The Population Genetics of Clonal and Partially Clonal Diploids

François Balloux,*,†,[1] Laurent Lehmann[‡] and Thierry de Meeûs[§]

**I.C.A.P.B., University of Edinburgh, Edinburgh EH9 3JT, Scotland, United Kingdom, †Department of Genetics, University of Cambridge, Cambridge CB2 3EH, United Kingdom, ‡Institute of Ecology (Zoology and Animal Ecology), University of Lausanne, 1015 Lausanne, Switzerland and §Centre d'Etude du Polymorphisme des Microorganismes, Equipe ESS, UMR 9926 CNRS-IRD, BP64501, 34394 Montpellier Cedex 5, France*

Manuscript received November 6, 2002
Accepted for publication April 11, 2003

ABSTRACT

The consequences of variable rates of clonal reproduction on the population genetics of neutral markers are explored in diploid organisms within a subdivided population (island model). We use both analytical and stochastic simulation approaches. High rates of clonal reproduction will positively affect heterozygosity. As a consequence, nearly twice as many alleles per locus can be maintained and population differentiation estimated as F_{ST} value is strongly decreased in purely clonal populations as compared to purely sexual ones. With increasing clonal reproduction, effective population size first slowly increases and then points toward extreme values when the reproductive system tends toward strict clonality. This reflects the fact that polymorphism is protected within individuals due to fixed heterozygosity. Contrarily, genotypic diversity smoothly decreases with increasing rates of clonal reproduction. Asexual populations thus maintain higher genetic diversity at each single locus but a lower number of different genotypes. Mixed clonal/sexual reproduction is nearly indistinguishable from strict sexual reproduction as long as the proportion of clonal reproduction is not strongly predominant for all quantities investigated, except for genotypic diversities (both at individual loci and over multiple loci).

THE essential feature of sexual reproduction is that genetic material from different ancestors is brought together in a single individual. If sexual reproduction is dominant in eukaryotic organisms (*e.g.*, Charlesworth 1989; West *et al.* 1999), many organisms of major medical or economical importance are known to reproduce mainly or strictly clonally (*e.g.*, Milgroom 1996; Taylor *et al.* 1999; Tibayrenc 1999). The presence or absence of a sexual process will crucially determine the genetics at both the individual and the population level and leads to several straightforward predictions. At the individual level, clonality will produce a strong correlation between alleles within individuals at different loci, as they share a common history within a clonal lineage. Sex on the other hand will break these associations, allowing for many more potential genetic combinations. Further, in diploids, absence of sex will promote divergence between alleles within loci, as the two copies will accumulate different mutations over time. This effect has been termed the "Meselson effect" and has recently been experimentally documented in bdelloid rotifers, which are believed to have been reproducing strictly clonally over long evolutionary time (Butlin 2000; Mark Welch and Meselson 2000, 2001). Heterozygosity is thus expected to increase indefinitely under clonal propagation (Birky 1996; Judson and Normak 1996). In another respect, theoretical considerations predict that the effective population size of clonal organisms should be lower than that of panmictic ones (*e.g.*, Orive 1993; Milgroom 1996). However, the few theoretical population genetics studies that we are aware of provide ambiguous conclusions on that topic (Orive 1993; Berg and Lascoux 2000) and numerous field observations support this ambiguity (*e.g.*, Butlin *et al.* 1998; Gabrielsen and Brochmann 1998; Cywinska and Hebert 2002). Thus, "whether organisms with clonal reproduction necessarily have lower genetic diversity is unclear" (Orive 1993, p. 337). These ambiguities illustrate what little is known on the population genetics consequences of clonal reproduction. In the absence of theoretical models providing clear expectations, estimating the rate of clonal reproduction in natural populations appears problematic (*e.g.*, Anderson and Kohn 1998) and even the detection of purely clonal populations is often controversial (*e.g.*, Tibayrenc 1997; Vigalys *et al.* 1997). Clonality is not just an academic matter (Tibayrenc 1997). Many diploid organisms believed to reproduce mainly or strictly clonally are of major medical, veterinary, and economical importance, including pathogenic fungi such as Candida or protozoans such as Trypanosoma. A better understanding of the reproductive system of such organisms might be crucial for planning successful long-term drug administration or vaccination programs (Tibayrenc *et al.* 1991; Milgroom 1996; Taylor *et al.* 1999).

Here we present both analytical and stochastic simula-

[1]*Corresponding author:* Department of Genetics, University of Cambridge, Downing St., Cambridge CB2 3EH, United Kingdom. E-mail: fb255@mole.bio.cam.ac.uk

tion results for the population genetics of clonally and partially clonally reproducing populations. We focus on a simple population subdivision model (island model) and restrict our work to neutral mutations. We derive the identities by descent, *F*-statistics, and mean coalescence times of alleles and genotypes for variable rates of clonal reproduction. We also investigate the allelic and genotypic diversities maintained under different rates of clonal reproduction.

MODEL ASSUMPTIONS AND GENETIC IDENTITIES

We consider a subdivided monoecious population of diploid individuals, which reproduce clonally with probability c, with sexual reproduction occurring at the complementary probability $(1 - c)$. Sexual reproduction in the model follows random union of gametes, self-fertilization occurs at a rate s, and a subpopulation is composed of N number of adults. In our model, individuals, rather than gametes, migrate following an island model (Wright 1951) at a rate m, implying that a migrant has an equal probability to reach any of the subpopulations. We further assume stable census sizes and population structure and no selection. The life cycle involves nonoverlapping generations and juvenile migration. The precise sequence goes as follows:

1. Adult reproduction and subsequent death
2. Juvenile dispersal
3. Regulation of juveniles, the survivors reaching adulthood

