

## ORIGINAL ARTICLE

# Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design

F Prugnolle<sup>1</sup> and T De Meeus<sup>1,2</sup>

<sup>1</sup>Laboratoire Génétique et Evolution des Maladies Infectieuses, UMR CNRS-IRD 2724, Montpellier, Cedex 5, France and <sup>2</sup>UMR 177 IRD-CIRAD, Laboratoire de Recherches et de Coordination sur les Trypanosomoses TA A-17/G, Centre CIRAD, Campus de Baillarguet, Montpellier, Cedex 5, France

Sampling design is of primary importance for empirical studies, in particular, population genetics. For parasitic organisms, a rather frequent way of sampling individuals from local populations is to collect and genotype only one randomly chosen parasite (or isolate) per host individual (or subpopulation), although each host (subpopulation) harbors a set of parasites belonging to the same species (that is, an infrapopulation). Here, we investigate, using simulations, the consequences of such sampling design regarding the estimates of linkage disequilibrium and departure from the Hardy–Weinberg expectations (H–WE) in clonal parasites with an acyclic life cycle. We show that collecting and genotyping only

one individual pathogen per host individual (or per subpopulation) and pooling them to form one ‘artificial’ subpopulation may generate strongly misleading patterns of genetic variations that may lead to false conclusions regarding their reproduction mode. In particular, we show that when subpopulations (or infrapopulations) are genetically differentiated, (i) the level of linkage disequilibrium is significantly reduced and (ii) the departure from the H–WE is strongly modified, sometimes giving a forged picture of a strongly recombining organism despite high levels of clonal reproduction. *Heredity* (2010) **104**, 135–140; doi:10.1038/hdy.2009.128; published online 7 October 2009

**Keywords:** linkage disequilibrium; clonality; sampling design; molecular markers; population genetics

## Introduction

For many organisms, in particular, very small ones such as parasites, the analysis of genetic variation at different hierarchical levels is often the only way to investigate population parameters such as gene flow, size of reproductive units or breeding strategies (for example, Criscione *et al.*, 2005). Interpreting the distribution of genetic variation in these terms is nevertheless often not a simple process. This is especially so, as the sampling design may generate artificial patterns that may lead to erroneous conclusions regarding the biology of the species concerned. A famous example is the Wahlund effect (Wahlund, 1928). When individuals belonging to two or more genetically differentiated populations are gathered into the same sample (considering that they belong to one unique population), this may artificially increase the observed homozygosity compared with the Hardy–Weinberg expectations (H–WE), even if the subpopulations themselves were in H–W equilibrium. This may thus lead to false biological interpretations, for example, the existence of self-fertilization in the studied species. The quality of the sampling design is therefore of crucial importance for biological inferences.

For parasitic organisms, including micropathogens, a rather frequent way of sampling from local populations is to collect and genotype only one randomly chosen parasite per host individual, even though each host harbors an infrapopulation. An infrapopulation refers to all individuals of a parasite species in or on an individual host at a particular time (Criscione *et al.*, 2005). All infrapopulations at a given place and time constitute the component population (Bush *et al.*, 1997). Within the component population, infrapopulations may be considered as demes or subpopulations. Indeed, most of the time, they constitute cohesive genetic units (in which genetic exchanges may potentially occur) that may persist, more or less structured (depending on the migration rate among infrapopulations at each generation and the rate of infrapopulation extinction), over time (Criscione and Blouin, 2005, 2006). *Plasmodium falciparum* oocysts present within each infected anopheline mosquito (Razakandrainibe *et al.*, 2005) as well as, the different bacterial strains of the same species that coexist within a host are examples of infrapopulations. In this latter case, the infrapopulation evolves for many generations within an individual host in which it is regulated by the immune system, competition for resources and/or space. Migration events occur from other infrapopulations. Genetic diversity inside the infrapopulation may also evolve under the effect of repeated mutations and recombinations occurring between strains. Infrapopulations thus behave just as true subpopulations. Variations of life cycles and transmission patterns from one

Correspondence: Dr F Prugnolle, Laboratoire Génétique et Evolution des Maladies Infectieuses, UMR CNRS-IRD 2724, 911 Av. Agropolis, BP 64501, 34394 Montpellier Cedex 5, France.

