Molecular Ecology (2004) **13**, 2859–2864 doi: 10.1111/j.1365-294X.2004.02273.x

Sex-specific correlation between heterozygosity and clone size in the trematode *Schistosma mansoni*

FRANCK PRUGNOLLE,*‡ MARC CHOISY,* ANDRÉ THÉRON,† PATRICK DURAND* and THIERRY DE MEEÛS*

*GEMI, Equipe ESS, UMR-2724 CNRS-IRD, centre IRD, 911 avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France, †CBETM, UMR 5555 CNRS-UP, Université de Perpignan, 52 Avenue. de Villeneuve, 66860 Perpignan, France

Abstract

The mode of reproduction (sexual and/or asexual) and the mating system determine the patterns of gene transmission and genotype formation across generations. *Schistosoma mansoni* is a dioecious trematode that necessarily alternates sexual and asexual reproduction during its life cycle. In a previous study of the distribution of *S. mansoni* genetic variability within and between definitive host individuals, we noticed that deleting multilocus genotypes from each infrapopulation so as to keep only one copy of each multilocus genotype, seemed to have a substantial effect on F_{IS} values. More precisely, female F_{IS} increased when repeated genotypes were removed whereas no effect was observed on male F_{IS}. This suggested that multilocus genotypes at high frequency tended to be more heterozygous. The aim of the present study is specifically to test and analyse this phenomenon. We demonstrate that the number of repetitions per clone correlates with individual heterozygosity. This effect is however, sex-specific: only female clone size correlates with heterozygosity. We discuss this phenomenon in relation to the heterozygosity–fitness relationship and the sex-specific response to inbreeding depression.

Keywords: clonality, clone size, genetic variability, heterozygosity, randomization tests

Received 8 March 2004; revision received 7 May 2004; accepted 28 May 2004

Introduction

Clonal reproduction describes the production of new individuals with the original parental genotype. In partially clonal species, the constitution of the population and the contribution of each clone to the next generation have significant importance in the maintenance and the distribution of genetic variation (Balloux *et al*. 2003).

Many parasite species exhibit complex life cycles with obligatory alternation of asexual and sexual reproduction. In trematodes, the asexual phase occurs within the first invertebrate intermediate host and thousands of genetically identical larvae are released into the environment. This phenomenon has long been considered only from a demographic point of view, as an adaptation to compensate for the considerable random loss of larvae during the free-living transmission phase and thus to increase the probability of meeting the downstream host (e.g. Combes 2001). To date, very few studies have investigated the impact of this clonal phase on the constitution and structure of parasite infrapopulations within definitive hosts and hence on the patterns of genetic variation (see, however, Mulvey *et al*. 1991).

Schistosoma mansoni is a dioecious blood helminth responsible for one of the most important human helminthiases (schistosomiasis) in tropical countries. In this species, clonal reproduction takes place within a freshwater snail. It produces thousands of free-living cercariae that are infective for the definitive vertebrate host in which sexual reproduction occurs. In Guadeloupe (French West Indies), *S. mansoni* has undergone a host shift from humans towards the murine host *Rattus rattus*, which appears to be its unique definitive host in the island (Sire *et al*. 2001). This schistosome–rat system offers a unique opportunity to investigate the impact of clonal larval reproduction on the distribution of the genetic diversity of adult *S. mansoni* because it allows the collection of adult parasites, which is impossible in humans.

Correspondence: F. Prugnolle. ‡Present address: Theoretical and Molecular Population Genetics Group, Department of Genetics, Downing street, Cambridge CB2 3EH, UK. Fax: +44 1223 333992; E-mail: fp235@mole.bio.cam.ac.uk

In a previous study of the distribution of *S. mansoni* genetic variability within and between infrapopulations (an infrapopulation corresponds to the parasite individuals present within one host individual), Prugnolle *et al*. (2002) noticed that deleting repeated multilocus genotypes from each infrapopulation to keep only one copy of each seemed to have a substantial effect on F_{IS} values. More precisely, this led to an increase in female F_{IS} but not in male F_{IS}. Therefore, these results suggested that female clones (i.e. multilocus genotypes repeated within the infrapopulation) were more heterozygous than expected by chance.

