

# ***F*-statistics under alternation of sexual and asexual reproduction: a model and data from schistosomes (platyhelminth parasites)**

FRANCK PRUGNOLLE,\* DENIS ROZE,† ANDRÉ THÉRON‡ and THIERRY DE MEEÛS\*

\*Equipe ESS, GEMI, UMR-2724, IRD de Montpellier, 911 av. Agropolis, BP 64501, 34394 Montpellier cedex 5, France, †Equipe ETE, CEPM, UMR-2724, IRD de Montpellier, 911 av. Agropolis, BP 64501, 34394 Montpellier cedex 5, France, ‡CBETM, UMR 5555 CNRS-UP, Université de Perpignan, 66860 Perpignan, France

## **Abstract**

Accurate inferences on population genetics data require a sound underlying theoretical null model. Nearly nothing is known about the gene dynamics of organisms with complex life cycles precluding any biological interpretation of population genetics parameters. In this article, we used an infinite island model to derive the expectations of those parameters for the life cycle of a dioecious organism obligatorily alternating sexual and asexual reproductions as it is the case for schistosomes (platyhelminth parasites). This model allowed us to investigate the effects of the degree of mixing among individuals coming from different subpopulations at each new generation (represented in the model by the migration rates before and after clonal reproductions) and the variance in the reproductive success of individuals during the clonal phase. We also consider the effects of different migration rates and degrees of clonal reproductive skew between male and female individuals. Results show that the variance in the reproductive success of clones is very important in shaping the distribution of the genetic variability both within and among subpopulations. Thus, higher variance in the reproductive success of clones generates heterozygous excesses within subpopulations and also increases genetic differentiation between them. Migration occurring before and after asexual reproduction has different effects on the patterns of  $F_{IS}$  and  $F_{ST}$ . When males and females display different degrees of reproductive skew or migration rates, we observe differences in their respective population genetic structure. While results of the model apply to any organism alternating sexual and clonal reproductions (e.g. all parasitic trematodes, many plants, and all aphididae), we finally confront some of these theoretical expectations to empirical data from *Schistosoma mansoni* infecting *Rattus rattus* in Guadeloupe.

*Keywords:* asexual reproduction, complex life cycle, *F*-statistics, heterozygote excesses, schistosomes

Received 13 September 2004; revision received 17 January 2005; accepted 11 February 2005

## **Introduction**

Most discussions on the genetics of populations started with the simplest description of a population as a very large single collection of randomly mating individuals (Hartl & Clark 1997; Provine 2001). Obviously, real populations do

not fit this model and population geneticists had to introduce more complex models to take into account that most natural populations are subdivided into subpopulations connected by migration, and that organisms often do not mate randomly within subpopulations.

Realization of the importance of population structure and mating systems served as the impetus for the derivation of fixation indices (Wright 1921, 1943, 1951, 1965). These indices, also called *F*-statistics, describe how genetic variance is apportioned at different hierarchical levels.  $F_{IS}$  describes the correlation of uniting gametes relative to gametes drawn at random from within a subpopulation;  $F_{ST}$

Correspondence: Franck Prugnolle, Present address: Theoretical and Molecular Population Group, Department of Genetics, Cambridge CB2 3EH, England. E-mail: fp235@mole.bio.cam.ac.uk Fax: +44 1223 333 992. Franck Prugnolle and Denis Roze equally participated to the realization of this article.

the correlation of gametes within subpopulations relative to gametes drawn at random from the entire population; and  $F_{IT}$ , the correlation between uniting gametes relative to gametes randomly drawn from the entire population. For the null model of infinitely sized, random mating subpopulations freely exchanging migrants,  $F_{IS} = F_{ST} = F_{IT} = 0$ . Therefore, deviations from these expected null values provide estimates of population structure and deviations from random mating within subpopulations (Hartl & Clark 1997).

Fixation indices are immensely popular in empirical population genetics and are also the most commonly used tool to describe the expected apportionment of genetic variance in theoretical studies. Thus, expectations were derived for variable migration patterns (Wright 1943; Rousset 1997; Berg *et al.* 1998; Vitalis 2002), variable mating systems (Wright 1965; Chesser 1991a, b; Balloux *et al.* 2003), and different demographic structure and history (Whitlock & McCauley 1990; Whitlock 1992a, b; Laporte & Charlesworth 2002). However, to our knowledge, no results about fixation indices have been derived for species displaying complex life cycles such as organisms alternating different reproductive strategies during their life cycle.

Many organisms reproduce sexually at some stage in their life cycle and asexually at other time (see Table 1) (Calow 1989). If sexuality generally precedes the colonization of new patches or a change in environmental conditions, clonality generally coincides with periods of stable environmental conditions favouring population growth, hence allowing the rapid exploitation of resources (Calow 1989). However, for certain organisms as many pathogens whose complex life cycle involves more than one obligate host, cloning is not only a means of exploiting a patch of resource;

perhaps more important, it is also a means of producing propagules in sufficiently larger quantities to compensate the high risk of failure when transferring from one host to another (Calow 1989).

Many parasite species, which are medically and veterinary very important, enter this category of organisms. Thus, all trematodes (8000 species; De Meeûs & Renaud 2002) as *Schistosoma* spp. (schistosomiasis), *Fasciola* spp. (fasciolosis), *Echinostoma* spp., some cestodes as *Echinococcus* sp., and all apicomplexan as *Plasmodium* spp. (malaria) or *Toxoplasma* spp. (toxoplasmosis), obligatorily alternate sexual and asexual reproductions during their life cycle. Modalities of both sexual and asexual phases can differ from one group to another (e.g. asexual reproduction occurs in diploid stages in trematodes but in haploid ones in apicomplexan). To date, nothing is known about the gene dynamics of such complex life cycles which precludes any good biological interpretation of population genetics parameters. Given the growing development of population genetics studies in parasites because of its use in both pathogen epidemiology and evolution (Constantine 2003), there is therefore a need of null models. This is the problem we address in this study, using trematode parasites and more specifically schistosomes as reference model.