E-mail: prugnoll@mpl.ird.fr

Received 25 March 2009; revised 11 August 2009; accepted 13 August 2009; published online 7 October 2009

generation to another among parasite species determine the level and patterns of genetic structure observed among infrapopulations (Prugnolle *et al.*, 2005a, 2005b; Criscione and Blouin, 2006; Criscione, 2008). A component population can therefore be considered and modeled as a classical subdivided population (the subpopulations being the infrapopulations).

The one individual per host (infrapopulation) sampling strategy has frequently been used, particularly, to investigate the biology and especially the reproductive mode of some parasitic fungi (Boerlin *et al.*, 1996; Mzilahowa *et al.*, 2007; Pujol *et al.*, 1993; reviewed in Nebavi *et al.*, 2006), some protozoa (MacLeod *et al.*, 2000; de Freitas *et al.*, 2006; Koffi *et al.*, 2007; Kuhls *et al.*, 2007), some helminths (Mulvey *et al.*, 1991; Nejsun *et al.*, 2005) or some bacteria (Maynard Smith *et al.*, 1993). Using simulations, we here investigate the effect of such sampling design on the observed distribution of genetic variability within populations of clonal and partially clonal organisms. For the sake of simplicity, we have limited our investigations to clonal parasites with an acyclic life cycle (de Meeus *et al.*, 2007). This life cycle is characterized by the fact that individuals in a population originate either from clonal or sexual reproduction. The frequency of sexually or asexually produced individuals can vary. This life cycle is found in a variety of parasites including some imperfect fungi (for example, *Candida albicans*), some Parabasalia (*Trichomonas vaginalis*), some Metamonadina (*Giardia lamblia*), some parasitic amoebas or some kinetoplastid parasites (*Leishmania*, *Trypanozoma*).

For the distribution of genetic variability, we have considered two aspects that are often used to make inferences on the nature of pathogen reproduction modes (Tibayrenc and Ayala, 2002; Balloux *et al.*, 2003): (i) linkage disequilibrium between loci and (ii) departure from H-WE. We show that collecting and genotyping only one individual pathogen per host individual (or per subpopulation) and pooling them, considering that they belong to one unique subpopulation, may generate strongly misleading patterns of genetic variations that may lead to false conclusions regarding their mode of reproduction. In particular, we show that (i) the level of linkage disequilibrium is significantly reduced compared with the one observed within each subpopulation (or infrapopulation) and that (ii) the departure from the H-WE may be strongly modified, sometimes giving the false picture of an apparent strongly recombining species, although highly clonal.

## Materials and methods

### Simulations

We used an individual-based simulation approach as implemented in the software EASYPOP (version 2.0.1) (updated from Balloux (2001)) to generate all data sets. For every data set, we simulated 50 subpopulations (this term is used here as a synonym of demes or infrapopulations) comprising a fixed number of 50, 100 or 1000 individuals. Three migration rates  $m$  (0.001, 0.01 and 0.1) and five rates of asexual reproduction  $c$  (1, 0.999, 0.990, 0.800 and 0 (panmixia)) were simulated. A total of 20 loci with a mutation rate of  $10^{-5}$  were used. Mutations had an equivalent probability of generating any of 99

possible allelic states. At the start of the simulation, genetic diversity was set to the maximum possible value. This means that at the first generation, all the 99 possible alleles were randomly assigned in all individuals of all demes. The simulations were then run for 10000 generations. For each parameter set, five replicates were run. At the end of each simulation, only 20 subpopulations were randomly sampled within which only 20 individuals were kept.

### Parameter estimates and sampling design

The multilocus linkage disequilibrium  $\bar{r}_D$  was estimated using the program Multilocus 1.2 (Agapow and Burt, 2001).  $\bar{r}_D$  is a variant of the association index  $I_A$  (Maynard Smith *et al.*, 1993), independent of the number of loci analyzed. It has a form similar to a correlation coefficient and has a maximum value of 1. The departure from the H-WE was estimated using the Weir and Cockerham's, (1984)  $f$  unbiased estimator for  $F_{IS}$  with the program FSTAT V. 2.9.3 (Goudet 2007, updated from Goudet, 1995).

