this implies very limited dispersal in these mosquitoes, when six populations with high F_{IS} values (M2, M10, T2, T6, C3 and C12) were tested they showed no signal of isolation by distance within the ∼500-m diameter sites.

Genetic differentiation among sites

Genetic differentiation was estimated for all pairs of populations and is summarized in Table 3. In general, although the F_{ST}-based estimates of population differentiation were not higher than R_{ST}-based estimates, far more of the F_{ST}-based tests showed significant differentiation. The relative lack of signal of population differ-

Table 2. Comparison of the number of heterozygotes generated compared with the numbers expected for five microsatellite loci from the same *Aedes aegypti* individuals using both the original primers designed by Slotman et al. (2007) and redesigned, exterior primer pairs.

Locus	New primers	Size range (bp) for new primers	Size range (bp) for original primers	Population	N_1	No. Hets observed with old and new primers (N_1)	No. Hets expected (N_1)	F_{IS} (N_1)	N_2	zF_{IS} (N_2)
AC4	Forward: 5′-TAAGCAAGCAGCATGTTTCG-3′	210–214	111–115	Yangon	24	12	14.7	0.186	48	0.062**
	Reverse: 5′-TCTCGTCTCACACGCATACAC-3′			Chiang Mai	30	18	18.9	0.053	43	0.056
CT2	Forward: 5′-ATGCTCTCCCAAACTCTTCG-3′	278–288	168–178	Yangon	23	16	12.4	−0.295	48	0.048**
	Reverse: 5′-GTGCGACCAAGGTTAGATCC-3′			Chiang Mai	30	18	18.5	0.024	43	0.176*
AG7	Forward: 5′-CCAAAGCTATGTGTTTAGTGGTAGG-3′	204–218	136–150	Yangon	21	12	12.9	0.073	48	0.05
	Reverse: 5′-ACGGGCGTGTATTACAGGAG-3′			Chiang Mai	27	20	22.3	0.103	43	0.15*
AC7	Forward: 5′-CCAGCAATAGGAAAGTCTTAGGC-3′	214–228	114–127	Yangon	23	13	14.4	0.099	48	0.082
	Reverse: 5′-TGTTTAACAATCTCATTGGACTCG-3′			Chiang Mai	30	11	14.3	0.232	43	0.201
AG3	Forward: 5′-CGGAGAGCAGGAAAGTTCAC-3′	233–243	146–156	Yangon	21	14	15.4	0.09	48	0.119*
	Reverse: 5′-TTGGCGGGACTCTATTGTG-3′			Chiang Mai	30	23	24	0.041	43	0.059**

F_{IS} values were estimated for the smaller number of individuals (N_1) genotyped with both primer pairs and for the final sample size (N_2) with the original primers.
N_1 = Number of individuals genotyped with modified primers.
N_2 = Number of individuals genotyped with original primers.
Level of significance: *$P < 0.05$; **$P < 0.01$.

Table 3. Average pairwise population-genetic differentiation of Aedes aegypti mosquitoes from Southeast Asia and Africa at a range of spatial scales.

	Within clusters		Within and between countries					Between countries across continents
	~5 km (~3–13 km)	~50 km (~20–70 km)	~500 km (~150–500 km)	>500 km (~500–1000 km)	>1000 km (~1000–1500 km)	<2000 km (~1500–2000 km)	>2000 km	(2000 to ~10000 km)
Average pairwise F_{ST}	0.026	0.032	0.039	0.043	0.045	0.047	0.063	0.066
Range	0–0.049	0.014–0.058	0.015–0.09	0.011–0.083	0.012–0.1	0.024–0.083	0.031–0.108	0.024–0.137
No. significant pairwise F_{ST} values	24/27	26/27	146/148	172/175	203/203	66/66	24/24	147/150
Average pairwise R_{ST}	0.041	0.023	0.044	0.041	0.042	0.052	0.052	0.041
Range	−0.008 to 0.123	−0.019 to 0.067	−0.006 to 0.214	−0.012 to 0.207	−0.009 to 0.158	−0.005 to 0.175	0.009 to 0.108	−0.117 to 0.265
No. significant pairwise F_{ST} values	11/27	3/27	42/148	45/175	57/203	19/66	16/24	41/150

entiation from R_{ST} is probably due to deviation from the assumed stepwise mutation model (Balloux and Lugon-Moulin 2002). The F_{ST}-based estimates revealed that even sites that were only ~5 km apart were significantly genetically differentiated, although the level of differentiation was low (average F_{ST} values of 0.026). This observation was repeated at all spatial scales and even pairwise comparisons among populations from Africa and Asia had only low levels of differentiation (average F_{ST} values of 0.066). It is possible that there are some false positives as 5% was used as the significance level for these tests, but these will have minimal effect on the overall findings as the majority of tests were positive (808 out of 820) and typically had high significance ($P \leq 0.0001$). As spatial scale increases, the average F_{ST} value and the level of significance also increases. (The three nonsignificant comparisons at the highest spatial scale involved comparisons with Kenya or Northeast India where sample sizes were very low.) This is consistent with the signal of isolation by distance found in the populations from mainland Southeast Asia (Fig. 1). The Mantel test shows that genetic and geographic distances are significantly correlated, although the level of this correlation is relatively low ($R^2 = 0.111$; $P = 0.036$).

Landscape factors underlying population structure

For the 38 populations (i.e. excluding the small Northeast Indian and Kenyan samples), GENELAND consistently estimated a highly peaked posterior probability distribution for the number of population clusters (Fig. 2) with a modal value of $k = 7$ and the second highest value for $k = 8$. There was also a small but significant posterior probability for $k = 38$, consistent with the small but significant level of structure detected between most populations (Table 3). For $k = 7$, the average F_{IS} value for each cluster was 0.14 and F_{ST} values between clusters ranged from 0.01 to 0.06. The four long runs of GENELAND with $k = 7$ generated largely consistent results, which are represented in Fig. 3. The Nigerian population was always reported as a distinct cluster as was M4 (the site 50 km from Yangon) and T8 (the site 50 km from the cluster of three populations in Ubon Ratchathani, eastern Thailand). There was also evidence of regional clustering. The Central and Upper Myanmar populations (M5–M12) were always clustered together and half of the time they were also grouped with Yangon populations (all individuals from M1 and M2 and nine individuals from M3). In addition, 11 of the Cambodian populations (except for C4) always form a cluster. Conversely, the Cambodian C4 population (50 km outside of Battambang) only clustered with the other Cambodian populations once and instead usually clustered with Chiang
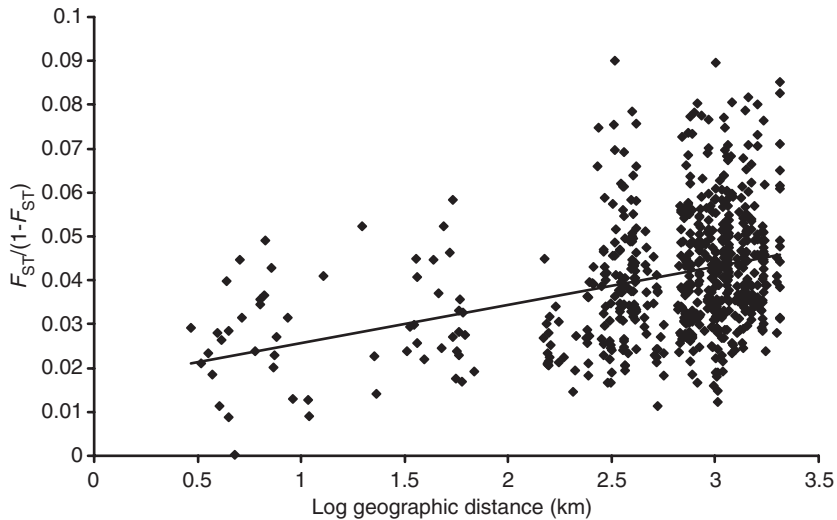
Figure 1 Scatterplot and regression line of genetic and geographic distance for all 36 populations of *A. aegypti* in mainland Southeast Asia (Mantel test: $R^2 = 0.111$, $P = 0.036$).
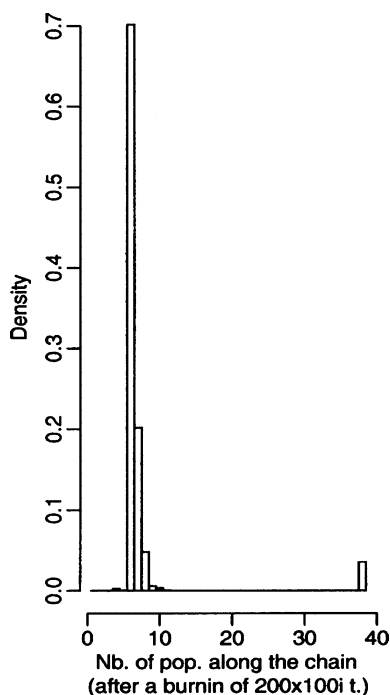

Figure 2 Distribution of posterior probability for the number of genetic clusters (k) estimated by GENELAND.

