Inferring Recent Outcrossing Rates Using Multilocus Individual Heterozygosity: Application to Evolving Wheat Populations

Jérôme Enjalbert* and Jacques L. David[†]

*INRA, Station de Génétique Végétale, Ferme du Moulon, 91190 Gif sur Yvette, France and
[†]INRA, Domaine de Melgueil, 34130 Mauguio, France

ABSTRACT

Using multilocus individual heterozygosity, a method is developed to estimate the outcrossing rates of a population over a few previous generations. Considering that individuals originate either from outcrossing or from n successive selfing generations from an outbred ancestor, a maximum-likelihood (ML) estimator is described that gives estimates of past outcrossing rates in terms of proportions of individuals with different n values. Heterozygosities at several unlinked codominant loci are used to assign n values to each individual. This method also allows a test of whether populations are in inbreeding equilibrium. The estimator's reliability was checked using simulations for different mating histories. We show that this ML estimator can provide estimates of outcrossing rates for the final generation outcrossing rate (t_0) and a mean of the preceding rates (t_p) and can detect major temporal variation in the mating system. The method is most efficient for low to intermediate outcrossing levels. Applied to nine populations of wheat, this method gave estimates of t_0 and t_p. These estimates confirmed the absence of outcrossing ($t_0 = 0$) in the two populations subjected to manual selfing. For free-mating wheat populations, it detected lower final generation outcrossing rates ($t_0 = 0$–0.06) than those expected from global heterozygosity ($t = 0.02$–0.09). This estimator appears to be a new and efficient way to describe the multilocus heterozygosity of a population, complementary to Fis and progeny analysis approaches.

THE effect of mating system on plant genetic diversity is a major theme in evolutionary genetics, as it plays a major role in structuring the genetic variation within and between populations. For example, self-fertilization correlates with lower within-population diversity and with higher between-population differentiation, and global polymorphism is generally lower in selfing species (Godt and Hamrick 1991).

Two strategies have been used to estimate the outcrossing rates of a mixed mating population from genetic markers: measuring the frequency of heterozygous genotypes and the analysis of progeny arrays (Ritland 1983). Assaying the frequency of heterozygotes in a population gives an estimate of the outcrossing rate assuming that (i) the selfing rate has been constant for a sufficient number of generations, (ii) the population is in inbreeding equilibrium, and (iii) selfing is the major cause of departure from Hardy-Weinberg frequencies (see Brown and Allard 1970). In such cases, a simple relation exists between Wright's within-population inbreeding coefficient f and the outcrossing rate t_f, $f = (1 - t_f)/(1 + t_f)$ (Wright 1969). This method is widely used when surveying marker diversity in natural populations, for example, in Centrosema (Penteado et al.

1996) or in *Bulinus truncatus* (Viard et al. 1997). However, estimates of t_f must be treated with caution when temporal variation exists in outcrossing rates.

The second method is based on progeny arrays. Morphological markers were used long ago to identify outcrossing vs. selfing events in progenies of known maternal genotypes (e.g., in wheat; Hayes 1918). As the detection of a single nonmaternal allele in a genotype proves its outbreeding origin, accurate estimates of the outcrossing rate in natural populations can be obtained using multilocus information on progenies (especially in conjunction with estimates of maternal genotypes). Powerful estimators of outcrossing rate based on progeny analysis have been proposed (Ritland and Jain 1981; Shaw et al. 1981) and widely used (Schoen and Brown 1991). Progeny arrays can also measure the variance in outcrossing rates between maternal individuals, e.g., as has been done in studies in *Eucalyptus regnans* (Moran et al. 1989) and *Acacia nilotica* (Mandal and Ennos 1995). They also yield estimates of both ovule and pollen allelic frequencies (Godt and Hamrick 1991). Additionally, differences between monolocus and multilocus estimates make it possible to infer the amount of inbreeding due to mating between relatives (Shaw and Allard 1982; Ritland 1984). Note that progeny arrays can also yield a Wright estimate t_f, using the inbreeding coefficient of mothers.

Environmental conditions can cause considerable temporal variation in mating behavior. Low tempera-

Corresponding author: Jérôme Enjalbert, INRA, Unité de Pathologie Végétale, BP01, 78850 Thiverval-grignon, France.
E-mail: enjalber@grignon.inra.fr

tures or light intensity can modify outcrossing in some selfing species, as documented in wheat (DEMOTES-MAINARD *et al.* 1995) and rice (LI *et al.* 1996). Low population density can reduce outcrossing, as demonstrated in Bombacaceous trees (MURAWSKI *et al.* 1990) and in *Cuphea laminuligera* (KRUEGER and KNAPP 1991). Studies over successive years are thus required to measure temporal variation in outcrossing rates (BARRETT *et al.* 1993; SPROULE and DANCIK 1996).