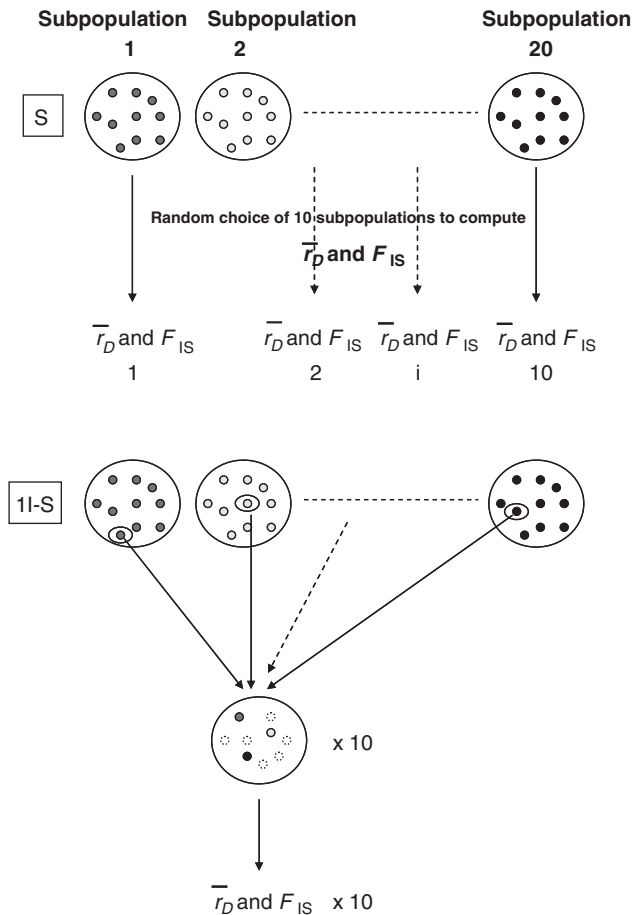
For each set of simulations, both parameters were estimated using two sampling designs: the 'Subpopulation (S)' design and the 'One-Individual per Subpopulation (1I-S)' design. For the first one, parameters were estimated from 10 randomly chosen subpopulations (and the 20 individuals present in each of the 10 subpopulations). For the second design, they were estimated from an artificially created subpopulation. This artificial subpopulation included 20 individuals, each originating from a different subpopulation (1 individual taken randomly in each of the 20 simulated subpopulations). This sampling was repeated 10 times to get 10 estimates of  $\bar{r}_D$  and  $f$  as obtained in the case of the 'S' design (cf. Figure 1).

### Statistical analyses

The effect of sampling design (S and 1I-S designs) on LD and  $F_{IS}$  was investigated using a Wilcoxon rank test for paired data using the R program (R Development-core-team, 2008). A linear model was then used to determine what explanatory variable affected the difference of LD or  $F_{IS}$  between the two designs. The model analyzed was  $\Delta\bar{r}_D = \bar{r}_{D(S)} - \bar{r}_{D(1I-S)}$  or  $\Delta F_{IS} = F_{IS(S)} - F_{IS(1I-S)} \sim Nm + \text{poly}(c, 2) + Nm:\text{poly}(c, 2)$ .  $Nm$  referred to the effective number of migrants,  $c$  to the clonal rate,  $\text{poly}(c, 2)$  to a quadratic function because it gave a better fit than simple linear relationships (data not shown), and ':' to the interaction between terms. We used a stepwise procedure to select the best model using the AIC (Akaike Information Criterion) (Akaike, 1974).

### Application to real data set

We applied similar sampling designs and then computed the linkage disequilibrium and the  $F_{IS}$  for the two different parasite organisms, *C. albicans* and *P. falciparum*. *C. albicans* is a diploid fungus that may be responsible for severe mucosal or systematic infections in immunocompromised persons (Beck-Sague and Jarvis, 1993; Boerlin *et al.*, 1996; Hull *et al.*, 2000; Berman and Sudbery, 2002; Bougnoux *et al.*, 2004). *P. falciparum* is a protozoan parasite that is responsible for the deadliest form of malaria in humans. Both pathogens reproduce clonally, at least at one stage of their life cycle.



