

## ORIGINAL ARTICLE

# Effective size of the hierarchically structured populations of the agent of malaria: a coalescent-based model

F Prugnolle<sup>1</sup>, P Durand<sup>1</sup>, F Renaud<sup>1</sup> and F Rousset<sup>2</sup>

<sup>1</sup>Laboratoire Génétique et Evolution des Maladies Infectieuses, UMR 2724 CNRS-IRD, IRD de Montpellier, Montpellier, Cedex 5, France and <sup>2</sup>Institut des Sciences de l'Évolution, Université Montpellier 2, CNRS, Place Eugène Bataillon, CC065, Montpellier, Cedex 5, France

Using the coalescence theory, we derived a simple expression for the asymptotic inbreeding effective population size of *Plasmodium falciparum*, the most malignant agent of malaria, in relationship to F-statistics at different hierarchical levels. We consider the effective size of malaria parasites, both for the intrinsic interest of the result for the study of this medically important organism and as an example illustrating general arguments that should clarify effective size calculations in a wide range of organisms with complex life cycles and a hierarchical population structure. We consider in this study a model with four hierarchical levels (villages, oocyst

infrapopulations, oocysts within infrapopulations and the oocyst). The derived expression is applicable to both island and isolation by distance models and is a function of three F-statistics: the genetic differentiation among villages ( $F_{VT}$ ), the genetic differentiation among oocyst infrapopulations ( $F_{MV}$ ) and, finally, the departure from panmixia ( $F_{IM}$ ) within oocyst infrapopulations. The logic of the derivation of effective size presented in this study is applicable to any organism showing the same levels of subdivision.

*Heredity* (2010) **104**, 371–377; doi:10.1038/hdy.2009.127; published online 7 October 2009

**Keywords:** hierarchical structure; effective size; coalescence; F-statistics; malaria

## Introduction

Various studies have shown how the concept of effective size could be used more or less as Sewall Wright initially envisioned it: to generalize in a straightforward manner many of the results of the Wright–Fisher model, with respect to the likelihood of samples (Nordborg, 2001), the distribution of allele frequencies or the fixation probability of mutants (Caballero and Hill, 1992; Roze and Rousset, 2003, 2004). However, the computation of the effective size of spatially structured populations has progressed slowly (Whitlock and Barton, 1997; Nunney, 1999; Rousset, 2003, 2004; Blythe, 2007). In this article, we consider the effective size of malaria parasites, both for the intrinsic interest of the result for the study of this medically important organism and as an example illustrating the general arguments that should clarify effective size calculations in a wide range of organisms with complex life cycles and a hierarchical population structure.

Parasite populations are particularly concerned by population subdivision. For example, malaria parasites

sampled within one mosquito will be more similar to each other than parasites sampled in different, geographically adjacent mosquitoes (for example, within the same house, (Razakandrainibe *et al.*, 2005)). This result owes in part to the co-transmission during the infection events of products of clonal reproduction. A multi-level population structure is relevant not only for malaria but also for many other parasite species (Criscione *et al.*, 2005; Prugnolle *et al.*, 2005b). To understand how their life cycles and population structure affect effective size, one option is to formalize effective size as a function of the parameters that determine the life cycle (for example, selfing rate, percentage of descendants produced by clonal reproduction, variance in reproductive success between individuals during clonal reproduction, migration rate between infrapopulations, migration rate between local populations, number of infrapopulations, number of local populations and regional populations, etc.) (Prugnolle *et al.*, 2005a). However, simple analytic expressions may become difficult to derive when the number of parameters in the model is high. An alternative approach, which is followed in this study, is to derive expressions in terms of Wright's F-statistics (Wang, 1997a,b; Whitlock and Barton, 1997; Nunney, 1999), some of which can easily be estimated using genetic markers. For example, Wright (1938) first derived the effective size of a structured population as

$$N_e = 2N_T / (1 - F_{ST}) \quad (1)$$

where  $N_T$  is the total population size and  $F_{ST}$  Wright's well-known measure of population structure (for example, Wright, 1951). This result has been extended to 'hierarchical'

