Molecular Ecology (2005) **14**, 2695–2702 doi: 10.1111/j.1365-294X.2005.02643.x

F-statistics of clonal diploids structured in numerous demes

THIERRY DE MEEÛS* and FRANÇOIS BALLOUX†
*Génétique et Evolution des Maladies Infectieuses, Equipe Evolution des Systèmes Symbiotiques, UMR 2724 CNRS-IRD, Centre IRD de Montpellier, 911 Avenue d'Agropolis, BP 64501, 34394 Montpellier cedex 5, France, †Department of Genetics, Downing Street, University of Cambridge, Cambridge CB2 3EH, UK

Abstract

The expected apportionment of genetic diversity in diploid clonal organisms structured in numerous subpopulations is explored. Under the specific assumptions considered, corresponding, for instance, to clonal pathogens infecting a large number of hosts, the co-ancestry between individuals within subpopulations is the only nontrivial quantity. Thus the population structure can be fully described either by F_{ST} or F_{IS}, as $F_{ST} = -F_{IS}/(1 - F_{IS})$. We show that, for most of the parameter space considered, including simulations where equilibrium is not reached and/or where homoplasy is high, the number of effective migrants is most accurately estimated as $Nm = -(1 + F_{IS})/4F_{IS}$. We further propose a criterion to test for the absence of cryptic sexual reproduction based on the *F*-statistics F_{IS} and F_{ST}, which is applied to three previously published empirical data sets.

Keywords: clonality, diploids, heterozygosity, parthenogenesis, population genetics, population structure

Received 24 February 2005; revision accepted 13 May 2005

Introduction

Over the last few years, there has been renewed interest in the theoretical population genetics of clonal or partially clonal organisms as testified by the number of recent papers addressing this issue (e.g. Birky 1996; Berg & Lascoux 2000; Balloux *et al*. 2003; Bengtsson 2003; Ceplitis 2003; De Meeûs & Balloux 2004). While these papers might have been useful at clarifying the expectations for effective population sizes and apportionment of genetic variation under varying rates of clonal reproduction, there is still an urgent need for theoretical studies evaluating what inferences can be drawn from population genetics data for clonal organisms. In this note we address the estimation of the number of migrants for populations of clonal diploid organisms subdivided in a large number of demes; a typical example being pathogenic organisms infecting a large number of hosts, such as *Candida albicans*, *Leishmania* or *Trypanosoma*. Under these specific assumptions $F_{ST} = -F_{IS}/(1 - F_{IS})$, and the number of migrants (Nm) are more accurately inferred from the *F*-statistic F_{IS} rather than F_{ST} under most of the parameter range. We also propose a criterion for testing for strict clonal reproduction based on the same equality: $F_{ST} = -F_{IS}/(1 - F_{IS})$. This criterion is straightforward to apply to population genetics data, and deviation from this equality can be considered as strongly suggestive of cryptic sexual reproduction. In order to illustrate the approach we apply this criterion to simulated data and to three previously published empirical data sets.

Correspondence: Thierry De Meeûs, Fax: (33) 467 41 62 99; E-mail: demeeus@mpl.ird.fr

Methods

The model

We consider a subdivided monoecious population of diploid individuals with nonoverlapping generations. Individuals reproduce clonally with probability c and sexually with the corresponding probability $(1 - c)$. Self-fertilization occurs at a rate s. There are n subpopulations, or demes, each composed of N individuals. Migration between the subpopulations follows an island model (Wright 1951) with a migration rate m. The mutation rate is u for all alleles and therefore the probability of two alleles identical by descent before mutation still being identical after mutation will be $\gamma = (1 - u)^2$. We further assume stable population sizes and no selection.

Because of the symmetry of the island model, only the following probabilities of identity by descent are needed

to describe the apportionment of genetic variation in a subdivided monoecious population.

F: The inbreeding coefficient, defining the probability that two alleles drawn at random from a single individual are identical by descent.

θ: Co-ancestry of individuals drawn at random from within the same subpopulation, defined as the probability that two randomly sampled alleles from two different individuals within a subpopulation are identical by descent.

α: Co-ancestry of individuals randomly drawn from different subpopulations. This is defined as the probability that two randomly sampled alleles from two individuals in different subpopulations are identical by descent.

These identities can be calculated in juveniles (before dispersal) or in adults (after dispersal). Throughout the paper we will refer to adult identities (F_A, θ_A, α_A).