The aim of the present study is specifically to test and analyse this phenomenon. We first test, via a randomization procedure, whether the observed effect of clonal amplification on F_{IS} can be obtained by chance. Then, we directly test whether the number of repetitions per clone (clone size) correlates with individual heterozygosity. We effectively demonstrate a significant effect of multilocus heterozygosity on clone size. This effect occurs in a sex-specific way, as only female clone size correlates with heterozygosity. We discuss this phenomenon in relation to the heterozygosity–fitness relationship and to the sex-specific response to inbreeding depression.

Materials and methods

Sampling

The study was performed in the Dans Fond locality, one of the numerous transmission sites located along the marshy forest of the Grande Terre island of Guadeloupe (see Sire *et al*. 2001 for detailed description). Six parasitized rats were captured during the dry season in June 1997 using traps baited with coconut. For each rat, adult schistosomes were recovered by a standard perfusion technique (Duvall & Dewitt 1967), carefully washed in a physiological saline solution and stored in 70% ethanol.

Genotyping

DNA extraction and protocol for DNA amplification are presented in Barral *et al*. (1996) and Durand *et al*. (2000). A total of 357 worms (180 males and 177 females) were genotyped at seven microsatellite loci (GenBank accession numbers: AF202965, AF202966, AF202967, AF202968, R95529, L46951, M85305, Durand *et al*. 2000). Acrylamide electrophoretic gels were read independently by two persons (yielding identical results). Individuals that raised scoring problems were genotyped at least twice.

Multilocus analyses

We determined the multilocus genotypic diversity as the ratio of the number of distinct multilocus genotypes (G) to the number of individuals (N) sampled within each infrapopulation.

Tests of linkage disequilibrium were realized on samples for which only one copy of each multilocus genotype was kept. Linkage disequilibrium between each pair of loci within each infrapopulation was tested using a randomization procedure under FSTAT version 2.3.9 (updated from Goudet 1995). In this software, the P-value of the test is obtained as follows. Genotypes at the two loci are associated at random a number of times (126 000) and a statistic (the statistic used is the log-likelihood ratio G-statistic) is recalculated on the randomized data set. The P-value is estimated as the proportion of statistics from randomized data sets that are larger or equal to the observed. Because of the strong dependence of these multiple tests, the significance of P-values at the 0.05 level was estimated after correction with a conservative method (Bonferroni method, e.g. Rice 1989).

Clonality and heterozygosity

To examine the association between heterozygosity and clone size we used two different approaches. The first one makes use of Monte Carlo simulations with the CLONALITY V.1. program (developed by F.P. and M.C. and available upon request) and the second one is a generalized linear model as used, for instance, in Coulson *et al*. (1998) and Hämmerli & Reusch (2003).

Monte Carlo simulations. F_{IS} measures the within-sample departures from Hardy–Weinberg equilibrium expectations (Wright 1965). This parameter was estimated by Weir & Cockerham's (1984) unbiased estimator f. We used the program CLONALITY V.1. to test whether or not the observed F_{IS} was expected under the null hypothesis of random sampling of clones by rats. Randomized data sets were generated as follows: (i) within the real data set (sample N), the program detects the number of different clones within each infrapopulation (i.e. the number of multilocus genotypes having copies) and measures their size (i.e. number of copies) and the true f is computed (Fig. 1a); (ii) only a single copy of each different multilocus genotype is retained (sample U) (Fig. 1b); (iii) a new sample is generated (sample R1) by amplifying randomly chosen genotypes of sample U so that the sample size and the amount of repetitions of multilocus genotypes are kept identical to those found in sample N, in each infrapopulation (Fig. 1c); (iv) the procedure is repeated 5000 times (samples R1 to R5000) and a corresponding f_{Ri} is computed each time (f_{R1} to f_{R5000}); (v) a P-value is obtained by computing the proportion of time $f_{Ri} \leq f$. 5000 simulations proved to be enough to reach asymptotic state. The analyses were performed on males and females separately.

