Available online at www.sciencedirect.com
SCIENCE DIRECT
International Journal for Parasitology 35 (2005) 255–263
www.parasitology-online.com

Population genetics of complex life-cycle parasites: an illustration with trematodes

Franck Prugnolle[a,*], Hua Liu[a], Thierry de Meeûs[b], François Balloux[a]

[a]*Theoretical and Molecular Population Genetics Group, Department of Genetics, Cambridge CB2 3EH, England*
[b]*GEMI, Equipe ESS, UMR CNRS-IRD 2724, IRD de Montpellier, 911 av. Agropolis, BP 64501, 34394 Montpellier cedex 5, France*

Received 28 July 2004; received in revised form 13 October 2004; accepted 18 October 2004

Abstract

Accurate inferences on population genetics data require a sound underlying theoretical null model. Organisms alternating sexual and asexual reproduction during their life-cycle have been largely neglected in theoretical population genetic models, thus limiting the biological interpretation of population genetics parameters measured in natural populations. In this article, we derive the expectations of those parameters for the life-cycle of monoecious trematodes, a group comprising several important human and livestock parasites that obligatorily alternate sexual and asexual reproduction during their life-cycle. We model how migration rates between hosts, sexual and asexual mutation rates, adult selfing rate and the variance in reproductive success of parasites during the clonal phase affect the amount of neutral genetic diversity of the parasite (effective population size) and its apportionment within and between definitive hosts (using *F*-statistics). We demonstrate, in particular, that variance in reproductive success of clones, a parameter that has been completely overlooked in previous population genetics models, is very important in shaping the distribution of the genetic variability both within and among definitive hosts. Within definitive hosts, the parameter F_{IS} (a measure of the deviation from random mating) is decreased by high variance in clonal reproductive success of larvae but increased by high adult self-fertilisation rates. Both clonal multiplication and selfing have similar effects on between-host genetic differentiation (F_{ST}). Migration occurring before and after asexual reproduction can have different effects on the patterns of F_{IS}, depending on values of the other parameters such as the mutation rate. While the model applies to any hermaphroditic organism alternating sexual and clonal reproduction (e.g. many plants), the results are specifically discussed in the light of the limited population genetic data on monoecious trematodes available to date and their previous interpretation. We hope that our model will encourage more empirical population genetics studies on monoecious trematodes and other organisms with similar life-cycles.

Keywords: *F*-statistics; Complex life-cycle; Asexual reproduction; Self-fertilisation; Effective population size; Trematodes

1. Introduction

The study of population genetic structure of living organisms is central to the understanding of micro-evolutionary processes (Nevo et al., 1984; Nadler, 1995). For small organisms and in particular parasites, the analysis of genetic variation at different hierarchical levels is often the only way to investigate natural population parameters such as gene flow, size of reproductive units and breeding strategies (Nadler, 1995). Population genetic structure can also constitute a powerful tool to investigate epidemiological patterns (Paterson and Viney, 2000).

However, interpreting measurements of genetic variation and its distribution in terms of biological parameters such as dispersal/transmission rates, breeding systems or effective population size is often very difficult. Two main reasons can be invoked to explain this difficulty: (i) a multiplicity of causes can explain specific patterns of genetic variation (e.g. heterozygote deficits can be explained by self-fertilisation, preferential mating between kin, Walhund effects, selection) (Hartl and Clark, 1997), (ii) the lack of clear expectations when the organisms under scrutiny display life-cycles that greatly depart from those used in theoretical population genetic models. Parasites with

* Corresponding author. Tel.: +44 1223 332584; fax: +44 1223 333992.
E-mail address: fp235@mole.bio.cam.ac.uk (F. Prugnolle).

0020-7519/$30.00
doi:10.1016/j.ijpara.2004.10.027

complex life-cycles such as digenean trematodes clearly belong to this category of organisms.

All trematodes (about 8000 species; e.g. De Meeûs and Renaud, 2002) are parasitic, many of them being medically and/or economically relevant (e.g. *Schistosoma* spp.; *Fasciola* spp.; *Paragonimus* spp. and *Clonorchis* spp.) (Combes, 2001). They are characterised by complex life-cycles with an obligate alternation of asexual and sexual reproduction during their lifespan. The asexual phase occurs within an invertebrate intermediate host (generally a mollusk) in which genetically identical larvae (clones) are produced whereas the sexual phase occurs in the vertebrate definitive host (aquatic or not). To date, no study has investigated the impact of this clonal phase on the genetic structure of parasites within definitive hosts or the importance of the obligate alternation of sexual and asexual reproduction. It has however been suggested that the severe deficits of heterozygotes observed in monoecious trematode populations might be produced by the aggregation of clones among hosts, the possibility of selfing as well as mating within clones (Lydeard et al., 1989; Mulvey et al., 1991; Vilas et al., 2003). Clonal reproduction could also contribute to the strong genetic differentiation observed between hosts and geographic locations (Mulvey et al., 1991; Theron et al., 2004). These arguments are strictly verbal and it remains largely unclear how and how much each of these factors shapes the distribution of genetic variation both within and between hosts thus precluding biological interpretation of population genetic parameters estimated from real data.

In this article, we use a finite island model to explore the effects of the alternation of sexual and asexual reproduction in monoecious trematodes on the partitioning of variance among and within definitive hosts. Apportionment of genetic variation at different hierarchical levels is described using the commonly used fixation indices (also called *F*-statistics: F_{IS} and F_{ST}; Wright, 1965). The model allows investigating how gene variation is affected by (i) the degree of mixing among parasites originating from different hosts at each new infection (represented in the model by the migration rates before and after clonal reproduction), (ii) sexual and asexual mutation rates, (iii) different self-fertilisation rates in adult parasites within definitive hosts and (iv) the variance in the reproductive success of parasites during the larval clonal phase (which may be seen as a measure of 'reproductive skew'). Finally, this model allows investigation of the effect of different population parameters on the effective population size of trematodes.