Recurrence equations

The recurrence equations for the different identities by descent among adults in a monoecious population with mixed clonal, and sexual reproduction in an island model were given in Balloux *et al*. (2003) (equation 5) as:

$$\begin{cases} F_{A(t+1)} = \gamma\left\{cF_{A(t)} + (1-c)\left[s\left(\dfrac{1+F_{A(t)}}{2}\right) + (1-s)\theta_{A(t)}\right]\right\} \\ \theta_{A(t+1)} = \gamma\left\{q_s\left[\dfrac{1}{N}\left(\dfrac{1+F_{A(t)}}{2}\right) + \left(1-\dfrac{1}{N}\right)\theta_{A(t)}\right] + (1-q_s)\alpha_{A(t)}\right\} \\ \alpha_{A(t+1)} = \gamma\left\{q_d\left[\dfrac{1}{N}\left(\dfrac{1+F_{A(t)}}{2}\right) + \left(1-\dfrac{1}{N}\right)\theta_{A(t)}\right] + (1-q_d)\alpha_{A(t)}\right\} \end{cases} \quad \text{(eqn 1)}$$

with:

$$\begin{cases} q_s = \dfrac{(1-m)[(1-m)N-1] + \dfrac{m(1-mN)^2}{1-mN(n-1)}}{n-1} \\ q_d = \dfrac{1-q_s}{n-1} \end{cases} \quad \text{(eqn 2)}$$

q_s being the probability that two individuals taken at random within the same subpopulation after migration were born in the same subpopulation and q_d the probability that two individuals sampled after migration in different subpopulations originated from the same subpopulation (Wang 1997). Under clonal reproduction, no gametes are produced, and for migration to occur, individuals themselves must be exchanged between populations. This leads to the relatively cumbersome expression for q_s as it involves sampling of individuals without replacement. If subpopulation size is reasonably large, one can apply the much more compact expression assuming sampling with replacement that also applies to gametic migration. This simpler expression can be obtained by taking the limit of q_s as N tends to infinity: $\lim_{N\to\infty} q_s = (1-m)^2 + m^2/(n-1)$. The difference between the two sampling schemes, and hence between zygotic and gametic migration is function of migration, subpopulation size and is also marginally affected by the number of subpopulations. However, the difference between the two sampling schemes rapidly becomes trivial with increasing subpopulation size (Nagylaki 1983), and is negligible over most the parameter space. Assuming a population subdivided into 1000 demes counting 10 individuals each and with migration rate of 0.001, the relative difference between q_s under gametic and zygotic migration is of the order of 0.01%.

Simulations

We used an individual-based simulation approach as implemented in the software EASYPOP (version 1.8) (Balloux 2001) to generate all population genetics data sets. For all simulations, we used 20 loci with a mutation rate of 10^{-5}. Mutations had an equivalent probability to generate any of the 99 (weak homoplasy), five or two (strong homoplasy) possible allelic states. At the start of the simulation, genetic diversity was set to the maximum possible value at the first generation and the simulations were then run for 2000 (short), 10 000 (medium) or 20 000 (long) generations. All simulations were run for a 100% rate of clonal reproduction (no sex), a large number of demes (n = 1000), small subpopulation size (N = 10) and restricted migration rate (m = 0.001). Additional simulations were run to test the effect of initial genetic variability and deviations from an island model of migration. One set of simulations was run for 2000 generations (short) with minimal initial genetic diversity (i.e. all individuals in the population being homozygote for the same allele at all loci out of the 99 possible allelic states). Simulations deviating from the island model comprised one-dimensional and two-dimensional stepping-stone models. For these geographically explicit simulations, all three levels of homoplasy (2, 5 and 99 allelic states) were considered. For the two-dimensional stepping-stone model, the lattice comprised 1024 subpopulations (32 × 32 lattice) rather than the 1000 considered in all other simulations. For each parameter set, we performed 10 replicates. *F*-statistics were then estimated on a subsample of 50 subpopulations comprising 10 individuals each, using Weir & Cockerham's (1984) method implemented in FSTAT version 2.9.3.2 (Goudet 1995).

Data sets analysed

We restricted ourselves to studies of organisms for which an island model of migration might be considered reasonably realistic, and where the sampling strategy employed should not have generated strong Wahlund effect.