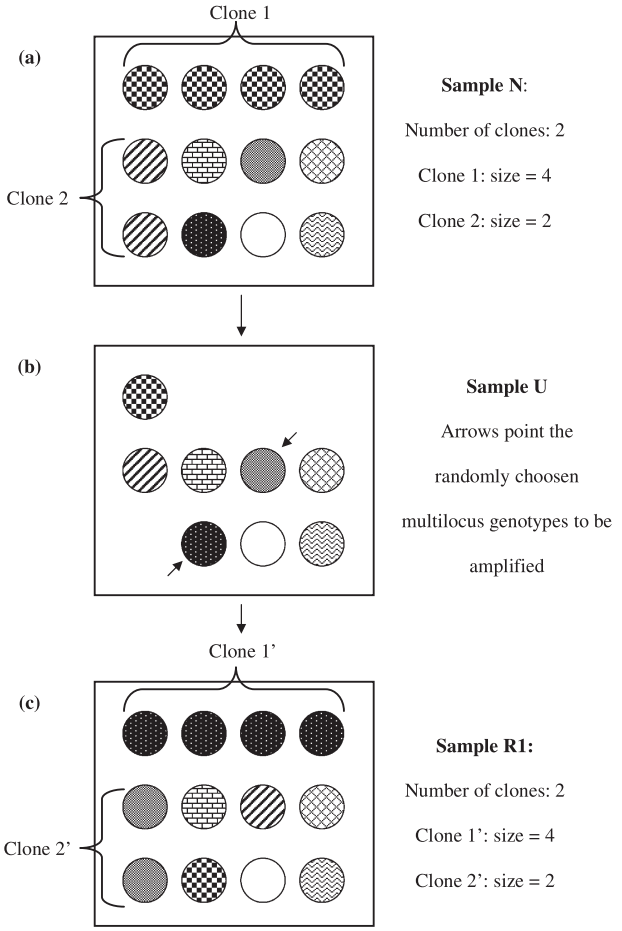

Fig. 1 Schematic representation of the randomization procedure applied by the program CLONALITY V.1. In this scheme, a square represents an infrapopulation. Circles within the square represent the different individuals. Circles displaying different draws display different multilocus genotypes. (a) Sample N: original sample; (b) Sample U: sample where copies of clones were deleted; (c) Sample R1: first random sample. The expected f_{R1} is computed from this later sample.

For multiple independent tests, the exact binomial test was used to assess when the proportion of significant *P*-values was significantly higher than expected ($\alpha = 0.05$) by chance (unilateral test performed by S-PLUS 2000, Professional release 1, MathSoft, Inc). This test was preferred to a sequential Bonferroni correction (Rice 1989) to avoid losing too much statistical power. The latter is known to be too conservative when the statistical power of the test used is already weak (for instance, when the sample size is low, as is the case in our study).

Generalized linear model. We quantified the multilocus heterozygosity (H) as the number of typed loci for which an individual was heterozygous. The sex (S)-specific effect of heterozygosity (H) on clone size (CS) was investigated with the following analysis of covariance model structure

$$CS = L + R + S + H + S \times H + \text{Constant}.$$

which corrects for the schistosome load (L) within host individuals (independently measured for males and females, see Table 1) as well as for a possible rat effect (R). It is to be noted that the rat effect is collinear to the schistosome load. All other independent variables are orthogonal and $S \times H$ denotes the interaction term between sex (S) and heterozygosity (H). The distribution of the response variable was Poisson-like (typical of count processes) and its variance (1.3577) was close to its mean (1.2570). The analysis was thus performed in a generalized linear model framework with Poisson error structure and the natural log-link function.

Results

Multilocus analyses

Among the 357 worms analysed, 284 different multilocus genotypes were discriminated (131 and 153 in females and males, respectively). The genotypic diversity (G/N) value for each infrapopulation studied is given in Table 1. Over all samples, each infrapopulation harboured clones (multilocus genotypes having repetitions) with a mean ± standard error] of 3.15 ± 0.47 copies per clone.

No linkage disequilibrium was observed between loci within each infrapopulation except for males in rats 5 and 17 after Bonferroni corrections.

Table 1 Per rat schistosome load (*Nt*), number of genotyped schistosomes (*N*) and number of different multilocus genotypes per analysed sample (*G*). Values are given for female (*F*) and male (*M*) schistosomes. Rat 1 to Rat 17: samples collected in Dans Fond 1997

	Rat 1		Rat 5		Rat 8		Rat 10		Rat 14		Rat 17	
	F	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*
Nt	50	51	62	54	499	610	179	209	127	146	82	62
N	30	30	30	30	30	30	27	30	30	30	30	30
G	21	24	27	27	27	27	23	28	17	27	16	20
G/N	0.7	0.8	0.9	0.9	0.9	0.9	0.85	0.93	0.57	0.9	0.53	0.66

Table 2 Comparison between the observed (*Obs*) and the expected (*Exp*) dataset of the mean *f* overall loci and infrapopulations, for each locus overall infrapopulations and for each infrapopulation overall loci. Values are given for males and females separately