Correspondence: Dr F Prugnolle, Laboratory of Genetic and Evolution of Infectious Diseases, UMR 2724 CNRS-IRD, IRD de Montpellier, 911 Av. Agropolis, 34394 Montpellier, Cedex 5 BP 64501, France.  
E-mail: prugnoll@mpl.ird.fr or Dr F Rousset, Institut des Sciences de l'Évolution, Université Montpellier 2, CNRS, Place Eugène Bataillon, CC065, 34095 Montpellier Cedex 5, France.  
E-mail: francois.rousset@univ.montp2.fr  
Received 13 February 2009; revised 25 May 2009; accepted 17 August 2009; published online 7 October 2009

models with several levels of structure. An example is given by partial selfing, where the effective size is (under slightly different conditions in Nunney, 1999; Rousset, 2004):

$$N_e = 2N_T / [(1 - F_{ST})(1 + F_{IS})] \quad (2)$$

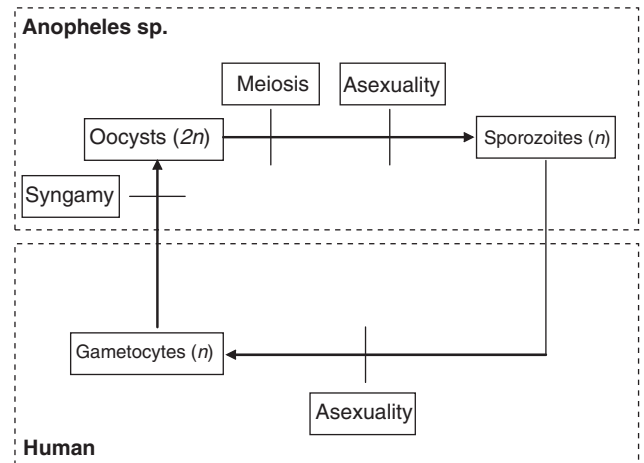
where  $F_{IS}$  is also a well-known measure of covariance of allelic types of homologous genes within individuals. Such results can be obtained with minimal algebra (Rousset, 2004, pp 161, 163) or by simple coalescent arguments (Wakeley and Aliacar, 2001; Rousset, 2003). The purpose of this article is to illustrate the latter approach under a complex parasitic life cycle.

### Life cycle of *Plasmodium falciparum*

The life cycle of *P. falciparum* is complex. Infection in humans begins with the bite of an infected female Anopheline mosquito. The female mosquito injects infective haploid sporozoites into the human host. These are transported through the bloodstream into the liver where they invade hepatocytes. Here, they undergo asexual mitotic replication and give rise to an exoerythrocytic schizont. This process takes 5–7 days and is asymptomatic. Rupture of the liver schizonts releases thousands of merozoites into the bloodstream. Merozoites invade erythrocytes and develop again through asexual mitotic replication into erythrocytic schizonts, containing up to 20 daughter merozoites. On rupture of the schizonts, merozoites can invade fresh erythrocytes and give rise to new schizonts resulting in a cyclical (often synchronous) pattern of blood-stage infection. For *P. falciparum*, each cycle takes up to 48 h. Schizont rupture is typically associated with bouts of rigours and fevers. Some merozoites will go on to differentiate into male (microgametocyte) and female (macrogametocyte) gametocytes. It is to be noted that the same merozoite clone can differentiate into both male and female gametocytes. Gametocytes are taken up by mosquitoes and migrate to the midgut. Male gametocytes undergo a series of changes, including three rounds of genome replication and mitotic division, resulting in the release of eight highly motile, flagellated gametes within only 10 min. The female gametocyte apparently undergoes few obvious morphological changes at this stage and is ready for fertilization. Sexual replication occurs in the mosquito's midgut involving fusion of a female (macro) and male (micro) gamete to form a motile zygote (ookinete), penetration of the midgut wall and formation of oocysts. Meiosis occurs within the oocyst leading to the development of haploid infective sporozoites, which then replicate through a set of mitotic division and migrate to the salivary glands (see <http://highered.mcgraw-hill.com/olc/dl/120090/bio44.swf> for an animated view of the life cycle and Figure 1 for a schematized view of this life cycle highlighting its main steps).

### Model assumptions and formalization

The data considered in this study are oocyst genotypes. Starting from an oocyst (diploid stage), the events of the life cycle that are important here are meiosis, human infection by sporozoites, asexual reproduction, mosquito infection by gametocytes, syngamy and oocyst formation. It is shown below that a generation is an iteration of



**Figure 1** Schematized *Plasmodium falciparum* life cycle. Diploid stages are denoted by  $2n$ , whereas haploid stages are denoted by  $n$ .

this life cycle (see Figure 1 for a schematic view of this life cycle).

The total population is subdivided into  $n$  different villages. Within each village, the parasite local population is again subdivided into  $N_{im}$  infrapopulations of oocysts (an infrapopulation of oocysts corresponds to the group of oocysts present within one individual mosquito).  $N_{im}$  simply corresponds to the number of infected female mosquitoes. The mean number of oocysts per mosquito is considered fixed and denoted by  $N_0$ . When it is variable from one mosquito to another, the harmonic mean should be considered.

