

# Do malaria parasites mate non-randomly in the mosquito midgut?

T. J. C. ANDERSON\*, R. E. L. PAUL†, C. A. DONNELLY AND K. P. DAY

Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, Oxford University, South Parks Road, Oxford OX1 3PS, UK

(Received 19 March 1999 and in revised form 26 August 1999 and 1 November 1999)

## Summary

Polymerase chain reaction (PCR)-based genotyping of oocysts dissected from mosquito midguts has previously been used to investigate overall levels of inbreeding within malaria parasite populations. We present a re-analysis of the population structure of *Plasmodium falciparum* malaria using diploid genotypes at three antigen-encoding loci in 118 oocysts dissected from 34 mosquitoes. We use these data to ask whether mating is occurring at random within the mosquito midgut, as is generally assumed. We observe a highly significant deficit of heterozygous oocysts *within mosquitoes* at all three loci, suggesting that fusion of gametes occurs non-randomly in the mosquito gut. A variety of biological explanations, such as interrupted feeding of mosquitoes, positive assortative mating and outcrossing depression, could account for this observation. However, an alternative artefactual explanation – the presence of non-amplifying or null alleles – can account for the observed data equally well, without the need to invoke non-random mating. To evaluate this explanation further, we estimate the frequencies of null alleles within the oocyst population using maximum likelihood, by making the assumption that non-amplifying oocysts at any of the three loci are homozygous for null alleles. Observed levels of visible heterozygotes fit closely with those expected under random mating when non-amplifying oocysts are accounted for. Other lines of evidence also support the artefactual explanation. Overall inbreeding coefficients have been recalculated in the light of this analysis, and may be considerably lower than those estimated previously. In conclusion, we suggest that the deficit of heterozygotes observed is unlikely to indicate non-random mating within the mosquito gut and is better explained by misscoring of heterozygotes as homozygotes.

## 1. Introduction

The level of inbreeding and the rate at which recombination breaks up associations between genes is critically important for many aspects of malaria parasite biology and has been the subject of intense recent discussion. Inbreeding may influence the persistence of clonal genotypes (Paul *et al.*, 1995; Hastings & Wedgewood-Oppenheim, 1997), the maintenance of antigenically distinct ‘strains’ (Gupta *et al.*, 1996), sex

ratio (Read *et al.*, 1992; Dye & Godfray, 1993) and the spread of drug resistance (Dye & Williams, 1997; Hastings, 1997; MacKinnon, 1997; Hastings & MacKinnon, 1998).

Measurement of inbreeding in malaria parasites is difficult. These protozoa have a complex life cycle involving both mitotic division of haploid parasites in the vertebrate host and sexual fusion of macro- and micro-gametes to form a diploid zygote in the mosquito host. Sexual fusion is an obligate part of the life cycle. The transient diploid stage (the ookinete) attaches to the mosquito midgut wall where it develops into an oocyst. These contain the meiotic products, the sporozoites. Inbreeding can be measured within malaria populations by examining levels of heterozygosity in oocysts. This technique has been used to

\* Corresponding author. Department of Genetics, Southwest Foundation for Biomedical Research, PO Box 760549, San Antonio, TX 78245-0549, USA.

† Current address: Biochimie et Biologie Moléculaire des Insectes, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France. e-mail: topotito@pasteur.fr

measure inbreeding in *Plasmodium falciparum* populations in Tanzania (Babiker *et al.*, 1994) and Papua New Guinea (PNG) (Paul *et al.*, 1995) and both studies detected high levels of heterozygote deficit. Overall inbreeding coefficients ( $F$ ) were estimated at 0.34 and 0.9 respectively in Tanzania and PNG. It has been suggested that the differences observed are related to levels of malaria transmission in the two locations. In Tanzania infective bites are common (Smith *et al.*, 1993) and people are frequently infected with multiple malaria clones. Feeding mosquitoes ingest malaria gametocytes of different genotypes and consequently the malaria population is relatively outbred. In contrast, in PNG infective bites are rarer (Burkot *et al.*, 1988a), fewer individuals have multiple infections, and selfing levels are much higher.

In both Tanzania and PNG, analyses were conducted on single oocysts from each mosquito gut. The heterozygote deficits observed were assumed to be due to the partitioning of the parasite population into discrete subpopulations within the human and mosquito hosts. To phrase this in terms of inbreeding coefficients (following the notation of Weir & Cockerham, 1984), the overall inbreeding coefficient ( $F$ ) was assumed to equal  $\theta_m$ , the component due to subdivision of the total population into subpopulations within mosquitoes. Micro- and macrogametes were assumed to mix randomly within the mosquito gut, and therefore the component attributable to inbreeding *within* individual mosquitoes ( $f$ ) was assumed to be zero. However, non-random mating within the mosquito gut could also contribute to overall levels of inbreeding observed. Babiker *et al.* (1994) have investigated this *indirectly* by comparing observed oocyst heterozygosity with expectations derived from the distribution of parasite clones in the human host, and concluded that available data were consistent with random mating within mosquitoes. Unfortunately, their approach had little power to discriminate between random and non-random mating since single oocysts were genotyped from individual mosquitoes.

We reanalyse the data set collected by Paul *et al.* (1995) to investigate mating patterns *within* mosquitoes. Specifically we investigate levels of heterozygote deficit in the genotypes of oocysts from multiply infected mosquitoes. We show that there are high levels of *apparent* heterozygote deficit in oocysts *within* mosquitoes. Consequently, both  $\theta_m$ , the level of heterozygote deficit due to division of the total oocyst population into subpopulations (mosquitoes), and  $f$ , inbreeding within mosquitoes, contribute to overall levels of inbreeding ( $F$ ). However, we also show that the deficit of visible heterozygotes within mosquitoes is likely to be an artefact resulting from substantial numbers of non-amplifying alleles at all three loci typed, which result in heterozygotes being misscored

as homozygotes. We therefore re-evaluate measures of inbreeding coefficients from this data set.

## 2. Materials and methods

### (i) Collection of data

The data analysed were collected by Paul *et al.* (1995); here we briefly summarize the methods used. Indoor-resting catches of blood-fed mosquitoes of the species *Anopheles punctulatus*, *An. farauti* and *An. koliensis* were carried out for 10 consecutive nights in three coastal and three inland villages in Madang Province in the north coast of PNG. The mosquitoes were then held in cages for 6 days to allow oocysts to develop and then on the seventh day dissected for oocysts. Parasite DNA was extracted and N-terminal regions of *MSP-1* and *MSP-2* and *GLURP* were amplified in the polymerase chain reactions (PCR) as described previously (Ranford-Cartwright *et al.*, 1991; Paul *et al.*, 1995). Alleles were defined by polymorphic sequences to which DNA probes hybridized for *MSP-1* and -2 and by size for *GLURP* (Paul *et al.*, 1995).