For instance, we did not consider data sets where samples had been collected over several years. We retrieved three data sets of clonal organisms from the literature that seem to fit those criteria. One consists of microsatellite data of asexual populations of the aphid *Rhopalosiphum padi* (Delmotte *et al*. 2002) where *F*-statistics are given. Genotypes of *Trypanosoma brucei* (MacLeod *et al*. 2000) were downloaded from the *PNAS* website and *F*-statistics computed with FSTAT version 2.9.3.2 (Goudet 1995). For the sea anemone *Sargatia ornata* (Shaw *et al*. 1994), *F*-statistics were extrapolated from summary statistics given in the paper: H_O (observed heterozygosity), H_S [Nei's (1978) unbiased expected heterozygosity] and D_S [Nei's (1972) genetic distance]. F_{IS} was inferred as $F_{IS} = (H_O - H_S)/H_S$, and F_{ST} was inferred from D_S and H_S. Under the assumptions of biallelic loci (which is met in *S. ornata* allozymes) with small mutation rate and infinite-island model, we can rearrange Nei's (1972) formulation to extrapolate F_{ST} using the formula:

$$F_{ST} \approx \frac{(1 - H_S)(1 - e^{-D_S})}{H_S + (1 - H_S)(1 - e^{-D_S})} \qquad \text{(eqn 3)}$$

Results

Analytical solutions

When the number of subpopulations is large ($n \to \infty$), it can be seen from equation 2 that $q_s \approx (1 - m)^2$ and $q_d \approx 0$. Further assuming strict clonal reproduction ($c = 1$), the system of recurrence equations of the identities by descent given in equation 1 reduces to:

$$\begin{cases} F_{A(t+1)} = \gamma(F_{A(t)}) \\ \theta_{A(t+1)} = \gamma\left\{(1-m)^2\left[\frac{1}{N}\left(\frac{1+F_{A(t)}}{2}\right)+\left(1-\frac{1}{N}\right)\theta_{A(t)}\right]+[1-(1-m)^2]\alpha_{A(t)}\right\} \\ \alpha_{A(t+1)} = \gamma(\alpha_{A(t)}) \end{cases} \qquad \text{(eqn 4)}$$

Setting the system at equilibrium [$F_{A(t+1)} = F_{A(t)}$, $\theta_{A(t+1)} = \theta_{A(t)}$ and $\alpha_{A(t+1)} = \alpha_{A(t)}$] yields:

$$\begin{cases} \hat{F} = 0 \\ \hat{\theta} = \dfrac{(1 - u)^2 (1 - m)^2}{2N - 2(1 - u)^2(1 - m)^2(N - 1)} \\ \hat{\alpha} = 0 \end{cases} \qquad \text{(eqn 5)}$$

Wright's *F*-statistics (Wright 1965), the parameters most widely used to describe population structure (e.g. Nagylaki 1998), can be defined following Cockerham (1969, 1973) as:

$$\begin{cases} F_{IS} = \dfrac{F - \theta}{1 - \theta} \\ F_{ST} = \dfrac{\theta - \alpha}{1 - \alpha} \\ F_{IT} = \dfrac{F - \alpha}{1 - \alpha} \end{cases} \qquad \text{(eqn 6)}$$

Given the assumptions leading to the system of equation in (5), *F*-statistics reduce to:

$$\begin{cases} F_{IS} = \dfrac{-\hat{\theta}}{1 - \hat{\theta}} \\ F_{ST} = \hat{\theta} \\ F_{IT} = 0 \end{cases} \qquad \text{(eqn 7)}$$

It can be seen from equation 7 that whatever θ is, $F_{ST} = -F_{IS}/(1 - F_{IS})$. Note that under strict clonality $F_{IS} < 0$ and consequently $-F_{IS}/(1 - F_{IS}) > 0$. Thus, population structuring can be measured with either the parameter F_{IS} or F_{ST}. This stems from the fact that when reproduction is clonal and the number of subpopulations is large, the only nontrivial identity by descent coefficient is θ, the co-ancestry of individuals drawn at random from within the same subpopulation. Inbreeding, the probability of identity by descent within individuals (F) is always zero in clonal organisms and a large number of subpopulations will also reduce the co-ancestry between subpopulations (α) to zero. Under these conditions, the dynamics of alleles is entirely defined by the number of migrants and mutants within each subpopulation, and from equations 5 and 7 we can write:

$$F_{ST} = \frac{(1 - u)^2 (1 - m)^2}{2N - 2(N - 1)(1 - u)^2(1 - m)^2} \qquad \text{(eqn 8)}$$

Expanding the previous expression and assuming u and m sufficiently small so that terms of order u^2, m^2 and um can be ignored leads to:

$$\begin{cases} F_{ST} = \dfrac{1}{4N(m + u) + 2} \\ F_{IS} = -\dfrac{F_{ST}}{1 - F_{ST}} = -\dfrac{1}{4N(m + u) + 1} \end{cases} \qquad \text{(eqn 9)}$$

Note that for $m = 0$, F_{ST} tends towards 0.5 (providing u close to 0), which is thus the maximum possible value for genetic differentiation between populations of strictly clonal organisms (see also Balloux *et al*. 2003).