	Females			Males		
	Obs	*Exp*	*P*-value	*Obs*	*Exp*	*P*-value
Overall loci and infrapopulations						
	−0.11	−0.04	0.002	−0.045	−0.056	0.87
Per locus overall infrapopulations						
AF202966	−0.054	0.02	0.16	−0.064	−0.05	0.29
R95529	−0.075	−0.03	0.21	−0.042	−0.035	0.38
L46951	−0.080	0.070	0.007	−0.012	−0.083	0.97
AF202965	−0.25	−0.19	0.04	0.059	0.0006	0.97
M8535	−0.07	−0.046	0.37	−0.17	−0.15	0.26
AF202967	−0.04	0.110	0.09	0.23	0.16	0.91
AF202968	−0.17	−0.16	0.45	−0.12	−0.10	0.23
Per infrapopulation overall loci						
Rat 1	0.034	0.038	0.5	0.07	0.04	0.78
Rat 5	−0.10	−0.08	0.047	−0.18	−0.17	0.28
Rat 8	0.002	−0.003	0.72	−0.09	−0.07	0.09
Rat 10	0.015	0.035	0.13	0.050	0.052	0.41
Rat 14	−0.27	−0.10	0.02	−0.086	−0.12	0.94
Rat 17	−0.28	−0.11	0.03	−0.027	−0.06	0.87

P-value corresponds to the proportion of randomly obtained data set with equal or more extreme values than the observed one.

Monte Carlo simulations

Randomization tests demonstrated that, over all infrapopulations and loci, observed female *f* was significantly lower than expected under the null hypothesis of random sampling of clones by the hosts ($P = 0.002$) (Table 2). The majority of loci displayed the same tendency because the observed *f* was lower than the expected *f* for all loci (even if the difference was not always significant) (Table 2). The number of loci displaying significant *P*-values (two out of seven) was higher than expected by chance (*P*-binomial = 0.04). Moreover, over the six infrapopulations studied, three displayed significant results (*P*-binomial = 0.0022) (see Table 2).

There was no difference between the observed *f* and the expected *f* for males (Table 2).

Table 3 Analysis of deviance

Variable 1	Resid. Df	Resid. Dev	*P*-value (χ^2)
Constant	283	132.4414	
Load (*L*)	282	130.7916	0.1990
Rat (*R*)	277	124.7030	0.2977
Sex (*S*)	276	122.5005	0.1378
Heterozygosity (*H*)	275	117.2722	0.0223
$S \times H$	274	111.0460	0.0126

Variables are added sequentially from top to bottom and a likelihood ratio test is performed between adjacent nested models. Twice the difference in deviances (Resid. Dev.) of two adjacent nested models is compared to a χ^2 distribution with degree of freedom equal to the difference in degree of freedom (Resid. Df) of the two models under comparison. The last column gives the *P*-values associated with the tests.

Generalized linear model

An analysis of deviance of the full generalized linear model and its sequentially nested parts is presented in Table 3. The dispersion parameter was estimated by the ratio of the residual χ^2 statistic to the residual degrees of freedom (Venables & Ripley 1999) to be 0.6414. Simulation experiments and model simplification suggested that this under-dispersion could be the result of (i) the high number of very low counts and (ii) the collinearities between variables (Venables & Ripley 1999). We deliberately kept these collinearities to ascertain correction for both the parasite load and the rat effect. Other error structures we tried did not performed better. We tried a negative binomial error structure, estimating its dispersion parameter *k* inside the algorithm of likelihood maximization. The algorithm did not converge as *k* increased for ever, making the negative binomial tending toward a Poisson distribution. We also tried a quasi-likelihood, setting a variance equal to the mean, which unsurprisingly yielded the same results as the Poisson model.

The model showed that clone size significantly increased with heterozygosity for the female schistosomes whereas there was no significant effect of heterozygosity on the clone size for the male schistosomes (see Table 3 and Fig. 2).