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

The identities may be calculated in juveniles (F_J, θ_J, α_J), or adults (F_A, θ_A, α_A), or respectively before or after migration. In a first step, we express identities between adults one generation forward in time $(t + 1)$ as functions of juvenile identities $(t + \frac{1}{2})$. Adult identities are affected only by dispersal,

$$\begin{aligned} F_{A(t+1)} &= F_{J(t+1/2)} \\ \theta_{A(t+1)} &= (q_s\theta_{J(t+1/2)} + (1 - q_s)\alpha_{J(t+1/2)}) \\ \alpha_{A(t+1)} &= (q_d\theta_{J(t+1/2)} + (1 - q_d)\alpha_{J(t+1/2)}), \end{aligned} \tag{1}$$

with q_s being the probabilities that two individuals taken at random within the same subpopulation after migration were born in the same deme. The exact expression for q_s is relatively cumbersome (see Wang 1997). However, for relatively large values of N, q_s reduces to the much more compact form that we use throughout the article,

$$q_s \cong (1 - m)^2 + \frac{(m)^2}{n - 1}, \tag{2}$$

where m represents the migration rate and n the number of subpopulations. Now we can define q_d as the probability that two individuals sampled after migration in different subpopulations originated from the same deme:

$$q_d = \frac{1 - q_s}{n - 1}. \tag{3}$$

We then express juvenile identities as functions of adult identities in the previous generation. Here both mutation and the reproductive system will affect the genetic identities of juveniles. The mutation rate is u for all alleles and therefore the probability of two alleles that are identical by descent before mutation still being identical after mutation will be $\gamma \equiv (1 - u)^2$. In the absence of any mutation event, clonal reproduction occurring at rate c will produce offspring identical to its progenitor, so that the inbreeding coefficient of a clonally produced juvenile individual will be identical to its parent's. Selfing occurs with probability s, and in that case the coancestry will be $(1 + F_A)/2$. With a probability $1 - s$, nonselfing sexual reproduction occurs, the offspring will have two parents, and its inbreeding will be the parental coancestry (θ_A). This gives us the following juvenile identities as functions of adult identities:

$$\begin{aligned} F_{J(t+1/2)} &= \gamma\left(cF_{A(t)} + (1 - c)\left(s\left(\frac{1 + F_{A(t)}}{2}\right) + (1 - s)\theta_{A(t)}\right)\right) \\ \theta_{J(t+1/2)} &= \gamma\left(\frac{1}{N}\left(\frac{1 + F_{A(t)}}{2}\right) + \left(1 - \frac{1}{N}\right)\theta_{A(t)}\right) \\ \alpha_{J(t+1/2)} &= \gamma(\alpha_{A(t)}). \end{aligned} \tag{4}$$

Substituting Equation 4 in (1), we obtain the recurrence equations for describing the dynamics of identities among adults:

$$\begin{aligned} F_{A(t+1)} &= \gamma\left(cF_{A(t)} + (1 - c)\left(s\left(\frac{1 + F_{A(t)}}{2}\right) + (1 - s)\theta_{A(t)}\right)\right) \\ \theta_{A(t+1)} &= \gamma\left(q_s\left(\frac{1}{N}\left(\frac{1 + F_{A(t)}}{2}\right) + \left(1 - \frac{1}{N}\right)\theta_{A(t)}\right) + (1 - q_s)\alpha_{A(t)}\right) \\ \alpha_{A(t+1)} &= \gamma\left(q_d\left(\frac{1}{N}\left(\frac{1 + F_{A(t)}}{2}\right) + \left(1 - \frac{1}{N}\right)\theta_{A(t)}\right) + (1 - q_d)\alpha_{A(t)}\right). \end{aligned} \tag{5}$$

The recurrence equations for θ_A and α_A are identical to those given by Rousset (1996, Equation 2). Only F_A is affected by the variable amount of clonal reproduction and by the fact that we assume zygotic rather than gametic migration.

For analytical effectiveness, recurrence equations for identities by descent can be presented in matrix form,

$$\mathbf{Q}_{(t+1)} = \gamma\mathbf{G}\mathbf{Q}_t + \gamma\mathbf{D}, \quad (6)$$

where $\mathbf{Q}_t$ is a column vector of the probabilities of identities at generation $t + 1$. The transition matrix $\mathbf{G}$ defines the probabilistic changes of the vector variables, and $\mathbf{D}$ is the constant column vector. Solving (6) at equilibrium we obtain the identities from

$$\mathbf{Q} = \gamma(\mathbf{I} - \gamma\mathbf{G})^{-1}\mathbf{D} \quad (7)$$

with $\mathbf{I}$ being the identity matrix.

INDIVIDUAL-BASED SIMULATIONS

To obtain the variances of the quantities of interest, as well as multilocus behavior, we additionally performed stochastic individual-based simulation, as implemented in the software EASYPOP (version 1.7.4; Balloux 2001). 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 possible allelic states. This relatively high number of allelic states keeps the probability of obtaining indistinguishable alleles through different mutational events (homoplasy) low. At the start of the simulation, genetic diversity was set to the maximum possible value at the first generation and the simulation was then run for 10,000 generations, the point at which all statistics measured in EASYPOP (F_{IS}, F_{ST}, H_S, H_T, and the number of alleles) had reached equilibrium. All simulations were replicated 20 times.

F-STATISTICS

Deviations from random mating are generally expressed by means of *F*-statistics (Wright 1951). They are the most commonly used tools for describing gene flow and breeding structure in both theoretical and empirical studies (reviewed in Balloux and Lugon-Moulin 2002). *F*-statistics are defined as

$$F_{IS} = \frac{F - \theta}{1 - \theta}, \quad F_{ST} = \frac{\theta - \alpha}{1 - \alpha}, \quad F_{IT} = \frac{F - \alpha}{1 - \alpha} \quad (8)$$

(Cockerham 1969, 1973), where subscripts I, S, and T represent individuals, subpopulations, and the total population, respectively. F_{IS} can be thought of as a measure of the identity of alleles within individuals relative to the identity between alleles randomly drawn from two different individuals from within the same subpopulation. F_{ST} is the identity of alleles drawn randomly from within a subpopulation relative to alleles drawn from the entire population. F_{IT} is the identity of alleles within individuals relative to randomly drawn gametes from the entire population. In more biological terms F_{IS} is interpreted in terms of deviation from random mating, caused by the breeding system of the organism under study, and F_{ST} represents the heterozygote deficiency due to population subdivision. Finally F_{IT} is the measure of inbreeding taking into account both deviations from random mating within subpopulations and the effects of population subdivision. The relation linking the three coefficients can be expressed as

$$(1 - F_{IT}) = (1 - F_{ST})(1 - F_{IS}) \quad (9)$$

(Crow and Kimura 1970, p. 106).