Of the various definitions of effective size, the most directly useful definition may be the variance effective size as it is most relevant to modelling joint drift and selection processes (in particular, in diffusion approximations: Cherry and Wakeley (2003); Roze and Rousset (2003)). In such approximations, one expresses the variance of change in the frequency  $p$  of a given allele in the total population in the form  $p(1-p)/N_e$ . Such an expression is not necessarily valid, because the variance may depend on the distribution of the allele in the population, in a manner that cannot be summarized simply in terms of  $p$  (Ewens, 1982). In this respect, the traditional definition of effective size, as the size of an ideal population that would drift at the same rate as the population under study, may not be meaningful. As a result, various other definitions of effective size have been considered (Ewens, 1982; Whitlock and Barton, 1997). However, more or less recent papers have emphasized that the variance of change in allele frequency essentially takes the form  $p(1-p)/N_e$  in a range of models of population structure (Ethier and Nagylaki, 1988; Roze and Rousset, 2003), provided that enough subunits (demes) are considered. In such conditions, different definitions become equivalent, and we may focus on definitions that lead to easy computation by simple coalescent arguments. In particular, the asymptotic inbreeding effective size is defined as:

$$N_e = \lim_{t \rightarrow \infty} \left( \frac{1 - \sum_{k=1}^{t-1} C_{b,k}}{C_{b,t}} \right) \quad (3)$$

where  $C_{b,t}$  is the probability of coalescence at  $t$  for two homologous genes drawn in different demes in the

population. The latter specification ‘in different demes’ is not essential, but is adopted for ease of further exposition. For simplicity, too, we will consider an island model, but the same relationship of eigenvalue effective size with F-statistics can be obtained both for the island model and for the isolation by distance model by an argument valid for both cases (Rousset, 2004, pp 161, 163).

### Computation of $N_e$

We use an argument based on the separation of time scales (Wakeley and Aliacar, 2001; Rousset, 2003 for closely related arguments and Nagylaki, 1980; Ethier and Nagylaki, 1988; Hudson, 1998; Nordborg, 2001 for other early applications of the concept). The probability that two lineages coalesce in a given generation  $T$  is the sum over generations,  $t$ , that the two lineages last came back into the same deme  $T-t$  generations ago, times the probability that coalescence occurs at  $T$  given  $t$ . As the number of demes increases, the latter probability remains non-negligible only over a fixed number of generations. Over this fixed time span, the probability of the previous event  $t$  is essentially a constant,  $P_{\text{same}}$ , over generations. In other words,  $1/N_e$  may be computed as  $1/N_e = P_{\text{same}} \times P_{\text{coal}}$  where  $P_{\text{same}}$  corresponds to the probability that genes in oocysts collected from different villages were from the same village one generation earlier, and  $P_{\text{coal}}$  is the sum of the probabilities of coalescence over the distribution of  $t$ , and therefore is the probability that the first event back in the history of the incoming lineages is a coalescence event.  $P_{\text{coal}}$  is often described as a probability of identity by descent.

$P_{\text{same}}$  is inversely related to the number of villages,  $n$ . By contrast,  $P_{\text{coal}}$  can be approximated to any fixed degree of accuracy by a sum over a finite number (fixed independently of  $n$ ) of generations: it describes events occurring over a faster time scale than the events described by  $P_{\text{same}}$ , and  $P_{\text{coal}}$  may thus be said to describe the ‘instantaneous’ (relative to a time scale of  $n$  generations) coalescence for pairs of genes belonging to the same village.  $n/N_e$  may then be computed as the rate  $n \times P_{\text{same}}$  at which ancestral gene lineages gather in the same village in  $n$  generations, times the probability of ‘instant’ coalescence,  $P_{\text{coal}}$ .

### Coalescence time and F-statistics

In the infinite island model,  $P_{\text{coal}}$  may be well described by F-statistics relative to the total population (Hudson, 1998; Rousset, 2002):  $F_{\text{OT}}$  for two homologous genes within an oocyst,  $F_{\text{MT}}$  for genes from different oocysts in the same mosquito and  $F_{\text{VT}}$  for genes from oocysts in different mosquitoes from the same village. Here,  $F_{\text{MT}} \equiv 1 - (1 - F_{\text{MV}})(1 - F_{\text{VT}})$  where  $F_{\text{MV}}$  corresponds to the probability of identity for two genes drawn within an oocyst infrapopulation compared with the probability of identity for two genes drawn in oocysts from different mosquitoes within the same village.

We will now consider two different situations for computing  $P_{\text{coal}}$ . For exposition, we will first consider a model in which the variance in reproductive success of the different hierarchical levels (infrapopulations of oocysts, oocysts and sporozoites (gametes)) is that of a

Poisson distribution with a mean equalling 1. We will then allow greater variance at all levels.