### (ii) Estimation of allele and genotype frequencies

To ensure that oocysts of other malaria species, such as *Plasmodium vivax*, *P. malariae* and *P. ovale* were not included in the data set, only oocysts which amplified for at least one of the three loci typed were included. We conducted two analyses. In the first, oocysts which failed to amplify for any particular locus were regarded as experimental error and disregarded. This interpretation of the data is referred to as model 1 from here on. In this case the data at each locus were analysed as a two-allele system. In the second analysis (model 2), non-amplifying oocysts were treated as homozygotes for null alleles (Pemberton *et al.*, 1995). In this case the data were analysed as a three-allele system (two scoreable alleles and a non-scoreable null allele). For both analyses, maximum likelihood methods were used to estimate allele frequencies and corresponding genotype frequencies under the assumption of random mating, using a multinomial model for the data (see Appendix).

### (iii) Measuring goodness-of-fit to random mating within mosquitoes

Goodness-of-fit was judged using likelihood ratio (LR) statistics (see Appendix). The accuracy of asymptotic goodness-of-fit tests is known to suffer when the sample size (the number of oocysts within a mosquito) is small compared with the number of categories (genotypes) (Hirji, 1997). In this data set

many of the mosquitoes contain only a single oocyst, making such an approach inappropriate. We therefore estimated goodness-of-fit  $P$ -values empirically by comparing the observed LR statistic for all 34 mosquitoes combined with the distribution of LR statistics obtained from  $10^6$  parametric bootstrap realizations of the data. This was done by resampling genotypes from within each of the 34 mosquitoes and recalculating the LR statistics for each of the resampled oocyst populations. Resampling was done by randomly creating diploid genotypes from an oocyst population with allele frequencies as estimated by maximum likelihood. The LR statistics for the 34 oocyst populations were then summed. This procedure was repeated  $10^6$  times to give a distribution of simulated LR statistics, which were compared with the LR statistics of the observed data.

#### (iv) Calculation of inbreeding coefficients

We calculated inbreeding coefficients (Weir & Cockerham, 1984; Weir, 1996) for both model 1 (non-amplifying oocysts ignored) and model 2 (non-amplifying oocysts included) of the data. We used a three-level hierarchy. We measured inbreeding *within* mosquitoes ( $f$ ), inbreeding due to division of the oocyst population into individual mosquitoes ( $\theta_m$ ), and due to geographical location ( $\theta_g$ ) and overall levels of inbreeding ( $F$ ) which results from all these individual components. The programme GDA (Lewis & Zaykin, 1997) was used to calculate inbreeding coefficients. We calculated confidence intervals by bootstrap resampling of genotypes (for model 1) or alleles (for model 2). For model 1, we regenerated 1000 diploid data sets by resampling diploid genotypes from mosquitoes with the calculated genotype frequencies, while for model 2 we regenerated 1000 data sets by resampling alleles from mosquitoes with the maximum likelihood estimated allele frequencies.  $F$ -statistics were then calculated from each of these simulated data sets using GDA and the mean and 95% confidence intervals of the distribution of values are reported.

### 3. Results

The complete data set comprised 118 oocysts from 34 mosquito midguts, each of which amplified for at least one of the three loci. Twenty-two infected mosquitoes were collected from the three coastal villages, while the other 12 were collected from three villages 15–20 km inland (Paul *et al.*, 1995). Twenty oocysts failed to amplify for *MSP-2*, while 19 and 42 failed to amplify for *MSP-1* and *GLURP* respectively. IC1 and FC27 alleles were found at the *MSP-2* locus. At *MSP-*

1 three alleles (K1, Mad20 and Ro33) were detected. Ro33 was found only in a single oocyst and was grouped with the K1 allelic class for the analysis. *GLURP* alleles were arbitrarily divided into 'large' (> 880 bp) and 'small' (< 880 bp) allelic groups. The observed distribution of two-allele three-locus oocyst genotypes within the 34 mosquitoes is summarized in Table 1.

Expected genotype frequencies of the oocysts within each mosquito population are shown in Table 1 for model 1 (non-amplifying oocysts ignored). Goodness-of-fit was measured empirically using parametric bootstrapping of the data (see Section 2). There is a highly significant deficit of heterozygotes for each of the three loci (goodness-of-fit  $P$  values = 0.032, 0.002 and < 0.001 for *MSP-2*, *MSP-1* and *GLURP* respectively), which indicates that the data are not consistent with random mating if we assume non-amplifying oocysts are due to experimental error (see Table 3). The deficit is clearly observed by inspecting the column totals for each locus in Table 1. This deficit is mainly attributable to oocysts from just a few mosquitoes at each locus. For example, for *MSP-2*, mosquitoes 5, 12, 21, 30 and 33 make the largest contribution to the overall LR statistic.

Table 2 shows the expected allele and genotype frequencies of oocysts populations within each of the 34 mosquitoes for model 2 in which non-amplifying oocysts are included. Under this model of the data some oocysts which were classified as homozygotes will in fact be heterozygotes for non-amplifying alleles. For all three loci the observed data are consistent with random mating if we assume that non-amplifying alleles are present and responsible for oocysts which fail to amplify (goodness-of-fit  $P$  values are > 0.999, 0.991 and 0.949 for *MSP-2*, *MSP-1* and *GLURP* respectively: see Table 3). This suggests that non-amplifying alleles may provide a good explanation of the observed data. The close fit of expectations to observed data is clearly observed by inspecting the column totals at the base of Table 2.

Inbreeding coefficients were estimated for both model 1 and model 2 of the data (Table 4). For model 1 (non-amplifying oocysts ignored), inbreeding within mosquitoes ( $f$ ) is extremely high, confirming our previous analysis (Tables 1, 3). Both  $f$  and  $\theta_m$  contribute to the overall inbreeding coefficient ( $F$ ) of 0.94. The contribution due to subdivision of the oocysts among mosquitoes ( $\theta_m$ ) is only 0.60. In model 2,  $f$  is close to zero, since random mating within mosquitoes is assumed. In this case  $F$ , which is approximately equal to  $\theta_m$ , ranges from 0.44 to 0.56 for the three loci (mean 0.48). For both models of the data geographical subdivision ( $\theta_g$ ) makes no significant contribution to the overall heterozygote deficit.