2. Materials and methods

2.1. Assumptions and parameter definitions

As trematodes exploit different hosts over their life-cycle, we have to define separately the genetic structure of parasite populations in intermediate and definitive hosts. An Adult Parasite Subpopulation (thereafter abbreviated APS) is composed of all adult parasites within the same definitive host individual (sometimes referred as 'infrapopulation'). A Larval Parasite Subpopulation (LPS thereafter) is composed of parasite larvae reproducing asexually. As there is no genetic recombination at this stage, LPSs can be defined more loosely than APSs. One could for example consider a LPS as the parasites within the same intermediate host only, or more generally as all larvae found in a subpopulation of intermediate hosts (e.g. within a puddle). Due to the absence of genetic recombination at the larval stage, the exact definition of LPS does not affect the outcome of the model. The term 'population' corresponds to the hierarchical level encompassing all APSs or LPSs. To avoid any confusion, we will specify throughout the text whether we are referring to larval parasites or adult parasites.

We consider a monoecious population of trematodes subdivided into n APSs (finite island model), each consisting of N parasite individuals. The detailed life-cycle is represented in Fig. 1. As trematodes may reproduce by self-fertilisation (see e.g. Trouvé et al., 1996), we incorporate partial selfing in our model. Thus, for adults, reproduction within definitive host comprises selfing with probability s and random outbreeding with the corresponding probability $1-s$. For simplicity, we assume non-overlapping generations (the individuals reproduce and die). According to the island model, we assume that a proportion $1-m_1$ of the offspring produced in the same definitive host remains clustered and subsequently infects the same intermediate host (or subpopulation of intermediate hosts), while the remaining offspring m_1 infect other intermediate host (or subpopulation of intermediate hosts) at random. We further assume that, just after infection, each LPS contains exactly N parasite individuals. These N individuals produce a large number of asexual offspring within the intermediate host and die. The parameter c represents the amount of

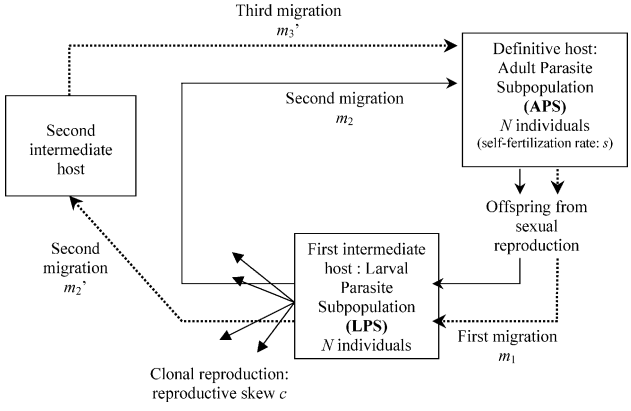

Fig. 1. Schematic representation of the life-cycle of monoecious trematodes with one (full line) or two intermediate hosts (dashed line). In our model, we followed the life-cycle with only one intermediate host. However, the second life-cycle can reduce to the first one by considering that m_2 is just $m_2' + m_3'$.

reproductive skew among larval parasites within a LPS, and thus measures the variance of the reproductive success during the clonal phase. Such variance is commonly observed in nature and results in the aggregation of clones among definitive hosts (e.g. Mulvey et al., 1991). Then, the asexual offspring are liberated and infect the definitive host (while it would be straightforward to introduce a second intermediate host in the model, as is generally the case in trematodes, this should not qualitatively affect our results). Again, we assume that a proportion $1-m_2$ of the offspring produced within the same LPS remains clustered and infect the same definitive host, while the others infect the remaining definitive host individuals at random.

Because of the symmetry of the island model, only the following probabilities of identity by descent are necessary to derive *F*-statistics.

Q_0 the inbreeding coefficient defined as the probability that two alleles at a given locus within an individual are identical by descent.

Q_1 coancestry of individuals within the same LPS or APS. It is defined as the probability that two alleles drawn at random at the same locus from two individuals from the same LPS or APS are identical by descent.

Q_2 coancestry of individuals drawn at random from different LPS or APS.

Wright's (1965) *F*-statistics can be derived from probabilities of identity by descent following (Cockerham, 1969, 1973):

$$F_{IS} = \frac{Q_0 - Q_1}{1 - Q_1}, \quad F_{ST} = \frac{Q_1 - Q_2}{1 - Q_2}, \quad F_{IT} = \frac{Q_0 - Q_2}{1 - Q_2} \quad (1)$$

where F_{IS} describes the departure from random mating within LPSs or APSs, F_{ST} the genetic differentiation between LPSs or APSs and F_{IT} the departure from random mating in the entire population. In what follows, we will only consider results about the parameters F_{IS} and F_{ST} for adult parasites (within the definitive host).

2.2. Recurrence equations for probabilities of identity by descent

2.2.1. After sexual reproduction before the first larval migration

We first consider the recurrence equation for the inbreeding coefficient $Q_0^*(t+1)$ in larvae just after adult sexual reproduction. Each zygote is formed by sampling one male gamete and one female gamete from parents within the same definitive host. A proportion, *s*, of the zygotes results from self-fertilisation (gametes come from the same parent) and a proportion, $1-s$, from the union of gametes from two different individuals. For selfing-derived zygotes, there is a probability (1/2) that the two alleles come from the same allele in the parent and (1/2) that they come from two different alleles at the same locus. In this latter case, there is a probability $Q_0(t)$ that these alleles are identical by descent. For non-selfing derived zygotes, the two alleles necessarily come from different parents and hence $Q_1(t)$ is the probability that they are identical by descent. Considering that the two alleles drawn have not mutated during sexual reproduction (with probability $\gamma_1=(1-u_1)^2$, where u_1 represents the mutation rate during sexual reproduction), after collecting terms, we obtain:

$$Q_0^*(t+1) = \gamma_1\left(s\left(\frac{1}{2}+\frac{1}{2}Q_0(t)\right)+(1-s)Q_1(t)\right) \quad (2)$$