Within-population deviations from random mating (F_{IS}): Replacing the solutions of Equation 7 in (8), we get F_{IS} after migration for subdivided populations with a mixed system of clonal and sexual reproduction (selfing set to $1/N$) and zygotic migration

$$F_{IS} = \frac{\gamma(q_s - c(\gamma(q_s - q_d) - 1) - 1)}{2N(1 - \gamma c)(\gamma(q_s - q_d) - 1) - \gamma(q_s - c(\gamma(q_s - q_d) - 1) - 1)}. \quad (10)$$

Neglecting mutation ($\gamma = 1$), but allowing for a mixed system of clonal reproduction with arbitrary selfing rate, we obtain

$$F_{IS} = \frac{Nn(1 - c)s - 1}{Nn(1 - c)(2 - s) + 1}. \quad (11)$$

The equation shows that F_{IS} is independent of the migration rate but sensitive to the total number of individuals in the population; this occurs because we assumed zygotic rather than gametic migration. Under random mating ($s = 1/N$) we further obtain

$$F_{IS} = \frac{(1 - c)n - 1}{(1 - c)(2N - 1)n + 1}. \quad (12)$$

When reproduction is strictly sexual ($c = 0$), Equation 12 reduces to the form

$$F_{IS} = \frac{n - 1}{(2N - 1)n + 1}. \quad (13)$$

For a strictly clonal population ($c = 1$), $F_{IS} = -1$. This reflects the fact that in the absence of sexual reproduction, all individuals are expected to be heterozygous at equilibrium $F = 0$, while $\theta = \frac{1}{2}$.

In Figure 1, we plot F_{IS} as obtained from Equation 10 against the rate of clonal reproduction. We also give values obtained from individual-based simulations. Analytical and stochastic simulation results are in excellent agreement. From Figure 1, it can be seen that for very high values of clonal reproduction, huge heterozygote excesses are obtained. However, as long as there is a small proportion of sexual reproduction, F_{IS} stays close to what is expected under panmixia; a significant excess of heterozygotes occurs only for extreme rates of asexuality. As long as there is mutation in the system, F_{IS} cannot reach -1 even for strict clonality. If the product

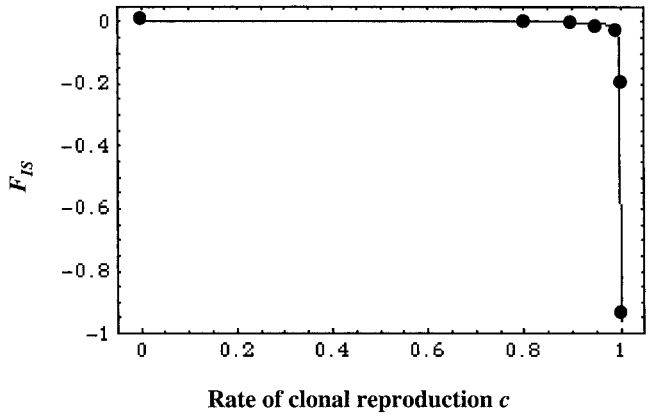

FIGURE 1.—F_{IS} as a function of the rate of clonal reproduction. Parameter values are as follows: number of populations, $n = 50$; number of individuals per population, $N = 50$; migration rate, $m = 0.1$; mutation rate, $u = 10^{-5}$; selfing rate is set to random mating, $s = 1/N$. The line represents analytical results and the solid circles simulation results.

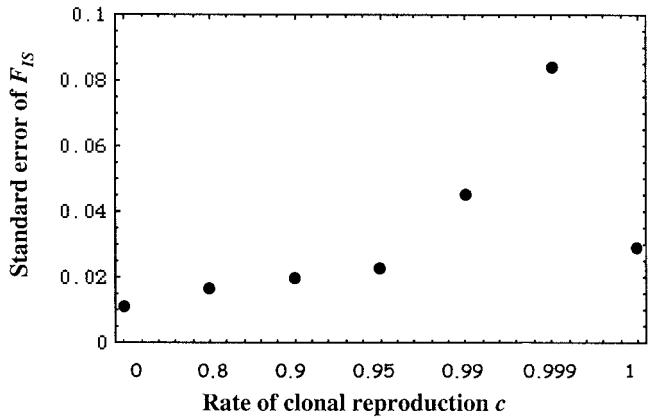

FIGURE 2.—Standard errors of F_{IS} as a function of the rate of clonal reproduction. The standard errors are computed on 20 physically unlinked loci. The simulation parameter values are as follows: number of populations, $n = 50$; number of individuals per population, $N = 50$; migration rate, $m = 0.1$; mutation rate, $u = 10^{-5}$; selfing rate is set to random mating, $s = 1/N$.

of the number of individuals in the complete population (nN) times the mutation rate is high, the F_{IS} value for complete clonality can be very much offset from -1. The reason for this can be seen from Equation 8. Under clonal reproduction all individuals will be heterozygous and this will not be changed by mutation, so $F = 0$, while θ decreases with increasing mutation rate.

The F_{IS} estimates from the stochastic simulations in Figure 1 are averaged over loci and replicates and do not reveal anything about the strong influence of the rate of clonal reproduction on the variance over loci. This huge variation among loci, in particular for low rates of sexual reproduction, is illustrated by standard errors in F_{IS} (Figure 2). The lowest variations are obtained with pure clonality and with <95% of clonality.

Population differentiation (F_{ST}): Again by replacing the solutions of Equation 7 in (8), we obtain F_{ST} for subdivided populations with a mixed system of clonal, selfing, and sexual reproduction after migration:

$$F_{ST} = \frac{\gamma(1 - c\gamma)(q_s - q_d)}{N(1 - \gamma(q_s - q_d))(2 + c(s - 2)\gamma - s\gamma) + \gamma(q_d(\gamma(c + 1) - 1) - q_s(\gamma(c + 1) - 2))}. \quad (14)$$

Neglecting mutation ($\gamma = 1$) leads to

$$F_{ST} = \frac{(1 - c)(q_s - q_d)}{N(1 - c)(2 - s)(1 - (q_s - q_d)) + c(q_d - q_s) + q_s}. \quad (15)$$

Finally, if only sexual reproduction is allowed ($c = 0$ and $s = 1/N$), we get

$$F_{ST} = \frac{(q_s - q_d)}{(2N - 1)(1 - (q_s - q_d)) + q_s}. \quad (16)$$

In Figure 3, we plot F_{ST} as obtained from Equation 14 against the proportion of clonal reproduction, as well as values obtained from the individual-based simulations.

The amount of clonal reproduction has a strong effect on population differentiation. Whereas even for very limited proportions of sex, there is no noticeable effect, when reproduction tends toward strict clonality, F_{ST} is strongly reduced. Note that in the absence of any mutation, F_{ST} would be defined but equal to 0, as all the genetic variance is within individuals and none between individuals and subpopulations. In all simulated cases the between-loci variance of F_{ST} strongly increases with the proportion of clonal reproduction (results not shown).