Table 1. *Observed and expected genotype frequencies at three loci for model 1 of the data, in which non-amplifying oocysts were disregarded*

Mosquito	Oocysts	Observed genotypes				Expected genotypes			
		I/I	F/F	I/F	Negative	I/I	F/F	I/F	LR
<i>MSP-2</i>									
1	2	1	1	0	0	0.5	0.5	1.0	2.8
2	1	1	0	0	0	1.0	0.0	0.0	0.0
3	1	0	0	0	1	nd	nd	nd	nd
4	11	9	0	0	2	9.0	0.0	0.0	0.0
5	6	3	1	0	2	2.3	0.3	1.5	4.5
6	1	1	0	0	0	1.0	0.0	0.0	0.0
7	8	8	0	0	0	8.0	0.0	0.0	0.0
8	7	6	0	1	0	6.0	0.0	0.9	0.1
9	11	8	0	0	3	8.0	0.0	0.0	0.0
10	1	0	0	0	1	nd	nd	nd	nd
11	1	1	0	0	0	1.0	0.0	0.0	0.0
12	7	1	4	0	2	0.2	3.2	1.6	5.0
13	1	0	1	0	0	0.0	1.0	0.0	0.0
14	1	0	1	0	0	0.0	1.0	0.0	0.0
15	1	1	0	0	0	1.0	0.0	0.0	0.0
16	1	0	0	1	0	0.3	0.3	0.5	1.4
17	1	1	0	0	0	1.0	0.0	0.0	0.0
18	1	1	0	0	0	1.0	0.0	0.0	0.0
19	1	1	0	0	0	1.0	0.0	0.0	0.0
20	6	3	0	0	3	3.0	0.0	0.0	0.0
21	7	4	2	0	1	2.7	0.7	2.7	7.6
22	1	1	0	0	0	1.0	0.0	0.0	0.0
23	1	1	0	0	0	1.0	0.0	0.0	0.0
24	8	7	0	1	0	7.0	0.0	0.9	0.1
25	1	1	0	0	0	1.0	0.0	0.0	0.0
26	1	1	0	0	0	1.0	0.0	0.0	0.0
27	3	0	2	0	1	0.0	2.0	0.0	0.0
28	1	0	1	0	0	0.0	1.0	0.0	0.0
29	4	2	0	0	2	2.0	0.0	0.0	0.0
30	6	1	3	0	2	0.3	2.3	1.5	4.5
31	2	0	1	1	0	0.1	1.1	0.8	0.3
32	3	0	2	1	0	0.1	2.1	0.8	0.2
33	7	5	2	0	0	3.6	0.6	2.9	8.4
34	3	0	2	1	0	0.1	2.1	0.8	0.2
Totals	118	69	23	6	20	64	18	16	35.1
		K/K	M/M	K/M	Negative	K/K	M/M	K/M	LR
<i>MSP-1</i>									
1	2	0	0	0	2	nd	nd	nd	nd
2	1	1	0	0	0	1.0	0.0	0.0	0.0
3	1	1	0	0	0	1.0	0.0	0.0	0.0
4	11	2	8	0	1	0.4	6.4	3.2	10.0
5	6	0	5	0	1	0.0	5.0	0.0	0.0
6	1	0	0	0	1	nd	nd	nd	nd
7	8	0	6	0	2	0.0	6.0	0.0	0.0
8	7	0	7	0	0	0.0	7.0	0.0	0.0
9	11	0	8	0	3	0.0	8.0	0.0	0.0
10	1	0	1	0	0	0.0	1.0	0.0	0.0
11	1	0	1	0	0	0.0	1.0	0.0	0.0
12	7	6	1	0	0	5.1	0.1	1.7	5.7
13	1	0	0	1	0	0.3	0.3	0.5	1.4
14	1	0	1	0	0	0.0	1.0	0.0	0.0
15	1	0	0	1	0	0.3	0.3	0.5	1.4
16	1	1	0	0	0	1.0	0.0	0.0	0.0
17	1	0	0	0	1	nd	nd	nd	nd
18	1	0	1	0	0	0.0	1.0	0.0	0.0
19	1	0	0	0	1	nd	nd	nd	nd
20	6	0	4	0	2	0.0	4.0	0.0	0.0
21	7	1	5	0	1	0.2	4.2	1.7	5.4

Table 1 (cont.)

Mosquito	Oocysts	Observed genotypes				Expected genotypes			
		K/K	M/M	K/M	Negative	K/K	M/M	K/M	LR
22	1	0	1	0	0	0-0	1-0	0-0	0-0
23	1	0	0	0	1	nd	nd	nd	nd
24	8	0	8	0	0	0-0	8-0	0-0	0-0
25	1	0	0	0	1	nd	nd	nd	nd
26	1	0	1	0	0	0-0	1-0	0-0	0-0
27	3	3	0	0	0	3-0	0-0	0-0	0-0
28	1	1	0	0	0	1-0	0-0	0-0	0-0
29	4	0	4	0	0	0-0	4-0	0-0	0-0
30	6	0	6	0	0	0-0	6-0	0-0	0-0
31	2	0	1	0	1	0-0	1-0	0-0	0-0
32	3	3	0	0	0	3-0	0-0	0-0	0-0
33	7	6	0	0	1	6-0	0-0	0-0	0-0
34	3	3	0	0	0	3-0	0-0	0-0	0-0
Totals	118	28	69	2	19	25	66	8	23.9
		L/L	S/S	L/S	Negative	L/L	S/S	L/S	LR
<i>GLURP</i>									
1	2	0	0	0	2	nd	nd	nd	nd
2	1	0	0	0	1	nd	nd	nd	nd
3	1	0	0	0	1	nd	nd	nd	nd
4	11	0	7	0	4	0-0	7-0	0-0	0-0
5	6	0	2	0	4	0-0	2-0	0-0	0-0
6	1	0	0	0	1	nd	nd	nd	nd
7	8	6	0	0	2	6-0	0-0	0-0	0-0
8	7	6	1	0	0	5-1	0-1	1-7	5-7
9	11	2	3	0	6	0-8	1-8	2-4	6-7
10	1	0	0	0	1	nd	nd	nd	nd
11	1	0	0	0	1	nd	nd	nd	nd
12	7	4	3	0	0	2-3	1-3	3-4	9-6
13	1	0	1	0	0	0-0	1-0	0-0	0-0
14	1	0	0	0	1	nd	nd	nd	nd
15	1	0	1	0	0	0-0	1-0	0-0	0-0
16	1	0	0	0	1	nd	nd	nd	nd
17	1	0	0	0	1	nd	nd	nd	nd
18	1	1	0	0	0	1-0	0-0	0-0	0-0
19	1	0	1	0	0	0-0	1-0	0-0	0-0
20	6	3	0	0	3	3-0	0-0	0-0	0-0
21	7	3	1	0	3	2-3	0-3	1-5	4-5
22	1	0	1	0	0	0-0	1-0	0-0	0-0
23	1	0	0	0	1	nd	nd	nd	nd
24	8	0	8	0	0	0-0	8-0	0-0	0-0
25	1	0	0	0	1	nd	nd	nd	nd
26	1	0	0	0	1	nd	nd	nd	nd
27	3	0	2	0	1	0-0	2-0	0-0	0-0
28	1	1	0	0	0	1-0	0-0	0-0	0-0
29	4	0	0	0	4	nd	nd	nd	nd
30	6	5	0	0	1	5-0	0-0	0-0	0-0
31	2	1	0	0	1	1-0	0-0	0-0	0-0
32	3	3	0	0	0	3-0	0-0	0-0	0-0
33	7	3	4	0	0	1-3	2-3	3-4	9-6
34	3	1	2	0	0	0-3	1-3	1-3	3-8
Totals	118	39	37	0	42	32	30	14	40

Likelihood ratio (LR) statistics describe the fit of the data with Hardy–Weinberg expectations. IC and FC27 alleles at *MSP*-2 are abbreviated to I and F; K1 and MAD20 alleles at *MSP*-1 are abbreviated to K and M; *GLURP* alleles are abbreviated to L (> 880 bp); and S (< 880 bp) alleles. Mosquitoes containing no amplifying oocysts for a particular locus were excluded from the data set and are marked 'nd'. Totals, summed overall mosquitoes, are shown for each locus.