Now we consider the recurrence equation for the coancestry coefficient $Q_1^*(t+1)$ for larvae before first migration. Two cases need to be considered to derive this equation. First, the two alleles come from the same parent with probability (1/N). In this case, there is a probability (1/2) that they come from the same allele and $(1/2)Q_0(t)$ from different alleles identical by descent. Second, the two alleles come from different parents from the same APS (probability $1-(1/N)$). Then, the two alleles are identical by descent with a probability $Q_1(t)$. Therefore, the recurrence equation for the coancestry coefficient for the newly produced larvae reads:

$$Q_1^*(t+1) = \gamma_1\left(\frac{1}{N}\left(\frac{1}{2}+\frac{Q_0(t)}{2}\right)+\left(1-\frac{1}{N}\right)Q_1(t)\right) \quad (3)$$

The coancestry of larvae within the total population is unaffected by reproduction but is dependent on mutation:

$$Q_2^*(t+1) = \gamma_1 Q_2(t) \quad (4)$$

2.2.2. After the first larval migration

Now we consider the first larval migration (before asexual reproduction), which influences the coancestry of alleles between individuals, Q_1 and Q_2, but not the within individual allele coancestry (inbreeding coefficient Q_0). To derive these equations, we need first to derive the probability that two individuals taken at random without replacement from the same LPS originate from the same APS before larval migration (denoted here as *a*) and the probability that two individuals sampled from different LPSs were born in the same APS (denoted as *b*).

The exact expression for *a* is relatively cumbersome (see e.g. Wang, 1997). However, for relatively large values of *N*, *a* reduces to the much more compact form that we use throughout the article:

$$a = (1-m_1)^2 + \frac{m_1^2}{n-1} \quad (5)$$

Now, *b*, the probability that two individuals sampled from different LPSs originate from the same APS can be written as:

$$b = \frac{1-a}{n-1} \quad (6)$$

Thus, using (5) and (6), we obtain:

$$Q_1^{m_1}(t+1) = aQ_1^*(t+1) + (1-a)Q_2^*(t+1) \quad (7)$$

$$Q_2^{m_1}(t+1) = bQ_1^*(t+1) + (1-b)Q_2^*(t+1) \quad (8)$$

2.2.3. After asexual reproduction before the second larval migration

After asexual reproduction, only the coancestry between larval individuals within LPSs is changed. Let P represent the probability that two asexually produced individuals, within the same LPS, have the same parent. P can be expressed as function of the mean and variance of the reproductive success of individuals during the asexual phase. Variance and mean can be condensed into a single parameter which may be considered as a measure of reproductive skew. Before asexual reproduction, N individuals are present within each LPS. We call μ and ν the mean and variance of the individual larval reproductive success. The probability that two individuals produced asexually within the same LPS have the same progenitor (e.g. Gale, 1990) is therefore:

$$P = \frac{\nu + \mu(\mu-1)}{\mu(\mu N - 1)} \quad (9)$$

Under the Wright-Fisher model assuming a Poisson distribution, the variance is:

$$\nu_{\mathrm{WF}} = \mu\left(1 - \frac{1}{N}\right) \quad (10)$$

while in the case where only one individual produces all offspring within the same LPS, the variance is:

$$\nu_{\mathrm{MAX}} = \mu^2(N-1) \quad (11)$$

We define a reproductive skew parameter c such that the variance of the individual reproductive success equals:

$$\nu = (1-c)\nu_{\mathrm{WF}} + c\,\nu_{\mathrm{MAX}} \quad (12)$$

Therefore ν increases linearly with c, and takes as extreme values ν_{WF} (when $c=0$) and ν_{MAX} (when $c=1$). Substituting Eq. (12) into Eq. (9) yields:

$$P = \frac{1 + c(N-1)}{N} \quad (13)$$

From Eq. (13) and assuming that the two randomly drawn alleles have not mutated during asexual reproduction (probability $\gamma_2=(1-u_2)^2$, where u_2 represents the mutation rate during asexual reproduction), we can now derive $Q_i^c \quad i \in \{0,1,2\}$:

$$Q_0^c(t+1) = \gamma_2 Q_0^*(t+1) \quad (14)$$

$$Q_1^c(t+1) = \gamma_2\left(\frac{P}{2}(1 + Q_0^*(t+1)) + (1-P)Q_1^{m_1}(t+1)\right) \quad (15)$$

$$Q_2^c(t+1) = \gamma_2 Q_2^{m_1}(t+1) \quad (16)$$

2.2.4. After the second larval migration

Equations for $Q_0^{m_2}, Q_1^{m_2}, Q_2^{m_2}$ are essentially identical to the first larval migration but the following terms Q_0^*, Q_1^*, Q_2^* have to be replaced by their respective counterparts Q_0^c, Q_1^c, Q_2^c.

2.3. Transition matrices and equilibrium solutions

The different expressions given above represent all the necessary tools to derive the complete recurrence equations for Q_i $i \in \{0,1,2\}$ and their equilibrium solutions for adult parasites just before sexual reproduction. These can be presented in a matrix form,

$$\mathbf{Q}_{t+1} = \gamma_1\gamma_2\mathbf{G}\mathbf{Q}_t + [\gamma_1\gamma_2(\mathbf{I}-\mathbf{G}) + (\gamma_2 - \gamma_1\gamma_2)(\mathbf{I}-\mathbf{H})]\vec{1} \quad (17)$$

where $\mathbf{Q}_{t+1}$ is a column vector of the three probabilities of identity by descent at generation $t+1$ (in adults before sexual reproduction) and $\mathbf{Q}_t$ the column vector of the three probabilities of identity by descent but at generation t.

$$\mathbf{Q}_{t+1} = \begin{pmatrix} Q_0(t+1) \\ Q_1(t+1) \\ Q_2(t+1) \end{pmatrix} \quad (18)$$