Mai and Songkhla (three out of four runs). The Thai populations from Chiang Mai (T1–T4) and Songkhla (T9–T12) always formed a cluster, even though they were more than 1000 km apart. This indicates that spatial distance alone was not responsible for the clusters detected. Similarly, the geographically distant Sri Lankan individuals were never found in a separate group but always clustered with three of the eastern Thai populations (T5–T7) and most of the individuals from the Yangon M3

population. The Yangon populations (M1–M3) were very heterogeneous compared with other populations. In addition to the clustering with the other Myanmar populations or with Sri Lanka and Eastern Thailand as outlined above, M1, M2 and some individuals from M3 also sometimes clustered (two out of four times) with the northern and southern Thai populations from northern Thailand (Chiang Mai, T1–T4) and southern Thailand (Songkhla, T8–T12).

When TESS was run with the spatial interaction parameter (ψ) set to 0 for $k = 3$–38, there were no clear clusters with the exception of Nigeria. When run like this, with no spatial information taken into account, TESS is equivalent to the STRUCTURE program of (Pritchard et al. 2000). With ψ set to 0.6, the likelihood of TESS increased with increase in k, reflecting the low but significant differentiation between all populations. As k_{max} was increased for repeated runs of TESS, the first point at which some estimates of k were lower than k_{max} was for $k = 7$ and $k = 8$, consistent with the GENELAND clustering results. However, it was only by $k_{max} = 16$–18 that k was consistently lower than k_{max} with $k = 16$ being the modal value. Figure 4 shows the cluster assignments from DISTRUCT for $k = 7$ and $k = 8$ (to provide a comparison with the GENELAND output) and for $k = 16$. At $k = 7$ and $k = 8$, the grouping structure is very comparable with that obtained from GENELAND, for example, the genetic distinctiveness of T8 in eastern Thailand and the genetic similarity of northern Thailand (Chiang Mai), southern Thailand (Songkhla) and C4 in Cambodia. In TESS at $k = 16$, these latter three regions become genetically distinct from each other. The M4 population on the outskirts of Yangon also becomes distinct (as in GENELAND) as does C11 in Cambodia. TESS and GENELAND also both identify the Yangon populations (M1–M3) as being genetically heterogeneous. In addition,
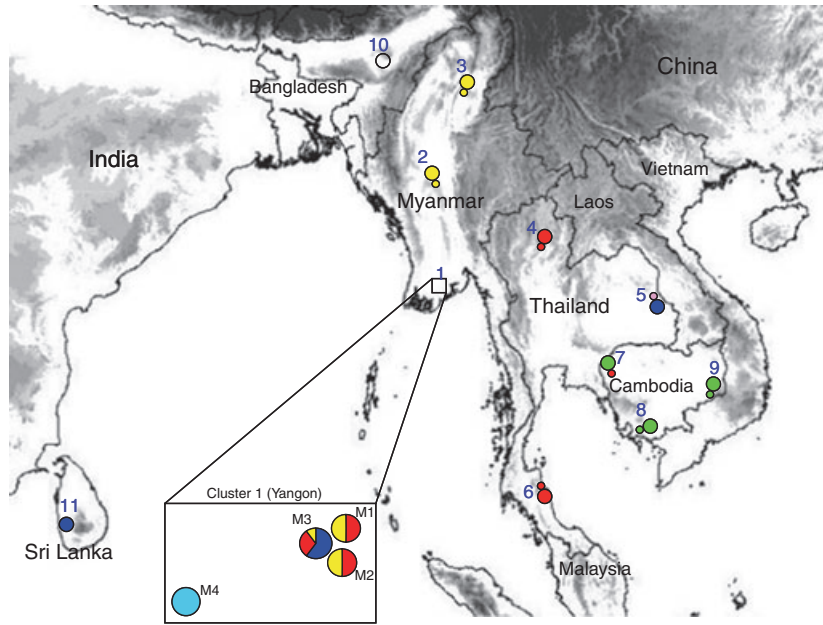
Figure 3 Relief map of Southeast Asia (and Sri Lanka) showing sample collection sites and genetic clusters detected by GENELAND. In mainland Southeast Asia, a large circle represents three sites spaced 5 km apart and a small circle represents the site 50 km distant from this. The Yangon sites are numbered according to Table 2. Each genetic cluster is represented by a different colour but the 7th cluster, Nigeria, is not shown.

the TESS assignments for $k = 7$ and $k = 8$ indicated that sites C1–C3 in Battambang, Cambodia, and sites T9–T11 in Songkla, southern Thailand and Sri Lanka were also very heterogeneous with individuals from these populations being assigned to several different genetic clusters. At $k = 16$, additional heterogeneity in the Cambodian and Nigerian samples also became apparent.

The three alternative, but not necessarily mutually exclusive, environmental scenarios that were fitted to the data using GESTE are shown in Table 4. The posterior probabilities were consistent across runs and are often high. However, the quality of fit to the models is actually very weak with wide HPDI and high σ^2 values so we cannot make reliable inferences from the model

Figure 4 Membership of individuals to 7, 8 and 16 genetic clusters estimated by TESS and CLUMPP.

Table 4. Posterior probabilities for 24 different models under the three environmental scenarios from the geste analysis.

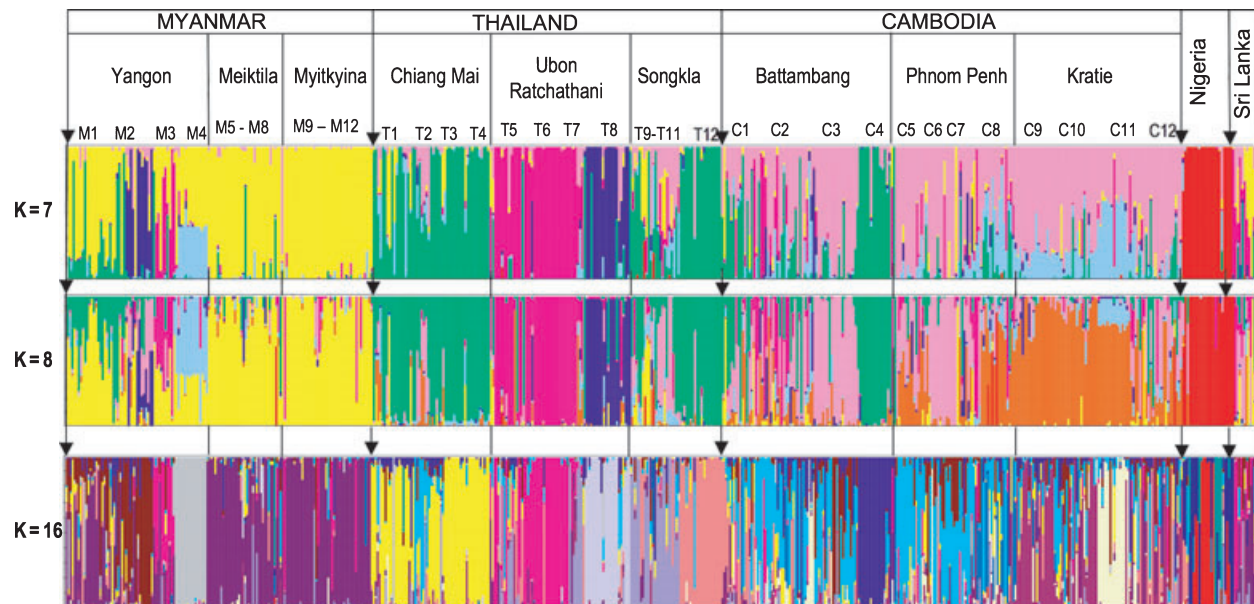| Environmental Scenario	Factors	Posterior probability
Connectivity	Constant	0.188
	Spatial distance	0.331
	Constant and spatial distance	0.170
	Size of city	0.337
	Constant and size of city	0.177
	Constant, spatial distance and size of city	0.160
	Spatial distance and size of city interaction	0.305
	All	0.305
Spatial range expansion	Constant	0.0333
	Longitude	0.0441
	Constant and longitude	0.0189
	Latitude	0.0514
	Constant and latitude	0.0261
	Constant, longitude and latitude	0.0252
	Longitude and latitude interaction	0.896
	All	0.896
Human transport routes	Constant	0.120
	Distance from port	0.210
	Constant and distance from port	0.102
	Size of city	0.206
	Constant and size of city	0.098
	Constant, distance from port and size of city	0.108
	Distance from port and size of city interaction	0.572
	All	0.572

Table 5. Tests of population expansion in *Aedes aegypti* using the *k*- and *g*-tests for 13 microsatellite loci.