Table 2. *Observed and maximum likelihood estimations of allele and genotype frequencies at three loci for model 2 of the data, in which it is assumed that non-amplifying oocysts are homozygous for null alleles*

		Maximum likelihood											
		Observed genotypes				Genotypes				Alleles			
Mosquito	Oocysts	I/I or I/O	F/F or F/O	I/F	O/O	I/I or I/O	F/F or F/O	I/F	O/O	O	I	F	LR
<i>MSP-2</i>													
1	2	1	1	0	0	0.7	0.7	0.4	0.2	0.3	0.3	0.3	1.6
2	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
3	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
4	11	9	0	0	2	9.0	0.0	0.0	2.0	0.4	0.6	0.0	0.0
5	6	3	1	0	2	2.7	0.7	0.3	2.2	0.6	0.3	0.1	0.8
6	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
7	8	8	0	0	0	8.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
8	7	6	0	1	0	6.0	0.0	0.9	0.0	0.0	0.9	0.1	0.1
9	11	8	0	0	3	8.0	0.0	0.0	3.0	0.5	0.5	0.0	0.0
10	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
11	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
12	7	1	4	0	2	0.6	3.7	0.4	2.3	0.6	0.1	0.4	1.0
13	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
14	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
15	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
16	1	0	0	1	0	0.3	0.3	0.5	0.0	0.0	0.5	0.5	1.4
17	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
18	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
19	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
20	6	3	0	0	3	3.0	0.0	0.0	3.0	0.7	0.3	0.0	0.0
21	7	4	2	0	1	3.3	1.3	0.8	1.5	0.5	0.4	0.2	2.4
22	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
23	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
24	8	7	0	1	0	7.0	0.0	0.9	0.0	0.0	0.9	0.1	0.1
25	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
26	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
27	3	0	2	0	1	0.0	2.0	0.0	1.0	0.6	0.0	0.4	0.0
28	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
29	4	2	0	0	2	2.0	0.0	0.0	2.0	0.7	0.3	0.0	0.0
30	6	1	3	0	2	0.7	2.7	0.3	2.2	0.6	0.1	0.3	0.8
31	2	0	1	1	0	0.1	1.1	0.8	0.0	0.0	0.3	0.8	0.3
32	3	0	2	1	0	0.1	2.1	0.8	0.0	0.0	0.2	0.8	0.2
33	7	5	2	0	0	4.1	1.0	1.3	0.6	0.3	0.5	0.2	4.9
34	3	0	2	1	0	0.1	2.1	0.8	0.0	0.0	0.2	0.8	0.2
Totals	118	69	23	6	20	67	21	8	22				13.7
		K/K or K/O	M/M or M/O	K/M	O/O	K/K or K/O	M/M or M/O	K/M	O/O	O	K	M	LR
<i>MSP-1</i>													
1	2	0	0	0	2	0.0	0.0	0.0	2.0	1.0	0.0	0.0	0.0
2	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
3	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
4	11	2	8	0	1	1.0	7.2	1.1	1.7	0.4	0.1	0.5	3.5
5	6	0	5	0	1	0.0	5.0	0.0	1.0	0.4	0.0	0.6	0.0
6	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
7	8	0	6	0	2	0.0	6.0	0.0	2.0	0.5	0.0	0.5	0.0
8	7	0	7	0	0	0.0	7.0	0.0	0.0	0.0	0.0	1.0	0.0
9	11	0	8	0	3	0.0	8.0	0.0	3.0	0.5	0.0	0.5	0.0
10	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
11	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
12	7	6	1	0	0	5.5	0.3	0.8	0.4	0.2	0.7	0.1	3.3
13	1	0	0	1	0	0.3	0.2	0.5	0.0	0.0	0.5	0.5	1.4
14	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
15	1	0	0	1	0	0.3	0.2	0.5	0.0	0.0	0.5	0.5	1.4
16	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
17	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0

Table 2 (cont.).

		Maximum likelihood											
		Observed genotypes				Genotypes				Alleles			
Mosquito	Oocysts	K/K or K/O	M/M or M/O	K/M	O/O	K/K or K/O	M/M or M/O	K/M	O/O	O	K	M	LR
18	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
19	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
20	6	0	4	0	2	0.0	4.0	0.0	2.0	0.6	0.0	0.4	0.0
21	7	1	5	0	1	0.5	4.6	0.5	1.3	0.4	0.1	0.5	1.5
22	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
23	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
24	8	0	8	0	0	0.0	8.0	0.0	0.0	0.0	0.0	1.0	0.0
25	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
26	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
27	3	3	0	0	0	3.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
28	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
29	4	0	4	0	0	0.0	4.0	0.0	0.0	0.0	0.0	1.0	0.0
30	6	0	6	0	0	0.0	6.0	0.0	0.0	0.0	0.0	1.0	0.0
31	2	0	1	0	1	0.0	1.0	0.0	1.0	0.7	0.0	0.3	0.0
32	3	3	0	0	0	3.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
33	7	6	0	0	1	6.0	0.0	0.0	1.0	0.4	0.6	0.0	0.0
34	3	3	0	0	0	3.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
Totals	118	28	69	2	19	27	68	3	20				11.1
		L/L or L/O	S/S or S/O	L/S	O/O	L/L or L/O	S/S or S/O	L/S	O/O	O	I	S	LR
<i>GLURP</i>													
1	2	0	0	0	2	0.0	0.0	0.0	2.0	1.0	0.0	0.0	0.0
2	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
3	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
4	11	0	7	0	4	0.0	7.0	0.0	4.0	0.6	0.0	0.4	0.0
5	6	0	2	0	4	0.0	2.0	0.0	4.0	0.8	0.0	0.2	0.0
6	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
7	8	6	0	0	2	6.0	0.0	0.0	2.0	0.5	0.5	0.0	0.0
8	7	6	1	0	0	5.5	0.3	0.8	0.4	0.2	0.7	0.1	3.3
9	11	2	3	0	6	1.7	2.7	0.3	6.3	0.8	0.1	0.1	0.7
10	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
11	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
12	7	4	3	0	0	2.9	1.8	1.5	0.8	0.3	0.4	0.3	5.6
13	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
14	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
15	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
16	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
17	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
18	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
19	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
20	6	3	0	0	3	3.0	0.0	0.0	3.0	0.7	0.3	0.0	0.0
21	7	3	1	0	3	2.8	0.8	0.3	3.2	0.7	0.2	0.1	0.6
22	1	0	1	0	0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
23	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
24	8	0	8	0	0	0.0	8.0	0.0	0.0	0.0	0.0	1.0	0.0
25	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
26	1	0	0	0	1	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
27	3	0	2	0	1	0.0	2.0	0.0	1.0	0.6	0.0	0.4	0.0
28	1	1	0	0	0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
29	4	0	0	0	4	0.0	0.0	0.0	4.0	1.0	0.0	0.0	0.0
30	6	5	0	0	1	5.0	0.0	0.0	1.0	0.4	0.6	0.0	0.0
31	2	1	0	0	1	1.0	0.0	0.0	1.0	0.7	0.3	0.0	0.0
32	3	3	0	0	0	3.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
33	7	3	4	0	0	1.8	2.9	1.5	0.8	0.3	0.3	0.4	5.6
34	3	1	2	0	0	0.5	1.6	0.6	0.3	0.3	0.2	0.5	2.2
Totals	118	39	37	0	42	35	33	5	45				18.1

Likelihood ratio (LR) statistics describe the fit of the data to expectations under the assumptions of random mating. Allele name abbreviations are as described in Table 1, while 'O' represents a null allele.