#### No variance in success greater than expected by a Poisson distribution among units

Genes within villages ‘instantaneously’ coalesce with probability  $F_{\text{VT}}$ , unless they come from the same infected mosquito (with probability  $1/N_{\text{im}}$ ), in which case they are from the same oocyst with probability  $1/N_0$ , and then coalesce with probability  $\frac{1}{2} + \frac{1}{2} F_{\text{OT}}$ , or else are from different oocysts and coalesce with probability  $F_{\text{MT}}$ . Hence

$$P_{\text{coal}} = F_{\text{VT}} + \frac{1}{N_{\text{im}}} \left( \frac{1 + F_{\text{OT}}}{2N_0} + \left( 1 - \frac{1}{N_0} \right) F_{\text{MT}} - F_{\text{VT}} \right) \quad (4)$$

where  $N_{\text{im}}$  is the number of infected mosquitoes per village.

$F_{\text{VT}}$  may be computed as the product of the probability that two gene lineages in a village are obtained from the same village one generation earlier (denoted as  $Z$  below), times the probability that the first event back in their ancestry is then a coalescence event, which is  $P_{\text{coal}}$  by definition:  $F_{\text{VT}} = ZP_{\text{coal}}$ . Both  $P_{\text{coal}}$  and  $Z$  are complex functions of mosquito dispersal, of human dispersal (although probably to a lesser extent) and of the details of clonal reproduction within mosquitoes and humans. However, we do not need to compute these quantities, as the following, previously unnoticed, argument shows. We view the dispersal of two gene lineages as a Markov chain with two states: state 1 = ‘lineages within the same village’ and state 2 = ‘lineages in different villages’. The backward probability transition matrix between these two states is  $G = \begin{pmatrix} Z & 1-Z \\ P_{\text{same}} & 1-P_{\text{same}} \end{pmatrix}$ . Provided that the gene lineages move independently, and that pairs of gene lineages in the same village have the same expected reproductive success as pairs of genes in different villages, the equilibrium distribution between the two states is  $\mathbf{u} = (1, n-1)/n$  and as this distribution must satisfy the equilibrium equation  $\mathbf{u} = \mathbf{u}G$ , then  $P_{\text{same}} = (1-Z)/(n-1)$  and

$$\begin{aligned} \frac{1}{N_e^{\text{Asymptotic}}} &= P_{\text{same}}P_{\text{coal}} = \frac{1-Z}{n-1}P_{\text{coal}} \\ &= \frac{1 - F_{\text{VT}}/P_{\text{coal}}}{n-1}P_{\text{coal}} = \frac{P_{\text{coal}} - F_{\text{VT}}}{n-1} \end{aligned} \quad (5)$$

The numerator  $P_{\text{coal}} - F_{\text{VT}}$  is simply the gain in ‘identity by descent’ when genes come from oocyst(s) in the same mosquito. It is to be noted that the simplifications in Equation (5) follow quite generally from the value of  $\mathbf{u}$ , and thus from the fact that gene lineages move independently. Otherwise, the argument has to be modified (for example, Rousset, 2003; eqs 23 and A.2; Rousset, 2004, p 182).

After rearranging F-statistics in Equation (5), we finally obtain

$$N_e^{\text{Asymptotic}} = \frac{N_{\text{imT}}}{(1 - F_{\text{VT}}) \left( F_{\text{MV}} + (1 - F_{\text{MV}}) \frac{1 + F_{\text{IM}}}{2N_0} \right)} \quad (6)$$

where  $N_{\text{imT}}$  is the number of infected mosquitoes in the total population, and  $F_{\text{IM}}$  corresponds to the probability of identity for two genes drawn within an oocyst compared with the probability of identity for two genes drawn in different oocysts from the same infrapopulation.

With variance in success greater than expected by a Poisson distribution among units

Hill (1972) derived a simple expression for the effect of variance in offspring number on coalescence probability and effective size, and we present in this study a simple generalization of this formula for complex life cycles.

Consider  $D$  descendants of  $P$  parents. The realized number of descendants of each parent  $i$  is denoted by  $d_i$ . Conditional on the  $d_i$ 's, the probability that two randomly chosen descendants are from the same parent is therefore:

$$\sum_i \frac{d_i(d_i - 1)}{D(D - 1)} \quad (7)$$

The expectation of this quantity over the distribution of the  $d_i$ 's conditional on the realized  $D$  value is

$$\frac{\sum_i E(d_i^2|D) - E(d_i|D)}{D(D - 1)} = \frac{\sum_i \text{Var}(d_i|D) - E^2(d_i|D) - E(d_i|D)}{D(D - 1)} \quad (8)$$

Assuming a constant metapopulation size over generations,  $P = D$  and  $E(d_i|D) = 1$ , this becomes

$$\frac{\sum_i \text{Var}(d_i|D)}{D(D - 1)} = \frac{\text{Var}(d_i)}{(D - 1)} \quad (9)$$

In the Wright–Fisher model,  $\text{Var}(d_i) = 1 - 1/D$  and thus the probability that two randomly chosen descendants are from the same parent is  $1/D$ , as expected. To write recursions for probabilities of identity in the presence of a given variance  $\sigma^2$  of offspring numbers, it then suffices to replace any  $1/D$  factor in the recursions (where  $D$  is a number of 'parents') by  $\sigma^2/(D - 1)$ .