Schistosomes are the most intensively studied trematode parasites (Snyder *et al.* 2001). They are responsible for one of the most important human parasitic disease (schistosomiasis, also known as bilharziasis) in tropical countries. Some 200 million people are infected, of which 20 million are thought to suffer severe consequences of infection (Chitsulo *et al.* 2000). Three species produce the most frequent clinical diseases: *Schistosoma haematobium* (found

**Table 1** Some examples of animal groups alternating sexual and asexual reproductions during their life cycle. These examples are taken from Calow (1989). The list here is incomplete and more examples can be obtained from Calow (1989)

Phylum	Frequency of clonality in taxon	Example
Cnidaria		
Hydrozoa	Common	Freshwater hydras
Scyphozoa	Universal	Medusa
Anthozoa	Common	Corals, anemones
Platyhelminthes		
Turbellaria	Common	Rhabdocoels
Trematoda	Universal	Schistosomatidae
Cestoda	Rare	<i>Echinococcus</i> sp.
Rotifera		
Monogononta	Universal	<i>Asplancha</i> sp.
Annelida		
Oligocheata	Common	Naididae and Aeslosomatidae
Polychaeta	Rare	<i>Dodecaceria</i> sp.
Echinodermata	Rare	Present in some Asteroidea, Ophiuridea and Holothuroidea
Arthropoda		
Insecta	Common	Typical in Aphids, Cynipidae and certain Cecidomyiidae
Crustacea	Common	Typical in cladocerans
Cycliophora	One known species	<i>Symbion pandora</i>

in 53 countries in Africa and the Middle East, including the islands of Madagascar and Mauritius); *Schistosoma mansoni* (in Egypt, northern and southern Africa, some West Indies islands, and the northern two-thirds of South America), and *Schistosoma japonicum* in China and Southeast Asia (Chitsulo *et al.* 2000). Environmental changes linked to climate changes, water resources development, and population movements and growth have led to the spread of the disease to previously low or non-endemic areas, particularly in sub-Saharan Africa (Chitsulo *et al.* 2000).

Among trematodes, schistosomes are particular for two main reasons: they are dioecious and have only two hosts (a freshwater snail as the intermediate host and a vertebrate as the definitive host) (Combes 2001). Sexual reproduction occurs in the definitive host. Eggs are laid out with the faeces of the definitive host in freshwater and hatch into miracidia, the free-living larvae that infect the intermediate host (snail). Inside the snail, an intense phase of clonal multiplication occurs and leads to the production of thousands of free-living cercariae that are infective for the definitive host. From a demographic perspective, this phenomenon has long been considered as an adaptation compensating for the considerable loss of larvae during the free-living transmission phase, thus increasing the probability of meeting the definitive host (Combes 2001). To date, very few studies have investigated the impact of this clonal phase on the genetic structure of the parasite's infrapopulation within definitive hosts (an infrapopulation corresponds to the population of parasites of the same species present within one individual host) and on the apportionment of genetic variability (see, however, Prugnolle *et al.* 2002; Theron *et al.* 2004).

Studies of the genetic structure of infrapopulation of *S. mansoni* infecting black rats (*Rattus rattus*) at a local scale in Guadeloupe have demonstrated a genetic differentiation between infrapopulations of parasites of the same site (Sire *et al.* 2001; Prugnolle *et al.* 2002), and an excess of heterozygotes within infrapopulations (Prugnolle *et al.* 2002). Moreover, multiple infections in rats by the same schistosome genotype are possible (Sire *et al.* 2001) and affect the estimates of *F*-statistics (Prugnolle *et al.* 2002; Theron *et al.* 2004). Recent models have shown that clonal reproduction leads to an excess of heterozygous individuals at neutral polymorphic loci, compared to the expectation under random mating (Balloux *et al.* 2003), but the influence of a complex life cycle on the distribution of genetic variation at different hierarchical levels remains unexplored. In this study, we used an infinite island model to explore the effects of the alternation of sexual and asexual reproductions in a dioecious parasite on the partitioning of variance among infrapopulations and among parasites within infrapopulations. This model allows us to investigate the effects of the degree of mixing among parasites coming from different hosts at each new infection (represented in the model by

the migration rates before and after clonal reproduction), and the variance in the reproductive success of parasites during the clonal phase (which may be seen as a measure of 'reproductive skew'). We will also consider the effects of different migration rates and degrees of clonal reproductive skew between male and female parasites. Although the island model may be considered as an overly simplified representation of the migration and infection patterns of trematode parasites, we see it as a first step in understanding the effects of complex life cycles on the partitioning of genetic variance at different levels. Finally, expectations of the model are tested by analysing a natural data set of genotypes sampled from infrapopulations of *S. mansoni* in Guadeloupe.

## Materials and methods

### Model assumptions and genetic identities

We assume a very large number of subpopulations, each representing the population of parasites within a single host. At the beginning of the life cycle (see Fig. 1), parasites are present in the definitive host, and are about to reproduce sexually. We assume that there are  $N$  parasites per host at this stage, half of them being male and the other female. These individuals reproduce sexually and die. According to the island model, we assume that a proportion  $1 - m_1$  of the offspring produced in the same host remains clustered and infects the same intermediate host (or the same population of intermediate host), whereas the other  $m_1$  infects other intermediate hosts at random. Again, this way of representing transmission between hosts is certainly an oversimplified version of real infection patterns; however, this is a simple way to account for the fact that parasites produced by different definitive hosts can be established in the same intermediate host or subpopulation of intermediate

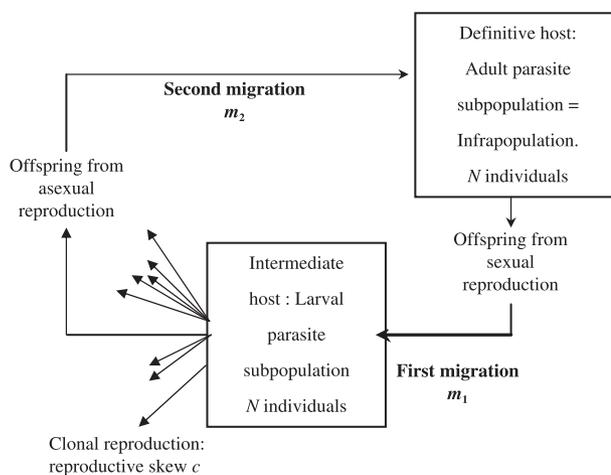


Fig. 1 Schematic representation of the life cycle of schistosomes.

hosts. This 'migration rate' may differ between male and female parasites, in which case we have two parameters  $m_1^M$  and  $m_1^F$ . We also assume that, just after infection, each intermediate host (or subpopulation of intermediate host) contains  $N$  parasites (here it would be easy to modify the model to have a different number, but this would not affect our results qualitatively). These  $N$  individuals produce a large number of asexual offspring within the intermediate host and die. As we will see later, we introduce two parameters  $c^M$  and  $c^F$  representing the amount of reproductive skew among male and female parasites, which measure the variance of the reproductive success in both sexes during the clonal phase. Then the asexual offspring are liberated and infect the definitive host; again, we assume that a proportion  $1 - m_2$  of the offspring produced within the same intermediate host (or subpopulation or intermediate host) remains clustered and infect the same definitive host, while the other  $m_2$  ( $m_2^M$  and  $m_2^F$  if we assume differences between males and females, respectively) infects other hosts at random.