Further assuming $m >> u$, we can equally extract the number of migrants as:

$$Nm = \frac{1 - 2F_{ST}}{4F_{ST}} \qquad \text{(eqn 10)}$$

or

$$Nm = -\frac{1 + F_{IS}}{4F_{IS}} \qquad \text{(eqn 11)}$$

From Fig. 1a it can be seen that the rate of clonal reproduction is critical for the accuracy of Nm measure (100% clonality is required), and the number of demes must be large (Fig. 1d), while the measures are much less sensitive to the number of individuals within subpopulations (Fig. 1c). It can also be seen from Fig. 1 that equation 11 (F_{IS}

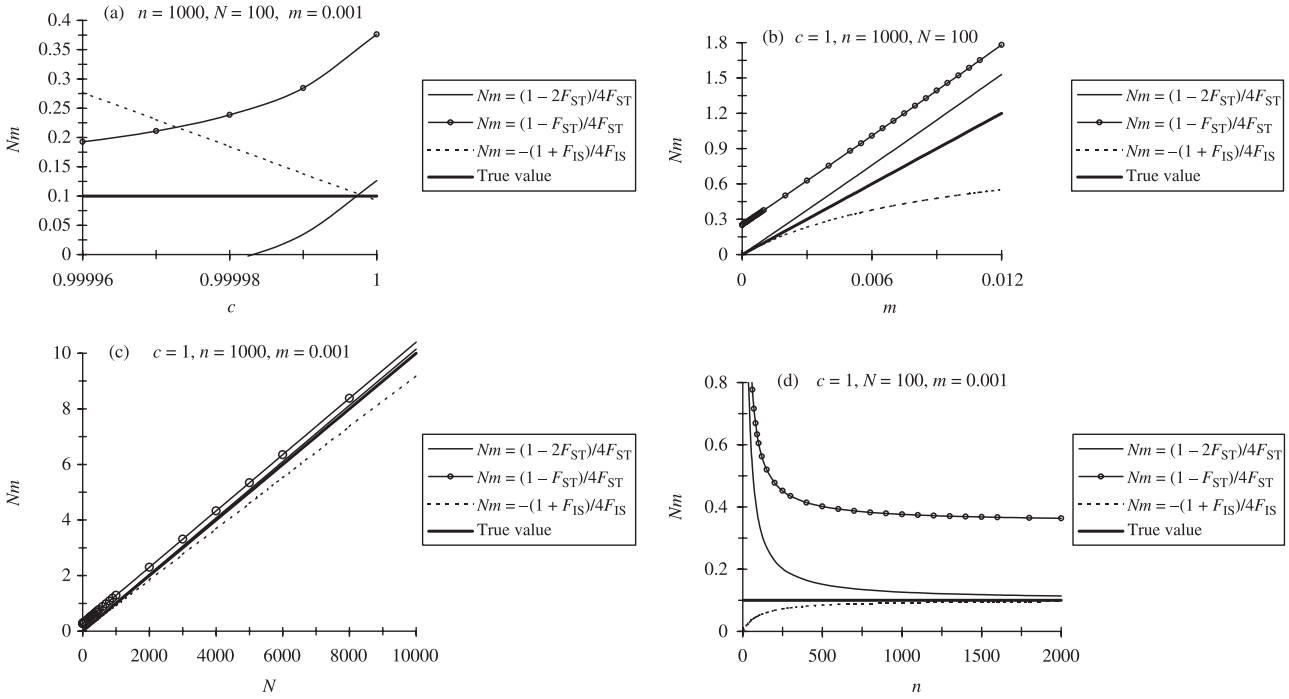

Fig. 1 Accuracy of Nm measures based on F_{ST} (equation 10) or F_{IS} (equation 11) methods with the true value and the classical $(1 - F_{ST})/4F_{ST}$ ($c = 0$) for different values of c (a), different values of m (b), different values of N (c) and different values of n (d). Note that F_{IS} and F_{ST} were themselves analytically computed from the general equations given in Balloux *et al*. (2003) (equations 10 and 14).