Care is required to apply this logic first to the contributions of different oocysts and then to the probability that two lineages coalesce, given that they come from the same oocyst. The probability that the genes come from the same oocyst is directly dependent on the variance of oocyst contribution as shown above. However, the probability of coalescence given that genes come from the same oocyst is an average, over the distribution of  $d_i$ , of the conditional probabilities of coalescence weighted by the probabilities of coming from the same oocyst:

$$\frac{E\left[d_i(d_i - 1)E\left(\sum_{j=1}^2 d_{ij}(d_{ij} - 1)/[d_i(d_i - 1)]|d_i\right)\right]}{E[d_i(d_i - 1)]} \quad (10)$$

where  $d_{ij}$  is the contribution of each homologous gene copy  $j$  of oocyst  $i$  and  $d_i$  the contribution of this oocyst. This reduces to simpler results when the  $d_{ij}$ 's are independent Poisson variables with mean  $\pi_{ij}$ , as then  $E[(d_{ij}(d_{ij} - 1)|d_i) = d_i(d_i - 1)(\pi_{ij}/\sum_j \pi_{ij})^2$ . The usual assumption for diploid organisms is that  $\pi_{ij} = 1/2$ , so that the above probability of coalescence is simply  $1/2$  as expected. However, fluctuations in  $\pi_{ij}$ 's will increase the probability of coalescence above  $1/2$  and, considering that each oocyst produces four sporozoite clones the sizes of which can be affected by various selection events within the mosquito host and then within humans, little can be assumed about such fluctuations and more generally about the joint distribution of the  $d_{ij}$ 's.

Furthermore, an additional hierarchical level may be considered, reflecting high among-mosquito variance in sporozoite transmission. The effects of multiple levels can be described as follows.

Up to Equation (9), no assumption has been made about the biological meaning either of 'parents' and 'descendants' or of details of the life cycle. Therefore, the formula applies to the number of copies, over a cycle from oocyst to oocyst, of any of the two 'gametes' (at a given locus) constitutive of the parental oocyst, as well as to contributions from the two gametes of the oocyst and to contributions of an infected mosquito. In other words, the probability of coalescence can be written as

$$c_1 = \frac{\text{Var}(ABC)}{2N_0N_{\text{im}} - 1} \quad (11)$$

where  $A$  is the total contribution of an infected mosquito to the next generation,  $B$  the fraction of this contribution from a given oocyst and  $C$  the fraction of the oocyst contribution from one of its gametes.  $N_0$  is the mean number of oocysts per mosquito, and  $N_{\text{im}}$  the number of infected mosquitoes per deme.

Similarly, the probability that two gene copies come from the same parental oocyst is

$$c_2 = \frac{\text{Var}(AB)}{N_0N_{\text{im}} - 1} \quad (12)$$

The probability that two gene copies come from the same parental infected mosquito is

$$c_3 = \frac{\text{Var}(A)}{N_{\text{im}} - 1} \quad (13)$$

and the total gain in identity because of the recent coalescence when gene copies come from the same deme (cf Equation (2)) can be written

$$P_{\text{coal}} = (1 - c_3)F_{\text{VT}} + (c_3 - c_2)(F_{\text{MT}}) + (c_2 - c_1)(F_{\text{OT}}) + c_1 \quad (14)$$

$$= F_{\text{VT}} + c_3(F_{\text{MT}} - F_{\text{VT}}) + c_2(F_{\text{OT}} - F_{\text{MT}}) + c_1(1 - F_{\text{OT}}) \quad (15)$$

Further analysis under specific model assumptions could use general formulas for the variance of a product, and in particular

$$\text{Var}\left(\prod_{i=1}^3 Y_i\right) = \prod_{i=1}^3 \text{Var}(Y_i) + \sum_{i=1}^3 E(Y_i^2) \prod_{j \neq i} \text{Var}(Y_j) + \sum_{i=1}^3 \text{Var}(Y_i) E(Y_{j \neq i}^2 Y_{k \neq (i,j)}^2).$$

In a statistical perspective, one would need to obtain estimates of each  $c_i$ , and for malaria transmission the variance of per mosquito contribution ( $A$ ) may be especially difficult to estimate.

Following the same logic as that in Equation (5), it turns out that:

$$N_e^{\text{Asymptotic}} = \frac{n - 1}{(1 - F_{\text{VT}})[c_3 F_{\text{MV}} + (1 - F_{\text{MV}})[c_2 F_{\text{IM}} + c_1(1 - F_{\text{IM}})]]} \quad (16)$$

which is of the same form as Equation (6) but for a more general situation. By considering a Poisson distribution of the realized number of descendants (at each hierarchical level), Equation (16) then equals Equation (6).