**Figure 1** Diagram showing the two different sampling strategies (the ‘Subpopulation’ (S) and the ‘One-Individual per Subpopulation’ strategy (1I-S) used to analyse the data. For the ‘S’ sampling strategy, linkage disequilibrium ( $\bar{r}_D$ ) and  $F_{IS}$  are computed independently for 10 randomly chosen subpopulations out of the 20 sampled. For the ‘1I-S’ sampling strategy, linkage disequilibrium ( $\bar{r}_D$ ) and  $F_{IS}$  are computed from an artificially created subpopulation. This subpopulation includes 20 individuals, each originating from a different simulated subpopulation (from the 20 subpopulations sampled). The creation of the artificial subpopulation is repeated 10 times to get 10 estimates of  $\bar{r}_D$  and  $F_{IS}$  as obtained in the case of the ‘S’ strategy.

For *C. albicans*, data come from Nebavi *et al.* (2006). In that study, 55 *C. albicans* infrapopulations were obtained from 42 HIV-positive patients (some patients were collected several times) with oral candidiasis from Abidjan and the suburbs (see Nebavi *et al.*, 2006 for further details). Each sample consisted of five isolates per patient, obtained by buccal swabbing. Each isolate was genotyped at 14 enzyme-coding loci.

For *P. falciparum*, the data come from Razakandrainibe *et al.* (2005). In that study, *Plasmodium* oocysts were obtained from a total of 145 infected anopheline mosquitoes. From these, we kept only the oocyst infrapopulations that displayed at least five genotyped oocysts (the total number of oocyst infrapopulations analyzed = 30).

For both parasites, linkage disequilibrium and  $F_{IS}$  were obtained as described in the section ‘Parameter estimates and sampling design’.