$\mathbf{I}$ is a identity matrix. $\mathbf{G}$ is the total transition matrix defining the probabilistic changes of the vector of variables (not shown here) and equals:

$$\mathbf{G} = \mathbf{M}_2\mathbf{G}_2\mathbf{M}_1\mathbf{G}_1 \quad (19)$$

where $\mathbf{M}_2, \mathbf{G}_2, \mathbf{M}_1, \mathbf{G}_1$ are the matrices making the transition for $Q_i \in \{0,1,2\}$ at each step. $\mathbf{G}_1$ represents the transition matrix for sexual reproduction, $\mathbf{M}_1$ for the first migration event, $\mathbf{G}_2$ for the clonal reproduction and $\mathbf{M}_2$ for the second migration event. $\mathbf{H}$ is equal to $\mathbf{M}_2\mathbf{G}_2$. $\vec{1}$ is a vector with elements 1. Note that it would be easy to derive the solutions for Q_i $i \in \{0,1,2\}$ when sampling parasites from intermediate hosts (before the parasite's asexual reproduction) by permuting the transition matrices in Eq. (17).

2.4. Effective population size

The effective population size (Wright, 1931) is the parameter summarising the amount of genetic drift to which a population is subjected. It is quantified as the number of idealised randomly mating individuals, which experience the same amount of random fluctuations at neutral locus as the population under scrutiny. The dynamics of idealised randomly mating individuals is described by the Wright-Fisher model, whose well known properties lead to different

definitions of the effective population size depending on whether the quantities of interest are the variance of change in allelic frequencies, inbreeding coefficients, or the rate of decline in heterozygozity (Whitlock and Barton, 1997). Here we use the coalescence effective size introduced by Balloux et al. (2003) which is defined as:

$$N_e = \bar{t} \tag{20}$$

where $\bar{t}$ is the expected time (number of generations back in time) it takes for two randomly sampled alleles in a population to coalesce to a common ancestor. For the Wright-Fisher model (and $m_1 = m_2 = 1$) $\bar{t} = n \times N$, so that the effective size reduces to the actual number of diploid individuals within the entire population.

There is a strict relationship between identities by descent probabilities and coalescence times (Rousset, 1996). The probability of identity of any pair of alleles is the probability that neither allele has undergone mutation since their most recent common ancestor (Hudson, 1990). Assuming the mutation rate to be identical during sexual and asexual reproduction, Eq. (17) reduces to:

$$\mathbf{Q} = (\mathbf{I} - \gamma^2\mathbf{G})^{-1}[(\mathbf{I} - \gamma^2\mathbf{G}) + (\gamma - \gamma^2)(\mathbf{I} - \mathbf{H})]\vec{\mathbf{I}} \tag{21}$$

where $\gamma = (1-u)^2$ and u is the mutation rate. A closer look reveals that the vector $\mathbf{Q}$ defines the probability generating functions of coalescence time at each level $i \in \{0,1,2\}$. These functions reduce to the calculations of expected coalescence times as in two-stage reproduction system:

$$\bar{t}_i = \frac{1}{2}\frac{\partial Q_i}{\partial \gamma}\Big|_{\gamma=1} \tag{22}$$

where $\bar{t}_i$ is the expected coalescence time (number of generations back in time) at level i and Q_i is the ith row of the equilibrium vector given by Eq. (21)

$$\begin{pmatrix} t_0 \\ t_1 \\ t_2 \end{pmatrix} = \frac{1}{2}(\mathbf{I} - \mathbf{G})^{-1}(\mathbf{I} + \mathbf{H})\vec{\mathbf{I}} \tag{23}$$

The mean coalescence time of two randomly sampled alleles is the expectation of the $\bar{t}_i$. In the finite island model this yields:

$$\bar{t} = \frac{1}{s}\left(\frac{1}{N}\bar{t}_0 + \left(1 - \frac{1}{N}\right)\bar{t}_1\right) + \left(1 - \frac{1}{s}\right)\bar{t}_2 \tag{24}$$

Substituting Eq. (24) into Eq. (20), we obtain the coalescence effective population size.

3. Results and discussion

3.1. Deviation from random mating within definitive hosts (F_{IS})

3.1.1. Effect of the reproductive system

Fig. 2A displays F_{IS} as a function of s (the adult self-fertilisation rate) and c (the reproductive skew parameter during larval clonal reproduction). Self-fertilisation and clonal reproductive skew have opposite effects. As expected F_{IS} increases under more frequent self-fertilisation (e.g. Wang, 1997), whereas it decreases under high reproductive skew during clonal reproduction. Opposite effects of