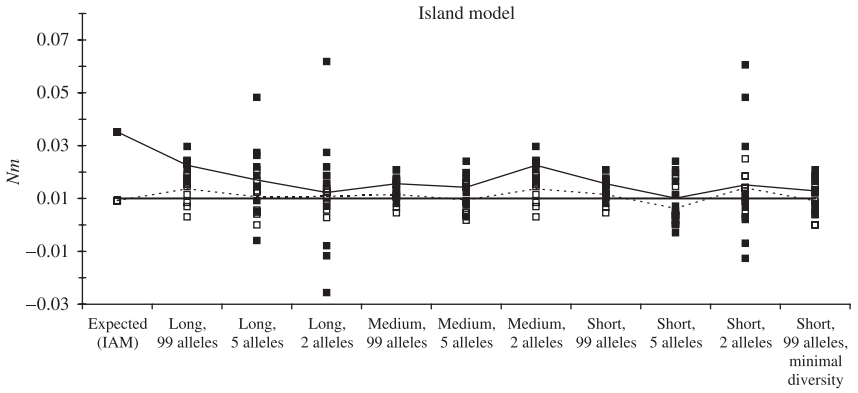

Fig. 2 Accuracy of Nm estimated from the 10 individuals of 50 sampled populations with the F_{ST} (equation 10) (full squares) and the F_{IS} (equation 11) (open squares) methods in our island model simulations (10 replicates per simulation type). Means among the replicates are represented by a thin solid line and a dotted line for F_{ST} and F_{IS} methods respectively. The true value for different simulation parameters (thick straight line) and the equilibrium values expected under the IAM (infinite allele model) are also presented. See Material and methods for further details.

based measure) provides the best Nm measures in most cases except for very high migration rates (Fig. 1b) and imperfect clonal reproduction (Fig. 1a). From Fig. 1d, it can be seen that the F_{IS} based method is much less sensitive (gives less biased Nm measures) than the F_{ST} based method to variation in the number of demes.

Simulations

In Fig. 2 we show that when the underlying assumptions of a large number of subpopulations exchanging a restricted number of migrants are met (here, $n = 1000$ and $m = 0.001$), with $N = 10$ and all 10 individuals sampled from 50 populations, the correspondence between F_{ST} and $-F_{IS}/(1 - F_{IS})$ is satisfactory over the entire parameter range considered. Somewhat surprisingly the fit between F_{ST} and $-F_{IS}/(1 - F_{IS})$ observed in our simulations is better than expected analytically from the general recurrence given in Balloux *et al*. (2003) (which gives slightly lower F_{ST} than in our simulations). This remains true even for simulations far from equilibrium, in particular those starting with minimal diversity, as well as for those with a low number of possible allelic states clearly violating the infinite allele model (IAM) assumptions. Under strict clonal reproduction,

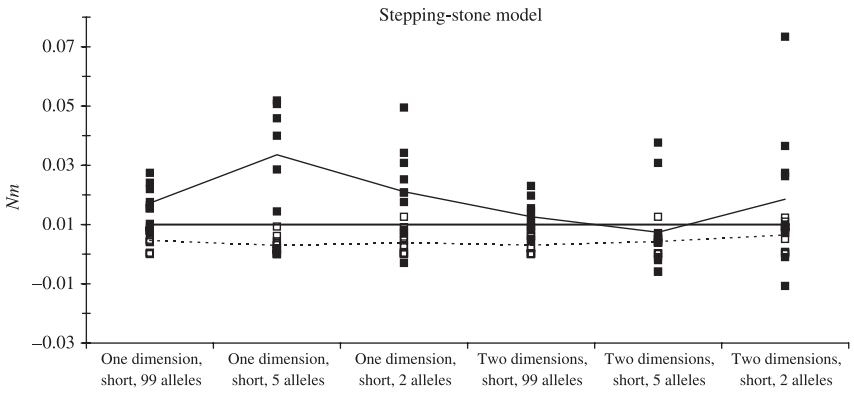

Fig. 3 Accuracy of *Nm* estimated from the 10 individuals of 50 sampled populations with the F_{ST} (equation 10) (full squares) and the F_{IS} (equation 11) (open squares) methods in our stepping-stone model simulations (10 replicates per simulation type). Means among the replicates are represented by a thin solid line and a dotted line for F_{ST} and F_{IS} methods, respectively. The true value for different simulation parameters (thick solid line) and the equilibrium values expected under the IAM (infinite allele model) are also presented. See Material and methods for further details.

Table 1 Results obtained for the three data sets re-analysed. *Nm* was only estimated when relevant. Ranges are the minimum and maximum observed over loci. See text and literature cited for more details