The partitioning of genetic variance is measured by Wright's  $F$ -statistics (Wright 1965).  $F_{IS}$  is defined as (Cockerham 1969, 1973):

$$F_{IS} = \frac{Q_0 - Q_1}{1 - Q_1} \quad (\text{eqn 1})$$

where  $Q_0$  is the probability that the two homologous alleles of an individual at a single locus are identical in state, and  $Q_1$  the probability that two alleles sampled from two different individuals from the same population are identical in state. We assume an infinite allele model, so that identity in state is equal to identity by descent (IBD); however,  $F$ -statistics tend to the same limit values under most mutational models as the mutation rate tends to zero (Rousset 1996). Here, we will consider the values of  $F$ -statistics at this limit.

$F_{ST}$  is defined as (Cockerham 1969, 1973):

$$F_{ST} = \frac{Q_1 - Q_2}{1 - Q_2} \quad (\text{eqn 2})$$

where  $Q_2$  is the probability that two alleles sampled from two different subpopulations are IBD. In the infinite island model,  $Q_2 = 0$ , and therefore we have simply  $F_{ST} = Q_1$ .

We will compute the expected values of  $F_{ST}$  and  $F_{IS}$  at the beginning of the life cycle (in the definitive host, just before sexual reproduction). If male and female parasites have different migration or reproductive skew parameters, or both, the values of these  $F$ -statistics can differ depending on whether alleles are sampled in males or females (e.g. Vitalis 2002). Therefore, we define  $Q_1^{MM}$ ,  $Q_1^{FF}$ , and  $Q_1^{MF}$  as the probabilities that two alleles sampled from two different males and two different females, and from one male and one female from the same subpopulation are IBD. We then have:

$$Q_1 = \frac{1}{4}Q_1^{MM} + \frac{1}{4}Q_1^{FF} + \frac{1}{2}Q_1^{MF} \quad (\text{eqn 3})$$

$$F_{ST}^{MM} = Q_1^{MM} \quad (\text{eqn 4})$$

$$F_{ST}^{FF} = Q_1^{FF} \quad (\text{eqn 5})$$

$$F_{ST}^{MF} = Q_1^{MF} \quad (\text{eqn 6})$$

$$F_{IS}^{MM} = \frac{Q_0 - Q_1^{MM}}{1 - Q_1^{MM}} \quad (\text{eqn 7})$$

$$F_{IS}^{FF} = \frac{Q_0 - Q_1^{FF}}{1 - Q_1^{FF}} \quad (\text{eqn 8})$$

Note that  $Q_0$  is the same for males and females and equals  $Q_1^{MF}$  at equilibrium. To calculate these  $F$ -statistics, we need to calculate the equilibrium value of the vector  $\mathbf{Q} = (Q_0, Q_1^{MM}, Q_1^{FF}, Q_1^{MF})$ . This can be done by writing recurrence equations, and calculating values at equilibrium. After sexual reproduction,  $\mathbf{Q}$  becomes:

$$\mathbf{Q}^s = \mathbf{S} \cdot \mathbf{Q} + \mathbf{S}_c, \quad (\text{eqn 9})$$

where  $\mathbf{S}$  is a  $4 \times 4$  matrix whose first line is  $(0,0,0,1)$ , while the three other lines are

$$\left[ \frac{1}{2N}, \frac{1}{4} \left( 1 - \frac{2}{N} \right), \frac{1}{4} \left( 1 - \frac{2}{N} \right), \frac{1}{2} \right],$$

and

$$\mathbf{S}_c = \left( 0, \frac{1}{2N}, \frac{1}{2N}, \frac{1}{2N} \right)$$

Indeed two genes sampled in two different individuals after sexual reproduction (whether males or females) come from the same parental gene with probability  $1/2N$ , from the two homologous genes of the same parent also with probability  $1/2N$ , from two different parents of the same sex with probability  $(1 - 2N)/4$ , and from two parents of different sexes with probability  $1/2$ . After transmission to the intermediate host, the vector of probabilities of identity becomes:

$$\mathbf{Q}^{m_1} = \mathbf{M}_1 \cdot \mathbf{Q}^s, \quad (\text{eqn 10})$$

where  $\mathbf{M}_1$  is a  $4 \times 4$  matrix with elements 1,

$$(1 - m_1^M)^2, (1 - m_1^F)^2 \text{ and } (1 - m_1^M)(1 - m_1^F)$$

on its diagonal and zeros at all other positions. After asexual reproduction within the intermediate host, we have:

$$\mathbf{Q}^a = \mathbf{A} \cdot \mathbf{Q}^{m_1} + \mathbf{A}_c \quad (\text{eqn 11})$$

$$\text{with } \mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ P^M/2 & 1 - P^M & 0 & 0 \\ P^F/2 & 0 & 1 - P^F & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (\text{eqn 12})$$

and  $A_C = (0, P^M/2, P^F/2, 0)$ . Here  $P^M$  represents the probability that two asexually produced males, within the same intermediate host, have the same father, whereas  $P^F$  is the probability that two asexually produced females, within the same intermediate host, have the same mother.  $P^M$  and  $P^F$  can be expressed as the functions of the mean and variance of the reproductive success of males and females during the asexual phase, which can be condensed into a single parameter which may be seen as a measure of reproductive skew as follows. Just before asexual reproduction,  $N/2$  males are present within each intermediate host, we call  $\mu^M$  and  $v^M$  the mean and variance, respectively, of the male reproductive success. The probability that two males produced asexually within the same intermediate host have the same father is (e.g. Gale 1990):

$$P^M = \frac{v^M + \mu^M(\mu^M - 1)}{\mu^M \left( \frac{\mu^M N}{2} - 1 \right)} \quad (\text{eqn 13})$$

In the case where male offspring are sampled with replacement from the parent males (Wright-Fisher model), we obtain:

$$v_{WF} = \mu^M \left( 1 - \frac{2}{N} \right) \quad (\text{eqn 14})$$

whereas in the case where only one male produces all male offspring within the intermediate host, we have:

$$v_{MAX} = \mu^{M^2} \left( \frac{N}{2} - 1 \right) \quad (\text{eqn 15})$$

We define a 'male reproductive skew' parameter  $c^M$  such that the variance of the male reproductive success equals:

$$v^M = (1 - c^M)v_{WF} + c^M v_{MAX} \quad (\text{eqn 16})$$

Therefore  $v^M$  increases linearly with  $c^M$ , and takes extreme values  $v_{WF}$  (when  $c^M = 0$ ) and  $v_{MAX}$  (when  $c^M = 1$ ). After replacing  $v^M$  in equation 13 with the expression given in equation 16, and using equations 14 and 15,  $\mu^M$  cancels out in the expression of  $P^M$  and one obtains:

$$P^M = \frac{2 + c^M(N - 2)}{N} \quad (\text{eqn 17})$$

Similarly, we define a parameter  $c^F$  that measures female reproductive skew and allows deriving the equation for female probability. This latter equation is the same as for males but replacing M with F.