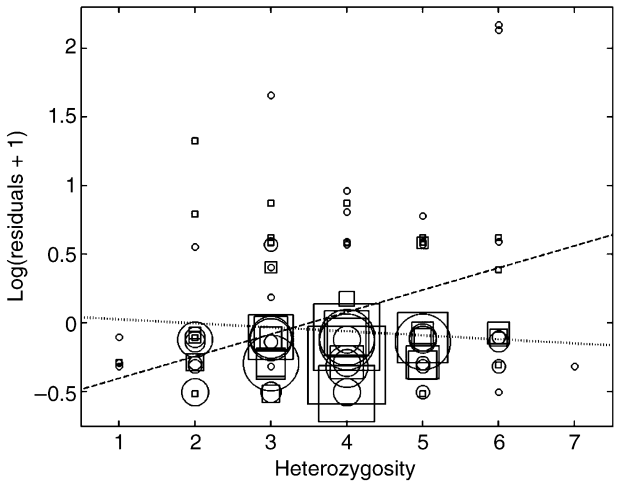

Fig. 2 Sex-specific relation between heterozygosity and clone size. Clone size has been corrected for both the parasite load and the rat effect. The natural logarithm of the residuals plus one is plotted against heterozygosity for females (circles) and males (squares). The sizes of the circles and squares are proportional to the number of data. The dashed line shows the predicted values from the generalized linear model for females. The slope is significantly different from 0 (likelihood ratio test between the models with and without the slope: $P = 0.0161$). The dotted line shows the predicted values from the generalized linear model for males. The slope is not significantly different from 0 (likelihood ratio test between the models with and without the slope: $P = 0.7185$).

Discussion

Our results clearly demonstrate a substantial effect of clonality in shaping the within infrapopulation genetic variability in *Schistosoma mansoni* infecting black rats in Guadeloupe. In particular, our Monte Carlo simulations show that the female excess of heterozygotes observed among infrapopulations and loci was significantly higher than expected under the null hypothesis of random sampling of clones by the hosts. For males, no difference was observed. These results suggested that female clones (but not male ones) were more heterozygous than expected by chance. This hypothesis was corroborated by the generalized linear model which demonstrated a female-specific correlation between multilocus heterozygosity and clone size (the number of repetitions per clone).

All these observations suggest that the success (recruitment and/or settlement within the definitive host) or production (within the intermediate host) of a female clone depends, at least in part, on its multilocus heterozygosity. Many studies have focused on the relationship between heterozygosity at a set of molecular markers and variation in fitness-associated traits (Hansson & Westerberg 2002) and some have demonstrated a positive correlation. Different hypotheses, reviewed in Hansson & Westerberg (2002), have been proposed to explain such a correlation in natural populations. Under the hypothesis of 'direct overdominance', observed variations in fitness are directly determined by the genotype at the genetic markers. While functional constraints on microsatellites cannot completely be ruled out (e.g. Kashi & Soller 1999; Li *et al*. 2002), there is no evidence that heterozygosity at microsatellite loci themselves may provide any selective advantage. Therefore, a direct overdominance at these markers can confidently be rejected to explain our observations. Alternative hypotheses, referred to as 'associative overdominance hypotheses', can be considered. According to these hypotheses ('local and general effect hypotheses'), a heterozygosity–fitness correlation results from a genetic association between the neutral marker loci and the loci under selection (David 1998). On the one hand, the 'local effect hypothesis' proposes that a heterozygosity–fitness correlation is the result of an apparent heterozygote advantage at the genetic markers that results from effects of heterozygosity/homozygosity at closely linked fitness loci. On the other hand, the 'general effect hypothesis' suggests that multilocus heterozygosity reflects variation in the inbreeding coefficient of individuals. It would thus be associated with fitness components because of the effects of homozygosity at genome-wide distributed loci (David 1998). Accordingly, we may suppose that relatively more heterozygous individuals are selected in natural populations of *S. mansoni* because of: (i) overdominance processes (e.g. heterozygous parasites might be able to exploit hosts more effectively, Agrawal & Lively 2001), or (ii) inbreeding depression resulting from the detrimental effects of recessive deleterious alleles in homozygotes (relatively inbred parasites might suffer inbreeding depression as a result of the expression of partly recessive deleterious alleles). Our data set does not allow us to distinguish between the two hypotheses and further studies need to be done to clarify this point.