Species and reference	Parameter	Method	Mean	Range
Rhopalosiphum padi (asexuals)	F_{IS}	From article	−0.50	[−0.89,0.43]*
Delmotte *et al*. (2002)	F_{ST}	From article	0.06	[0.00,0.06]*
	F'_{ST}	$-F_{IS}/(1-F_{IS})$	0.33	NA
	Loci	7 microsatellites		
Trypanosoma brucei (all)	F_{IS}	Genotypes re-analysed	−0.17	[−0.31,0.02]
MacLeod *et al*. (2000)	F_{ST}	Genotypes re-analysed	0.23	[−0.02,0.23]
	F'_{ST}	$-F_{IS}/(1-F_{IS})$	0.15	[−0.02,0.17]
Trypanosoma brucei (human)	F_{IS}	Genotypes re-analysed	−0.50	[−0.55,0.45]
MacLeod *et al*. (2000)	F_{ST}	Genotypes re-analysed	0.29	[0.22,0.42]
	F'_{ST}	$-F_{IS}/(1-F_{IS})$	0.34	[0.31,0.36]
	Nm	Equation 11	0.25	[0.20,0.26]
	Loci	3 minisatellites		
Sargatia ornata	H_O	From article	0.47	NA
Shaw *et al*. (1994)	H_S	From article	0.32	NA
	D_S	From article	0.16	NA
	F_{IS}	$(H_O-H_E)/H_E$	−0.47	NA
	F_{ST}	Equation 3	0.24	NA
	F'_{ST}	$-F_{IS}/(1-F_{IS})$	0.32	NA
	Nm	Equation 11	0.28	NA
	Loci	14 allozymes		

NA, not available; *computed without repeated genotypes.

equilibrium is hard to reach and 20 000 generations are not sufficient (Balloux & De Meeûs, unpublished); the number of generations seems however to have limited impact on the accuracy of the estimates. In fact, an analysis of the details of the simulations shows that satisfactory results are generated after about 50 and 200 generations for simulations starting with maximum and minimum diversity respectively. The accuracy of the estimates is only very marginally affected by the number of possible allelic states (homoplasy). In concordance with our analytical results, the most accurate estimates of *Nm* are obtained with the F_{IS}-based method. Interestingly, this conclusion is also rather robust to violation of the island model (i.e. stepping stone) as shown in Fig. 3. The F_{IS} based method underestimates *Nm* both in the one-dimensional or two-dimensional stepping-stone models. The F_{ST}-based method may provide better mean estimates in the two-dimensional stepping-stone model but at the cost of higher variance.

Data sets analysed

The results of the re-analysis of the three available data sets are presented in the Table 1. For *Rhopalosiphum padi* only asexual populations were tested for our criterion. For that

species F_{ST} is very far from $-F_{IS}/(1-F_{IS})$, the variance of F_{IS} across loci seems very large and F_{ST} is very low. For *Trypanosoma brucei* the correspondence between F_{ST} and $-F_{IS}/(1-F_{IS})$ is improved but is still quite unsatisfactory when both animal and human samples are considered together. When only human samples are considered, the correspondence is much improved, with a quite narrow range of F_{IS} across loci and a high level of population differentiation. Also, the range of estimated F_{IT} [(–0.17,0.16)] contains 0. For *Sargatia ornata*, the variation of F_{IS} over loci was not available but the correspondence between estimated F_{ST} and $-F_{IS}/(1-F_{IS})$ is reasonably good, with strong population differentiation (Table 1). Note that Shaw *et al*. (1994) described these populations as displaying 'extreme heterogeneity between samples'.

Discussion

We have shown that any population of diploid organisms reproducing strictly clonally and subdivided as an island model with numerous subpopulations will always converge to the following equilibrium values for *F*-statistics:

$$\begin{cases} F_{IS} = -\dfrac{1}{4Nm+1} \\ F_{ST} = -\dfrac{F_{IS}}{1-F_{IS}} \\ F_{IT} = 0 \end{cases} \qquad \text{(eqn 12)}$$

Clonal diploids subdivided in numerous demes should thus be characterized by an apportionment of genetic variance fitting the last two lines of equation 12, and in such a population, all the information is contained within a single *F*-statistic (F_{IS} or F_{ST}). This conclusion is robust as long as clonality is perfect and migration relatively low. More specifically, it can be shown (by iterative exploration) that the ratio m/n must stay below 10^{-5} for $-1 < R_{DF} < 1$, where $R_{DF} = [-F_{IS}/(1-F_{IS}) - F_{ST}]/F_{ST}$ is the relative difference between the two methods. Note that when $c = 1$, R_{DF} cannot take negative values. If R_{DF} lies outside the range [–1 … +1], the exchangeability between F_{ST} and $-F_{IS}/(1-F_{IS})$ is lost and Nm cannot be correctly measured by the F_{IS} method. When $m/n \gg 10^{-5}$ (say 10^{-3}), a value of $c < 1$ exists for which $F_{ST} = -F_{IS}/(1-F_{IS})$. This c value will always be large, but the corresponding *F*-statistics will also denote the presence of sex (see Balloux *et al*. 2003). For instance, for the extreme parameter values $n = 2$, $m = 0.1$ and $N = 10$, clearly violating the underlying assumptions of a large number of relatively isolated subpopulations, the crossing point would lie at $c = 0.82$. Under these parameters $F_{IS} = -0.08$ and $F_{ST} = 0.075$, which might be consistent with the criterion $F_{ST} = -F_{IS}/(1-F_{IS})$ but their absolute values do not fit those expected for a pure clonal and strongly structured population. In such a case F_{IS} is expected to be strongly negative and F_{ST} should be much higher (see Balloux *et al*. 2003 and equation 9 of the present study). In such clonal populations structured in numerous demes, the estimate of Nm is more accurately derived from F_{IS} than from F_{ST}. This is mostly because F_{IS} is less sensitive than F_{ST} to the structure of the population (Fig. 1b, d). The dependence of F_{IS} on migration rate is in variance to some of our earlier results where we showed F_{IS} to be independent to migration (Balloux *et al*. 2003). The reason behind this discrepancy lies in our previous derivations of *F*-statistics being based on coalescence theory and neglecting mutation, which in turn cancelled out migration terms, which are only retained when mutation is taken into account (Balloux 2004). Another interesting consequence of the pattern of population structure explored in this study is that $F_{IT} = 0$. This would mean that the heterozygote deficit measured on individuals irrespective of the subpopulation they come from would (erroneously) mimic panmixia. The sampling scheme is thus critical.