Finally, after the infection of the definitive host, the matrix of probabilities of identity equals:

$$\mathbf{Q}' = \mathbf{M}_2 \cdot \mathbf{Q}^a \quad (\text{eqn 18})$$

where  $M_2$  is a  $4 \times 4$  matrix with elements 1,

$$(1 - m_2^M)^2, (1 - m_2^F)^2 \text{ and } (1 - m_2^M)(1 - m_2^F)$$

on its diagonal and zeros at all other positions.

The probabilities of identity at equilibrium are therefore obtained by solving:

$$\mathbf{Q} = \mathbf{G} \cdot \mathbf{Q} + \mathbf{G}_c \quad (\text{eqn 19})$$

where  $G = M_2 \cdot AM_1 \cdot S$  and  $G_c = I_2 \cdot A \cdot M_1 S_c + M_2 \cdot A_c$ . The solutions of these equations are complicated and are not given here; however, some useful approximations can be obtained if one assumes that  $N$  is large (large number of parasites within hosts), and that the migration parameters are small (parasites have a strong tendency to remain clustered).

### F-statistics and clonal reproductive skew in natural populations of *Schistosoma mansoni*

*Parasites collection, sampling sites, and genotyping.* Samples collected in 2001 in the marshy forest focus of Guadeloupe have been used to test for the effect of clonality on the distribution of the genetic variability both within and among infrapopulations of parasites. Five sites separated by few kilometres were sampled. In each of these sites, four black rats (*Rattus rattus*) (definitive host) were sampled and parasites were recovered using a standard perfusion technique (Duvall & Dewitt 1967). Within each infrapopulation, 10 male and 10 female *Schistosoma mansoni* were genotyped at seven microsatellite markers (GenBank Accession nos: AF202965, AF202966, R95529, L46951, M85305; Durand *et al.* 2000; AF325695 and AF325697; Curtis *et al.* 2001). DNA extraction protocol is presented in details in Durand *et al.* (2000). Polymerase chain reaction (PCR) amplification processes are presented in Durand *et al.* (2000) and Curtis *et al.* (2001).

*Analyses.* In order to verify how variance of the reproductive success of clones affects F-statistics, we first computed the variance of the number of copies (VNC) of clones within each infrapopulation (one subpopulation) for males and females separately. This variance does not exactly represent the variance in the reproductive success because those individuals that do not produce any offspring after clonal reproduction are not taken into account in the computation of the variance, but it should nevertheless reflect the variance in the reproductive success of the clones. For the infrapopulations where only one clone was present, the VNC was impossible to compute, therefore such infrapopulations have been deleted from the data set.

$F_{IS}$  and  $F_{ST}$  parameters were estimated by Weir & Cockerham's (1984) unbiased estimators  $f$  (for  $F_{IS}$ ) and  $\theta$  (for  $F_{ST}$ ) using FSTAT version 2.9.3 (updated from Goudet 1995). A Spearman rank correlation test was performed to test for the correlation between VNC and  $F_{IS}$  using s-PLUS 2000 (MathSoft). To study the relationship between VNC and  $F_{ST}$ , we analysed the relationship between pairwise  $F_{ST}$  and the mean VNC computed between each infrapopulation pairs of the same transmission site. A Mantel randomization test was performed to test for the correlation between  $F_{ST}$  and the mean VNC using FSTAT version 2.9.3.

## Results

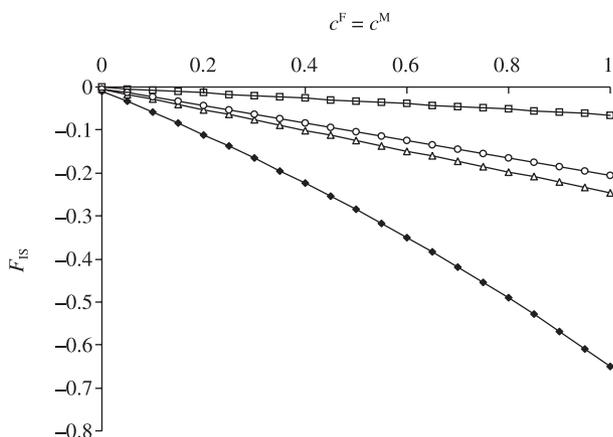
### Deviation from random mating within population ( $F_{IS}$ )

Figure 2 shows  $F_{IS}$  as a function of  $c^M = c^F$  (the reproductive skew parameters, here assumed to be equal in males and females). It can be seen that  $F_{IS} < 0$  and decreases when the variance of the reproductive success of clones increases. All things being equal,  $F_{IS}$  computed for females or for males is lower than  $F_{IS}$  for all individuals within the same subpopulation (Fig. 2).

Assuming all parameters being identical for male and female, an approximation for  $F_{IS}$  when  $N$  is large and the migration parameters small is

$$F_{IS} \approx -\frac{c(m_1 + m_2)}{(4 - c)m_1 + (4 + c)m_2} \quad (\text{eqn 20})$$

which decreases as  $m_1$  increases and increases as  $m_2$  increases; hence, migration occurring before and after



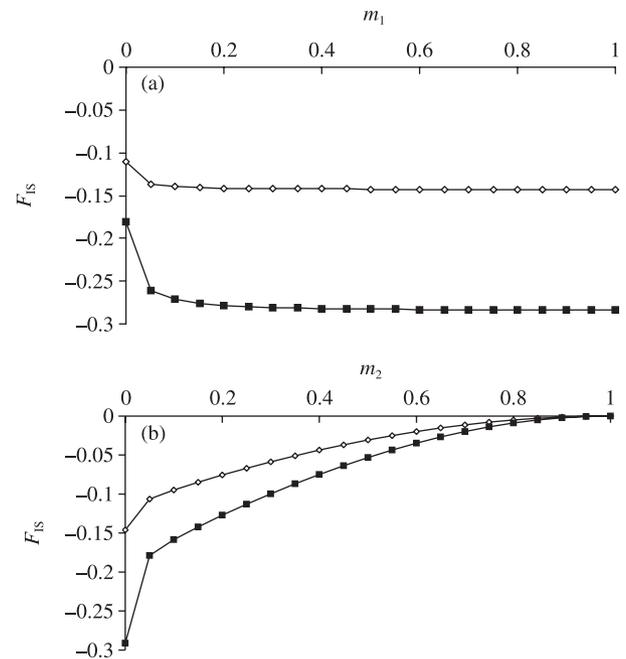
**Fig. 2** Equilibrium values of adult  $F_{IS}$  in relation to  $c$ , the parameter that makes the variance of the reproductive success of clones varying between the Wright–Fisher model ( $c = 0$ ) and the maximal variance (that is when only one individual of each sex gives all the descendants of the same sex after asexual reproduction) ( $c = 1$ ). For all relations,  $N = 100$ ;  $c^F = c^M$ ; and  $m_1^F = m_1^M = m_2^F = m_2^M = m$ . For total  $F_{IS}$ : open squares,  $m = 0.5$ ; open circles,  $m = 0.1$ ; and open triangles,  $m = 0.01$ . For sex-specific  $F_{IS}$ : black diamonds  $m = 0.01$ .

clonality have opposite effects on  $F_{IS}$ . From nonsimplified equations of  $F_{IS}$ , we observed the same result for arbitrary values of  $N$  and migration parameters, as illustrated in Fig. 3(a, b). Thus, for fixed values of  $c$ , when migration after sexual reproduction increases (whereas migration rate after asexual reproduction is fixed),  $F_{IS}$  values decrease (Fig. 3a). On the contrary, when the migration rate after asexual reproduction increases (whereas the migration rate after sexual reproduction is fixed),  $F_{IS}$  values increase (Fig. 3b).