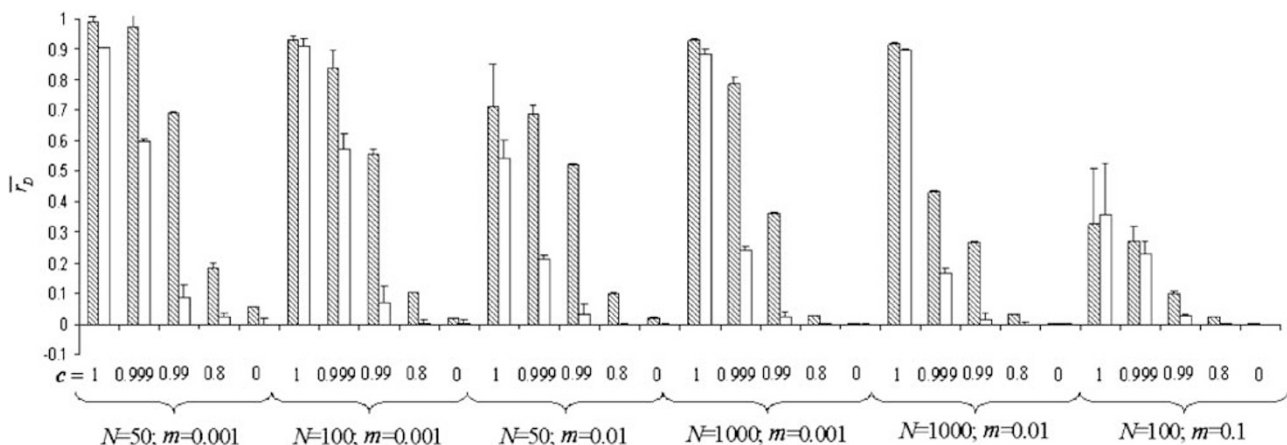
## Results

### Linkage disequilibrium

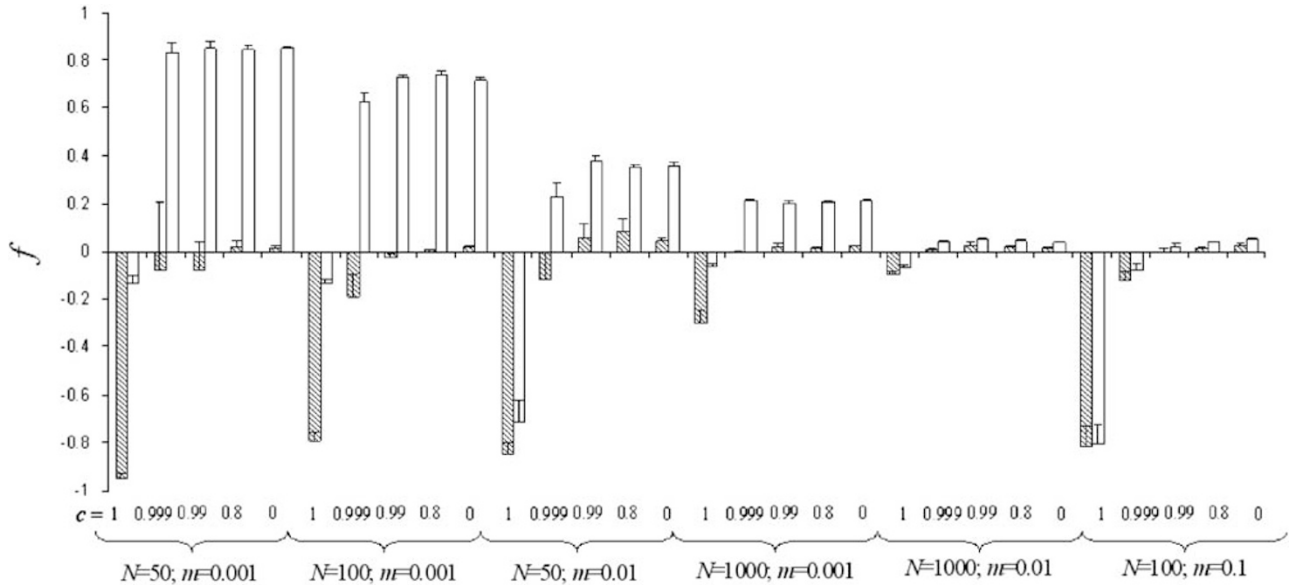
Results are presented in Figure 2. Over all sets of simulations, the level of linkage disequilibrium ( $\bar{r}_D$ ) is significantly different between the two kinds of sampling designs (Wilcoxon signed rank test  $P$ -value < 0.0001). We show, in particular, that it is significantly lower in the case of the 1I-S design than under the S design (Figure 1). The difference between the two estimates of LD ( $\Delta\bar{r}_D$ ) is significantly dependent on the parameters of the simulations. The difference decreases as  $Nm$  increases (coefficient estimate = -0.016,  $P$ -value = 0.0168), whereas it increases as does the clonal rate (coefficient estimates = 0.48 and 0.27,  $P$ -value = 0.0052). Interestingly though, the strongest differences are observed for clonal rates very close to 1 (0.99 or 0.999) but not 1.

### Departure from the H-WE

Results are presented in Figure 3. Over all simulations, the average  $F_{IS}$  estimated within subpopulations using the S design is significantly lower than the  $F_{IS}$  estimated using the 1I-S design (Wilcoxon signed rank test  $P$ -value < 0.0001). The difference observed between estimates is



**Figure 2** Variation of Agapow and Burt’s, (2001) estimates of LD ( $\bar{r}_D$ ) measured within samples using the S (hatched bars) and the 1I-S (white bars) sampling design under different simulation scenarios.  $c$  represents the rate of clonal reproduction within subpopulations,  $N$  the subpopulation size and  $m$  the migration rate.



**Figure 3** Variation of Weir and Cockerham's, (1984) estimates of  $F_{IS}$  ( $f$ ) measured within subpopulations using the S (hatched bars) and the II-S sampling (white bars) design under different simulation scenarios.  $c$  represents the rate of clonal reproduction within subpopulations,  $N$  the subpopulation size and  $m$  the migration rate.

explained by  $Nm$  (coefficient estimate = 0.055,  $P$ -value < 0.0001) but not by the clonal rates (coefficient estimate = -0.029 and -0.021,  $P$ -value = 0.9886).

#### Application to *C. albicans* and *P. falciparum* data set

The same patterns were observed for *C. albicans* and *P. falciparum* real data sets. For both, we observed a decline in the amount of LD when sampling individuals using the II-S design (*C. albicans*: 'S' design  $\bar{r}_D = 0.83$ ; 'II-S' design:  $\bar{r}_D = 0.12$ ; *P. falciparum*: 'S' design  $\bar{r}_D = 0.27$ ; 'II-S' design:  $\bar{r}_D = 0.14$ ); and an increase in  $F_{IS}$  (*C. albicans*: 'S' design  $f$  (95% confidence interval obtained by bootstrap over loci) = -0.85 (-0.90; -0.73); 'II-S' design  $f$  (95% confidence interval obtained by bootstrap over loci) = 0.07 (-0.11; 0.34); *P. falciparum*: 'S' design  $f$  (95% confidence interval obtained by bootstrap over loci) = 0.10 (0.049; 0.146); 'II-S' design:  $f$  (95% confidence interval obtained by bootstrap over loci) = 0.49 (0.41, 0.58). Note that for *C. albicans*, the confidence interval obtained by bootstrap over loci of the  $F_{IS}$  comprised 0, which is characteristic of panmixia.