Our simulations fit well to the analytical results, even in those cases where homoplasy was high and simulations were far from equilibrium. Strong deviation from the island model (one- and two-dimensional stepping-stone designs) only weakly affected the results, suggesting that our approach could be of use for a wide range of empirical data sets. We then applied our results to empirical data from the literature. We limited ourselves to organisms for which an island model of migration might be considered reasonably realistic. The sea anemone *Sargatia ornata*, an obligate clonal diploid, reasonably fits our predictions [extrapolated $F_{ST} = 0.24$ and $-F_{IS}/(1-F_{IS}) = 0.32$] (Shaw *et al*. 1994). This would be consistent with 0.28 migrants per generation exchanged between numerous subpopulations at the scale investigated (southwestern Great Britain). In contrast, the estimates are highly divergent for the aphid *Rhopalosiphum padi*, for which $F_{ST} = 0.06$ and $-F_{IS}/(1-F_{IS}) = 0.33$ (Delmotte *et al*. 2002). Migration rates are probably high in this system, as suggested by the low levels of differentiation between samples. Moreover, the assumption of pure clonality is probably violated (Halkett *et al*. 2005a). The high variance of F_{IS} among loci that can be observed (see table 2 in Delmotte *et al*. 2002) is indeed a criterion for the existence of low rates of sex (see Balloux *et al*. 2003; De Meeûs & Balloux 2004). As a consequence, the clonal lineages of *R. padi* are probably not ancient asexuals but are more likely experiencing cryptic episodes of sex (Halkett *et al*. 2005a, b). In *Trypanosoma brucei* (MacLeod *et al*. 2000), we estimated $F_{ST} = 0.23$ and $-F_{IS}/(1-F_{IS}) = 0.15$ for the complete data set. The fit is better than in the previous example, but far from perfect. This may be due to cryptic sexual reproduction or to an overestimate of F_{IS} due to a Wahlund effect (admixture of *T. brucei brucei* and *T. brucei rhodeziense*). When only human parasites are considered (*T. brucei rhodeziense*) the fit is much improved [F_{ST} =

0.29 and $-F_{IS}/(1 - F_{IS}) = 0.34$]. This strongly suggests that *T. brucei rhodeziense* reproduces strictly clonally (as stated by MacLeod *et al*. 2000) and further suggests that this pathogen is structured in numerous subpopulations (villages) exchanging an equivalent of 0.25 migrants per generations at the scale investigated (Kenya–Uganda–Zambia). The narrowness of F_{IS} range obtained across loci provides some confidence, even though more than three loci would of course be preferable. With more loci (at least five), confidence intervals could be obtained by bootstrapping over loci. Unfortunately, this could not be done for any of the data sets either because the raw data were not available (*S. ornata* and *R. padi*) or only three loci were available (*T. brucei*).