Here we develop a method to estimate the outcrossing rates for a few previous generations using a single generation survey. It is based on the analysis of multilocus heterozygosity within individuals. In mixed mating species, individuals can either originate from outcrossing between individuals or be derived from a varying number of selfing generations from outbred individuals. Hence individuals display varying levels of heterozygosity in their genomes, as initial heterozygosity resulting from the founder outcrossing event is halved by each successive selfing generation (*i.e.*, for two independent loci, recently produced genotypes are more likely to be double heterozygotes, whereas more and more double homozygotes accumulate with each selfing generation). The proportion of individuals exhibiting a high level of heterozygosity should thus give information about the outcrossing rate in the most recent generation, while the proportions of individuals with varying levels of homozygosity could be used to estimate outcrossing rates in the previous generations. Consequently, a population can be partitioned into classes of individuals sharing the same number of selfing generations since the last outcrossing event in their genealogy. Each class presents distinct expected levels of multilocus individual heterozygosity (MIH), and class frequencies are a result of the mating history of the population. Using both probability formulas, we developed a maximum-likelihood estimator of outcrossing rates on the basis of multilocus patterns observed in a random sample of individuals drawn from an infinite population. This approach allows a test of whether a population had a constant selfing rate in the past, *i.e.*, if it is in inbreeding equilibrium.

Using simulated populations, the estimator properties are evaluated for various mating histories as functions of the sample size, the number of loci used to estimate MIH, and their Nei diversity. The MIH method is tested on nine populations of wheat (*Triticum aestivum*) derived from a pilot program of dynamic management of genetic resources. These populations have known contrasting mating histories. We show that this approach can be used when progeny analysis is not possible for practical reasons, or when temporal variance in outcrossing rates is suspected.

ANALYTICAL MODEL

Inbreeding classes distribution: Our model assumes that populations evolve with temporal variation in out-

crossing rates with no genetic drift, no selection (particularly no heterotic selection), and with the random mating of outcrossing gametes. In populations of hermaphroditic species, individuals may originate from a cross between two different parents or from the selfing of a single individual, which itself could have resulted from selfing or outcrossing. Each individual can then be indexed on n ($0 \leq n < \infty$), the number of selfing generations since the last outcrossing event in its genealogy. Such an individual is said to belong to class S_n. Then any population observed at a given generation is composed of an infinity of classes of individuals $\{S_0, \ldots, S_\infty\}$ in proportions $\{Q_0, \ldots, Q_\infty\}$, with of course $\Sigma_{i=0}^{\infty} Q_i = 1$.

Individuals from all classes are assumed to outcross at the same rate. If t_0 is the outcrossing rate that occurred in the production of the present generation, then individuals that were in class S_n one generation before produced progeny in class S_0 at a rate of t_0 and in S_{n+1} at a rate of $(1 - t_0)$. It is therefore clear that

$$Q_{0,0} = \sum_{n=0}^{\infty} t_0 Q_{n,1} = t_0, \qquad (1)$$

with $Q_{n,1}$ the proportions of the S_n existing at the previous generation.

If we use t_{-n} to denote the outcrossing rate in the population n generations before the one observed, then, given that individuals in class S_n in the present population derive from outbred ancestors created n generations previously, we have $Q_{1,0} = Q_{0,1}(1 - t_0) = (1 - t_0)\Sigma_{n=0}^{\infty} t_{-1} Q_{n,2} = t_{-1}(1 - t_0)$.

By recurrence, the proportion of individuals descending from n generations of selfing in the generation observed is thus

$$Q_{n,0} = Q_{n-1,1}(1 - t_0) = Q_{0,n} \prod_{i=0}^{n-1}(1 - t_{-i}) = t_{-n} \prod_{i=0}^{n-1}(1 - t_{-i}). \qquad (2)$$