Table 3. *Fit of observed data to random mating within mosquitoes*

Locus	Model 1 (ignoring non-amplifying oocysts)		Model 2 (including non-amplifying oocysts)	
	Observed LR	GOF <i>P</i> value	Observed LR	GOF <i>P</i> value
<i>MSP-2</i>	35.06	<b>0.032</b>	13.74	> 0.999
<i>MSP-1</i>	23.93	<b>0.002</b>	11.10	0.991
<i>GLURP</i>	39.91	< <b>0.001</b>	18.07	0.949

Goodness-of-fit (GOF) *P* values were estimated empirically for both model 1 (ignoring non-amplifying oocysts) and model 2 (including non-amplifying oocysts) by comparing the observed LR with the distribution of LR statistics obtained from  $10^6$  bootstrap realizations of the data. Significant *P* values are in **bold**.

Table 4. *Inbreeding coefficients estimated using two different models of the observed data*

Model	Locus	$f$ (within mosquitoes)	$\theta_m$ (between mosquitoes)	$\theta_g$ (between locations)	$F$ (overall)
Model 1 (ignoring nulls)	<i>GLURP</i>	1.000 (1.000, 1.000)	0.518 (0.449, 0.750)	-0.086 (-0.130, -0.015)	1.000 (1.000, 1.000)
	<i>MSP-1</i>	0.827 (0.405, 0.861)	0.747 (0.676, 0.922)	0.047 (-0.034, 0.121)	0.956 (0.950, 0.957)
	<i>MSP-2</i>	0.697 (0.406, 0.839)	0.534 (0.393, 0.744)	0.051 (-0.042, 0.247)	0.859 (0.741, 0.936)
	Overall	0.855 (0.725, 0.904)	0.599 (0.567, 0.740)	0.001 (-0.044, 0.068)	0.942 (0.907, 0.965)
Model 2 (including nulls)	<i>GLURP</i>	0.083 (-0.113, 0.277)	0.393 (0.322, 0.471)	0.012 (-0.032, 0.067)	0.444 (0.327, 0.565)
	<i>MSP-1</i>	0.081 (-0.155, 0.305)	0.517 (0.467, 0.577)	0.024 (-0.014, 0.068)	0.557 (0.454, 0.663)
	<i>MSP-2</i>	0.065 (-0.131, 0.258)	0.402 (0.326, 0.474)	0.009 (-0.029, 0.058)	0.442 (0.311, 0.561)
	Overall	0.077 (-0.039, 0.197)	0.439 (0.401, 0.477)	0.015 (-0.010, 0.044)	0.482 (0.414, 0.549)

Estimates are based on data in Table 1 (model 1) and Table 2 (model 2). Ninety-five per cent confidence intervals are derived by parametric bootstrapping of the data (see text).

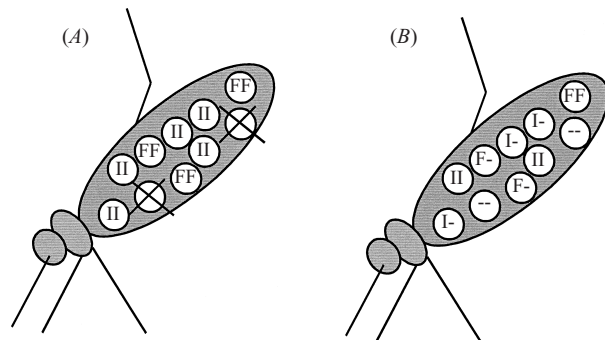


Fig. 1. Two alternative interpretations of the observed data. Oocysts are shown as circles in the abdomen of a mosquito, while the lettering within each oocyst denotes the genotype at a locus with two alleles (I and F). In (A) oocytes which fail to amplify have been ignored (model 1 of the data). In this case, it is assumed that both alleles have been amplified from each oocyst and there is a deficit of heterozygotes relative to Hardy-Weinberg expectations. In (B) oocysts that fail to amplify are assumed to be homozygous for a 'null' allele (-), and many of the other oocysts are heterozygous for the null allele (model 2 of the data). In this case, the observed distribution of genotypes is consistent with Hardy-Weinberg expectations.

#### 4. Discussion

Significant deficits of heterozygotes were observed in oocyst populations within mosquitoes (Tables 1, 3). There are two possible explanations for these deficits.

They may result from non-random fertilization of malaria gametes within the mosquito midgut. Alternatively, they may result from experimental artefacts, which result in heterozygotes being scored incorrectly as homozygotes. These alternative explanations for the observed data are illustrated in Fig. 1. We first discuss possible biological explanations for the observed deficits. We then evaluate the evidence for an artefactual explanation. Finally, we re-evaluate inbreeding coefficients for malaria populations in this region of PNG.

##### (i) Biological explanations

A variety of biological explanations could explain the deficit of heterozygotes within mosquitoes. Four are provided below:

(a) *Positive assortative mating.* If parasites preferentially mate with others of the same clone this would result in heterozygote deficits similar to those observed. Oocysts resulting from genetic crosses between different parasite genotypes have been analysed for *P. falciparum* (Rosenberg *et al.*, 1992; Ranford-Cartwright *et al.*, 1993) and for *P. chabaudi* (Taylor *et al.*, 1997). In none of these cases have deviations from expected genotype frequencies been observed. However, laboratory crosses involving a



limited selection of 'domesticated' parasite clones may provide a rather misleading impression of mating patterns in the field situation.

(b) *Reduced viability of heterozygotes.* Outcrossing depression has been observed in a wide variety of organisms (Waser & Price, 1994; Schneller, 1996; Schierup & Christiansen, 1996; for a review see Templeton, 1994). Outcrossing may break up favourable gene complexes and reduce fitness in organisms that characteristically have high levels of inbreeding. Laboratory crosses between two different clonal malaria genotypes have revealed genetic hyper-variability in telomere-related sequences and chromosome rearrangements in heterologous crosses but not in 'inbred' crosses (Vernick *et al.*, 1988). Such genomic instability during meiotic recombination could have significant effects on the survival of outcrossed ookinetes. This could conceivably result in deficits of heterozygous oocysts in the midgut wall. However, the absence of deviations from Hardy-Weinberg expectations in laboratory crosses argues against this.