EFFECTIVE POPULATION SIZE

Effective population size: The effective population size (WRIGHT 1931) is the parameter summarizing the amount of genetic drift to which a population is subjected. It is quantified as the number of idealized randomly mating individuals that experience the same amount of random fluctuations at a neutral locus as the population under scrutiny. The dynamics of idealized randomly mating individuals are described by the Wright-Fisher model, whose well-known properties lead to different definitions of the effective population size depending on whether the quantities of interest are the variance of change in allelic frequencies, inbreeding coefficients, or the rate of decline in heterozygosity (EWENS 1982; WHITLOCK and BARTON 1997). Here we introduce a new definition of effective size called the coalescence effective size,

$$N_e = \frac{1}{2}\bar{t}, \quad (17)$$

where $\bar{t}$ is the expected time it takes for two randomly sampled alleles in a population to coalesce to a common

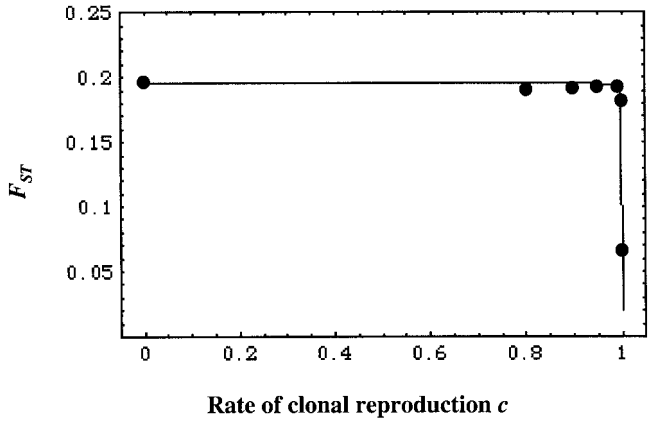

FIGURE 3.—F_{ST} as a function of the rate of clonal reproduction. Parameter values are as follows: number of populations, $n = 50$; number of individuals per population, $N = 100$; migration rate, $m = 10^{-2}$; mutation rate, $u = 10^{-5}$; selfing rate is set to random mating, $s = 1/N$. The line represents analytical results and the solid circles simulation results.

ancestor. For the Wright-Fisher model $\bar{t} = 2N$, so that the effective size reduces to the actual number of diploid individuals. This definition of effective size allows us to disentangle the allelic effective size (all classical definitions) from the genotypic effective size (see below), and we can further obtain their variance.

There is a strict relationship between identity-by-descent probabilities and coalescence times (SLATKIN 1991; ROUSSET 1996). The probability of identity of any pair of alleles is the probability that neither allele has undergone mutation since their most recent common ancestor (HUDSON 1990). Recalling Equation 7,

$$\mathbf{Q} = \gamma(\mathbf{I} - \gamma\mathbf{G})^{-1}\mathbf{D}. \tag{18}$$

The matrix $\mathbf{G}$ is diagonalizable for $c \neq 1$. We can represent the vector $\mathbf{D}$ on the basis of the right eigenvectors of the matrix $\mathbf{G}$ as $\mathbf{D} = \Sigma_j a_j r_j$, where j is the number of columns of $\mathbf{G}$, a_j the coefficient determined by the preceding system of equations, and $\mathbf{r}_j \equiv (r_{1j}, \ldots, r_{kj})^T$ the jth right eigenvector of $\mathbf{G}$. Using the fact that the jth eigenvalue of the matrix $(\mathbf{I} - \gamma\mathbf{G})^{-1}$ is $1/(1 - \gamma\lambda_j)$, and its associated right eigenvector is $\mathbf{r}_j$, where λ_j is the jth eigenvalue of $\mathbf{G}$, we can express Equation 18 following ROUSSET (2002) as

$$\mathbf{Q} = \sum_j \frac{\gamma a_j r_j}{1 - \gamma\lambda_j} = \sum_{t=1}^{\infty} \gamma^t \sum_j \lambda_j^{t-1} a_j r_j. \tag{19}$$

The second equality is obtained by using the property of geometric series. Then

$$c_i(t) = \sum_j \lambda_j^{t-1} a_j r_{ij} \tag{20}$$

is the coalescence probability of alleles at time t at any hierarchical level i (where i stands for F, θ, and α). From this, we can obtain the expected coalescence times by classical tools. However, after substituting Equation 20 into Equation 19, a closer look reveals that the vector $\mathbf{Q}$ defines the probability-generating functions of coalescence time at each level i. These functions reduce to the calculations of expected coalescence times as

$$\bar{t}_i = \left.\frac{\partial Q_j}{\partial \gamma}\right|_{\gamma=1}, \tag{21}$$

where $\bar{t}_i$ is the expected coalescence time at level i and Q_j is the ith row of the equilibrium vector given by Equation 20 and their variances as

$$\sigma^2(\bar{t}_i) = \left[\frac{\partial^2 Q_j}{\partial \gamma^2} + \frac{\partial Q_j}{\partial \gamma} - \left(\frac{\partial Q_j}{\partial \gamma}\right)^2\right]_{\gamma=1}. \tag{22}$$

At this point we have all necessary tools to obtain the mean coalescence times of alleles in a subdivided population with arbitrary rates of clonal reproduction. Writing the recurrence equations (5) under the form given in Equation 18 and using Equation 21 yields the following mean coalescence times,

$$\bar{t}_F = \frac{2(1 + (1 - c)nN(1 - s))}{1 - c}$$

$$\bar{t}_\theta = \frac{1 + (1 - c)nN(2 - s)}{1 - c}$$

$$\bar{t}_\alpha = \frac{n - 1}{2m} + \frac{4 + (1 - c)n(3 + 4n(2 - s))}{4(1 - c)} + \frac{n^2 m}{8(n - 1)}, \tag{23}$$

where $\bar{t}_\alpha$ is a low migration limit obtained by a first-degree Taylor expansion. The mean coalescence time of two randomly sampled alleles is the expectation of the $\bar{t}_i$; in the finite-island model this yields

$$\bar{t} = \frac{1}{n}\left(\frac{1}{N}\bar{t}_F + \left(1 - \frac{1}{N}\right)\bar{t}_\theta\right) + \left(1 - \frac{1}{n}\right)\bar{t}_\alpha. \tag{24}$$

Substituting Equation 24 into (17), we obtain the coalescence effective population size. Note that this effective size captures the loss of allelic diversity in the population and we refer to it as $2N_e$, the *allelic effective population size*, which is equal to $\bar{t}$ in our model. In Figure 4, we plot the effective population size as a function of clonal reproduction and selfing rate (the union of gametes within individuals). Increasing the rate of clonal reproduction has no noticeable effect on most of the parameter space. However, when the reproductive system tends toward complete asexual reproduction, the effective population size suddenly tends toward infinite values. This slightly counter-intuitive result simply reflects that the genetic diversity within individuals cannot be lost in clonal organisms. Doubling of N_e compared to random mating is observed approximately when the rate of sexual reproduction is in the order of 10^{-4} with the simulation parameters used in Figure 4. Contrarily, increased rates of selfing decrease effective population size. This effect is linear and the effective population size ranges