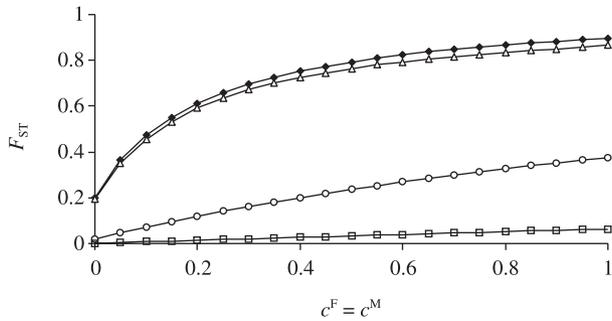
### Population differentiation ( $F_{ST}$ ) between subpopulations

In Fig. 4, we plot  $F_{ST}$  against  $c$  (given that  $c^M = c^F = c$ ) and consider that migration is equal between males and females and before and after asexual reproductions. The variance in the reproductive success of clones has a strong effect on population differentiation. When it increases,  $F_{ST}$  also increases.

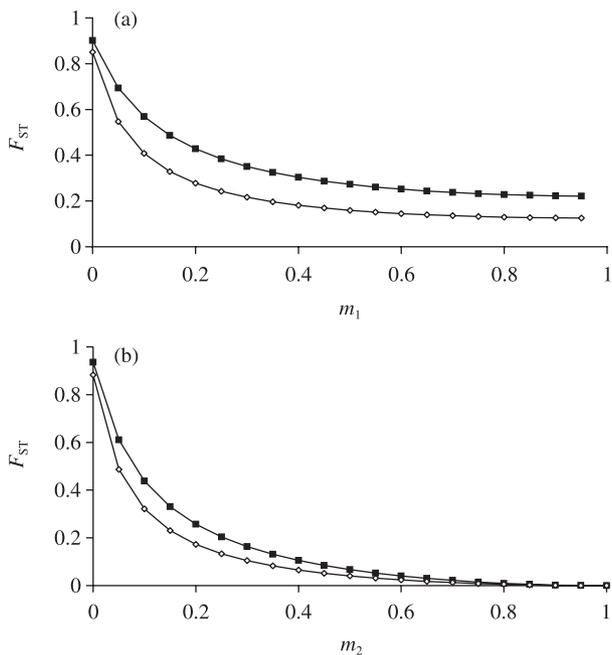
As expected, increasing migration either before or after clonal reproductions decreases the value of  $F_{ST}$  (Fig. 5a, b).  $F_{ST}$  tends to zero as migration after clonal reproduction ( $m_2$ ) tends to 1, and to a positive value when migration before clonal reproduction ( $m_1$ ) tends to 1 (provided that  $m_2 < 1$ ); indeed, if  $m_2 < 1$  and  $c > 0$ , genes sampled in two males or two females from the same deme can coalesce, even if  $m_1 = 1$ .



**Fig. 3** Equilibrium values of total adult  $F_{IS}$  in relation to the migration rate occurring (a) before ( $m_1$ ) [ $N = 100$ ;  $c^F = c^M = c$ ;  $m_2^F = m_2^M = m_2 = 0.01$ ] and (b) after asexual reproduction ( $m_2$ ) [ $N = 100$ ;  $c^F = c^M = c$ ; and  $m_1^F = m_1^M = m_1 = 0.01$ ]. In a and b: open circles  $c = 0.5$  and black squares  $c = 0.1$ .



**Fig. 4** Equilibrium values of  $F_{ST}$  in relation to  $c$ . For all relations,  $N = 100$ ;  $c^F = c^M$ ; and  $m_1^F = m_1^M = m_2^M = m_2^F = m$ . For total  $F_{ST}$ : open squares,  $m = 0.5$ ; open circles,  $m = 0.1$ ; and open triangles,  $m = 0.01$ . For sex-specific  $F_{ST}$ : black diamonds:  $m = 0.01$ .



**Fig. 5** Equilibrium values of total adult  $F_{ST}$  in relation to the migration rate occurring (a) before ( $m_1 = m_1^F = m_1^M$ ) [ $N = 100$ ;  $c^F = c^M = c$ ; and  $m_2^F = m_2^M = 0.01$ ] and (b) after asexual reproduction ( $m_2 = m_2^F = m_2^M$ ) [ $N = 100$ ;  $c^F = c^M = c$ ; and  $m_1^F = m_1^M = 0.01$ ]. In (a) and (b): open circles,  $c = 0.5$  and black squares,  $c = 0.1$ .

*Male–female differences*

Different migration parameters, or different variances, or both in clonal reproductive success between males and females will lead to different values for  $F_{ST}$  and  $F_{IS}$  depending on whether genes are sampled in males or females. We define:

$$\Delta F_{ST} = \frac{F_{ST}^{MM} - F_{ST}^{FF}}{F_{ST}} \tag{eqn 21}$$

Assuming that  $N$  is large and the migration parameters small, if  $m_1$  differs among males and females, all other things being equal, equation 21 reduces to:

$$\Delta F_{ST} \approx 2(1 - c)(m_1^F - m_1^M) \tag{eqn 22}$$

Similarly, if  $m_2$  differs between males and females, all else being equal, equation 21 reduces to:

$$\Delta F_{ST} \approx 2(m_2^F - m_2^M) \tag{eqn 23}$$

Finally, if  $c$  differs between males and females, all else being equal, we obtain

$$\Delta F_{ST} \approx \frac{8(c^M - c^F)(m_1 + m_2)}{c^M + c^F} \tag{eqn 24}$$

Approximations for

$$\Delta F_{IS} = (F_{IS}^{MM} - F_{IS}^{FF})/F_{IS}$$

are more complicated; however, they indicate that  $\Delta F_{IS}$  should have the same sign as  $\Delta F_{ST}$ . Indeed, we always have

$$Q_0 < (Q_1^{MM}, Q_1^{FF})$$

in our model (as will be discussed later). From the definitions of the different  $F$ -statistics (equation. 1 to 8), one shows easily that if

$$Q_1^{MM} > Q_1^{FF}$$

(that is, if  $\Delta F_{ST} > 0$ ) then

$$F_{IS}^{MM} > F_{IS}^{FF},$$

which means

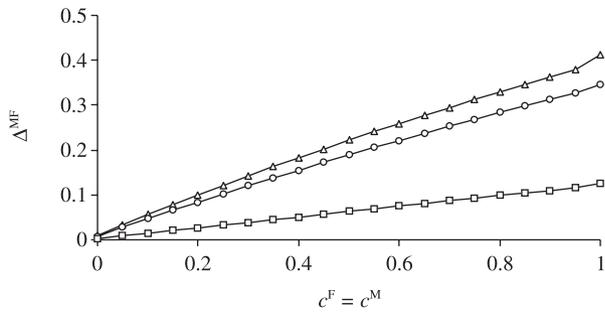
$$\Delta F_{IS} > 0 \text{ (} F_{IS}^{MM}, F_{IS}^{FF} \text{ and } F_{IS} \text{ are always negative).}$$