(c) *Unequal development rates of meiotic products.* Meiosis results in four meiotic products. These meiotic products divide mitotically to form the sporozoites which fill the oocysts. It is generally assumed that the meiotic products develop and divide at equal rates and are therefore equally represented in the mature oocyst. However, if rates of cell division are unequal, even by a small degree, some meiotic products will be poorly represented in the oocyst. Heterozygous ookinetes could therefore appear to be homozygous when genotyped at the mature oocyst stage, if unequal development rates occur. Once again, the absence of heterozygote deficits in data from laboratory crosses suggests that this explanation is unlikely (Rosenberg *et al.*, 1992; Ranford-Cartwright *et al.*, 1993; Taylor *et al.*, 1997).

(d) *Multiple feeds.* Heterozygote deficits would also be expected if mosquitoes frequently take multiple feeds from different infectious individuals. In this case homozygotes for different alleles might result from blood meals from different people. This explanation is rather unlikely as only 2–3% of infections are gametocyte-producing at any one time in this region of PNG (Graves *et al.*, 1988) and only 10% of mosquitoes take multiple blood meals in this region (Burkot *et al.*, 1988b). Therefore, it is a reasonable assumption that all the oocysts within a single mosquito result from a single blood meal.

## (ii) Artefactual explanations

Four lines of evidence suggest that PCR artefacts may provide the most likely explanation for the apparent deficit of heterozygotes within mosquitoes.

(a) When we make the assumption that non-amplifying oocysts are homozygous for null alleles and re-estimate the frequency of oocyst genotypes, we observe a remarkably good fit to the random mating model (Tables 2, 3; Fig. 1).

(b) If this model of the data is correct, then we might expect levels of heterozygote deficit to be higher in mosquitoes containing non-amplifying oocysts than in mosquitoes in which all oocysts amplify. For *MSP-2*, we can compare the distribution of non-amplifying oocysts and heterozygotes in the data set. Eighteen additional *P. falciparum* oocysts were found in the six mosquitoes containing heterozygous oocysts and no non-amplifying oocysts were recorded among these. However, among the 94 oocysts typed from 28 mosquitoes in which no heterozygous oocysts were recovered, there were 20 non-amplifying oocysts. The difference is significant (Fisher's exact test (two-tailed)  $P = 0.0392$ ), suggesting that non-amplifying oocysts and heterozygote deficit are associated. For *MSP-1* and *GLURP*, 2 and 0 heterozygotes respectively were found, so similar comparisons were not possible for these loci (Table 1).

(c) An artefactual model of the data would also be supported if the loci with the greatest number of non-amplifying oocysts also show the largest heterozygote deficit. In this data set there are 20, 19, and 42 non-amplifying oocysts for *MSP-2*, *MSP-1* and *GLURP* respectively. Six, two and zero heterozygotes were observed for these three loci, and deficits were highest in *GLURP*. *GLURP* has the most oocysts which fail to amplify, as well as fewest heterozygotes, which is once again consistent with an artefactual explanation.

(d) Reanalysis of genotypic data for the same three loci from blood-stage parasites from the same region of PNG (for details and data see Paul *et al.*, 1995 and Paul, 1995) also supports an artefactual explanation. The alleles detected in the blood-stage data were binned in the same way as the oocyst data and the statistical methods of Hill & Babiker (1995) were used to calculate the mean number of clones within individuals for data sets derived from all three loci. These estimates should be similar for all three loci if we assume selective neutrality. However, support intervals for clonal carriage rates estimated from the three loci are non-overlapping in this data set. Assuming a conditional Poisson distribution of clones per person, estimated mean numbers of clones were 1.576 (95% CI: 1.371, 1.848) for *MSP-1*, 2.43 (95% CI: 2.285, 3.122) for *MSP-2* and 1.041 (95% CI: 1.006, 1.131) for *GLURP*. This contrasts with another data set from Tanzania in which both *MSP-1* and *MSP-2* data from blood-stage parasites give similar estimates of clonal carriage rates and show good concordance with observed levels of oocyst heterozygosity (Hill & Babiker, 1995). The discordance observed between clonal carriage rates estimated for

the three loci in the PNG data set suggests either strong selection on some of the loci or, alternatively, considerable differences in the efficiencies of the PCR techniques for each locus. Interestingly, the pattern of non-amplifying loci and mixed infections observed for the three loci in the blood-stage data set closely parallels that observed in the oocyst data set. In the blood-stage data sets *GLURP* has the highest number of amplification failures (32/134), followed by *MSP-1* (16/134) and *MSP-2* (4/134), and the highest clonal carriage rates are found in *MSP-2*, followed by *MSP-1* and *GLURP*. Likewise, in the oocyst data, the highest number of heterozygotes are observed in *MSP-2*, followed by *MSP-1* and *GLURP*.

Alleles may fail to amplify for a variety of reasons. If alleles contain mutations in the primer sites these may prevent primer annealing and eliminate or reduce the efficiency with which the allele is amplified. Amplification of all three loci discussed here involves a two-stage, nested PCR. Hence, primers must match template at four different primer sites. The relevant regions of *MSP-1* and -2 have been sequenced from multiple malaria clones. For *MSP-2* no mismatches have been detected in any of the primer sites. However, in *MSP-1*, a polymorphism (G → A) occurs one base from the 3' end of one of the internal primers in some isolates. However, it is not known how alleles bearing this mismatch amplify when competing with wild-type alleles, as may occur in heterozygous oocysts. For *GLURP* only a single sequence has been published, so it is quite possible that variation occurs in the primer sequences for this locus. Other features of DNA sequences, such as secondary structure and length, may also influence the efficiency of amplification, even when primer sequences are conserved, resulting in loss of alleles from heterozygotes.

The phrase 'allelic dropout' has recently been coined (Gagneux *et al.*, 1997), to describe the stochastic loss of one allelic product from heterozygotes. This may occur when template is limiting, such as with amplification from single hairs (Gagneux *et al.*, 1997). Template DNA prepared from malarial oocysts in this study may also be at limiting concentration. Mosquitoes were kept for 6 days following capture to allow maturation of oocysts. Numbers of sporozoites in mature oocysts vary considerably, ranging from 1359–4445 (Rosenberg & Runsiwongse, 1991) to just under 10000 (Pringle, 1965). Sporozoites are not visible after 6 days, and mitotic division is still occurring at this time. If we assume a constant geometric increase in the numbers of cells per oocyst, a maturation time of 9–10 days (Sinden & Strong, 1978) and 1000–10000 sporozoites in each mature oocyst (Pringle, 1965; Rosenberg & Runsiwongse, 1991), then numbers of cells per oocyst at day 6 range from 200 to 600. The DNA prepared from each oocyst were re-eluted in 25  $\mu$ l and 2  $\mu$ l used

in each PCR reaction. Hence, as few as 16 genome equivalents may have been used as template in each PCR reaction and loss during DNA preparation may reduce this figure still further. Allelic dropout has recently been observed empirically for *MSP-1* and -2 in experiments in which different haploid clones were mixed and low template concentrations used for amplification (Marian Bruce, personal communication). If such stochastic errors cause the observed heterozygote deficits, three replicate amplifications per oocyst for each locus should result in 95% efficiency in recovering the correct genotype (Gagneux *et al.*, 1997). Unfortunately, insufficient oocyst material was left over to investigate whether further heterozygotes could be recovered using this technique.