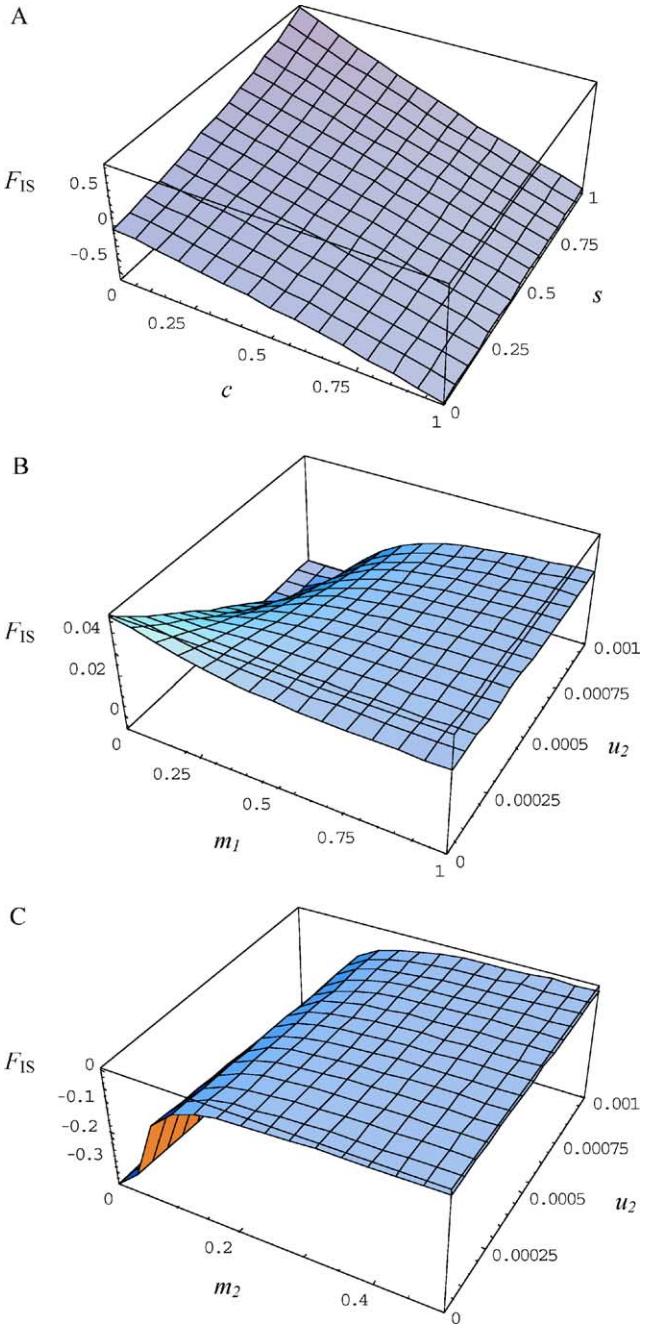

Fig. 2. Relationship between F_{IS} in definitive hosts and: (A) the adult self-fertilisation rate (s) and c, the parameter that makes the variance of the reproductive success of clones to vary between the variance of the Wright-Fisher model ($c = 0$) and the maximal variance ($c = 1$). Here, $N = 20$, $n = 100$, $m_1 = 10^{-3}$, $m_2 = 10^{-3}$, $u_1 = 10^{-3}$, $u_2 = 10^{-3}$; (B) the first migration rate (m_1) and the second mutation rate (u_2). Here, $N = 20$, $n = 100$, $c = 0.7$, $s = 0.1$, $m_2 = 0.1$, $u_1 = 10^{-5}$; (C) the second migration rate (m_2) and the second mutation rate (u_2). Here, $N = 20$, $n = 100$, $c = 0.7$, $s = 0.1$, $m_1 = 10^{-3}$, $u_1 = 10^{-3}$.

self-fertilisation and clonal variance of the reproductive success on F_{IS} can be explained through their respective effects on the inbreeding and coancestry coefficients. Just before sexual reproduction, the two homologous alleles sampled from two individuals of the same APS can coalesce (i.e. be descended from the same ancestral allele) during the previous event of clonal reproduction with a probability that increases with increasing variance in reproductive success; namely when the variance in reproductive success increases, so does Q_1 thus reducing F_{IS}. On the other hand, the two homologous alleles of an individual cannot descend from the same ancestral allele taken before clonal reproduction but can descend from the same ancestral allele taken (back in time) before sexual reproduction whenever a proportion of offspring resulted from self-fertilisation. Therefore, when the self-fertilisation rate increases, so does Q_0 and therefore F_{IS}.

One important result of this model is therefore the effect of the variance of the reproductive success between clones on F_{IS}. Contrarily to what has been previously proposed (e.g. Mulvey et al., 1991; Vilas et al., 2003), the aggregation of clones within definitive hosts cannot account in itself for positive F_{IS} (deficit in heterozygotes). Thus, even if identical multilocus genotypes (clones) are present within the same definitive host, we do not expect heterozygote deficits in offspring (parasite larvae) as long as adult parasites reproduce at random. Therefore, heterozygote deficits (positive F_{IS}) are only generated by self-fertilisation. Clonal multiplication only affects F_{IS} within definitive hosts relative to F_{IS} observed in intermediate hosts by modifying the probability of identity by descent between individuals (Q_1) in adult parasites. The variance in reproductive success between clones (translating in the variance of the number of copies per multilocus genotype within definitive hosts) affects F_{IS} nearly linearly. Thus, even if all offspring were produced through self-fertilisation ($s=1$), strong heterozygote excesses could still be observed in adult parasites if, during larval clonal reproduction, the variance is high (see Fig. 2A). In order to clarify this counterintuitive result, it might be useful to consider an extreme example consisting of adult parasites with high self-fertilisation rates. Under this situation, we expect strong heterozygote deficits in larvae after sexual reproduction. Let us now consider that, during clonal reproduction, the variance of the reproductive success between clones is maximal ($c=1$) so that only one clone infects each single definitive host. Then, within each definitive host, a fraction of the loci will be homozygous for the same allele in all parasites and hence F_{IS} will be undetermined; whereas the remaining loci will be heterozygous in all parasites with an associated F_{IS} of -1. As a consequence, the expected F_{IS} overall loci will only be affected by the heterozygote loci and will be equal to -1 despite the high self-fertilisation rate.

This leads to the consequence that whenever there is variance in the reproductive success during clonal multiplication, F_{IS} estimated within APSs will not be indicative of the rate of self-fertilisation. Selfing rates can however still be roughly estimated from adult F_{IS} if this later is computed on samples where only one copy of each different multilocus genotype within each definitive host is kept since deleting repeated multilocus genotypes will remove the effect of the variance in the reproductive success between clones. Obviously, the self-fertilisation rate can also be estimated from larval F_{IS} estimates, when sampling is made before larval asexual reproduction.

3.1.2. Effects of migration

In nature, the first migration rate (m_1) is determined by multiple parameters such as the behavior of the definitive host, the active ability of miracidia (first free-living larval stage) to swim and disperse, the presence of passive dispersal of the miracidia before entering the first intermediate host as well as the behavior of the intermediate host after infection. The second migration rate m_2 (after clonal reproduction) is determined by the abilities of cercariae and metacercariae to disperse (actively or passively), the dispersal behavior of the first and the second intermediate host (when it exists), and is also dependent on the behavior of the definitive host (as for the first migration rate). What are the effects of these two dispersal rates on the distribution of genetic variability within adult parasites?