*Genetic differentiation between males and females within subpopulations ( $\Delta^{MF}$ )*

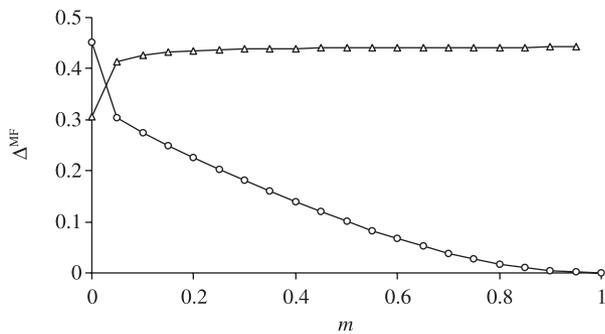
As will be discussed later, both the clonal phase and differences in male and female migration rates can cause some degree of genetic differentiation between males and females of the same subpopulation (before sexual reproduction). This differentiation, that we call  $\Delta^{MF}$ , may be measured as a difference in the probability of identity of two genes, depending on whether they are sampled in different individuals of the same sex or in individuals of different sexes, from the same subpopulation:

$$\Delta^{MF} = \frac{(Q_1^{MM} + Q_1^{FF})/2 - Q_1^{MF}}{1 - Q_1^{MF}} \tag{eqn 25}$$

As shown in Fig. 6,  $\Delta^{MF}$  increases as the variance of the reproductive success of clones increases (everything else being equal). Assuming identical migration rates for males



**Fig. 6** Equilibrium values of  $\Delta^{MF}$  in relation to  $c$ . For all relations,  $N = 100$ ;  $c^F = c^M$ ;  $m_1^F = m_1^M = m_2^F = m_2^M = m$ ; open squares,  $m = 0.5$ ; open circles,  $m = 0.1$ ; and open triangles,  $m = 0.01$ .



**Fig. 7** Equilibrium values of  $\Delta^{MF}$  in relation to the migration rate occurring before (open triangles,  $N = 100$ ;  $c^F = c^M = 0.1$ ;  $m_2^F = m_2^M = 0.01$ ) and after asexual reproduction (open circles,  $N = 100$ ;  $c^F = c^M = 0.1$ ).

and females, migration occurring before asexual reproduction increases  $\Delta^{MF}$ , whereas migration occurring after asexual reproduction decreases it (Fig. 7). When  $N$  is large and migration rates small, we obtain the following approximation:

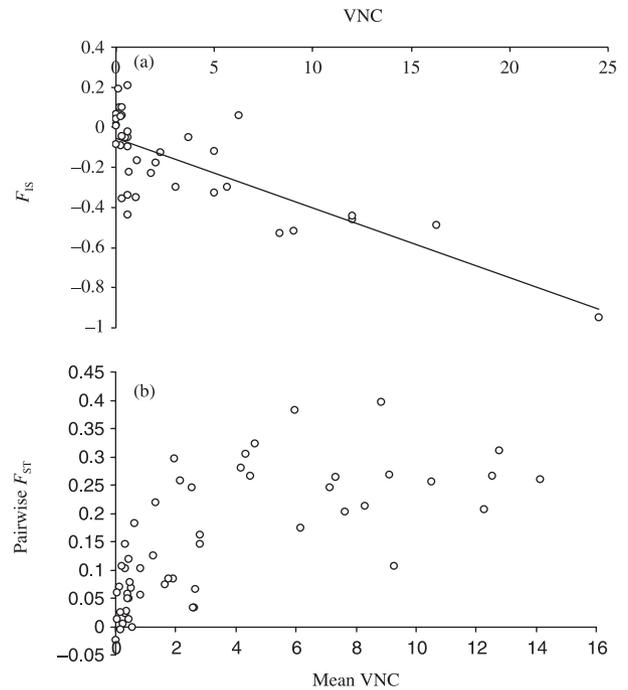
$$\Delta^{MF} \approx \frac{c(m_1 + m_2)}{2m_1 + (2 + c)m_2} \quad (\text{eqn 26})$$

#### $F_{IS}$ and $F_{ST}$ against VNC in natural infrapopulations

As shown in Fig. 8(a, b), variation of  $F_{IS}$  and  $F_{ST}$  against VNC of *S. mansoni* clones in Guadeloupe follows the same pattern as expected under our model.  $F_{IS}$  decreases when VNC increases within each infrapopulation (Spearman rank correlation test:  $R = -0.67$ ;  $P = 0.0001$ ); pairwise  $F_{ST}$  increases when the mean VNC of clones computed between pairs of infrapopulations increases (Mantel correlation test:  $R = 0.75$ ;  $P = 0.0005$ ).

## Discussion

We used an analytical approach to investigate the dynamics of genetic variance in subdivided populations of a complex



**Fig. 8** Relationship between the variance in the number of copies per clone (VNC) observed in natural infrapopulations of *Schistosoma mansoni* and (a)  $F_{IS}$  and (b) pairwise  $F_{ST}$ .

life cycle parasite with an obligate alternation between sexual and asexual reproductions. Different asexual reproduction patterns were defined in our model by different variances in the reproductive success of clones. Results show that the variance in the reproductive success of clones is very important in shaping the distribution of the genetic variability either within and among infrapopulations of adult parasites. While this model was primarily performed on *Schistosoma mansoni* as a reference life cycle model, all results concerning the effect of the variance in the reproductive success of clones can nevertheless also apply to any diploid organism alternating sexual and asexual reproductions during its life cycle (see Table 1).