In our analysis we have assumed that non-amplifying oocysts are homozygous for null alleles and calculated expected genotypes for a three-allele system (two typeable alleles and one null) for each locus. In the case of 'allelic dropout', where alleles are lost stochastically rather than due to differences in amplification efficiency, such an analysis is inappropriate. Furthermore, it is not possible to generate explicit predictions about expected heterozygote deficits in this case. However, we might expect allelic dropout to be a problem when template concentrations are low, and the presence of non-amplifying oocysts within mosquitoes is likely to be indicative of this.

### (iii) *Re-estimation of inbreeding coefficients*

The observation of high levels of heterozygote deficit in oocysts within mosquitoes and the suggestion that this results from non-amplifying alleles calls for a re-evaluation of inbreeding coefficients for this data set. Estimates of inbreeding coefficients for both models of the data are shown in Table 4. Paul *et al.* (1995) calculated overall inbreeding coefficients ( $F$ ) to be 0.9. This figure was calculated by resampling single oocysts from each mosquito. We have recalculated inbreeding coefficients using the whole data set, which allows us to evaluate the contribution of heterozygote deficits *within mosquitoes* to overall inbreeding coefficients. For model 1 (ignoring non-amplifying oocysts), our analysis gives similar values of  $F$  to Paul *et al.* (1995), and also reveals little evidence for geographical variation ( $\theta_g \approx 0$ ). However, subdivision of the oocyst population among mosquitoes ( $\theta_m$ ) accounts for only 0.60 of the total reduction in heterozygosity; the rest results from high levels of inbreeding within mosquitoes ( $f = 0.86$ ).

We can also calculate inbreeding coefficients for model 2 (including non-amplifying oocysts) using the maximum likelihood estimates of non-amplifying alleles in Table 2. In this case mating is assumed to be random within mosquitoes. For this model of the data,  $F$  is substantially lower ( $F = 0.48$ ). This value is

almost half that estimated for model 1 and only slightly higher than that observed by Babiker *et al.* (1994) in Tanzania. The important point here is that even low levels of non-amplifying alleles within the malaria population may severely bias inbreeding coefficients. Until we can eliminate non-amplifying alleles as a cause of heterozygote deficits observed within mosquitoes, this estimate should be viewed as a lower bound for overall levels of inbreeding in this data set. If 'allelic dropout' rather than null alleles causes the heterozygote deficit then overall inbreeding coefficient ( $F$ ) will be somewhat higher. In this case  $F$  will be approximate the value of  $\theta_m$  for model 1 (0.60).

It is worth noting that our statistical methods may have underestimated numbers of null alleles in some oocyst populations and therefore our estimates of heterozygosity and outbreeding may be conservative. This may occur as a result of the low numbers of oocysts in many mosquitoes. The problem is that when we have few oocysts per mosquito and all are positive for one allele, then these may be either homozygous or heterozygous for a non-amplifying allele. We have no information to suggest which is correct, but the maximum likelihood estimated genotypes will be exclusively homozygotes.

In conclusion, the data do not provide strong evidence for non-random fertilization of micro- and macro-gametes within the mosquito gut. A number of lines of evidence suggest that the observed deficits of heterozygotes are the result of technical limitations of the PCR techniques used, which result in heterozygotes being scored incorrectly as homozygotes. We suggest that in future work on oocysts genotypes, multiple oocysts per mosquito should be characterized as done by Paul *et al.* (1995), since this allows evaluation of all components of inbreeding as well as allowing possible sources of error in the data to be identified. More generally, we urge considerable caution when using PCR techniques to score heterozygosity, particularly in situations when levels of template are limiting.

#### Appendix. Calculation of likelihood ratio statistics

Likelihood ratio statistics (LR) are calculated as two times the difference between model log likelihoods and 'saturated' likelihoods. If  $p$  and  $q$  (where  $q = 1 - p$ ) are allele frequencies at a particular locus which can combine to form three genotypes – A ( $p^2$ ), B ( $2pq = 2p(1 - p) = 2(p - p^2)$ ) and C ( $q^2$ ) – and  $n$  is the numbers of oocysts, then the model log likelihood is calculated from the expected genotype frequencies as follows:

$$\begin{aligned} \text{Model log likelihood} &= A \times \ln(p^2) \\ &+ C \times \ln(q^2) \\ &+ B \times \ln(2(p - p^2)). \end{aligned}$$

The saturated log likelihoods are calculated from the observed genotype frequencies as follows:

$$\begin{aligned} \text{Saturated likelihoods} &= A \times \ln(A/n) \\ &+ B \times \ln(B/n) + C \times \ln(C/n). \end{aligned}$$

Hence for the oocyst population in mosquito 1 (see Table 1), the model log likelihood is  $-2.8$ , the saturated likelihood is  $-1.4$  and the Likelihood ratio is  $2(-1.4 - (-2.8)) = 2.8$ . For model 2 (Table 2) we looked at whether the data are consistent with random mating if we include null alleles. In this case oocysts from which only a single allele have been amplified could either be homozygotes, or heterozygotes for a 'null' allele. Here we have a three-allele system with two alleles at frequency  $p$  and  $q$ , while  $r$  is the frequency of a null allele (so that  $p + q + r = 1$ ). There are four possible 'phenotypes' and six genotypes – A ( $pp + 2pr$ ), B ( $qq + 2qr$ ), C ( $2pq$ ) or Z ( $rr$ ) – of which only genotypes C and Z can be unambiguously scored. In this case we estimate the model log likelihood and saturated likelihoods in a similar way to above:

$$\begin{aligned} \text{Model log likelihood} &= A \times \ln(p^2 + 2pr) \\ &+ B \times \ln(q^2 + 2qr) \\ &+ C \times \ln(2pq) + Z \times \ln(r^2), \end{aligned}$$

while

$$\begin{aligned} \text{Saturated log likelihood} &= A \times \ln(A/n) \\ &+ B \times \ln(B/n) \\ &+ C \times \ln(C/n) \\ &+ Z \times \ln(Z/n). \end{aligned}$$

This work was supported by the Wellcome Trust. We thank W. G. Hill and J. T. Williams for assistance with the analysis of blood-stage data, and three anonymous referees for critical comments.