## Discussion

As shown in Equations (6) and (16), the effective population size of *P. falciparum* may be seen as a function of the following five variables: (1) the number of infected female mosquitoes, (2) the harmonic mean of the number of oocysts per infected mosquito, (3) the genetic differentiation measured between oocyst infrapopulations, (5) the departure within infrapopulations from the Hardy–Weinberg expectations and finally (5) the genetic differentiation observed between populations from different villages. As shown in Figure 2, each of these variables has a different effect on the effective population size. It decreases as  $F_{MV}$  and  $F_{IM}$  increases, but increases as  $F_{VT}$  increases. Obviously, when  $N_{imT}$  and  $N_0$  increase, so does  $N_e$ . It must be noted that if  $F_{IM}$  equals 0 (no departure from the Hardy–Weinberg expectations) and  $F_{MV}$  is null (no genetic differentiation between oocyst infrapopulations), the effective population size of the total population simply reduces to  $N_e = 2nN_{im}N_0/(1-F_{VT})$ , which corresponds to the classic result of the effective population size for a finite island model (Equation (1)). Similarly, Equation (2) for partial selfing in the island model can also be recovered from Equation (6).

It is worth emphasizing that in our model, as in other standard formulations (Maruyama, 1977; Nagylaki, 1983; Rousset, 2004), the realized number of settled offspring of any given adult is a random variable and so is the realized number of settled offspring from any given deme and hierarchical units. This is similar in all models that assume population regulation after dispersal. Thus, Equation (2), and therefore Equations (6) and (16), holds under the classical assumption of regulation after dispersal (Rousset, 2004), rather than under the assumption that sub-populations contribute equally to the next generation (Nunney, 1999, reproduced in Wang and Caballero, 1999) if contributions are interpreted as the realized number of settled offspring. A model with population regulation only before dispersal would need to assume non-independent dispersal of individuals from different demes to keep fixed deme sizes. The model of Nunney (1999) makes other assumptions distinct from the present models (for example, separate sexes) and could benefit from being reconsidered with the methods of this paper.

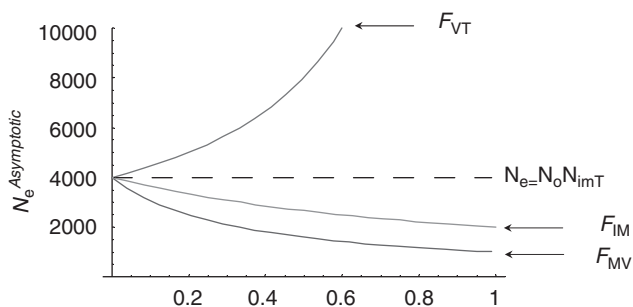
In the context of malaria biology, Equations (6) and (16) would not be exactly true when vectors can die

during dispersal: pairs of gene lineages in the same village have some low probability of being harboured by the same mosquito and then of dying simultaneously if the mosquito dies during emigration, whereas the future contribution of pairs of gene lineages in two different mosquitoes is affected only by the probability that two mosquitoes die. However, the correlation in dispersal events could be substantial only if very few mosquitoes dispersed between villages.

In our model,  $F_{IM}$  measures the departure from panmixia within oocyst infrapopulations. The few studies that have analysed it within natural oocyst infrapopulations reported high positive values (Paul et al., 1995; Razakandrainibe et al., 2005; Annan et al., 2007; Mzilahowa et al., 2007), meaning that infrapopulations harboured more homozygous individuals than expected if gametes had met at random. Two main hypotheses are generally proposed to explain a positive  $F_{IS}$  (in our model,  $F_{IM}$ ) in oocyst infrapopulations: self-fertilization and/or Wahlund effects. Self-fertilization may occur, for instance, because gametocytes (that give both male and female gametes) generally circulate when aggregated in the bloodstream, which may favour amphimixia between gametes from the same parent. Regarding the Wahlund effect, it may occur if the vector takes two successive infective blood meals in two different infected hosts sufficiently spaced to prevent gametes from the two meals fusing. As mosquitoes rarely do so, however (Koella et al., 1998), the positive  $F_{IS}$  observed in oocyst infrapopulations is generally believed to be mainly the consequence of self-fertilization (Razakandrainibe et al., 2005; see however Mzilahowa et al., 2007). In our model, taking into account the potential existence of Wahlund effects within oocyst infrapopulations would require a slightly different formalization. The rate of coalescence within oocyst infrapopulations should then be computed as a function of the probabilities that genes taken in two random oocysts may come from the same or different mosquitoes one generation earlier.