Sites	*k*-Test (number of positives)	*g*-Test
Cluster 1 (Myanmar)	7/13 NS	0.64
Cluster 2 (Myanmar)	7/13 NS	0.84
Cluster 3 (Myanmar)	6/13 NS	0.65
Total (Myanmar)	5/13 NS	0.70
Cluster 4 (Thailand)	7/13 NS	0.70
Cluster 5 (Thailand)	5/13 NS	0.62
Cluster 6 (Thailand)	4/13 NS	0.81
Total (Thailand)	3/13*	0.63
Cluster 7 (Cambodia)	8/13 NS	0.67
Cluster 8 (Cambodia)	8/13 NS	0.47
Cluster 9 (Cambodia)	8/13 NS	0.86
Total (Cambodia)	7/13 NS	0.66
Total (Southeast Asia)	6/13 NS	0.62
Northeast India	9/13 NS	1.07
Sri Lanka	7/13 NS	0.69
Kenya	8/13 NS	1.14
Nigeria	10/13 NS	0.42

*k-test (P-value = 0.037).

results. We show the data here for two reasons. First, these data give some indication of factors that are worthy of further investigation as they appear to indicate a role for human communication routes (via ports and roads that connect large towns), spatial distance and spatial range expansion in determining genetic population structure in this species. Secondly, the data indicate how future sampling could be carried out to ensure that reliable inferences could be made from the modelling. Although the relatively large number of populations used here should enable robust model determination (Foll and Gaggiotti 2006), most of the populations have similar and very low population-specific F_{ST} values providing little information for the modelling. There is a small proportion of populations with higher F_{ST} values (notably M8 and M10 in Central and Upper Myanmar, T7 in Northern Thailand and C7 in Phnom Penh, Cambodia) that have experienced greater genetic drift relative to the others. It is these distinctive populations that are largely determining the regression, but as they are few in number there is great uncertainty in parameter estimation. As these distinctive populations are both more inland and also tend to be more isolated, future sampling would need to include sites where environmental factors were not confounded.

Tests of population expansion

The results of the k- and g-tests for population bottlenecks are shown in Table 5. In Southeast Asia, the clusters of four sites each have about half the loci (from 4 to 8 out of 13 loci) with positive values of the k-test statistic, which indicates that the allele frequency distributions are no more peaked than expected by chance. The number of loci that have positive values of k decreases when these clusters are pooled into country or the whole of Southeast Asia (from 3 to 6 out of 13 loci) with only Thailand having a significantly low number of positive k-values. Sri Lanka, with 7 out of 13 loci positive, fits in with the other Southeast Asian sites whereas Nigeria appears to have a larger number of positive k test values (10 out of 13). All the g-values for these groups of populations are also positive indicating that a model of constant population size could not be rejected for any of the population groupings.

Discussion

Restricted dispersal on a small spatial scale
Although the overall level of genetic differentiation was low, populations that were only 5 km apart were significantly different from each other. Even within our collection areas that were no more than 500 m in diameter,

estimates of the inbreeding coefficient (F_{IS}) were often significant. The generation of the same genotypes with alternative primers indicates that this inbreeding is unlikely to be explained solely, if at all, by the presence of null alleles. Together, therefore, these data indicate that genetic structuring on a very small spatial scale (<500 m) is a general phenomenon in *A. aegypti* in Southeast Asia, in agreement with previous localized studies (e.g. Paupy et al. 2004). The lack of detection of a signal of isolation by distance within a collection site indicates that structuring is not due to limited dispersal in a continuous population. For some sites, preliminary analyses using the TESS software showed some indication of genetic clustering within sites but, before any conclusions can be firmly drawn on this, more loci and more individuals are needed to provide the necessary resolution at this small spatial scale.

Data from mark–release recapture studies shows that *A. aegypti* mosquitoes have a limited flight range but can move up to a few hundred metres around their larval habitats (Reiter et al. 1995). Although such dispersal distances could be sufficient to genetically homogenize clusters within a 500-m area, there are several factors that may prevent this. First, dispersal rates are reduced where oviposition sites are abundant (Edman et al. 1998). Secondly, the frequency distribution of dispersal distances is highly skewed with the vast majority dispersing extremely low distances (Harrington et al. 2005). Thirdly, in *A. aegypti*, mating takes place in swarms near to the host, in and around houses (Hartberg 1971; Cabrera and Jaffe 2007). Consequently, the small-scale genetic clustering of *A. aegypti* may be due to the clustering of oviposition sites and hosts around human habitation coupled with low dispersal. It is possible that these genetic clusters correspond to a house or group of closely situated houses and their immediate environs as studies of the distribution of *A. aegypti* have shown that they tend to be highly spatially clustered with a house typically acting as the unit of clustering (Getis et al. 2003). However, at present, we do not know the exact spatial scale of genetic clusters nor what environmental features may form the barriers between them.

The levels of genetic differentiation here were notably lower than those detected in previous studies. Average values of F_{ST} in this study were 0.026 or 0.032 on spatial scales of 5 and 50 km (Table 3) respectively. In comparison, other microsatellite-based studies on a similar spatial scale (comparing sites within cities) had overall F_{ST} values of 0.056 (Huber et al. 2002) and 0.053 (Paupy et al. 2004) with many individual values being substantially larger (>0.1). Another study on a larger spatial scale (14 samples from three cities in Thailand, Vietnam and Cambodia) had an overall F_{ST} value of 0.117 (Huber et al.

2004) compared with an average value of 0.045 here for a similar spatial scale. This could be due to differences in polymorphism level of the microsatellites used in different studies (Hedrick 1999). However, as the markers appear to have similar diversity, at least part of the reason for the lower levels of genetic differentiation in this study could be that our 500 m sampling sites encompass multiple demes due to the highly clustered structure of *A. aegypti* (Wright 1921). In some other studies, larvae were collected from small areas of two to three (Huber et al. 2004) or four to five houses (Huber et al. 2002). Despite the small collection area, these studies reported some very high levels of F_{IS} for sampling sites (up to 0.661 in the later study and 0.579 in the earlier one). This may indicate that the sampling area contains only a small number of families with high levels of inbreeding. Alternatively, if the high F_{IS} is due to a sampling effect, for example, the pooling of a relatively small number of larval collections containing siblings, this would result in over-inflated estimates of population differentiation. Despite these differences between studies associated with differences in sampling strategy, it is clear that *A. aegypti* has a clustered distribution and restricted dispersal on a very small spatial scale.

Factors underlying large spatial scale population structure

Although a signal of isolation by distance is detected, here it is low (Fig. 1), indicating that the restricted mosquito dispersal detected at a small spatial scale does not explain larger-scale population structure. Further, the Nigerian sample has relatively high genetic distinctiveness and substantially higher allelic richness than the Southeast Asian populations. This is consistent with Africa being the ancestral region of *A. aegypti* and suggests there has been a founder effect in its colonization of Southeast Asia. This provides evidence to support the suggestion made by several authors (Smith 1956; Tabachnick and Powell 1979; Failloux et al. 2002) that these mosquitoes have spread from Africa to Southeast Asia via shipping. As Southeast Asian *A. aegypti* most likely originated from east, rather than west, Africa (Tabachnick 1991), it will be necessary to confirm the founder effect by determining the genetic diversity of *A. aegypti* in East Africa, which was not possible here due to the small sample size.

Despite the higher genetic diversity found in Africa, there was no evidence of a substantial population expansion in Southeast Asia. There are several reasons why the tests may be unable to detect an expansion even if one has occurred, e.g. too few loci used, too long since the time of expansion (Reich et al. 1999). However, it is also possible that the lack of an expansion signal indicates

there have been multiple introductions into Southeast Asia from several genetically differentiated sources. This would result in a high variance among loci in the variance of allele frequency sizes and generate multimodal allele size frequency distributions, resulting in negative g- and k-values respectively. Even if there has been colonization from multiple sources, the overall numbers of mosquitoes introduced may have been insufficient to capture all the genetic diversity present in Africa. This certainly seems to be a likely scenario given that, over the last two to three centuries, there have been large numbers of ships coming into many Southeast Asian ports from different locations in Africa.