#### References

- Babiker, H. A., Ranford-Cartwright, L. C., Currie, D., Charlwood, J. D., Billingsley, P., Teuscher, T. & Walliker, D. (1994). Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. *Parasitology* **109**, 413–421.
- Burkot, T. R., Graves, P. M., Paru, R., Wirtz, R. A. & Heywood, P. F. (1988a). Human malaria transmission studies in the *Anopheles punctulatus* complex in Papua New Guinea: Sporozoite rates, inoculation rates, and sporozoite densities. *American Journal of Hygiene and Tropical Medicine* **39**, 135–144.
- Burkot, T. R., Graves, P. M., Paru, R. & Lagog, M. (1988b). Mixed blood feeding by the malaria vectors in the *Anopheles punctulatus* complex (Diptera: Culicidae). *Journal of Medical Entomology* **25**, 205–213.
- Dye, C. & Godfray, H. C. (1993). On sex ratio and inbreeding in malaria parasite populations [letter]. *Journal of Theoretical Biology* **161**, 131–134.
- Dye, C. & Williams, B. G. (1997). Multigenic drug resistance among inbred malaria parasites. *Proceedings of the Royal Society of London, Series B* **264**, 61–67.

- Gagneux, P., Boesch, C. & Woodruff, D. (1997). Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair samples. *Molecular Ecology* **6**, 861–868.
- Graves, P. M., Burkot, T. R., Carter, R., Cattani, J. A., Lagog, M., Parker, J., Brabin, B. J., Gibson, F. D., Bradley, D. J. & Alpers, M. P. (1988). Measurement of malarial infectivity of human populations to mosquitoes in the Madang area, Papua New Guinea. *Parasitology* **96**, 215–264.
- Gupta, S., Maiden, M. C. J., Feavers, I. M., Nee, S., May, R. M. & Anderson, R. M. (1996). The maintenance of strain structure in populations of recombining infectious agents. *Nature Medicine* **2**, 437–442.
- Hastings, I. M. (1997). A model for the origins and spread of drug resistant malaria. *Parasitology* **115**, 133–141.
- Hastings, I. M. & MacKinnon, M. (1998). The emergence of drug-resistant malaria. *Parasitology* **117**, 411–417.
- Hastings, I. & Wedgewood-Oppenheim, B. (1997). Sex, strains and virulence. *Parasitology Today* **13**, 375–383.
- Hill, W. G. & Babiker, H. A. (1995). Estimation of numbers of malaria clones in blood samples. *Proceedings of the Royal Society of London, Series B* **262**, 249–257.
- Hill, W. G., Babiker, H. A., Ranford-Cartwright, L. C. & Walliker, D. (1995). Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genetical Research* **65**, 53–61.
- Hirji, W. (1997). A comparison of algorithms for exact goodness-of-fit tests for multinomial data. *Communications in Statistics – Simulations and Computations, Series B* **26**, 1197–1227.
- Lewis, P. O. & Zaykin, D. (1997). Genetic Data Analysis: computer program for the analysis of allelic data. Version 1.0. Free program available from GDA website (<http://chee.unm.edu/gda/>).
- MacKinnon, M. (1997). Survival probability of drug resistant mutants in malaria parasites. *Proceedings of the Royal Society of London, Series B* **264**, 53–59. [Also see erratum: MacKinnon, M. (1997). *Proceedings of the Royal Society of London, Series B* **264**, 1849.]
- Paul, R. E. L. (1995). Unpublished PhD thesis, Oxford University.
- Paul, R. E. L., Packer, M. J., Walmsley, M., Lagog, M., Ranford-Cartwright, L. C., Paru, R. & Day, K. P. (1995). Mating patterns in malaria parasite populations of Papua New Guinea. *Science* **269**, 1709–1711.
- Pemberton, J. M., Slate, J., Bancroft, D. R. & Barrett, J. A. (1995). Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology* **4**, 249–252.
- Pringle, G. (1965). A count of the sporozoites in an oocyst of *Plasmodium falciparum*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **59**, 289–290.
- Ranford-Cartwright, L. C. (1995). Fit for fertilization: mating in malaria parasites. *Parasitology Today* **11**, 154–157.
- Ranford-Cartwright, L. C., Balfe, P., Carter, R. & Walliker, D. (1991). Genetic hybrids of *Plasmodium falciparum* identified by amplification of genomic DNA from single oocysts. *Molecular and Biochemical Parasitology* **49**, 239–244.
- Ranford-Cartwright, L. C., Balfe, P., Carter, R. & Walliker, D. (1993). Frequency of cross-fertilization in the human malaria parasite *Plasmodium falciparum*. *Parasitology* **107**, 11–18.
- Read, A. F., Narara, A., Nee, S., Keymer, A. E. & Day, K. P. (1992). Gametocyte sex ratios as indirect measures of outcrossing rates in malaria. *Parasitology* **104**, 387–395.
- Rosenberg, R. & Rungsiwongse, J. (1991). The number of sporozoites produced by individual malaria oocysts. *American Journal of Tropical Medicine and Hygiene* **45**, 574–577.
- Rosenberg, R., Rungsiwongse, J., Kangsadalampai, S., Sattabongkot, J., Suwanabun, N., Chaiyaroj, S. C. & Mongkolsuk, S. (1992). Random mating of natural *Plasmodium populations* demonstrated from individual oocysts. *Molecular and Biochemical Parasitology* **53**, 129–133.
- Schierup, M. H. & Christiansen, F. B. (1996). Inbreeding depression and outbreeding depression in plants. *Heredity* **77**, 461–468.
- Schneller, J. J. (1996). Outbreeding depression in the fern *Asplenium ruta-muraria* L.: evidence from enzyme electrophoresis, meiotic irregularities and reduced spore viability. *Biological Journal of the Linnean Society* **59**, 281–295.
- Sinden, R. E. & Strong, K. (1978). An ultrastructural study of the sporogonic development of *Plasmodium falciparum* in *Anopheles gambiae*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **72**, 477–491.
- Smith, T., Charlwood, J. D., Kihonda, J., Mwankusye, S., Billingsley, P., Meuwissen, J., Lyimo, E., Takken, W., Teuscher, T. & Tanner, M. (1993). Absence of seasonal variation in malaria parasitaemia in an area of intense seasonal transmission. *Acta Tropica* **54**, 55–72.
- Taylor, L. H., Walliker, D. & Read, A. F. (1997). Mixed genotype infections of malaria parasites: within-host dynamics and transmission success of competing clones. *Proceedings of the Royal Society of London, Series B* **264**, 927–935.
- Templeton, A. R. (1994). Coadaptation, local adaptation, and outbreeding depression. In *Principles of Conservation Biology* (ed. G. K. Meffe & C. R. Carroll), pp. 1–600. Sunderland, MA: Sinauer Associates.
- Vernick, K. D., Walliker, D. & McCutchan, T. F. (1988). Genetic hypervariability of telomere-related sequences is associated with meiosis in *Plasmodium falciparum*. *Nucleic Acids Research* **16**, 6973–6985.
- Waser, N. M. & Price, M. V. (1994). Crossing-distance effects in *Delphinium nelsonii*: outbreeding and inbreeding depression in progeny fitness. *Evolution* **48**, 842–852.
- Weir, B. S. (1996). *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates.
- Weir, B. S. & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.