$F_{MV}$  measures the covariance of genes from different oocysts within a mosquito relative to genes from oocysts in different mosquitoes from the same village. Again, the few studies that have analysed the genetic variability of oocyst infrapopulations found strong genetic differentiation between them (Razakandrainibe et al., 2005; Annan et al., 2007). Several aspects of parasite biology or of their hosts may explain the amount of genetic differentiation observed between oocyst infrapopulations. For instance, for the pathogen, the clonal reproduction that occurs during blood stages may induce a strong drift within the human host, especially if there is a strong variance in the reproductive success of the different blood-circulating clones. For the vector, the fact that they rarely take successive blood meals from several infected individuals at the same time (Koella et al., 1998) may limit migration among infrapopulations and thus may also participate in the high differentiation observed.

Finally,  $F_{VT}$  measures the similarity of parasites among mosquitoes within villages relative to the total population. The  $F_{VT}$  relevant to describe the variance of change in allele frequency in the total population can in principle be adequately estimated only from samples from the total population at stationarity. When these unrealistic conditions do not hold, the quantities usually estimated



**Figure 2** Effect of hierarchical structure on the effective population size ( $N_e$ ).  $F_{IM}$  measures the deviation from the Hardy–Weinberg expectations within oocyst infrapopulations,  $F_{MV}$  measures the genetic differentiation between oocyst infrapopulations, and  $F_{VT}$  measures the genetic differentiation between villages.

as  $F_{ST}$  may give poor information about the relevant  $F_{VT}$ . In practice, although, when spatial patterns of isolation by distance are weak, as is often observed and theoretically expected under a wide range of conditions,  $F_{ST}$  measured over small spatial scales could provide reasonably more accurate approximations than the long distance ones (which are more affected by past demographic events, mutations, etc.; Hutchison and Templeton, 1999). Herein,  $F_{VT}$  is mainly dependent on the level of migration occurring between villages, and on the local (within a village) population size (or effective density under isolation by distance), which is also given by the formula for the total effective size, without its among-village component (Rousset, 1999), and is thus again dependent on the local number of infected mosquitoes, the average number of oocysts, self-fertilization and population structure occurring between intrapopulations. Regarding the migration of the parasite between villages, it is necessarily host dependent for *P. falciparum*. Therefore, it depends on both the mobility of humans and that of the vector. In regions of high transmission, the analysis of genetic differentiation among different villages or sites distant by several kilometres up to hundreds or thousands of kilometres have reported very small  $F_{VT}$  (reported as  $F_{ST}$  estimates corrected for the differentiation occurring at lower levels (between intrapopulations) in Annan *et al.* (2007) and Prugnolle *et al.* (2008) and as traditional  $F_{ST}$  in Anderson *et al.* (2000). In contrast, in regions of low transmission, where the local effective size is likely to be lower (due to lower parasite prevalence), other studies reported strong  $F_{VT}$  even between close sites (Bogreau *et al.*, 2006).

## Conclusions

In conclusion, we herein propose simple expressions (Equations (6) or (16)) of the asymptotic inbreeding effective population size of *P. falciparum*. In the case of Equation (6), which is a particular case of Equation (16), the realized number of 'descendants' for each hierarchical unit follows a Poisson distribution. As a consequence, it turns out that all parameters of the equation could be easily measurable on the field and so does  $N_e$ . However, Equation (6) has to be taken with lots of precautions. The model leading to Equation (6) was mainly exposed to facilitate the understanding of the more general model. It is indeed likely that certain hypotheses of this model (in particular, the Poisson distribution of 'descendants') do not hold *in natura*. When the number of realized descendants does not follow this distribution (Equation (16)), this leaves a difficult problem of estimation if applied to real data, especially because the variance of per mosquito contribution may be rather difficult to estimate *in natura*, even after adopting a proper temporal sampling strategy.

Although the equations proposed in this study were initially developed for the specific case of *P. falciparum*, their derivation follows a sufficiently general model applicable to any organisms displaying the same kind of hierarchical structure.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

We thank the two referees for very useful suggestions. We also thank CNRS and IRD for financial support. This research was financed by the ANR 07 SEST 012 'MGANE'.