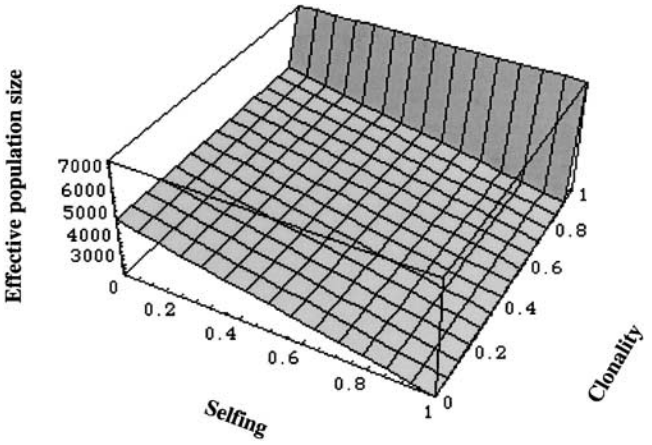

FIGURE 4.—Effective population size as a function of selfing (front axis) and clonal reproduction (right axis). Parameter values are as follows: number of populations, $n = 50$; number of individuals per population, $N = 100$; migration rate, $m = 0.1$.

between N_e under absence of selfing to $N_e/2$ for strict selfing.

Genotypic and allelic effective population size: We have shown that increased rates of clonal reproduction will increase the allelic effective population size, and thus clonal populations are expected to maintain more alleles at neutral loci than are sexually reproducing ones. We can go a step further and address the issue of how clonal reproduction will affect the number of different genotypes maintained. The coalescence approach allows us to capture qualitatively these trends by calculating the *genotypic effective population size*. To obtain this quantity, we need, in addition to F and θ, ρ, the probabilities that three alleles randomly sampled in two different individuals are identical. These three variables are necessary to calculate the probability Δ that two genotypes are identical. However, these higher-order coefficients are complicated and we therefore limit ourselves to a non-subdivided monoecious population without mutation. We follow the approach of COCKERHAM (1971, pp. 243–244) to calculate the dynamics of these four variables. Collecting the identities given in APPENDIX A leads to the following system of recurrence equations:

$$F_{(t+1)} = cF_{(t)} + (1-c)\left(\frac{1}{N}\left(\frac{1+F_{(t)}}{2}\right) + \left(1-\frac{1}{N}\right)\theta_{(t)}\right)$$

$$\theta_{(t+1)} = \frac{1}{N}\left(\frac{1+F_{(t)}}{2}\right) + \left(1-\frac{1}{N}\right)\theta_{(t)}$$

$$\rho_{(t+1)} = c\left(\frac{1}{N}F_{(t)} + \frac{N-1}{N}\rho_{(t)}\right) + (1-c)\left(\frac{1}{N^2}\left(\frac{1}{4}+\frac{3}{4}F_{(t)}\right) + \frac{3(N-1)}{2N^2}\theta_{(t)} + \frac{(2N-1)(2N-2)}{4N^2}\rho_{(t)}\right)$$

$$\Delta_{(t+1)} = c^2\left(\frac{1}{N} + \left(1-\frac{1}{N}\right)\Delta_{(t)}\right) + 2c(1-c)\left(\frac{1}{N^2}\left(\frac{1}{2}+\frac{1}{2}F_{(t)}\right) + \frac{3(N-1)}{N^2}\left(\frac{1}{2}\theta_{(t)} + \frac{1}{4}\rho_{(t)} + \frac{1}{4}\Delta_{(t)}\right) + \frac{(N-1)(N-2)}{N^2}\Delta_{(t)}\right) + (1-c)^2\left(\frac{1}{4N^2} + \frac{3}{4N^3}F_{(t)} + \frac{2((2N)^2-1)}{(2N)^3}\theta_{(t)} + \frac{4(2N-1)(2N-2)}{(2N)^3}\rho_{(t)} + \frac{(2N-1)(2N-2)(2N-3)}{(2N)^3}\Delta_{(t)}\right). \quad (25)$$

Note that when $c = 0$, $F = \theta$, and the recurrence equations reduce to COCKERHAM's (1971) model. Substituting these equations into a transition matrix **G** and a column vector of constants **D** following Equation 18 allows us to obtain the mean coalescence times:

$$\bar{t}_F = \frac{2(c + N(1-c))}{1-c}$$

$$\bar{t}_\theta = \frac{c + 2N(1-c)}{1-c}$$

$$\bar{t}_\rho = \frac{N(8N + 2c^2(N-1) + c(9-10N) - 3)}{(1-c)(3N - c(N-1) - 1)}$$

$$\bar{t}_\Delta = \begin{cases} \dfrac{c^3a_0 + c^2a_1 + ca_2 + a_3}{c^3b_0 + c^2b_1 + cb_2 + b_3}, & c \neq 1 \\ N, & c = 1. \end{cases} \quad (26)$$

$\bar{t}_\Delta$ is the mean coalescence time for genotypes and we refer to it as the genotypic effective size. This quantity is undefined for $c = 1$ as the matrix **G** is not diagonalizable in this case. However, we know that when $c = 1$, the genotypic identity is independent of F, θ, and ρ and thus follows a dynamic similar to haploid genetics, with mean coalescence time N. When $c \neq 1$, the mean genotypic coalescence time $\bar{t}_\Delta$ takes the form of a relatively complex polynomial (the coefficients are given in APPENDIX B). Note that in the absence of clonal reproduction, there is a very compact approximation for the genotypic coalescence time $\bar{t}_\Delta \approx 3N$ (see APPENDIX B). Mean coalescence time for alleles in a nonsubdivided population can be obtained as $\bar{t} = (1/N)\bar{t}_F + (1 - 1/N)\bar{t}_\theta$ and thus reads

$$\bar{t} = \frac{c(1+N) + 2N^2(1-c)}{N(1-c)}. \quad (27)$$

Note that we could have obtained the allelic effective size directly from Equation 24 assuming no migration, one subpopulation, and a selfing rate of $1/N$. In Figure 5, we give mean coalescence times for both alleles and genotypes. It can be seen that contrarily to what is observed at the allelic level, genotypic effective size decreases with increasing clonality. The decrease is relatively smooth over the complete parameter range of c and reaches N for strict clonality. The intuitive reason behind this is that when there is no segregation at all, the two alleles within a diploid individual behave as a single haploid locus. The rate of clonal reproduction has thus an antagonistic effect on the variability of alleles and genotypes.