The spatial clustering analyses (TESS and GENELAND) generated similar results giving greater confidence in their findings. Both methods show that although there is some regional clustering with similarities among populations within countries, genetic clusters do not correspond obviously to spatial distance; some populations that are very close to each other are highly divergent (for example, the M4 population 50 km from the main Yangon Cluster and the T8 population 50 km from the eastern Thailand cluster) whereas some populations that are far from each other were genetically similar. Some sites that are geographically distant yet genetically similar (e.g. T9–T12 in Songkla, southern Thailand, and T1–T4, Chiang Mai, northern Thailand) are connected by major roads. In addition, many of the locations with high genetic heterogeneity are ports (Colombo in Sri Lanka, Yangon in Myanmar, Songkla in Thailand) or major cities (Phnom Penh, Cambodia). Together, these findings indicate that there is some long-distance dispersal of *A. aegypti* facilitated by major human transportation routes. This evidence for passive migration supports previous similar suggestions based on population-genetic studies in the southern USA (Merrill et al. 2005) and in Southeast Asia (Huber et al. 2004). Passive dispersal likely involves the movement of immature stages of mosquito as well as adults; eggs, larvae and pupae could easily occur in water containers transported by people and the eggs can withstand desiccation for several months (Christophers 1960). Much of the genetic structuring caused by passive migration could be the result of the initial colonization process, when population sizes were smaller and there were empty niches to expand into. Although contemporary passive migration may have a smaller influence on structure, there is every reason to suspect that such passive dispersal is ongoing.

The historical processes of colonization and range expansion in this species appear to have played a major role in shaping large-scale population structure in *A. aegypti* in Southeast Asia. This therefore makes it more difficult to infer contemporary factors that are of relevance for vector control, i.e. active and passive dispersal using conventional population-genetics methods; the limitations of traditional population-genetics methods that assume equilibrium between migration–mutation–drift equilibrium for the inference of gene flow are well recognized (e.g. Nichols and Beaumont 1996). In this context, it is worth noting that the clustering approaches used here provided valuable information on genetic structure that was not apparent from the traditional population-genetics methods.

Implications for vector control

The very restricted dispersal of *A. aegytpi* on a small spatial scale has several implications for vector control. Conventional control measures such as insecticides or the removal of larval habitats in and around houses are often implemented following a dengue outbreak (WHO 2006). The limited dispersal of *A. aegypti* indicates that this approach should be effective in removing infective populations and preventing their spread. (Of course, this does not prevent the spread of the virus by humans.) A major problem for vector control is insecticide resistance (Gubler 2002). Restricted dispersal of *A. aegypti* could make the implementation of a stable zone strategy to delay the spread of insecticide resistance genes difficult. In the stable zone strategy, the area treated with insecticide needs to be small relative to the scale of dispersal in order to allow the reinvasion of fitter, nonresistant genotypes into the treated area so preventing the establishment of the resistance genes (Lenormand and Raymond 1998). The restricted dispersal of *A. aegypti* would also make the strategy of introducing and driving refractory genes through large geographical areas (James et al. 2006) extremely difficult. On the other hand, the use of sterile insect technique (SIT) could be very effective in localized areas. In SIT, large numbers of sterile males are released and reduce population sizes when they mate with local females (Thomas et al. 2000 and references therein). This method is most effective when dispersal is limited and sufficient numbers of sterile males are released to generate a travelling wave of extinction (Lewis and Van Den Driessche 1993). As the extent of passive large-scale dispersal would affect all of these control measures, there is clearly a need to better understand the extent and means by which this takes place.

Conclusion

Our finding that Southeast Asian populations of *A. aegypti* are characterized by genetic structuring on a very small spatial scale indicates the need for further fine-scale studies. Such studies need to determine the exact

spatial scale of genetic clusters and to identify the factors determining active dispersal, particularly which environmental features may form barriers to dispersal. An effective approach to this will be detailed landscape genetic studies (using methods such as those used in this study) in which individuals are sampled continuously and on a fine scale throughout urban, peri-urban and rural environments. It is also important to determine the role of environmental factors, particularly those leading to passive migration, in shaping large-scale population structure and the extent to which these are historical or contemporary effects. The GESTE-modelling approach has the potential to be extremely valuable for this if more loci are used coupled with a more extensive geographical sampling programme to give greater power to distinguish between alternative models. As all analytical methods undoubtedly have their specific limitations, the future use of a variety of methodologies with correspondingly careful interpretation is the approach most likely to lead to a meaningful understanding of population structure and gene flow in *A. aegypti*.

Acknowledgements

Literature cited

Balloux, F., and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. Molecular Ecology **11**:155–165.

Bilgin, R. 2007. Kgtests: a simple Excel Macro program to detect signatures of population expansion using microsatellites. Molecular Ecology Notes **7**:416–417.

Bosio, C. F., B. J. Beaty, and W. C. Black. 1998. Quantitative genetics of vector competence for Dengue-2 virus in *Aedes aegypti*. American Journal of Tropical Medicine and Hygiene **59**:965–970.

Bosio, C. F., L. C. Harrington, J. W. Jones, R. Sithiprasasna, D. E. Norris, and T. W. Scott. 2005. Genetic structure of

Aedes aegypti populations in Thailand using mitochondrial DNA. American Journal of Tropical Medicine and Hygiene **72**:434–442.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368**:455–457.

Cabrera, M., and K. Jaffe. 2007. An aggregation pheromone modulates lekking behaviour in the vector mosquito *Aedes aegypti* (Diptera: Culicidae). Journal of the American Mosquito Control Association **23**:1–10.

Chen, C., E. Durand, F. Forbes, and O. François. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes **7**:747–756.

Christophers, S. R. 1960. *Aedes aegypti* (L.), the Yellow Fever Mosquito. Cambridge University Press, London.

Dieringer, D., and C. Schlötterer. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes **3**:167–169.

Edman, J. D., T. W. Scott, A. Costero, A. C. Morrison, L. C. Harrington, and G. G. Clark. 1998. *Aedes aegypti* (Diptera: Culicidae) movement influenced by availability of oviposition sites. Journal of Medical Entomology **35**:578–583.

Excoffier, L., G. Laval, and F. Schneider. 2006. An integrated software package for population genetics data analysis. User manual, Computational and Molecular Population Genetics Lab (CMPG), Institute of Zoology, University of Berne, Berne.

Failloux, A. B., M. Vazeille, and F. Rodhain. 2002. Geographic genetic variation in populations of the dengue virus vector *Aedes aegypti*. Journal of Molecular Evolution **55**:653–663.

Foll, M., and O. Gaggiotti. 2006. Identifying the environmental factors that determine the genetic structure of populations. Genetics **174**:875–891.

Francois, O., S. Ancelet, and G. Guillot. 2006. Bayesian clustering using hidden Markov random fields in spatial population genetics. Genetics **174S**:805–816.

Getis, A., A. C. Morrison, K. Gray, and T. W. Scott. 2003. Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru. American Journal of Tropical Medicine and Hygiene **69**:494–505.

Gooding, R. 1996. Genetic variation in arthropod vectors of disease-causing organisms: obstacles and opportunities. Clinical Microbiology Review **9**:301–320.

Gorrochotegui-Escalante, N., C. Gomez-Machorro, S. Lozano-Fuentes, L. Fernandez-Salas, M. De Lourdes Munoz, J. A. Farfan-Ale, J. Garcia-Region *et al.* 2002. Breeding structure of *Aedes aegypti* populations in Mexico varies by region. American Journal of Tropical Medicine and Hygiene **66**:213–222.

Gubler, D. J. 1998a. Dengue and Dengue Hemorrhagic Fever. Clinical Microbiology Review **11**:480–496.

Gubler, D. J. 1998b. Resurgent vector-borne diseases as a global health problem. Emerging Infectious Diseases **4**:442–450.

Gubler, D. J. 2002. The global emergence/resurgence of arboviral diseases as public health problems. Archives of Medical Research 33:330–342.

Gubler, D. J., and G. G. Clark. 1995. Dengue/dengue hemorrhagic fever: the emergence of a global health forum. Emerging Infectious Diseases 1:55–57.

Gubler, D. J., S. Nalim, R. Tan, H. Saipan, and J. S. Saroso. 1979. Variation in susceptibility to oral infection with dengue viruses among geographic strains of *Aedes aegypti*. American Journal of Tropical Medicine and Hygiene 28:1045–1052.

Guillot, G., A. Estoup, F. Mortier, and J. F. Cosson. 2005. A spatial statistical model for landscape genetics. Genetics 170:1261–1280.

Guzman, M. G., and G. Kouri. 2002. Dengue: an update. Lancet Infectious Diseases 2:33–42.

Hammon, W. M., A. Rudnick, G. Sather, K. D. Rogers, and L. J. Morse. 1960. New hemorrhagic fevers of children in the Philippines and Thailand. Transactions of the Association of American Physicians 73:140–155.