## References

- Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R *et al.* (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol* **17**: 1467–1482.
- Annan Z, Durand P, Ayala FJ, Arnathau C, Awono-Ambene P, Simard F *et al.* (2007). Population genetic structure of *Plasmodium falciparum* in the two main African vectors, *Anopheles gambiae* and *Anopheles funestus*. *Proc Natl Acad Sci USA* **104**: 7987–7992.
- Blythe RA (2007). The propagation of a cultural or biological trait by neutral genetic drift in a subdivided population. *Theor Popul Biol* **71**: 454–472.
- Bogreau H, Renaud FO, Bouchiba H, Durand P, Assi SB, Henry MC *et al.* (2006). Genetic diversity and structure of African *Plasmodium falciparum* populations in urban and rural areas. *Am J Trop Med Hyg* **74**: 953–959.
- Caballero A, Hill WG (1992). Effects of partial inbreeding on fixation rates and variation of mutant-genes. *Genetics* **131**: 493–507.
- Cherry JL, Wakeley J (2003). A diffusion approximation for selection and drift in a subdivided population. *Genetics* **163**: 421–428.
- Criscione CD, Poulin R, Blouin MS (2005). Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Mol Ecol* **14**: 2247–2257.
- Ethier SN, Nagylaki T (1988). Diffusion approximations of Markov-Chains with 2 time scales and applications to population-genetics. 2. *Adv Appl Probab* **20**: 525–545.
- Ewens WJ (1982). On the concept of the effective population size. *Theor Popul Biol* **21**: 373–378.
- Hill WG (1972). Effective size of populations with overlapping generations. *Theor Popul Biol* **3**: 278–289.
- Hudson RR (1998). Island models and the coalescent process. *Mol Ecol* **7**: 413–418.
- Hutchison DW, Templeton AR (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**: 1898–1914.
- Koella JC, Sorensen FL, Anderson RA (1998). The malaria parasite, *Plasmodium falciparum*, increases the frequency of multiple feeding of its mosquito vector, *Anopheles gambiae*. *Proc R Soc Lond B Biol Sci* **265**: 763–768.
- Maruyama T (1977). *Stochastic Problems in Population Genetics Lecture Notes in Biomathematics*, Vol. 17. Springer: Berlin.
- Mzilahowa T, McCall PJ, Hastings IM (2007). 'Sexual' population structure and genetics of the malaria agent *P. falciparum*. *PLoS ONE* **2**: e613.
- Nagylaki T (1980). The strong-migration limit in geographically structured populations. *J Math Biol* **9**: 101–114.
- Nagylaki T (1983). The robustness of neutral models of geographical variation. *Theor Popul Biol* **24**: 268–294.
- Nordborg M (2001). Coalescent theory. In: Balding DJ, Bishop M, Cannings C (eds). *Handbook of Statistical Genetics*. Wiley: Chichester, UK. pp 179–212.
- Nunney L (1999). The effective size of a hierarchically structured population. *Evolution* **53**: 1–10.
- Paul REL, Packer MJ, Walmsley M, Lagog M, Ranfordcartwright LC, Paru R *et al.* (1995). Mating patterns in malaria parasite populations of Papua-New-Guinea. *Science* **269**: 1709–1711.

- Prugnolle F, Durand P, Jacob K, Razakandrainibe F, Arnathau C, Villarreal D et al. (2008). A comparison of *Anopheles gambiae* and *Plasmodium falciparum* genetic structure over space and time. *Microbes Infect* **10**: 269–275.
- Prugnolle F, Liu H, de Meeus T, Balloux F (2005a). Population genetics of complex life-cycle parasites: an illustration with trematodes. *Int J Parasitol* **35**: 255–263.
- Prugnolle F, Theron A, Pointier JP, Jabbour-Zahab R, Jarne P, Durand P et al. (2005b). Dispersal in a parasitic worm and its two hosts: consequence for local adaptation. *Evolution* **59**: 296–303.
- Razakandrainibe FG, Durand P, Koella JC, De Meeus T, Rousset F, Ayala FJ et al. (2005). ‘Clonal’ population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. *Proc Natl Acad Sci USA* **102**: 17388–17393.
- Rousset F (1999). Genetic differentiation in populations with different classes of individuals. *Theor Popul Biol* **55**: 297–308.
- Rousset F (2002). Inbreeding and relatedness coefficients: what do they measure? *Heredity* **88**: 371–380.
- Rousset F (2003). Effective size in simple metapopulation models. *Heredity* **91**: 107–111.
- Rousset F (2004). *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press: Princeton, NJ.
- Roze D, Rousset F (2003). Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics* **165**: 2153–2166.
- Wakeley J, Aliacar N (2001). Gene genealogies in a metapopulation. *Genetics* **159**: 893–905.
- Wang JL (1997a). Effective size and F-statistics of subdivided populations. 1. Monoecious species with partial selfing. *Genetics* **146**: 1453–1463.
- Wang JL (1997b). Effective size and F-statistics of subdivided populations. 2. Dioecious species. *Genetics* **146**: 1465–1474.
- Wang JL, Caballero A (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82**: 212–226.
- Whitlock MC, Barton NH (1997). The effective size of a subdivided population. *Genetics* **146**: 427–441.
- Wright S (1938). Size of population and breeding structure in relation to evolution. *Science* **87**: 430–431.
- Wright S (1951). The genetical structure of populations. *Ann Eugen* **15**: 323–354.