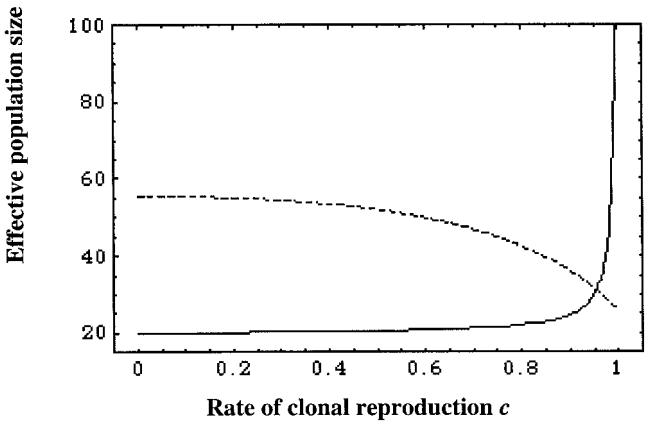

FIGURE 5.—Allelic effective population size (solid line) and genotypic effective population size (dashed line) in a nonsubdivided population. The values given are for a population of 20 individuals.

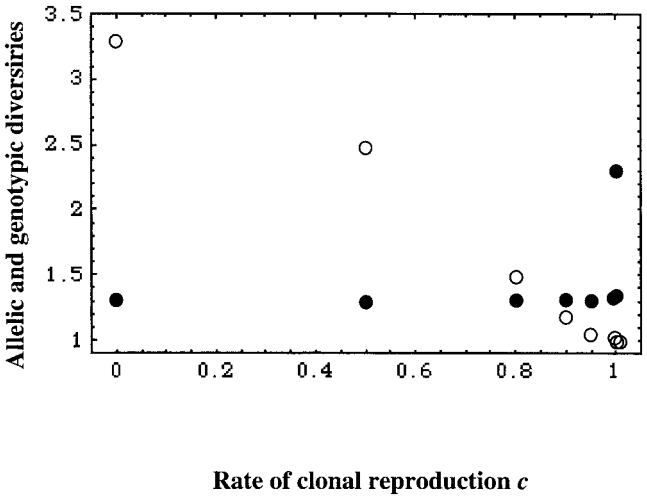

FIGURE 6.—Effective number of alleles (solid circles) and genotypes (open circles) as functions of the rate of clonal reproduction. The effective number of genotypes is computed for 20 physically unlinked loci. Parameter values are as follows: number of populations, $n = 50$; number of individuals per population, $N = 100$; migration rate, $m = 10^{-3}$; mutation rate, $u = 10^{-5}$; selfing rate is set to random mating, $s = 1/N$.

GENETIC DIVERSITIES

We can now take a closer quantitative look at how genetic diversity is distributed between alleles and genotypes with the stochastic simulations. Allelic diversity can be expressed as the effective number of alleles, n_e, corresponding to the number of equally frequent alleles needed to observe a given genetic diversity, which is $1/(\Sigma p_i^2)$, where p_i is the frequency of the ith allele. Similarly, we can express the effective number of genotypes as $G_e = 1/(\Sigma g_i^2)$, where g_i is the frequency of the ith genotype. In Figure 6 we plot both the effective number of alleles and genotypes within a subpopulation. The number of alleles maintained is strongly positively affected when the reproductive system tends to be completely asexual. This effect is generated by fixed heterozygosity (*i.e.*, under strict clonal reproduction in diploids the two alleles at each locus are behaving as two haploid loci). In contrast to allelic diversity, clonal reproduction decreases the effective number of genotypes steadily (Figure 6). To summarize, populations of clonal or subclonal organisms can maintain more allelic diversity at each single locus but fewer distinct multilocus genotypes.

DISCUSSION

We used both an analytical approach and stochastic individual-based simulations to describe the dynamics of genetic variance in subdivided populations, characterized by various levels of clonal reproduction. Higher rates of asexual reproduction will increase heterozygosity and decrease population differentiation. Diversity at single loci will be higher in clonal organisms than in sexuals, whereas the opposite is true for genotypic diversity. At the exception of genotypic diversity (both at single loci and over multiple loci), which decreases at a constant rate with increasing rates of asexual reproduction, all other quantities investigated are significantly affected only when sexual reproduction becomes rare.

Our results thus suggest that strict clonality may easily be detected in diploid populations due to heterozygote excess. Furthermore, very low levels of sex (cryptic sex) may also be revealed by on average low F_{IS} values with very important variance among loci, though DNA alterations may also lead to a similar pattern in a strictly clonal population. For instance, *Candida albicans* is known to undergo mitotic recombinations including chromosomal translocation (LOTT *et al.* 1999). Much effort has been put into testing for evidence of strict clonal reproduction with traditional population genetics (*e.g.*, TIBAYRENC *et al.* 1991) or through testing for the Meselson effect (high divergence at the two alleles of a single locus within individuals; reviewed in BUTLIN 2000). Extreme genetic divergence at single loci within individuals has been documented in bdelloid rotifers, which are believed to be ancient asexuals (MARK WELCH and MESELSON 2000, 2001). The Meselson effect could, however, not be detected in other potentially old asexual lineages (SCHÖN *et al.* 1998; NORMARK 1999). Whether this is due to rare sex or perhaps to extremely frequent gene conversion events (the copy of the DNA sequence of one chromosome on the other) is an unresolved issue to date.

Empirical data on genetic variation and its apportionment by means of *F*-statistics in clonal lineages, as compared to sexually reproducing populations of the same species, are rare. Furthermore, studies using dominant genetic markers (*e.g.*, rapidly amplified polymorphic DNAs) do not properly allow for the disentanglement between genetic variation within loci and within geno-

types. Indeed, as can be seen from Figure 6, the absolute genetic diversity (the sum of allelic and genotypic variability) does not provide any clear prediction on the rate of clonal reproduction. Another potential problem stems from the difficulty in ruling out the presence of rare sexual reproduction. However, a recent study by Delmotte *et al.* (2002) comparing eight sexual with five asexual populations of the aphid *Rhopalosiphum padi* could provide a test for our model. Their empirical results are overall in good agreement with our analytical expectations. As we expect, Delmotte *et al.* (2002) report increased excess in expected heterozygotes (F_{IS}) for asexuals and lower differentiation (F_{ST}) between asexual populations than between sexual populations. They also report lower genotypic variation and lower allelic variation in asexuals than in sexuals. This relatively good agreement between our model and their data suggests that these asexual populations have not experienced sexual reproduction in recent times. The discrepancy in allelic diversity could be due to different factors (*e.g.*, sampling, extinction-recolonization dynamics). However, even if we assumed all else being equal between the sexual and asexual aphid populations, selection could still reduce genetic diversities more effectively in the asexual populations. Mutations under strong directional selection make linked loci behave as if they were evolving under smaller effective population sizes (Robertson 1961). Due to the complete absence of recombination in strictly clonal organisms, any strongly deleterious dominant mutation will drive the lineage, where it appeared, to extinction. New beneficial mutations will also reduce the effective population sizes of clones, as lineages with a new beneficial mutation will displace other lineages.