Harrington, L. C., T. W. Scott, K. Lerdthusnee, R. C. Coleman, A. Costero, G. G. Clark, J. J. Jones *et al.* 2005. Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. American Journal of Tropical Medicine and Hygiene 72:209–220.

Hartberg, W. K. 1971. Observations on the mating behaviour of *Aedes aegypti* in nature. WHO Bulletin 45:847–850.

Hedrick, P. W. 1999. Highly variable loci and their interpretation in evolution and conservation. Evolution 53:313–318.

Herrera, F., L. Urdaneta, J. Rivero, N. Zoghbi, J. Ruiz, G. Carrasquel, J. A. Martínez *et al.* 2006. Population genetic structure of the dengue mosquito *Aedes aegypti* in Venezuela. Memórias do Instituto Oswaldo Cruz, Rio de Janeiro 101:625–633.

Hlaing, T., W. Tun-Lin, P. Somboon, D. Socheat, T. Setha, S. Min, M. S. Chang *et al.* 2009. Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. BMC Genetics 10:11.

Honório, N. A., W. da Costa Silva, P. J. Leite, J. M. Gonçalves, L. P. Lounibos, and R. Lourenço-de-Oliveira. 2003. Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. Memórias do Instituto Oswaldo Cruz, Rio de Janeiro 98:191–198.

Huber, K., L. Le Loan, T. H. Hoang, S. Ravel, F. Rodhain, and A. B. Failloux. 2002. Genetic differentiation of the dengue vector, *Aedes aegypti* (Ho Chi Minh City, Vietnam) using microsatellite markers. Molecular Ecology 11:1629–1635.

Huber, K., L. Le Loan, N. Chantha, and A. B. Failloux. 2004. Human transportation influences *Aedes aegypti* gene flow in Southeast Asia. Acta Tropica 90:23–29.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

James, A. A., M. Q. Benedict, G. K. Christophides, M. Jacobs-Lorena, and K. E. Olson. 2006. Evaluation of drive mechanisms (including transgenes and drivers) in different environmental conditions and genetic backgrounds. In B. G. J. Knols, and C. Louis, eds. Bridging Laboratory and Field Research for Genetic Control of Disease Vectors, pp. 149–155. Springer, Dordrecht, The Netherlands.

Kittayapong, P. 2006. Malaria and dengue vector biology and control in Southeast Asia. In B. G. J. Knols, and C. Louis, eds. Bridging Laboratory and Field Research for Genetic Control of Disease Vectors, pp. 111–127. Springer, Dordrecht, The Netherlands.

Lenormand, T., and M. Raymond. 1998. Resistance management: the stable zone strategy. Proceedings of the Royal Society of London, Series B: Biological Sciences. 265:1985–1990.

Lewis, M. A., and P. Van Den Driessche. 1993. Waves of extinction from sterile insect release. Mathematical Biosciences 116:221.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology and Evolution 18:189–197.

Merrill, S. A., F. B. Ramberg, and H. H. Hagedorn. 2005. Phylogeography and population structure of *Aedes aegypti* in Arizona. American Journal of Tropical Medicine and Hygiene 72:304–310.

Mousson, L., M. Vazeille, S. Chawprom, S. Prajakwong, F. Rodhain, and A. B. Failloux. 2002. Genetic structure of *Aedes aegypti* populations in Chiang Mai (Thailand) and relation with dengue transmission. Tropical Medicine and International Health 7:865–872.

Muto, R. 1998. Summary of dengue situation in WHO Western Pacific region. WHO Dengue Bulletin 22:12–22.

Nicholls, J. A., M. C. Double, D. M. Rowell, and R. D. Magrath. 2000. The evolution of cooperative and pair breeding in thornbills *Acanthiza* (*Pardalotidae*). Journal of Avian Biology 31:165–176.

Nichols, R. A., and M. A. Beaumont. 1996. Is it ancient or modern history that we can read in our genes? In M. E. Hochberg, and R. Barbault, eds. Aspects of the Genesis and Maintenance of Biological Diversity, pp. 69–87. Oxford UP, Oxford.

Pasteur, N., M. Marquine, F. Rousset, A. B. Failloux, C. Chevillon, and M. Raymond. 1995. The role of passive migration in the dispersal of resistance genes in *Culex pipiens quinquefasciatus* from French Polynesia. Genetical research 66:139–146.

Paupy, C., N. Chantha, K. Huber, N. Lecoz, J. M. Reynes, F. Rodhain, and A. B. Failloux. 2004. Influence of breeding sites features on genetic differentiation of *Aedes aegypti* populations analyzed on a local scale in Phnom Penh

Municipality of Cambodia. American Journal of Tropical Medicine and Hygiene **71**:73–81.

Paupy, C., C. Brengues, B. Kamgang, J. P. Hervé, D. Fontenille, and F. Simard. 2008. Gene flow between domestic and sylvan populations of *Aedes aegypti* (Diptera: Culicidae) in North Cameroon. Journal of Medical Entomology **45**:391–400.

Powell, J. R., W. J. Tabachnick, and J. Arnold. 1980. Genetics and the origin of a vector population, *Aedes aegypti*, a case study. Science **208**:1385–1387.

Prasittisuk, C., A. G. Andjaparidze, and V. Kumar. 1998. Current status of dengue/dengue haemorrhagic fever in WHO Southeast Asia region. WHO Dengue Bulletin **22**:1–15.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945–959.

Rattanarithikul, R., and P. Panthusiri. 1994. Illustrated keys to the medically important mosquitoes in Thailand. Southeast Asian Journal of Tropical Medicine and Public Health **25**:58–59.

Ravel, S., N. Monteny, D. V. Olmos, J. E. Verdugo, and G. Cuny. 2001. A preliminary study of the population genetics of *Aedes aegypti* (Diptera: Culicidae) from Mexico using microsatellite and AFLP markers. Acta Tropica **78**:241–250.

Ravel, S., J. P. Herve, S. Diarrassouba, A. Kone, and G. Cuny. 2002. Microsatellite markers for population genetic studies in *Aedes aegypti* (Diptera: Culicidae) from Cote d'Ivoire: evidence for a microgeographic genetic differentiation of mosquitoes from Bouake. Acta Tropica **82**:39–49.

Reich, D. E., M. W. Feldman, and D. B. Goldstein. 1999. Statistical properties of two tests that use multilocus data sets to detect population expansions. Molecular Biology and Evolution **16**:453–466.

Reiter, P., M. A. Amador, R. A. Anderson, and G. G. Clark. 1995. Dispersal of *Aedes aegypti* in an urban area after blood feeding as demonstrated by rubidium-marked eggs. American Journal of Tropical Medicine and Hygiene **52**:177–179.

Sambrook, J., and D. W. Russell. 2001. Molecular Cloning: A Laboratory Manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York.

Scarpassa, V. M., T. B. Cardoza, and R. P. J. Cardoso. 2008. Population genetics and phylogeography of *Aedes aegypti* (Diptera: Culicidae) from Brazil. American Journal of Tropical Medicine and Hygiene **78**:895–903.

Scott, T. W., W. Takken, B. G. J. Knols, and C. Boete. 2002. The ecology of genetically modified mosquitoes. Science **298**:117–119.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science **236**:787–792.

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

Slotman, M. A., N. B. Kelly, L. C. Harrington, S. Kitthawee, J. W. Jones, T. W. Scott, A. Caccone et al. 2007. Polymorphic microsatellite markers for studies of *Aedes aegypti* (Diptera: Culicidae), the vector of dengue and yellow fever. Molecular Ecology Notes **7**:168–171.

Smith, C. E. G. 1956. The history of dengue in tropical Asia and its probable relationship to the mosquito *Aedes aegypti*. Journal of Tropical Medicine and Hygiene **59**:243–251.

Storfer, A., M. A. Murphy, J. S. Evans, C. S. Goldberg, S. Robinson, S. F. Spear, R. Dezzani et al. 2006. Putting the 'landscape' in landscape genetics. Heredity **98**:128.

Tabachnick, W. J. 1991. Evolutionary genetics and arthropod borne disease: the yellow fever mosquito. American Entomologist (USA) **37**:14–24.

Tabachnick, W. J., and J. R. Powell. 1979. Worldwide survey of genetic variation in the yellow fever mosquito, *Aedes aegypti*. Genetical Research **34**:215–229.

Thomas, D. D., C. A. Donnelly, R. J. Wood, and L. S. Alphey. 2000. Insect population control using a dominant, repressible, lethal genetic system. Science **287**:2474–2476.

Thu, H. M., K. Lowry, T. T. Myint, T. N. Shwe, A. M. Han, K. K. Khin, K. Z. Thant et al. 2004. Myanmar dengue outbreak associated with displacement of serotypes 2, 3, and 4 by dengue 1. Emerging Infectious Diseases **10**:593–597.