## Discussion

Pathogens are frequently studied using a particular sampling strategy. Although each individual host may harbor an infrapopulation of the parasite, it is frequent that only one individual parasite per host is sampled (Pujol *et al.*, 1993; Boerlin *et al.*, 1996; Mzilahowa *et al.*, 2007; reviewed in Nebavi *et al.*, 2006). In this paper, we have investigated the effect of this sampling strategy on two aspects of genetic variability that are commonly analyzed to detect clonal reproduction: linkage disequilibrium and deviations from the H-WE (Tibayrenc and Ayala, 2002). For this study, we have specifically focused on diploid acyclic life cycle parasites.

Our results clearly show a strong effect of such design on the distribution of genetic variability at both levels. We show that sampling only one individual per

subpopulation (or infrapopulation) (II-S design) artificially decreases the level of linkage disequilibrium observed between loci and increases the  $F_{IS}$ . These modifications of LD and  $F_{IS}$  can be rather drastic as shown in Figures 2 and 3. For example, for the case where  $N = 50$ ,  $m = 0.001$  and  $c = 0.99$  (99% of individuals are clonally produced within each subpopulation), the average LD measured using the S design is about 10 times higher ( $\bar{r}_D = 0.68$ ) than the one measured using the II-S design ( $\bar{r}_D = 0.086$ ). Similarly,  $F_{IS}$  estimates are strongly increased under the II-S design ( $f = 0.85$ ) compared with that under S design ( $f = -0.0756$ ). This means, therefore, that depending on the sampling design, the analysis of genetic variability may give a very different picture of the factors that shaped population structure in the species under study. Thus, in the latter example, although under the S design all elements (LD and  $F_{IS}$ ) are congruent with a species being highly clonal (Balloux *et al.*, 2003; De Meeus *et al.*, 2006), the results obtained under the II-S design would be more consistent with a sexual species displaying high rates of self-fertilization.

One may obviously argue that LD and  $F_{IS}$  are not the only elements that may allow one to conclude that a species (or a population) is highly clonal. Tibayrenc and Ayala (2002) argue that other genetic criteria may constitute good indices of clonal reproduction. The occurrence of identical multilocus genotypes within subpopulations is one of them. Unfortunately, this criterion is also strongly affected by the sampling design. For  $N = 50$ ,  $m = 0.01$  and  $c = 0.8$ , we determined the number of clones per subpopulation and the number of identical multilocus genotypes per clone under the S and II-S designs using the program CLONALITY V 0.4 (Prugnolle *et al.*, 2008). As expected, the average number of clones (multilocus genotypes or haplotypes presenting replicates) observed within subpopulations and the average number of replicates per clone was higher under the S design (average number of clones = 2.9 and average

number of replicates per clone = 2.3) than under the 1I-S design (no clone was found).

Why does LD decrease ( $F_{IS}$  increase) under the 1I-S design compared with the LD ( $F_{IS}$ ) observed under the S design? When pooling individuals coming from genetically differentiated subpopulations of sexual organisms, it is traditional to observe an increase in linkage disequilibrium (Nei and Li, 1973). In the case of clonal organisms, the effect is opposite to that shown here. One obvious explanation for this opposite trend lies in the diversity of haplotypes (multilocus genotypes) occurring within subpopulations. When this diversity is low (for example, when the clonal rate is high), the amount of LD is very high within each subpopulation under the S design as shown in Figure 2. Pooling individuals coming from several genetically differentiated subpopulations (the 1I-S design) artificially increases the diversity of haplotypes occurring in the sample, which in turn reduces the amount of LD observed between loci. This phenomenon is congruent with the fact that when the effective migration rate increases (and so does the diversity of haplotypes maintained within each subpopulation), the level of LD decreases under the S design (Figure 2). Regarding the departure from the H-WE, the increase of  $F_{IS}$  occurs simple because of a classical Wahlund effect (Wahlund, 1928). A particular aspect of the Wahlund effect observed here for clonal organisms is that it may generate  $F_{IS}$  values close to 0 or above 0, although within each subpopulation the  $F_{IS}$  values are negative, indicative of an excess of heterozygotes compared with the H-WE.