Indeed our model does not include natural selection, so that our results apply strictly to neutral genetic variability or more generally to relatively weakly selected polymorphisms subject to genetic drift. Genetic drift is the main force driving allele frequencies as long as the selection differential s between alleles is not much above the inverse of effective population size ($1/N_e$). For higher selection differentials, the effect of genetic drift becomes negligible. However, our predictions should hold even for relatively important selection differentials in clonal and nearly clonal organisms, as the efficacy of selection acting simultaneously at linked sites is considerably reduced (Hill and Robertson 1966).

We assumed identical fitness (in both mean and variance) for clonally and sexually produced offspring. The rate of clonal reproduction is not a heritable trait in our model, as it is a fixed property of the population (clonally produced individuals do not have a higher chance to reproduce clonally themselves). Therefore, different fecundities for sexually or clonally produced offspring would result only in increasing the variance in reproductive success and thus would decrease the effective population size. Our results are thus qualitatively robust to reasonable differences in relative fitness between clonally and sexually produced offspring.

Finally, our model could lead to the development of new approaches to infer the rate of clonal reproduction. Our results show that all estimators based on identities by descent (including linkage disequilibrium approaches) are expected to be rather insensitive to the rate of clonal reproduction as long as it does not become strongly predominant. It is therefore doubtful that such estimators will allow precise inferences on the actual rate of clonal reproduction unless it is very close or equal to 1. As genotypic diversity decreases smoothly with the rate of clonal reproduction, one promising alternative approach would be to build estimators of clonal reproduction as functions of the relative genotypic and allelic identities.

We thank Nathalie Charbonnel, Sylvain Gandon, Jerôme Goudet, Andy Overall, Franck Prugnolle, François Renaud, Max Reuter, Denis Roze, Michel Tibayrenc, and two anonymous referees for very inspiring conversations and comments; François Rousset for having given access to unpublished material; and Sylvain de l'Hérault for his strong support. F.B. was supported by the Biotechnology and Biological Sciences Research Council and by grant 823A-067616 from the Swiss National Science Foundation.

LITERATURE CITED

Anderson, J. B., and L. M. Kohn, 1998 Genotyping, gene genealogies and genomics bring fungal population genetics above ground. Trends Ecol. Evol. **13:** 444–449.

Balloux, F., 2001 EASYPOP (Version 1.7): a computer program for population genetics simulations. J. Hered. **92:** 301–302.

Balloux, F., and N. Lugon-Moulin, 2002 The estimation of population differentiation with microsatellite markers. Mol. Ecol. **11:** 155–165.

Berg, L. M., and M. Lascoux, 2000 Neutral genetic differentiation in an island model with cyclical parthenogenesis. J. Evol. Biol. **13:** 488–494.

Birky, Jr., C. W., 1996 Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. Genetics **144:** 427–437.

Butlin, R. K., 2000 Virgin rotifers. Trends Ecol. Evol. **15:** 389–390.

Butlin, R. K., I. Schön and K. Martens, 1998 Asexual reproduction in nonmarine ostracods. Heredity **81:** 473–480.

Charlesworth, B., 1989 The evolution of sex and recombination. Trends Ecol. Evol. **9:** 264–267.

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

Cockerham, C. C., 1971 Higher order probability functions of identity of alleles by descent. Genetics **69:** 235–246.

Cockerham, C. C., 1973 Analysis of gene frequencies. Genetics **74:** 679–700.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory.* Harper & Row, New York.

Cywinska, A., and P. N. D. Hebert, 2002 Origin of clonal diversity in the hypervariable asexual ostracode. J. Evol. Biol. **15:** 134–145.

Delmotte, F., N. Leterme, J. P. Gauthier, C. Rispe and J. C. Simon, 2002 Genetic architecture of sexual and asexual populations of the aphid Rhopalosiphum padi based on allozyme and microsatellite markers. Mol. Ecol. **11:** 711–723.

Ewens, W. J., 1982 On the concept of the effective population size. Theor. Popul. Biol. **21:** 373–378.

Gabrielsen, T. M., and C. Brochmann, 1998 Sex after all: high levels of diversity detected in the arctic clonal plant *Saxifraga cernua* using RAPD markers. Mol. Ecol. **7:** 1701–1708.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