The first migration rate (m_1) can have opposite effects on adult F_{IS}. This depends on values of the other parameters such as the mutation rate during clonal reproduction, the migration rate after clonal reproduction, the clonal reproductive skew as well as the self-fertilisation rate, the number and the size of APSs/LPSs. Therefore, no simple analytical solution for the effect of m_1 on F_{IS} can be obtained and hence no clear general prediction can be made. In order to illustrate this phenomenon, we present below the effect of the first migration rate in relation to (i) the mutation rate during larval clonal reproduction and (ii) the migration rate after asexual reproduction.

Fig. 2B shows the relation between F_{IS} in adults and the migration rate occurring before asexual reproduction (m_1) plus the second mutation rate. We see that for fixed values of other parameters, the effect of m_1 on F_{IS} changes direction when the value of the mutation rate occurring during asexual reproduction increases. More specifically, when m_1 increases and u_2 (the mutation rate) is very low, F_{IS} decreases. Conversely, when m_1 increases and u_2 is high, F_{IS} increases as well.

Fig. 3 illustrates that this effect is also dependent on the value of m_2. Indeed, for a fixed value of u_2, one can see in Fig. 3 that the response of F_{IS} to an increase of m_1 changes when m_2 increases. In this figure, u_2 is fixed to 10^{-7}. When the second migration rate is low, F_{IS} increases when m_1 increases. On the other hand, when m_2 is high, F_{IS} decreases when m_1 increases. Such phenomena can be observed in the absence of any clonal reproductive skew (Wright-Fisher model).

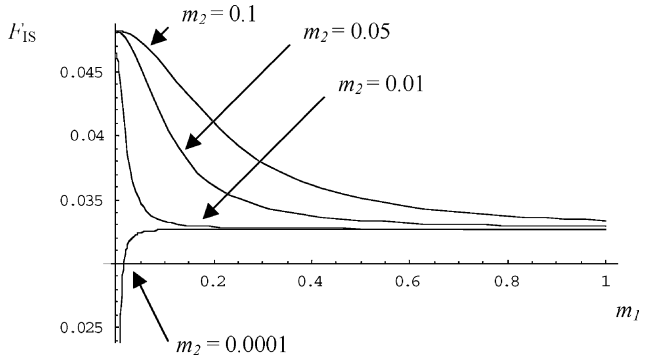

Fig. 3. Relationship between F_{IS} in definitive hosts and the first migration rate (m_1) for fixed values of mutation rates but different values of second migration rates (m_2). Here, $N=20$, $n=100$, $c=0.7$, $s=0.1$, $u_1=10^{-7}$, $u_2=10^{-7}$.

Such complicated effects on F_{IS} are not observed for m_2: when m_2 increases, so does F_{IS} independently of the values given to the other parameters (Fig. 2C). This is due to the fact that as m_2 increases, Q_1 decreases and converge to 0 whereas the value of Q_0 remains greatly determined by the self-fertilisation rate.

3.2. Differentiation of parasites between definitive hosts (F_{ST})

The effect of the aggregation of clones within definitive hosts on the values of F_{ST} was empirically demonstrated in two studies using different approaches. Mulvey et al. (1991) demonstrated, through simulations designed to compare the observed differentiation between *Fascioloides magna* APSs with that expected under a null model of no aggregation, that aggregation of clones accounted for around 50% of the observed differentiation (F_{ST}) between fluke APSs whereas Theron et al. (2004) showed that F_{ST} computed between adult APSs of *Schistosoma mansoni* (dioecious trematode) was significantly lower when estimated on samples where only one single copy of each different multilocus genotype was retained within each APS. However, although both approaches demonstrated an effect of clonal reproduction on F_{ST}, they did not explicitly identify which clonality-associated effects were important in determining F_{ST}. In other words, what exactly do these authors mean by 'aggregation of clones'? Our model provides an answer: F_{ST} of parasites between definitive hosts increases with the variance of clonal reproductive success. Indeed, this variance generates a higher frequency of genetically identical parasites within definitive hosts, thus leading to a higher probability that two alleles sampled from different individuals within the same host are identical by descent (that is, Q_1 increases) (Fig. 4).

It is noteworthy that, whenever the variance in reproductive success between clones is different from that of the Wright-Fisher model (and $m_2 \neq 1$), genetic differentiation between adult parasites can be observed even if the migration rate before asexual reproduction is maximal

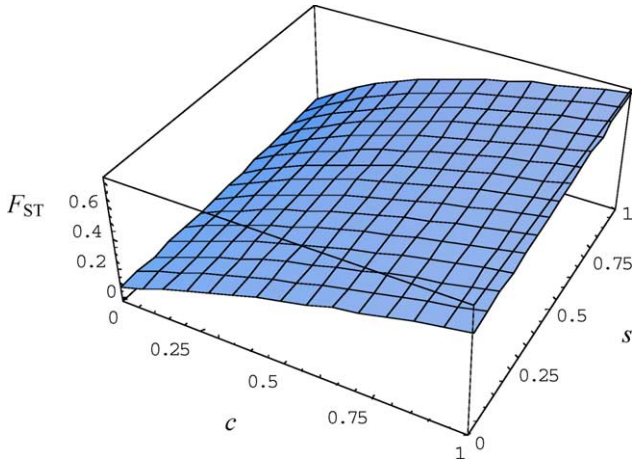

Fig. 4. Relationship between F_{ST}, the self-fertilisation rate (s) and c, the parameter that makes the variance of the reproductive success to vary between the variance of the Wright-Fisher model ($c=0$) and the maximal variance ($c=1$). Here, $N=200$, $n=100$, $m_1=0.5$, $m_2=0.5$, $u_1=10^{-5}$, $u_2=10^{-5}$.