In conclusion, we show that collecting and genotyping only one individual per host individual (per infrapopulation) in acyclic life cycle pathogens and pooling them to form one 'artificial' subpopulation may generate strongly misleading patterns of genetic variations that may lead to erroneous conclusions regarding their reproduction mode. We limited our investigations to clonal organisms. However, the same effect occurs in highly self-fertilizing sexual organisms, namely a reduction of the level of linkage disequilibrium between loci and an increase in  $F_{IS}$  (data not shown) when applying the same kind of sampling design. Our study suggests therefore that, for clonal pathogens, sampling has to be done with lots of precautions, especially when local populations are constituted by several highly differentiated subpopulations (for example, infrapopulations). As shown here, a sampling design that does not take into account the smallest possible level of substructure may lead to erroneous interpretations sometimes giving the false picture of an organism highly recombining although highly clonal. Obviously, our study goes further beyond the clonal parasites alone and applies to all clonal organisms that could present highly structured populations. In particular, estimation of recombination from sequence data, often employed to study prokaryotic organisms (McVean *et al.*, 2002), may also be concerned by this important caveat. For example, the intestinal habitat of mammals has been shown to offer multiple niche for parasites that colonize it (Haukisalmi and Henttonen, 1993) and also for bacterial species (Swidsinski *et al.*, 2005). If different clonal lineages segregating at different intestine sections are sampled on a one isolate per host basis it is probable that the recombination will be strongly overestimated in the bacterial group under study.

## Acknowledgements

We thank CNRS and IRD for financial support. This paper is financed in part by the ANR 07 SEST 012 'MGANE' and ANR 06 SEST 20-01 IAEL. We thank Patrick Durand for very useful discussions. We thank François Renaud and François Nébavi for providing the *P. falciparum* and *C. albicans* data, respectively. We thank C. Criscione for comments on a previous version of the paper.

## References

- Agapow PM, Burt A (2001). Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* **1**: 101–102.
- Akaike H (1974). A new look at the statistical model identification. *IEEE Trans Automat Contr* **19**: 716–723.
- Balloux F (2001). EASYPOP (version 1.7): a computer program for population genetics simulations. *J Hered* **92**: 301–302.
- Balloux F, Lehmann L, de Meeus T (2003). The population genetics of clonal and partially clonal diploids. *Genetics* **164**: 1635–1644.
- Beck-Sague CM, Jarvis WR (1993). Secular trends in the epidemiology of nosocomial fungal infections in the United-States, 1980-1990. *J Infect Dis* **167**: 1247–1251.
- Berman J, Sudbery PE (2002). *Candida albicans*: a molecular revolution built on lessons from budding yeast. *Nat Rev Genet* **3**: 918–930.
- Boerlin P, BoerlinPetzold F, Goudet J, Durussel C, Pagani JL, Chave JP *et al.* (1996). Typing *Candida albicans* oral isolates from human immunodeficiency virus-infected patients by multilocus enzyme electrophoresis and DNA fingerprinting. *J Clin Microb* **34**: 1235–1248.
- Bougnoux ME, Aanensen DM, Morand S, Théraud M, Spratt BG, d'Enfert C (2004). Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. *Inf Genet Evol* **4**: 243–252.
- Bush AO, Lafferty KD, Lotz JM, Shostak AW (1997). Parasitology meets ecology on its own terms: Margolis *et al.* revisited. *J Parasitol* **83**: 575–583.
- Criscione CD (2008). Parasite co-structure: broad and local scale approaches. *Parasite* **15**: 439–443.
- Criscione CD, Blouin MS (2005). Effective sizes of macroparasite populations: a conceptual model. *Trends Parasitol* **21**: 212–217.
- Criscione CD, Blouin MS (2006). Minimal selfing, few clones, and no among-host genetic structure in a hermaphroditic parasite with asexual larval propagation. *Evolution* **60**: 553–562.
- Criscione CD, Poulin R, Blouin MS (2005). Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Mol Ecol* **14**: 2247–2257.
- de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Gonçalves VF, Teixeira SMR *et al.* (2006). Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS Pathog* **2**: 226–235.
- De Meeus T, Lehmann L, Balloux F (2006). Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. *Inf Genet Evol* **6**: 163–170.
- de Meeus T, Prugnolle F, Agnew P (2007). Asexual reproduction: genetics and evolutionary aspects. *Cell Mol Life Sci* **64**: 1355–1372.
- Goudet J (1995). Fstat (vers.1.2): a computer program to calculate F-statistics. *J Hered* **86**: 485–486.
- Haukisalmi V, Henttonen H (1993). Coexistence in Helminths of the Bank Vole *Clethrionomys-Glaeolus* .2. Intestinal distribution and interspecific interactions. *J Anim Ecol* **62**: 230–238.
- Hull CM, Raisner RM, Johnson AD (2000). Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. *Science* **289**: 307–310.