Hudson, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–44.
Judson, O. P., and B. B. Normak, 1996 Ancient asexual scandals. Trends Ecol. Evol. **11:** 41–46.
Lott, T. J., B. P. Holloway, D. A. Logan, R. Fundyga and J. Arnold, 1999 Towards understanding the evolution of the human commensal yeast Candida albicans. Microbiology **145:** 1137–1143.
Mark Welch, M. D., and M. S. Meselson, 2000 Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science **288:** 1211–1215.
Mark Welch, M. D., and M. S. Meselson, 2001 Rates of nucleotide substitution in sexual and anciently asexual rotifers. Proc. Natl. Acad. Sci. USA **98:** 6720–6724.
Milgroom, M. G., 1996 Recombination and the multilocus structure of fungal populations. Annu. Rev. Phytopathol. **34:** 457–477.
Normark, B. B., 1999 Evolution in a putatively ancient asexual aphid lineage: recombination and karyotype change. Evolution **53:** 1458–1469.
Orive, M. E., 1993 Effective population size in organisms with complex life-histories. Theor. Popul. Biol. **44:** 316–340.
Robertson, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. **2:** 189–194.
Rousset, F., 1996 Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics **142:** 1357–1362.
Rousset, F., 2002 Inbreeding and relatedness coefficients: What do they measure? Heredity **88:** 371–380.
Slatkin, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. **58:** 167–175.
Schön, I., R. K. Butlin, H. I. Griffiths and K. Martens, 1998 Slow molecular evolution in an ancient asexual ostracod. Proc. R. Soc. Lond. Ser. B **265:** 235–264.
Taylor, J. W., D. M. Geiser, A. Burt and V. Koupopanou, 1999 The evolutionary biology and population genetics underlying fungal strain typing. Clin. Microbiol. Rev. **12:** 126–146.
Tibayrenc, M., 1997 Are Candida albicans natural populations subdivided? Trends Microbiol. **5:** 253–254.
Tibayrenc, M., 1999 Toward an integrated genetic epidemiology of parasitic protozoa and other pathogens. Annu. Rev. Genet. **33:** 449–477.
Tibayrenc, M., F. Kjellberg, J. Arnaud, B. Oury, F. Brénière *et al.*, 1991 Are eukaryotic microorganisms clonal or sexuals? A population genetics vantage. Proc. Natl. Acad. Sci. USA **88:** 5129–5133.
Vigalys, R., Y. Gräser and W. Presber, 1997 Response from Vigalys et al. Trends Microbiol. **5:** 254–256.
Wang, J., 1997 Effective size and *F*-statistics of subdivided populations I. Monoecious species with partial selfing. Genetics **146:** 1453–1463.
Weir, B. S., and C. C. Cockerham, 1969 Pedigree mating with two unlinked loci. Genetics **61:** 923–940.
West, S. A., C. M. Lively and A. F. Read, 1999 A pluralist approach to sex and recombination. J. Evol. Biol. **12:** 1003–1012.
Whitlock, M. C., and N. H. Barton, 1997 The effective size of a subdivided population. Genetics **146:** 2427–2441.
Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.
Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Communicating editor: M. K. Uyenoyama

APPENDIX A: GENOTYPIC PROBABILITIES OF IDENTITIES BY DESCENT

We follow the same rationale as Cockerham (1971, pp. 242–243), but add the dynamics of θ. When one offspring is produced clonally, his two alleles are not independent. When we sample alleles and look back to their common parent, the two genes of a clone always stem from the same individual. Two clones are randomly sampled with probability c^2; the four genes stem either from the same parent or from two different ones. The identity between genotypes and three alleles reads

$$\Delta_{t+1} = \frac{1}{N} + \left(1 - \frac{1}{N}\right)\Delta_t$$
$$\rho_{t+1} = \frac{1}{N}F_t + \left(1 - \frac{1}{N}\right)\rho_t \qquad \text{(A1)}$$

(two from one offspring, the third from the other). When one offspring is produced clonally and one through random mating, the identity among three alleles will differ if two or only one allele stem from the clonal offspring, the former case occurs with probability $c(1 - c)$ and the identity is as in Equation A1. The second case also occurs with probability $c(1 - c)$, and the three genes then all stem from the same parent with probability P^3 and share identity ρ^3. The three genes might as well stem from two different parents with probability P^{21} and then have identity ρ^{21}; finally, they can stem from three different parents with probability P^{111} and their identity is ρ^{111}. The three types of identities can be expressed as

$$\rho^3_{t+1} = \frac{1}{4} + \frac{3}{4}F_t, \quad \rho^{21}_{t+1} = \frac{1}{2}\theta_t + \frac{1}{2}\rho_t, \quad \rho^{111}_{t+1} = \rho_t. \qquad \text{(A2.1)}$$

For the identity between genotypes, we have with probability $2c(1 - c)$:

$$\Delta^3_{t+1} = \frac{1}{2} + \frac{1}{2}F_t, \quad \Delta^{21}_{t+1} = \frac{1}{2}\theta_t + \frac{1}{4}\rho_t + \frac{1}{4}\Delta_t, \quad \Delta^{111}_{t+1} = \Delta_t. \qquad \text{(A2.2)}$$

When we sample two offspring issued from random mating with probability $(1 - c)^2$, the identities for ρ are given by Equation A2.1. The genotypic identity is obtained by summing Δ^4, Δ^{22}, Δ^{31}, Δ^{211}, and Δ^{1111}, each identity weighted by its corresponding probability P^4, P^{22}, P^{31}, P^{211}, and P^{1111}. The five identities are written

$$\Delta^4_{t+1} = \frac{1}{4} + \frac{3}{4}F_t, \qquad \Delta^{22}_{t+1} = \frac{1}{12} + \frac{4}{12}\theta_t + \frac{4}{12}\rho_t + \frac{3}{12}\Delta_t,$$
$$\Delta^{31}_{t+1} = \frac{1}{2}\theta_t + \frac{1}{2}\rho_t, \quad \Delta^{211}_{t+1} = \frac{1}{6}\theta_t + \frac{2}{6}\rho_t + \frac{3}{6}\Delta_t,$$
$$\Delta^{1111}_{t+1} = \Delta_t. \qquad \text{(A3)}$$

The different probabilities of gamete origins are given in Weir and Cockerham (1969) and read as

$$P^3 = \frac{1}{N^2}, \quad P^{21} = \frac{3(N-1)}{N^2}, \qquad P^{111} = \frac{(N-1)(N-2)}{N^2},$$
$$P^4 = \frac{1}{N^3}, \quad P^{22} = \frac{3(N-1)}{N^3}, \qquad P^{31} = \frac{4(N-1)}{N^3},$$
$$P^{211} = \frac{6(N-1)(N-2)}{N^3}, \qquad P^{1111} = \frac{(N-3)(N-2)(N-1)}{N^3}. \qquad \text{(A4)}$$

APPENDIX B: COEFFICIENT OF EQUATION 23

$$
\begin{aligned}
a_0 &= 2(N-1)^2(6N^2-1)\\
a_1 &= -2(N-1)(N(5N(2N+3)-19)+2)\\
a_2 &= -2(1+N(3N-4)(N(14N-17)+6))\\
a_3 &= 2N(N(50N^2-66N+31)-5) \qquad \text{(B1.1)}\\
b_0 &= (N-1)^2(2N+3)\\
b_1 &= -(19N^2+28N-9)\\
b_2 &= -(30N^3-65N^2+44N-9)\\
b_3 &= 36N^3-45N^2+20N-3. \qquad \text{(B1.2)}
\end{aligned}
$$

If $c = 0$, then the mean coalescence time for two genotypes in the Wright-Fisher setting reduces to

$$\bar{t}_\Delta = \frac{2N(N(50N^2-66N+31)-5)}{(3N-1)(N(12N-11)+3)}. \qquad \text{(B2)}$$

Performing a Taylor expansion of first degree under large population size and substituting some close integers yields the approximation for Equation B2:

$$\bar{t}_\Delta \approx 3N. \qquad \text{(B3)}$$