#### Heterozygote excess (negative $F_{IS}$ values)

If heterozygote deficiencies are commonly observed in living organisms (e.g. Hartl & Clark 1997), the presence of heterozygote excesses is relatively rare. To date, two main types of hypotheses have been proposed to explain these heterozygote excesses: (i) causes that are relative to selection processes (overdominance and associative overdominance, Hartl & Clark 1997) and (ii) causes that are relative to particular population biological characteristics. This is the case, for example, in clonal populations showing extreme rates of clonality (Balloux *et al.* 2003). This is also the case in dioecious species when male and female, before

reproduction, are genetically differentiated, that is, when they display different allelic frequencies within subpopulations. Such deviations in allelic frequencies between sexes can occur particularly in two situations: (i) when the effective number of breeders is low (the difference between male and female allelic frequencies is then the result of binomial sampling error and is responsible for heterozygote excesses in offspring) (Luikart & Cornuet 1999) and (ii) when migration before reproduction is sex-specific that is when males and females migrate at different rates between subpopulations (Prout 1981; Berg *et al.* 1998). Finally, inbreeding avoidance is also expected to generate heterozygote excesses in offspring (Schwartz & Armitage 1980).

In our model, the interaction between two factors contributes to the observed heterozygote excesses (negative  $F_{IS}$  values), which are the variance in the reproductive success of clones and dioecy. The fact that  $F_{IS}$  is negative means that the two homologous genes of the same individual are on average less similar than two genes sampled randomly from the same subpopulation. Here, this is because of the fact that two homologous alleles sampled from two different males, or two different females, and from the same subpopulation just before sexual reproduction, can coalesce during the previous event of clonal reproduction (with a probability that increases as the degree of reproductive skew, or the variance in reproductive success, increases). The two homologous alleles of an individual, however, cannot coalesce during the previous event of clonal reproduction, or during the previous event of sexual reproduction, because individuals are dioecious (in fact, they can coalesce only during a sexual reproduction event anterior to the previous one). The fact that there is some degree of genetic differentiation between males and females of the same subpopulation also contributes to the decrease of probability that the two homologous genes of an individual are identical.

This effect of clonal reproductive skew on  $F_{IS}$  is confirmed in natural infrapopulations of parasites as when the variance in the number of copies per clone increases,  $F_{IS}$  decreases. Besides, when only one clone is present within infrapopulations,  $F_{IS}$  equals  $-1$  (some loci are fixed homozygous and  $F_{IS}$  is not defined for them, whereas for loci with fixed heterozygosity,  $F_{IS}$  equals  $-1$ ).

#### *Migration before and after asexual reproduction: effects on $F_{IS}$*

*Before asexual reproduction ( $m_1$ ).* When migration before asexual reproduction increases,  $F_{IS}$  decreases in adult populations (becomes more negative). This is because of the fact that migration in juveniles before clonality strongly decreases the IBD between two alleles sampled in one male and one female from the same subpopulation of adults ( $Q_1^{MF}$ ); indeed, these alleles can coalesce only if they were

in the same deme during the previous event of sexual reproduction. As a consequence, the identity between homologous genes of the same individual ( $Q_0$ ) decreases (as one comes from a male and the other from a female adult of the previous generation), and tends to 0 when the migration rate is maximal ( $m_1 = 1$ ). Identities between genes sampled in two different adult males, or two different adult females from the same subpopulation ( $Q_1^{MM}$  and  $Q_1^{FF}$ ) also decrease, but converge towards positive values determined by the variance of the reproductive success of clones. Indeed, two genes sampled from two males or two females from the same subpopulation can coalesce during the previous event of clonal reproduction, even if  $m_1 = 1$ .  $Q_0$  decreases faster than  $Q_1$  as  $m_1$  increases, leading to more negative  $F_{IS}$  values.

*After asexual reproduction ( $m_2$ ).* We saw that when migration occurring after asexual reproduction increases,  $F_{IS}$  increases and converges to 0 when  $m_2 = 1$ . This is because of the fact that as  $m_2$  increases, both  $Q_0$  and  $Q_1$  decrease and converge to 0, the difference between  $Q_0$  and  $Q_1$  being progressively reduced (when  $m_2 = 1$ , no coalescence is possible between two genes sampled from the same subpopulation, whether they come from the same individual or from two different individuals).

#### *Variance of the reproductive success of clones and $F_{ST}$*

Increasing the variance in the reproductive success of clones increases the probability that the two alleles sampled from two males, or two females of the same subpopulation just before sexual reproduction coalesce during the previous event of clonal reproduction (i.e.  $Q_1$  increases). Therefore, as the variance increases, so does  $F_{ST}$ . Obviously,  $F_{ST}$  is also affected by migration. This effect of clonal reproductive skew on  $F_{ST}$  is confirmed in natural infrapopulations of parasites. When the mean variance in the number of copies of clones increases (VNC), pairwise  $F_{ST}$  computed between *S. mansoni* infrapopulations within the same site also increases.

#### *Conclusions for natural populations of schistosomes ...*

This model demonstrates the importance of clonal reproduction in shaping the genetic variability both within and between infrapopulations of parasites. We have seen that the variance in reproductive success of clones and also migration rates occurring before and after asexual reproduction have important effects. Results from field studies support these conclusions.

Prugnolle *et al.* (2004) demonstrated that the clonal reproductive skew of males and females was different in *Schistosoma mansoni* infecting *Rattus rattus* in Guadeloupe because of a potential inbreeding depression process

affecting more, specifically females. This pattern is responsible, in part, for the sex-specific genetic structure observed in the parasite at a local scale in Guadeloupe and is characterized by the fact that female infrapopulations display significantly more heterozygote excesses and are more genetically differentiated than male infrapopulations. However, the signal of a sex-specific genetic structure is conserved when copies of clones are deleted from the data set when any effect of the variance of the reproductive success of clones on the distribution of the genetic variability is removed (Prugnolle *et al.* 2002). This implies that other factors such as a difference in the effective migration rates ( $m_1$  or  $m_2$ ) between male and female larval stages (Prugnolle *et al.* 2002, 2003) are also responsible for this pattern. *In natura*, migration before and after asexual reproduction is determined by different phenomena: the mobilities of definitive hosts within transmission sites, of miracidia/cercariae, and of molluscs infected by miracidia/emitting cercariae. We therefore would need data from both adults and larvae to be able to infer  $m_1$  and  $m_2$  to test this hypothesis. The results presented here underline the importance of pursuing studies coupling theoretical approaches and real data analysis for improving the understanding of the population biology of complex life cycle organisms as schistosomes. In this respect, the model presented here, even if it certainly constitutes an overly simplified version of the real infection patterns in schistosomes, may constitute a basis for the interpretation of genetic data obtained from natural populations. This model could also constitute a framework for further studies dealing with other complex life cycle organisms of interest by incorporating different transmission, migration patterns or different reproductive systems.