Estimates of the effective number of migrants (Nm) from genetic data should be interpreted with caution (e.g. Whitlock & McCauley 1998), because F_{ST} is not a linear function of Nm, migration is rarely completely random between demes, equilibrium is rarely reached and the sampling variances are usually high. The advantage of the criterion we propose here is that F_{IS} is less sensitive to the structure of the population and sampling than F_{ST}. Furthermore, it can be repeatedly and independently estimated from different subpopulations (and not only from different loci). Within strongly structured metapopulations F_{IS} will be nearly linearly dependent on Nm, and under such circumstances F_{IT} is always expected to lie around 0 if individuals reproduce strictly clonally. The neutral assumption is also important because in asexuals all loci are physically linked and thus all affected by selective events (hitch-hiking). Selection is however, more likely to affect F_{ST} than F_{IS}, and will mainly generate nonequilibrium situations, a problem to which our method does not seem too sensitive. Nevertheless, conclusions should not rely solely on this criterion we propose, particularly when $F_{ST} \neq -F_{IS}/(1 - F_{IS})$. Possible patterns suggestive of natural selection should be checked for (for a discussion dealing with those issues, see Barraclough *et al*. 2003). The use of sufficient replication, both in terms of the number of subpopulations and loci analysed, will again be critical.

Precise recommendations on optimal sampling design is difficult at this stage and will anyway always be constrained by the means available, but we would recommend a minimum of 7–10 loci and 20 subsamples of 10 individuals each. Such a sampling strategy would also allow evaluating confidence intervals over loci and populations, which would be a great asset. Unfortunately this was not possible with the three data sets we could retrieve from the literature. Nevertheless, our analytical and simulation results, as well as those obtained from the three empirical data sets that were not designed for our purpose, suggest our criterion could be useful for the analysis of clonal populations, at least in conjunction with other criteria (see De Meeûs & Balloux 2004).

Acknowledgements

We would like to thank D. Roze for very useful discussions, C. Chevillon for a critical reading of an early manuscript, A. Porter and one anonymous referee for the quality of their suggestions that considerably helped improve the present manuscript. TDM is supported by the CNRS and IRD. FB acknowledges support from the BBSRC.

References

Balloux F (2001) EASYPOP (version 1.7): a computer program for population genetics simulations. *Journal of Heredity*, **92**, 301–302.

Balloux F (2004) Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution*, **58**, 1891–1900.

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

Barraclough TG, Birky CW, Burt A (2003) Diversification in sexual and asexual organisms. *Evolution*, **57**, 2166–2172.

Bengtsson BO (2003) Genetic variation in organisms with asexual reproduction. *Journal of Evolutionary Biology*, **16**, 189–199.

Berg LM, Lascoux M (2000) Neutral genetic differentiation in an island model with cyclical parthenogenesis. *Journal of Evolutionary Biology*, **13**, 488–494.

Birky CW Jr (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics*, **144**, 427–437.

Ceplitis A (2003) Coalescence times and the Meselson effect in asexual eukaryotes. *Genetical Research, Cambridge*, **82**, 183–190.

Cockerham CC (1969) Variance of gene frequencies. *Evolution*, **23**, 72–84.

Cockerham CC (1973) Analysis of gene frequencies. *Genetics*, **74**, 679–700.

De Meeûs T, Balloux F (2004) Clonal reproduction and linkage disequilibrium in diploids: a simulation study. *Infection, Genetics and Evolution*, **4**, 345–351.

Delmotte F, Leterme N, Gauthier JP, Rispe C, Simon JC (2002) Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, **11**, 711–723.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Halkett F, Plantegenest M, Prunier-Leterme N *et al*. (2005a) Admixed sexual and facultatively asexual aphid lineages at mating sites. *Molecular Ecology*, **14**, 325–336.

Halkett F, Simon JC, Balloux F (2005b) Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, **20**, 194–201.

MacLeod A, Tweedie A, Welburn SC *et al*. (2000) Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proceedings of the National Academy of Sciences, USA*, **97**, 13442–13447.

Nagylaki T (1983) The robustness of neutral models of geographical variation. *Theoretical Population Biology*, **24**, 268–294.

Nagylaki T (1998) Fixation indices in subdivided populations. *Genetics*, **148**, 1325–1332.

Nei M (1972) Genetic distance between populations. *American Naturalist*, **106**, 283–292.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Shaw P, Ryland JS, Beardmore JA (1994) Population genetic parameters within a sea anemone family (Sagartiidae) encompassing clonal, semiclonal and aclonal modes of reproduction. In: *Genetics and Evolution of Aquatic Organisms* (ed. Beaumont AR), pp. 351–358. Chapman & Hall, London.

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

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

Whitlock MC, McCauley DE (1998) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

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

Thierry de Meeûs is a researcher at the Centre National pour la Recherche Scientifique (CNRS) in Montpellier. His main fields of interest are the population genetics and the evolutionary ecology of host–parasite systems. See more details at http://gemi.mpl.ird.fr/cepm/SiteWebESS/GB/deMeeus/TdeMeeusGB.html. François Balloux is an assistant professor at Cambridge University. He is interested in various questions relating to population genetics. More details can be found at http://www.gen.cam.ac.uk/newdept/research/labs/balloux.htm

---