Males and females can respond differentially to inbreeding and outbreeding (Coulson *et al*. 1999). For instance, in red deer fawns, outbred females survive better than inbred ones but outbred males have a lower survival than inbred ones. Differences in strategies or life-history traits in males and females can generate such sex-dependent responses to inbreeding (Coulson *et al*. 1999). Besides, some authors found that inbreeding depression can be more severe under stressful conditions such as when competition is strong (e.g. Meagher *et al*. 2000; Christen *et al*. 2002). This seems to be the case in the tapeworm *Schistocephalus solidus* where the advantage of outcrossing becomes apparent only in a competitive situation (Christen *et al*. 2002). Accordingly, different levels of competition between males and between females could generate sex-dependent responses to inbreeding depression in *S. mansoni*. Under this hypothesis, competition would be higher between females than between

males and relatively 'outbred' females would be stronger competitors within definitive hosts than their relatively 'inbred' counterparts. Controlled infestations of experimental definitive hosts with more or less heterozygous female clones should provide a simple test of this hypothesis. Such experiences would also allow understanding of which life-history trait is affected by multilocus heterozygosity (e.g. ability to settle within the definitive host).

In conclusion, in *S. mansoni*, the fittest clones are the most heterozygous (at least for females). This represents an unexpected consequence of clonality which probably leads to increased maintenance of polymorphism in these parasite populations.

Acknowledgements

We thank J. Goudet for providing parts of the FSTAT V.2.9 program, a pivotal subroutine of CLONALITY V.1. We also thank O. Gimenez, F. Balloux, C. Chevillon, F. Renaud, N. Perrin, P. Jarne, P. David, J.P. Pointier and the referees for very useful discussions and suggestions. A. Théron, P. Durand and T. de Meeûs are supported by the CNRS (PNDBE), F. Prugnolle by a MENRT, and Marc Choisy by a BDI CNRS/Région Languedoc-Roussillon. This work received financial support from the French Ministry of Ecology and Sustainable Development (Contract no. CV 02000071, MEDD, programme 'Ecosystèmes Tropicaux').

References

Agrawal AF, Lively CM (2001) Parasites and the evolution of self-fertilization. *Evolution*, **55**, 869–879.

Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.

Barral V, Morand S, Pointier JP, Théron A (1996) Distribution of schistosome genetic diversity within naturally infected *Rattus rattus* detected by RAPD markers. *Parasitology*, **113**, 511–517.

Christen M, Kurtz J, Milinski M (2002) Outcrossing increases infection success and competitive ability: experimental evidence from a hermaphrodite parasite. *Evolution*, **56**, 2243–2251.

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

Coulson T, Pemberton J, Albon Set al. (1998) Microsatellites reveal heterosis in red deer. *Proceedings of the Royal Society of London-Series B*, **256**, 489–495.

Coulson T, Albon S, Slate J, Pemberton J (1999) Microsatellite loci reveal sex-dependent response to inbreeding and outcrossing in red deer calves. *Evolution*, **53**, 1951–1960.

David P (1998) Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity*, **80**, 531–537.

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.

Goudet J (1995) Fstat (vers. 1.2). A computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Hämmerli A, Reusch TBH (2003) Inbreeding depression influences genet size distribution in a marine angiosperm. *Molecular Ecology*, **12**, 619–629.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, **11**, 2467–2474.

Kashi Y, Soller M (1999) Functional roles of microsatellites and minisatellites. In: *Microsatellites: Evolution and Applications* (eds Goldstein DB, Schlotterer C), pp. 11–22. Oxford University Press, New York.

Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, **11**, 2453–2466.

Meagher S, Penn DJ, Potts WK (2000) Male–male competition magnifies inbreeding depression in wild house mice. *Proceedings of the National Academy of Sciences of the USA*, **97**, 3324–3329.

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

Prugnolle F, De Meeûs T, Durand P, Sire C, Theron A (2002) Sex-specific genetic structure in *Schistosoma mansoni*: evolutionary and epidemiological implications. *Molecular Ecology*, **11**, 1231–1238.

Rice WR (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

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 for Parasitology*, **31**, 1609–1616.

Venables WN, Ripley BD (1999) *Modern Applied Statistics with S-PLUS*, 3rd edn. Springer Verlag, Berlin.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

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 Cambridge University. He completed the present work during his PhD on the co-evolution between *Schistosoma mansoni* and its intermediate (*Biomphalaria glabrata*) and definitive (*Rattus rattus*) hosts in Guadeloupe. Marc Choisy is a PhD student working on epidemiological statistics. André Théron is a senior research scientist working on the ecology and evolution of schistosomes and other parasites. Patrick Durand is a research engineer in molecular biology. Thierry De Meeûs is a researcher in evolutionary ecology and population genetics in host–parasite systems.