($m_1=1$; miracidial subpopulations not being genetically differentiated). This means therefore that observing significant F_{ST} between APSs will not be indicative in any case of the existence of genetic differentiation between miracidial subpopulations. Only the analysis of genetic differentiation between definitive hosts after deleting multiple copies of clones from each APS can bring an answer of whether there is genetic differentiation between miracidial subpopulations. Indeed, if miracidial subpopulations are not genetically differentiated, no significant genetic differentiation will be found between definitive hosts when analysing the dataset once multiple copies of clones have been deleted, as each APS will then only represent a binomial sub-sample of the total larval undifferentiated pool. On the other hand, a positive F_{ST} will be found if miracidial subpopulations are genetically differentiated.

Finally, additional factors can affect F_{ST}. For instance, when self-fertilisation increases, so does F_{ST} as, at equilibrium, Q_1 increases within adult subpopulations (Fig. 4). Obviously, F_{ST} is also affected by migration before and after clonal reproduction: when migration increases, F_{ST} decreases.

3.3. Effective population size (N_e)

The effective population size (N_e) is an important quantity in evolutionary biology, as it describes the amount of genetic drift to which a population is subjected. As a consequence, it conditions both the probability and the time needed for a new advantageous allele to reach fixation (Whitlock and Barton, 1997). This is of obvious practical interest for pathogens, since beneficial alleles may include drug and vaccine resistance genes. Below we explore how effective populati\on size is affected by the life-cycle of trematodes or other organisms sharing similar life histories.

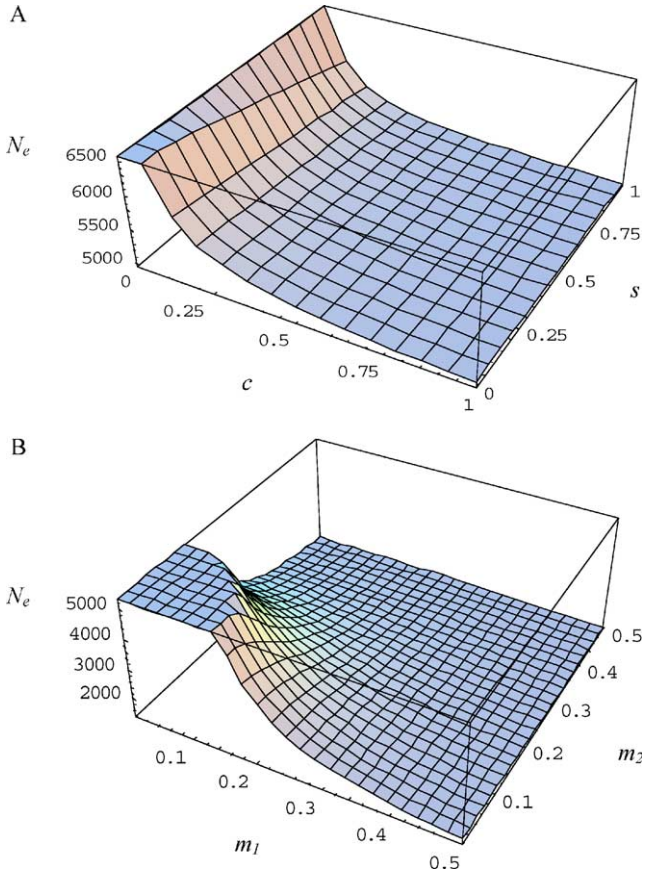

Fig. 5. Relationship between the effective population size (N_e) and: (A) the self-fertilisation rate (s) and c, the parameter that makes the variance of the reproductive success of clones to vary between the variance of the Wright-Fisher model ($c=0$) and the maximal variance ($c=1$). Here, $N=50$, $n=100$, $m_1=0.1$, $m_2=0.1$; (B) both migration rates (before and after asexual reproduction). Here, $N=50$, $n=100$, $c=0.1$, $s=0.1$.

3.3.1. Effect of the reproductive system

The effective population size is dependent on the rate of reproductive skew c (variance of the reproductive success between clones in intermediate hosts) as well as the adult self-fertilisation rate. Obviously, as the reproductive skew and the self-fertilisation rate increase the effective size decreases (Fig. 5A). The degree of reproductive skew has a few noticeable effects on most of the parameter space but when it tends to that of the Wright-Fisher model (c tends to zero) the effective population size suddenly surges to very large values and can even be higher than the real census size ($N\times n$), similar to what is expected when subpopulations are strongly genetically differentiated (Whitlock and Barton, 1997). The effective size linearly decreases with increasing self-fertilisation, for any value of c.

3.3.2. Effects of migration

Increasing the migration rate decreases the effective population size because it reduces the coalescent time for two alleles drawn at random from two different parasites coming from different definitive hosts. From Fig. 5B, it can be seen that the two migration rates approximately have the same effect on the effective population size.

3.4. Conclusions for population genetic studies of monoecious trematodes and other parasites with similar life-cycle

The analysis of parasite genetic structure is important for the understanding of basic population parameters such as reproductive strategies of parasites, transmission rates between hosts as well as effective population size. Such information is particularly relevant when attempting to understand how traits such as genetic resistance to anti-parasitic agents can spread through the parasite population (Anderson and May, 1992). In monoecious trematodes but also for other parasites displaying similar life-cycle as *Echinococcus* spp. (cestodes), the apportionment of genetic variation within and between APSs is the result of multiple factors, some of which have opposite effects. Our model demonstrates, in particular, that larval clonal reproduction and variance in the reproductive success between clones, a parameter that has never been considered in previous population genetics models, can strongly affect the distribution of genetic variation in adult parasites both within and between definitive hosts.

In comparison to the complexity of trematode life-cycles and life history traits, the current model is obviously a crude simplification of reality. We made in particular two strong assumptions by considering non-overlapping generations and keeping the number and size of parasite subpopulations constant both in the larval and adult stages. Many trematodes are long-lived so that generations are certainly overlapping. Including overlapping generations into the present model would however translate into a massive increase in mathematical complexity and would require the construction of a completely new model with many more parameters, while it is unlikely this would significantly affect our main conclusions, at least qualitatively. The assumption of invariant number and size of parasite throughout the life-cycle could be relaxed but again should not qualitatively affect our results.