#### ... and other organisms alternating sexual and asexual reproductions

Many organisms, other than schistosomes, reproduce sexually at some stages, but asexually at others (some example are given in Table 1). This is the case for parasites such as parasitic trematodes (8000 species) and some cestodes (e.g. *Echinococcus* sp.). For free-living species, the alternation between sexual and clonal reproductions is common in unicellular organisms, fungi, plants, rotifers, cladocerans, and insects (e.g. aphids, gall wasps and cecidomyid flies). While we acknowledge that our model was initially performed specifically on fit schistosome life cycle, we are confident that our results concerning the effect of the variance on the reproductive success of clones (or variance of clone size) apply to any organism displaying some form of asexual reproduction. In particular, we think that, even if their life cycle can slightly differ from that of schistosomes (hermaphroditism, several round of successive asexual reproductions), each time a population is composed

by clones of different size, the effects of clonality on  $F_{IS}$  and  $F_{ST}$  should qualitatively converge to that presented here (obviously if individuals are sampled before sexual reproduction): the higher is the variance of the clone size, the lower the  $F_{IS}$ , and the higher the  $F_{ST}$ . Therefore, our model and results should also constitute a sound basis for the interpretation of genetic data obtained from natural populations of other organisms displaying sexual reproduction at some stage of their life cycle and asexual reproduction at other stages.

#### Acknowledgements

We thank C. Chevillon and F. Balloux for very useful discussions and suggestions. A. Théron and T. de Meeûs are supported by the CNRS (PNDBE) and F. Prugnolle and D. Roze by MENRT. FP is also supported by a Lavoisier Fellowship from the Ministère Français des Affaires Etrangères. This work received financial support from the French Ministry of Ecology and Sustainable Development (contract no. CV 02000071, MEDD, programme 'Ecosystèmes Tropicaux').

#### References

- Balloux F, Lehmann L, De Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.
- Berg LM, Lascoux M, Pamilo P (1998) The infinite island model with sex-differentiated gene flow. *Heredity*, **81**, 63–68.
- Calow P (1989) *A Functional Biology of Clonal Animals*. Chapman & Hall, London.
- Chesser RK (1991a) Influence of gene flow and breeding tactics on gene diversity within populations. *Genetics*, **129**, 573–583.
- Chesser RK (1991b) Gene diversity and female philopatry. *Genetics*, **127**, 437–447.
- Chitsulo L, Engels D, Montresor A, Savioli L (2000) The global status of schistosomiasis and its control. *Acta Tropica*, **77**, 41–51.
- Cockerham CC (1969) Variance of gene frequencies. *Evolution*, **23**, 72–84.
- Cockerham CC (1973) Analyses of gene frequencies. *Genetics*, **74**, 697–700.
- Combes C (2001) *Parasitism: The Ecology and Evolution of Intimate Interactions*. University of Chicago Press, Illinois.
- Constantine CC (2003) Importance and pitfalls of molecular analysis to parasite epidemiology. *Trends in Parasitology*, **19**, 346–348.
- Curtis J, Sorensen RE, Page LK, Minchella DJ (2001) Microsatellite loci in the human blood fluke *Schistosoma mansoni* and their utility for other schistosome species. *Molecular Ecology Notes*, **1**, 143–145.
- De Meeûs T, Renaud F (2002) Parasites within the new phylogeny of Eucaryotes. *Trends in Parasitology*, **18**, 247–251.
- Durand P, Sire C, Theron A (2000) Isolation of microsatellite markers in the digenetic trematode *Schistosoma mansoni* from Guadeloupe island. *Molecular Ecology*, **9**, 997–998.
- Duvall RH, Dewitt WB (1967) An improved perfusion technique for recovering adult schistosomes from laboratory animals. *American Journal of Tropical Medicine and Hygiene*, **16**, 483–486.

- Gale JS (1990) *Theoretical Population Genetics*. Unwin-Hyman, London.
- Goudet J (1995) FSTAT version 1.2: a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.
- Hartl DL, Clark AG (1997) *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts.
- Laporte V, Charlesworth B (2002) Effective population size and population subdivision in demographically structured populations. *Genetics*, **162**, 501–519.
- Luikart G, Cornuet JM (1999) Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics*, **151**, 1211–1216.
- Prout T (1981) A note on the island model with sex dependent migration. *Theoretical and Applied Genetics*, **59**, 327–332.
- Provine WR (2001) *The Origins of Theoretical Population Genetics: with a New Afterword*. University of Chicago Press, Illinois.
- Prugnolle F, Choisy M, Theron A, Durand P, De Meeùs T (2004) Sex-specific correlation between heterozygosity and clone size in the trematode *Schistosoma mansoni*. *Molecular Ecology*, **13**, 2859–2864.
- Prugnolle F, De Meeus T, Durand P, Sire C, Theron A (2002) Sex-specific genetic structure in *Schistosoma mansoni*: evolutionary and epidemiological implications. *Molecular Ecology*, **11**, 1231–1238.
- Prugnolle F, Durand P, Theron A, Chevillon C, De Meeùs T (2003) Sex-specific genetic structure: new trends for dioecious parasites. *Trends in Parasitology*, **19**, 171–174.
- Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
- Schwartz OA, Armitage KB (1980) Genetic variation in social mammals: the marmot model. *Science*, **207**.
- Sire C, Durand P, Pointier JP, Theron A (2001) Genetic diversity of *Schistosoma mansoni* within and among individual hosts (*Rattus rattus*): infrapopulation differentiation at microspatial scale. *International Journal of Parasitology*, **31**, 1609–1616.
- Snyder SD, Loker ES, Johnston DA, Rollinson D (2001) The schistosomatidae: advances in phylogenetics and genomics. In: *Interrelationships of the Platyhelminthes* (eds Littlewood DTJ, Bray RA), pp. 356. Taylor & Francis, London and New York.
- Theron A, Sire C, Rognon A, Prugnolle F, Durand P (2004) Molecular ecology of *Schistosoma mansoni* transmission inferred from the genetic composition of larval and adult infrapopulations within intermediate and definitive hosts. *Parasitology*, **129**, 571–585.
- Vitalis R (2002) Sex-specific genetic differentiation and coalescence times: estimating sex-biased dispersal rates. *Molecular Ecology*, **11**, 125–138.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Whitlock MC (1992a) Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution*, **46**, 608–615.
- Whitlock MC (1992b) Nonequilibrium population structure in forked fungus beetles — extinction, colonization, and the genetic variance among populations. *American Naturalist*, **139**, 952–970.
- Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction — genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.
- Wright S (1921) Systems of mating. *Genetics*, **6**, 111–178.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.
- Wright S (1965) The interpretation of population structure by *F*-statistics with special regards to systems of mating. *Evolution*, **19**, 395–420.

---

Franck Prugnolle is a Post Doctoral fellow in François Balloux's laboratory at the University of Cambridge. He completed the present work during his PhD on the co-evolution between *Schistosoma mansoni* and its intermediate (*Biomphalaria glabrata*) and definitive host (*Rattus rattus*) in Guadeloupe. Denis Roze is a Post Doctoral fellow in Nick Barton's laboratory at the University of Edinburgh. He completed the present work during his PhD on the effect of selection in subdivided populations. André Théron is a senior research scientist working on the ecology and evolution of schistosomes and other parasites. Thierry De Meeùs is a researcher in evolutionary ecology and population genetics in host-parasite systems.

---