Wallis, G. P., W. J. Tabachnick, and J. R. Powell. 1983. Macrogeographic genetic variation in a human commensal: *Aedes aegypti*, the yellow fever mosquito. Genetical Research **41**:241–258.

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

WHO. 2006. Situation of dengue/DHF in the Southeast Asia region: prevention and control status in SEA countries http://www.searo.who.int/en/Section10/Section332_1099.htm (accessed 15 January 2010).

WHO. 2007. Mobilizing research to halt exponential growth of dengue. TDR news **77**:8–11.

WHO. 2008. Dengue and dengue haemorrhagic fever fact sheet No. 117 (Revised in May 2008). http://www.who.int/mediacentre/factsheets/fs117/en/ (accessed 15 January 2010).

Wright, S. 1921. Systems of mating: the effects of inbreeding on the genetic composition of a population. Genetics **6**:124–143.

Appendix

Statistics for microsatellite diversity. Allelic richness (R_S), observed heterozygosity (H_O) and population inbreeding coefficients (F_{IS}) with associated probability are given for each locus for each population. Additionally, R_S, H_O and gene diversity are averaged over all loci (for each population) and F_{IS} is averaged over all populations (for each locus). The numbers of significant deviations from Hardy–Weinberg Equilibrium (HWE) are given as proportions of the total number of loci and total number of populations.

Locus	Myanmar populations											
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
AC4												
R_S	3	2	4	4	2	2	2	2	3	2	3	3
H_O	0.56	0.44	0.68	0.63	0.46	0.50	0.50	0.50	0.54	0.45	0.45	0.50
F_{IS}	0.06	0.18	0.05	0.14	0.06	0.02	0.04	0.02	0.12	0.13	0.29	0.18
P-values	0.00	0.52	0.69	0.10	1.00	1.00	1.00	1.00	0.59	1.00	0.08	0.25
AG4												
R_S	5	5	6	4	6	5	4	5	4	4	6	5
H_O	0.50	0.72	0.64	0.56	0.61	0.61	0.58	0.63	0.58	0.55	0.65	0.65
F_{IS}	0.10	0.10	0.13	0.22	0.18	0.25	0.21	0.19	0.19	0.28	0.05	0.13
P-values	0.04	0.07	0.31	0.00	0.25	0.39	0.49	0.32	0.61	0.34	0.92	0.16
CT2												
R_S	3	4	4	3	3	4	4	4	4	4	3	2
H_O	0.54	0.64	0.68	0.56	0.50	0.67	0.50	0.50	0.58	0.55	0.50	0.50
F_{IS}	0.05	0.12	0.09	0.14	0.13	0.12	0.25	0.32	0.11	0.26	0.10	0.04
P-values	0.00	0.19	0.00	0.14	0.24	0.37	0.20	0.19	0.82	0.48	0.83	1.00
AG1												
R_S	4	5	6	6	6	4	4	3	4	4	3	5
H_O	0.65	0.62	0.73	0.50	0.71	0.72	0.67	0.63	0.54	0.73	0.50	0.74
F_{IS}	0.06	0.09	0.04	0.09	0.07	0.07	0.11	0.10	0.25	0.08	0.27	0.05
P-values	0.66	0.80	0.02	0.65	0.26	0.26	0.83	0.52	0.00	0.79	0.41	0.83
AC5												
R_S	6	7	5	5	8	6	5	7	6	4	5	5
H_O	0.58	0.74	0.64	0.78	0.75	0.44	0.58	0.69	0.67	0.64	0.80	0.74
F_{IS}	0.26	0.09	0.20	0.02	0.02	0.36	0.27	0.21	0.06	0.09	0.01	0.07
P-values	0.00	0.00	0.00	0.53	0.72	0.00	0.24	0.01	0.36	0.45	0.91	0.26
AG5												
R_S	6	6	4	5	7	4	4	4	6	6	5	6
H_O	0.63	0.69	0.45	0.53	0.57	0.56	0.58	0.69	0.58	0.73	0.75	0.68
F_{IS}	0.19	0.16	0.33	0.35	0.14	0.04	0.28	0.06	0.25	0.16	0.05	0.14
P-values	0.00	0.00	0.01	0.00	0.16	0.16	0.27	0.03	0.09	0.05	0.68	0.17
AG7												
R_S	8	7	9	10	10	7	6	11	10	4	7	12
H_O	0.65	0.64	0.82	0.75	0.82	0.78	0.58	0.81	0.71	0.55	0.80	0.79
F_{IS}	0.05	0.10	0.05	0.17	0.04	0.10	0.15	0.09	0.14	0.23	0.05	0.10
P-values	0.40	0.17	0.23	0.01	0.06	0.66	0.04	0.08	0.14	0.16	0.08	0.48
AC7												
R_S	6	5	6	6	6	7	3	8	7	5	6	8
H_O	0.56	0.56	0.73	0.66	0.75	0.72	0.42	0.75	0.75	0.64	0.75	0.79
F_{IS}	0.08	0.05	0.08	0.08	0.08	0.13	0.40	0.09	0.06	0.21	0.08	0.04
P-values	0.17	0.77	0.67	0.70	0.29	0.30	0.17	0.24	0.78	0.27	0.01	0.09
AG3												
R_S	6	5	6	5	7	4	4	5	5	2	6	6
H_O	0.67	0.62	0.55	0.66	0.61	0.50	0.67	0.56	0.71	0.36	0.65	0.71
F_{IS}	0.12	0.15	0.28	0.18	0.17	0.31	0.11	0.24	0.03	0.33	0.10	0.11
P-values	0.00	0.05	0.01	0.02	0.38	0.08	0.21	0.18	0.19	0.54	0.26	0.50
AC1												
R_S	5	4	7	4	6	4	3	5	6	3	5	6

Appendix (continued)

Locus	Myanmar populations											
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
H_O	0.73	0.59	0.68	0.56	0.68	0.67	0.42	0.69	0.71	0.55	0.70	0.79
F_{IS}	0.06	0.20	0.19	0.28	0.08	−0.05	0.42	0.12	0.08	0.21	0.09	0.02
P-values	0.00	0.01	0.28	0.00	0.30	0.79	0.08	0.80	0.52	0.79	0.45	0.61
AG2												
R_S	11	10	11	10	12	10	12	9	11	9	9	11
H_O	0.77	0.54	0.77	0.53	0.86	0.89	0.83	0.81	0.75	0.91	0.85	0.79
F_{IS}	0.10	0.32	0.14	0.34	0.04	−0.01	0.12	0.06	0.13	−0.01	0.04	0.11
P-values	0.00	0.00	0.08	0.00	0.37	0.35	0.39	0.05	0.41	0.94	0.12	0.11
AT1												
R_S	8.00	5.00	4.00	5.00	5.00	5.00	4.00	7.00	7.00	4.00	7.00	7.00
H_O	0.81	0.64	0.68	0.69	0.71	0.78	0.75	0.81	0.79	0.64	0.75	0.76
F_{IS}	0.02	0.22	0.07	0.17	0.04	−0.01	0.02	0.06	0.04	0.14	0.10	0.08
P-values	0.01	0.02	0.11	0.00	0.00	0.99	0.86	0.00	0.19	0.58	0.70	0.13
AC2												
R_S	3	4	4	6	3	6	4	4	4	3	4	4
H_O	0.67	0.59	0.59	0.56	0.57	0.78	0.67	0.75	0.67	0.55	0.55	0.65
F_{IS}	0.04	0.22	0.15	0.25	0.18	0.08	0.15	0.02	0.08	0.14	0.21	0.15
P-values	0.00	0.00	0.76	0.01	0.07	0.85	0.34	0.93	0.97	0.67	0.25	0.00
Average R_S	5.69	5.31	5.85	5.62	6.23	5.23	4.54	5.69	5.92	4.15	5.31	6.15
Average H_O	0.64	0.62	0.66	0.61	0.66	0.66	0.60	0.68	0.66	0.60	0.67	0.70
No. significant departures from HWE	10/13	7/13	5/13	8/13	1/13	1/13	1/13	3/13	1/13	1/13	1/13	1/13
Mean gene diversity over all loci	0.70	0.73	0.76	0.75	0.73	0.74	0.73	0.76	0.74	0.71	0.74	0.77
±SD	0.36	0.37	0.39	0.38	0.37	0.39	0.39	0.40	0.38	0.38	0.38	0.39