While we acknowledge that our model constitutes an overly simplified version of real infection patterns and

Table 1
Summary of the effects of an increase of the self-fertilisation rate (s), the clonal reproductive skew (Clonal Variance), the migration before (m_1) and after (m_2) clonal reproduction on the apportionment of parasite genetic variation both within (F_{IS}) and between definitive hosts (F_{ST}) and on the effective population size (N_e)

	F_{IS}	F_{ST}	N_e
s	+	+	−
m_1	+ or −	−	−
Clonal Variance	−	+	−
m_2	+	−	−

+, stands for an increase and − for a decrease.

transmission in trematodes, we are confident it nevertheless constitutes a sound basis for the interpretation of genetic data obtained from natural populations (see Table 1). Obviously, biologically insightful inferences can be made only if clear hierarchical sampling is undertaken, that is if parasites from each definitive host are clearly identified. In other words, as far as possible, it is important to avoid pooling parasites from different definitive hosts to be able to retrieve maximal biological information from genetic data. Moreover, as noted before, genetic analyses on samples with and without copies of clones can provide different biological information.

In this model, we deliberately did not consider dioecious trematodes such as schistosomatidae because parameterisation of their life-cycle must consider the dynamics of alleles in males and females separately. Nevertheless, the effects of model parameters such as migration and clonality on F_{IS} and F_{ST} should qualitatively converge to that presented here. However, such a model is still needed as it would additionally allow investigating the effect of differential life-history traits between males and females such as sex-biased dispersal rates (Prugnolle et al., 2002, 2003) or sex-specific variance in the clonal reproductive success (Prugnolle et al., 2004).

Acknowledgements

We thank Denis Roze for his help in modelling clonal reproduction and two anonymous Referees for very useful comments. F.P is supported by a Lavoisier Fellowship from the Ministère Français des Affaires Etrangères. T. de M. is supported by the CNRS. L.H is supported by a grant from the Cambridge Overseas Trust.

References

Anderson, R.M., May, R.M., 1992. Infectious Diseases of Humans. Oxford University Press, Oxford.

Balloux, F., Lehmann, L., De Meeus, T., 2003. The population genetics of clonal and partially clonal diploids. Genetics 164, 1635–1644.

Cockerham, C.C., 1969. Variance of gene frequencies. Evolution 23, 72–84.

Cockerham, C.C., 1973. Analyses of gene frequencies. Genetics 74, 697–700.

Combes, C., 2001. Parasitism. The Ecology and Evolution of Intimate Interactions. Chicago Press.

De Meeûs, T., Renaud, F., 2002. Parasites within the new phylogeny of eucaryotes. Trends Parasitol. 18, 247–251.

Gale, J.S., 1990. Theoretical Population Genetics. Unwin Hyman, London.

Hartl, D.L., Clark, A.G., 1997. Principles of Population Genetics, Third ed. Sinauer Associates p. 542.

Hudson, R.R., 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 7, 1–44.

Lydeard, C., Mulvey, M., Aho, J.M., Kennedy, P.K., 1989. Genetic-variability among natural populations of the Liver Fluke *Fascioloides magna* in White-Tailed Deer, *Odocoileus virginianus*. Can. J. Zool. 67, 2021–2025.

Mulvey, M., Aho, J.M., Lydeard, C., Leberg, P.L., Smith, M.H., 1991. Comparative population structure of a parasite (*Fascioloides magna*) and its definitive host. Evolution 45, 1628–1640.

Nadler, S.A., 1995. Microevolution and the genetic structure of parasite populations. J. Parasitol. 81, 395–403.

Nevo, E., Bieles, A., Schlomo, B., 1984. The evolutionary significance of genetic diversity: ecological, demographic and life-history consequences, in: Mani, G.S. (Ed.), The Evolutionary Significance of Genetic Diversity: Ecological, Demographic and Life-History Consequences. Springer, Berlin.

Paterson, S., Viney, M.E., 2000. The interface between epidemiology and population genetics. Parasitol. Today 16, 528–532.

Prugnolle, F., De Meeus, T., Durand, P., Sire, C., Theron, A., 2002. Sex-specific genetic structure in *Schistosoma mansoni*: evolutionary and epidemiological implications. Mol. Ecol. 11, 1231–1238.

Prugnolle, F., Durand, P., Theron, A., Chevillon, C., De Meeus, T., 2003. Sex-specific genetic structure: new trends for dioecious parasites. Trends Parasitol. 19, 171–174.

Prugnolle, F., Choisy, M., Theron, A., Durand, P., De Meeus, T., 2004. Sex-specific correlation between heterozygosity and clone size in the trematode *Schistosoma mansoni*. Mol. Ecol. 13, 2859–2864.

Rousset, F., 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142, 1357–1362.

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.

Trouvé, S., Renaud, F., Durand, P., Jourdane, J., 1996. Selfing and outcrossing in a parasitic hermaphrodite helminth (Trematoda Echinostomatidae). Heredity 77, 1–8.

Vilas, R., Paniagua, E., Sanmartin, M.L., 2003. Genetic variation within and among infrapopulations of the marine digenetic trematode *Lecithochirium fusiforme*. Parasitology 126, 465–472.

Wang, J., 1997. Effective size and *F*-statistics of subdivided populations. I. Monoecious species with partial selfing. Genetics 146, 1453–1463.

Whitlock, M.C., Barton, N.H., 1997. The effective size of a subdivided population. Genetics 146, 427–441.

Wright, S., 1931. Evolution in mendelian populations. Genetics 16, 97–159.

Wright, S., 1965. The interpretation of population structure by *F*-statistics with special regards to systems of mating. Evolution 19, 395–420.