- Koffi M, Solano P, Barnabé C, De Meeûs T, Bucheton B, N'Dri L *et al.* (2007). Genetic characterisation of *Trypanosoma brucei* ssp by microsatellite typing: new perspectives for the molecular epidemiology of human African trypanosomiasis. *Inf Genet Evol* **7**: 675–684.
- Kuhls K, Keilonat L, Ochsenreither S, Schaar M, Schweynoch C, Presber W *et al.* (2007). Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. *Microbes Infect* **9**: 334–343.
- MacLeod A, Tweedie A, Welburn SC, Maudlin I, Turner CMR, Tait A (2000). Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proc Natl Acad Sci USA* **97**: 13442–13447.
- Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993). How clonal are bacteria? *Proc Natl Acad Sci USA* **90**: 4384–4390.
- McVean G, Awadalla P, Fearnhead P (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.
- Mulvey M, Aho JM, Lydeard C, Leberg PL, Smith MH (1991). Comparative population genetic structure of a parasite (*Fascioloides magna*) and its definitive host. *Evolution* **45**: 1628–1640.
- Mzilahowa T, McCall PJ, Hastings IM (2007). 'Sexual' population structure and genetics of the malaria agent *P.falciparum*. *PLoS ONE* **2**: e613.
- Nebavi F, Ayala FJ, Renaud F, Bertout S, Eholie S, Moussa K *et al.* (2006). Clonal population structure and genetic diversity of *Candida albicans* in AIDS patients from Abidjan (Cote d'Ivoire). *Proc Natl Acad Sci USA* **103**: 3663–3668.
- Nei M, Li WH (1973). Linkage disequilibrium in subdivided populations. *Genetics* **75**: 213–219.
- Nejsum P, Frydenberg J, Roepstorff A, Parker Jr ED (2005). Population structure in *Ascaris suum* (Nematoda) among domestic swine in Denmark as measured by whole genome DNA fingerprinting. *Hereditas* **142**: 7–14.
- Prugnolle F, Choisy M, de Meeus T (2008). Clonality V.0.4: a randomization-based program to test for heterozygosity-genet size relationships in clonal organisms. *Mol Ecol Resour* **8**: 954–956.
- Prugnolle F, Liu H, de Meeus T, Balloux F (2005a). Population genetics of complex life-cycle parasites: an illustration with trematodes. *Int J Parasitol* **35**: 255–263.
- Prugnolle F, Roze D, Theron A, DE Meeûs T (2005b). F-statistics under alternation of sexual and asexual reproduction: a model and data from schistosomes (platyhelminth parasites). *Mol Ecol* **14**: 1355–1365.
- Pujol C, Reynes J, Renaud F, Raymond M, Tibayrenc M, Ayala FJ *et al.* (1993). The yeast *Candida albicans* has a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients. *Proc Natl Acad Sci USA* **90**: 9456–9459.
- R Development-core-team (2008). R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org>, ISBN 3-900051-07-0.
- Razakandrainibe FG, Durand P, Koella JC, De Meeus T, Rousset F, Ayala FJ *et al.* (2005). 'Clonal' population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. *Proc Natl Acad Sci USA* **102**: 17388–17393.
- Swidsinski A, Loening-Baucke V, Lochs H, Hale L-P (2005). Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence *in situ* hybridization study in mice. *World J Gastroenterol* **11**: 1131–1140.
- Tibayrenc M, Ayala FJ (2002). The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol* **18**: 405–410.
- Wahlund S (1928). Zusammensetzung von populationen und korrelation-serscheiunungen von standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**: 65–106.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.