Locus	Thailand populations											
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
AC4												
R_S	2	2	2	2	3	3	3	3	3	4	2	2
H_O	0.47	0.50	0.50	0.51	0.61	0.57	0.54	0.63	0.62	0.50	0.40	0.36
F_{IS}	0.06	0.03	0.02	0.13	0.01	0.13	0.18	0.02	0.01	0.25	0.12	0.28
P-values	0.76	1.00	1.00	0.14	0.03	0.37	0.00	0.00	0.50	0.02	0.50	0.16
AG4												
R_S	8	6	6	5	6	5	6	5	5	8	6	5
H_O	0.77	0.79	0.69	0.67	0.78	0.71	0.67	0.78	0.71	0.71	0.60	0.70
F_{IS}	0.06	0.09	0.18	0.16	0.00	0.13	0.12	0.00	0.06	0.15	0.15	0.03
P-values	0.27	0.01	0.08	0.13	0.98	0.05	0.18	0.04	0.91	0.11	0.00	0.24
CT2												
R_S	3	2	3	2	4	3	3	5	2	5	3	2
H_O	0.51	0.50	0.63	0.47	0.59	0.61	0.42	0.67	0.43	0.61	0.31	0.45
F_{IS}	0.18	0.03	0.02	0.06	0.05	0.16	0.32	0.14	0.18	0.04	0.39	0.07
P-values	0.01	1.00	0.88	0.76	0.70	0.00	0.05	0.04	0.66	0.33	0.00	0.73
AG1												
R_S	6	5	4	5	7	6	6	5	5	5	5	5
H_O	0.67	0.57	0.63	0.69	0.73	0.68	0.83	0.57	0.71	0.64	0.55	0.73
F_{IS}	0.10	0.19	0.07	0.12	0.12	0.19	0.00	0.19	0.10	0.17	0.27	0.08
P-values	0.28	0.77	0.73	0.67	0.00	0.50	0.31	0.01	0.29	0.43	0.06	0.07
AC5												
R_S	8	7	7	4	7	6	8	7	7	8	7	6
H_O	0.79	0.57	0.75	0.71	0.61	0.71	0.79	0.72	0.81	0.68	0.76	0.79

Appendix (continued)

Locus	Thailand populations											
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
F_{IS}	0.06	0.31	0.07	0.07	0.25	0.15	0.07	0.15	0.05	0.15	0.08	0.02
P-values	0.16	0.00	0.78	0.08	0.00	0.47	0.09	0.00	0.83	0.05	0.00	0.02
AG5												
R_S	7	5	7	7	6	6	7	9	7	6	8	6
H_O	0.79	0.64	0.69	0.76	0.73	0.71	0.75	0.72	0.62	0.57	0.60	0.70
F_{IS}	0.05	0.04	0.08	0.04	0.09	0.15	0.14	0.18	0.12	0.32	0.23	0.12
P-values	0.00	0.74	0.27	0.00	0.18	0.07	0.00	0.00	0.32	0.00	0.00	0.34
AG7												
R_S	9	3	5	10	12	12	11	9	5	6	5	7
H_O	0.70	0.64	0.56	0.60	0.71	0.79	0.79	0.67	0.67	0.75	0.64	0.64
F_{IS}	0.15	0.10	0.23	0.16	0.20	0.12	0.10	0.19	0.13	0.08	0.13	0.10
P-values	0.04	0.63	0.37	0.00	0.01	0.62	0.03	0.02	0.56	0.00	0.02	0.09
AC7												
R_S	4	4	5	4	7	6	6	8	5	5	3	4
H_O	0.47	0.57	0.44	0.67	0.73	0.68	0.75	0.65	0.62	0.61	0.50	0.67
F_{IS}	0.20	0.18	0.27	0.03	0.05	0.18	0.06	0.20	0.17	0.12	0.04	0.04
P-values	0.07	0.50	0.04	0.02	0.49	0.04	0.00	0.05	0.23	0.16	0.53	0.18
AG3												
R_S	5	2	4	5	5	7	5	5	5	6	5	6
H_O	0.74	0.50	0.69	0.71	0.66	0.79	0.67	0.70	0.62	0.64	0.57	0.70
F_{IS}	0.06	0.03	0.03	0.09	0.15	0.04	0.16	0.12	0.18	0.19	0.22	0.13
P-values	0.00	1.00	0.87	0.05	0.05	0.35	0.69	0.11	0.37	0.12	0.02	0.23
AC1												
R_S	4	4	4	5	6	7	5	3	4	7	3	5
H_O	0.74	0.57	0.69	0.76	0.73	0.68	0.75	0.63	0.62	0.61	0.50	0.70
F_{IS}	0.06	0.16	0.09	0.03	0.08	0.15	0.10	0.10	0.19	0.19	0.30	0.10
P-values	0.00	0.83	0.14	0.03	0.56	0.19	0.04	0.00	0.71	0.00	0.00	0.42
AG2												
R_S	9	6	5	9	7	11	10	11	9	10	5	13
H_O	0.63	0.71	0.63	0.71	0.63	0.86	0.75	0.72	0.76	0.82	0.40	0.82
F_{IS}	0.13	0.12	0.05	0.06	0.11	0.03	0.08	0.08	0.08	0.08	0.41	0.07
P-values	0.19	0.99	0.53	0.44	0.01	0.05	0.41	0.71	0.41	0.00	0.00	0.75
AT1												
R_S	6	5	4	7	5	4	5	4	6	8	5	9
H_O	0.70	0.64	0.56	0.76	0.76	0.71	0.75	0.72	0.67	0.68	0.76	0.79
F_{IS}	0.16	0.21	0.24	0.07	0.02	0.06	0.05	0.05	0.17	0.18	0.08	0.04
P-values	0.00	0.43	0.34	0.05	0.88	0.94	0.96	0.00	0.16	0.00	0.00	0.13
AC2												
R_S	6	3	3	3	4	4	3	5	3	3	3	4
H_O	0.70	0.57	0.50	0.69	0.71	0.71	0.58	0.63	0.62	0.64	0.64	0.52
F_{IS}	0.04	0.20	0.26	0.05	0.01	0.03	0.16	0.21	0.07	0.03	0.09	0.23
P-values	0.01	0.84	0.48	0.00	0.02	0.59	0.03	0.00	0.70	0.00	0.00	0.25
Average R_S	5.92	4.15	4.54	5.23	6.08	6.15	6.00	6.08	5.08	6.23	4.62	5.69
Average H_O	0.67	0.60	0.61	0.67	0.69	0.71	0.70	0.68	0.65	0.65	0.56	0.66
No. significant departures from HWE	7/13	2/13	1/13	5/13	7/13	3/13	6/13	11/13	0	7/13	10/13	1/13
Mean gene diversity over all loci	0.74	0.68	0.68	0.72	0.76	0.79	0.78	0.77	0.73	0.72	0.76	0.68
±SD	0.38	0.36	0.36	0.37	0.39	0.41	0.40	0.39	0.38	0.37	0.39	0.35

Appendix (continued)

Locus	Cambodia populations											
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
AC4												
R_S	3	3	3	2	2	2	3	3	2	2	2	3
H_O	0.48	0.53	0.51	0.49	0.44	0.43	0.62	0.56	0.45	0.40	0.44	0.52
F_{IS}	0.24	0.12	0.13	0.03	0.12	0.15	0.11	0.04	0.09	0.21	0.15	0.13
P-values	0.10	0.12	0.00	1.00	0.70	0.50	1.00	0.01	0.75	0.21	0.53	0.10
AG4												
R_S	5	6	7	8	5	7	4	6	6	6	5	7
H_O	0.67	0.76	0.65	0.77	0.63	0.69	0.62	0.62	0.75	0.75	0.68	0.68
F_{IS}	0.13	0.04	0.20	0.09	0.14	0.15	0.20	0.17	0.05	0.08	0.13	0.11
P-values	0.50	0.69	0.00	0.09	0.09	0.00	0.86	0.22	0.81	0.52	0.24	0.12
CT2												
R_S	3	4	4	2	5	4	4	7	5	5	6	4
H_O	0.52	0.58	0.56	0.46	0.67	0.57	0.54	0.67	0.63	0.63	0.63	0.55
F_{IS}	0.17	0.16	0.19	0.04	0.14	0.19	0.26	0.04	0.12	0.19	0.20	0.17
P-values	0.48	0.03	0.05	1.00	0.24	0.15	0.42	0.60	0.12	0.16	0.13	0.17
AG1												
R_S	5	6	6	5	5	5	6	6	5	5	4	5
H_O	0.52	0.74	0.77	0.56	0.63	0.57	0.77	0.64	0.65	0.73	0.56	0.66
F_{IS}	0.32	0.04	0.05	0.21	0.20	0.26	0.06	0.15	0.09	0.05	0.17	0.14
P-values	0.03	0.09	0.03	0.00	0.01	0.03	0.12	0.01	0.02	0.47	0.02	0.07
AC5												
R_S	6	8	7	6	6	6	4	7	5	8	9	6
H_O	0.57	0.76	0.65	0.74	0.70	0.74	0.62	0.78	0.55	0.68	0.76	0.64
F_{IS}	0.27	0.12	0.16	0.07	0.06	0.09	0.18	0.04	0.25	0.16	0.11	0.15
P-values	0.29	0.02	0.11	0.00	0.15	0.00	0.34	0.00	0.00	0.03	0.03	0.03
AG5												
R_S	6	7	5	5	7	6	5	6	8	8	6	9
H_O	0.67	0.76	0.58	0.67	0.81	0.63	0.62	0.71	0.73	0.75	0.71	0.75
F_{IS}	0.19	0.01	0.25	0.16	0.03	0.25	0.22	0.14	0.05	0.05	0.15	0.13
P-values	0.45	0.67	0.00	0.02	0.01	0.00	0.04	0.02	0.48	0.68	0.00	0.18
AG7												
R_S	7	9	10	5	10	7	6	13	11	11	13	12
H_O	0.67	0.68	0.63	0.44	0.56	0.74	0.62	0.73	0.75	0.68	0.71	0.66
F_{IS}	0.17	0.18	0.12	0.21	0.29	0.09	0.24	0.14	0.09	0.20	0.14	0.15
P-values	0.01	0.00	0.10	0.09	0.06	0.15	0.71	0.01	0.36	0.00	0.02	0.04
AC7												
R_S	6	7	6	5	4	5	3	8	5	7	6	8
H_O	0.52	0.66	0.63	0.56	0.56	0.54	0.62	0.67	0.48	0.73	0.59	0.73
F_{IS}	0.30	0.14	0.03	0.13	0.13	0.16	0.14	0.13	0.20	0.02	0.20	0.05
P-values	0.02	0.15	0.08	0.20	0.45	0.19	0.89	0.04	0.16	0.28	0.10	0.00
AG3												
R_S	5	7	5	6	6	5	6	6	6	7	8	6
H_O	0.71	0.71	0.65	0.79	0.59	0.63	0.77	0.73	0.75	0.80	0.63	0.77
F_{IS}	0.08	0.10	0.14	−0.01	0.20	0.21	0.09	0.10	0.05	0.03	0.20	0.02
P-values	0.47	0.44	0.06	0.00	0.00	0.15	0.59	0.03	0.04	0.27	0.06	0.03
AC1												
R_S	5	6	6	3	5	5	4	8	6	5	6	3
H_O	0.48	0.71	0.70	0.49	0.67	0.71	0.54	0.80	0.75	0.58	0.56	0.61
F_{IS}	0.31	0.08	0.13	0.24	0.09	0.11	0.19	0.03	0.05	0.27	0.30	0.09
P-values	0.02	0.19	0.00	0.03	0.41	0.26	0.01	0.01	0.37	0.01	0.00	0.79
AG2												
R_S	12	12	12	7	11	10	6	14	11	15	14	10
H_O	0.71	0.68	0.70	0.56	0.74	0.77	0.62	0.84	0.65	0.65	0.80	0.68
F_{IS}	0.12	0.15	0.15	0.20	0.12	0.06	0.17	0.06	0.15	0.17	0.06	0.14

Appendix (continued)

| Locus | Cambodia populations | | | | | | | | | | | |
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
P-values	0.24	0.00	0.03	0.04	0.06	0.36	0.22	0.28	0.19	0.04	0.12	0.03
AT1												
R_S	6.00	6.00	7.00	4.00	6.00	8.00	4.00	8.00	6.00	6.00	8.00	10.00
H_O	0.67	0.66	0.74	0.67	0.59	0.69	0.77	0.78	0.60	0.63	0.73	0.75
F_{IS}	0.21	0.20	0.05	0.12	0.30	0.18	0.04	0.04	0.19	0.19	0.14	0.11
P-values	0.07	0.00	0.31	0.00	0.00	0.00	0.96	0.01	0.41	0.00	0.02	0.47
AC2												
R_S	4	4	3	4	4	3	2	4	5	4	4	5
H_O	0.67	0.53	0.67	0.51	0.48	0.46	0.46	0.73	0.55	0.48	0.63	0.59
F_{IS}	0.12	0.30	0.08	0.27	0.30	0.41	0.24	0.07	0.19	0.36	0.14	0.20
P-values	0.49	0.00	0.00	0.00	0.01	0.00	0.07	0.00	0.03	0.00	0.06	0.07
Average R_S	5.62	6.54	6.23	4.77	5.85	5.62	4.38	7.38	6.23	6.85	7.00	6.77
Average H_O	0.60	0.67	0.65	0.59	0.62	0.63	0.63	0.71	0.64	0.65	0.65	0.66
No. significant departures from HWE	4/13	6/13	8/13	8/13	5/13	6/13	2/13	0/13	4/13	6/13	6/13	5/13
Mean gene diversity over all loci	0.75	0.76	0.74	0.68	0.74	0.76	0.73	0.78	0.72	0.76	0.77	0.75
±SD	0.39	0.39	0.38	0.35	0.38	0.39	0.39	0.40	0.37	0.39	0.39	0.38

| Locus | Other populations | | | | No. significant departures from HWE in all 40 populations | Average F_{IS} in all 40 populations |
	Kenya	Nigeria	Sri Lanka	NE India		
AC4						
R_S	2	4	2	3	7/40	0.11
H_O	0.75	0.63	0.59	0.50		
F_{IS}	−0.11	0.05	0.05	0.23		
P-values	1.00	0.13	0.48	0.39		
AG4						
R_S	3	7	6	3	8/40	0.11
H_O	0.75	0.85	0.71	0.67		
F_{IS}	0.05	−0.02	0.11	−0.22		
P-values	1.00	0.00	0.51	1.00		
CT2						
R_S	5	8	3	2	9/40	0.12
H_O	1.00	0.83	0.65	0.67		
F_{IS}	−0.17	−0.02	0.05	−0.13		
P-values	0.66	0.01	0.66	1.00		
AG1						
R_S	3	4	4	3	13/40	0.11
H_O	1.00	0.70	0.65	0.67		
F_{IS}	−0.33	0.05	0.10	0.02		
P-values	1.00	0.01	0.12	1.00		
AC5						
R_S	5	8	6	5	18/40	0.12
H_O	1.00	0.85	0.76	0.67		
F_{IS}	−0.17	0.02	0.02	0.23		
P-values	0.67	0.50	1.00	0.06		

Appendix (continued)

| | Other populations | | | | | |
Locus	Kenya	Nigeria	Sri Lanka	NE India	No. significant departures from HWE in all 40 populations	Average F_{IS} in all 40 populations
AG5						
R_S	3	6	4	2	20/40	0.15
H_O	0.50	0.65	0.71	0.50		
F_{IS}	0.22	0.18	0.10	0.06		
P-values	1.00	0.00	0.55	1.00		
AG7						
R_S	6	11	6	4	15/40	0.12
H_O	1.00	0.89	0.82	1.00		
F_{IS}	−0.08	0.01	0.00	−0.25		
P-values	1.00	0.33	0.43	0.76		
AC7						
R_S	2	10	5	3	10/40	0.14
H_O	0.25	0.61	0.63	0.67		
F_{IS}	0.46	0.29	0.23	0.14		
P-values	1.00	0.00	0.30	1.00		
AG3						
R_S	3	9	3	3	14/40	0.14
H_O	0.75	0.58	0.56	0.33		
F_{IS}	−0.11	0.31	0.14	0.55		
P-values	1.00	0.00	0.84	0.02		
AC1						
R_S	3	11	5	3	17/40	0.13
H_O	0.75	0.61	0.71	1.00		
F_{IS}	0.00	0.31	0.11	−0.40		
P-values	0.65	0.00	0.34	0.49		
AG2						
R_S	4	10	10	3	14/40	0.11
H_O	0.75	0.74	0.76	0.83		
F_{IS}	0.00	0.14	0.17	−0.22		
P-values	1.00	0.00	0.00	0.79		
AT1						
R_S	4.00	11.00	6.00	3.00	17/40	0.10
H_O	1.00	0.81	0.71	0.67		
F_{IS}	−0.22	0.08	0.14	0.08		
P-values	1.00	0.00	0.27	0.58		
AC2						
R_S	3	6	4	3	19/40	0.16
H_O	0.75	0.60	0.64	0.33		
F_{IS}	−0.24	0.22	0.16	0.58		
P-values	1.00	0.00	0.68	0.33		
Average R_S	3.54	8.08	4.92	3.08		
Average H_O	0.79	0.72	0.68	0.65		
No. significant departures from HWE	0	10/13	1/13	1/13		
Mean gene diversity over all loci	0.71	0.80	0.77	0.67		
±SD	0.41	0.41	0.41	0.37		