Expected multilocus individual heterozygosity: In diploid species, most individuals that result from an outcrossing event (S_0) can be detected since they are heterozygous at many loci (BENNETT and BINET 1956). Individuals in class S_{n+1} have half the mean heterozygosity of individuals in the previous class S_n. We note that $P(ht/S_n)$ and $P(hm/S_n)$ are the probabilities of individuals from class S_n being heterozygous and homozygous, respectively, at a given locus. In the generation observed, $P(ht/S_0)$, *i.e.*, the probability of being heterozygous after an outcrossing event is

$$P(ht/S_0) = 1 - \sum_{i=1}^{k} p_i^2 = D, \qquad (3)$$

where k is the number of alleles at the locus; p_i, the frequency of the ith allele (assumed to be constant over the generations); and D, Nei's gene diversity (NEI 1978) at this locus. As previously mentioned, heterozygosity is halved by each selfing generation, and thus the probabil-

ity of the locus being heterozygous after n generations of selfing is

$$P(ht/S_n) = \frac{1}{2}P(ht/S_{n-1}) = \frac{D}{2^n},$$

and, reciprocally, $\quad P(hm/S_n) = 1 - \frac{D}{2^n}.$

Now consider G_x, the multilocus heterozygosity pattern of an individual x at all the loci genotyped. Using the variable a_l, where $a_l = 1$ if x is heterozygous at locus l, and $a_l = 0$ otherwise, then for L unlinked loci in linkage equilibrium,

$$P(G_x/x \in S_n) = \prod_{l=1}^{L} P(ht_l/S_n)^{a_l} P(hm_l/S_n)^{1-a_l}$$

$$= \prod_{l=1}^{L} \left(\frac{D_l}{2^n}\right)^{a_l} \left(1 - \frac{D_l}{2^n}\right)^{1-a_l}, \quad (4)$$

with D_l the gene diversity of locus l, and ht_l and hm_l the probability to be heterozygous and homozygous, respectively, at locus l.

This product only holds at linkage equilibrium; otherwise disequilibrium measures have to be introduced (not developed here). Note that the correlation between heterozygous states of independent loci in a mixed mating population observed by BENNETT and BINET (1956) results from combining of the very first selfing classes, mainly composed of multiple heterozygous individuals, with highly selfed classes composed of multiple homozygotes. This correlation is absent within a given selfing class S_n and has no effect on S_0 formation as the disequilibrium is zygotic rather than gametic. Any heterozygosity correlation appearing in S_0 by departure from our hypothesis (nonrandom mating, population of small size . . .) will be rapidly lost while heterozygosity diminishes during the successive selfings.

Maximum-likelihood estimation of recent outcrossing rates: From the previous expressions, the likelihood of G_x, the genotype of individual x, is

$$L(x) = P(G_x) = \sum_{n=0}^{\infty} P(G_x \cap x \in S_n)$$

$$= \sum_{n=0}^{\infty} P(G_x/x \in S_n)P(x \in S_n). \quad (5)$$

When n is large (threshold value of $n = 20$ in our numerical procedure) the probability of any locus being heterozygous approaches zero (that of being homozygous is one) and this makes simplification of the summation of $P(G_x/x \in S_n)$ possible,

$$L(x) = \sum_{n=0}^{20} P(G_x/x \in S_n)P(x \in S_n) + R, \quad (6)$$

where R is the approximated value of $\sum_{n=21}^{\infty} P(G_x/x \in S_n)P(x \in S_n)$, detailed in APPENDIX A.

As $P(x \in S_n) = Q_{n,0}$, using Equations 1, 2, and 4, the

first term of $L(x)$ depends both on gene diversities and outcrossing rates for successive generations and, hence,

$$L(x) = t_0 \prod_{l=1}^{L} D_l^{a_l}(1 - D_l)^{1-a_l}$$

$$+ \sum_{n=1}^{N} \left(t_{-n} \prod_{i=0}^{n-1} (1 - t_{-i}) \prod_{l=1}^{L} \left(\frac{D_l}{2^n}\right)^{a_l} \left(1 - \frac{D_l}{2^n}\right)^{1-a_l} \right) + R.$$

$$(7)$$

The mating behavior is estimated for z previous generations, and the outcrossing rates of all generations before generation z are assumed to be constant and equal to t_p. As developed in APPENDIX C, when outcrossing rates vary before generation z, t_p is a function of (t_{-z}, $t_{-z-1}, \ldots, t_{-\infty}$), mainly depending on the very first terms (recent outcrossing rates).

Thus $z + 1$ outcrossing rates ($t_0, \ldots, t_{-z+1}, t_p$) must be jointly estimated by the maximum-likelihood technique. A complete expression for $L(x)$ is presented in APPENDIX A.

The likelihood of a sample of X individuals $L(1, .x., X)$ of genotype (G_1, G_x, G_X) is the product of the individual likelihoods, on the assumption of independence between sampled individuals, $L(1, .x., X) = \prod_{i=1}^{X} L(i)$. Values $\{\hat{t}_0, \ldots, \hat{t}_{-z+1}, \hat{t}_p\}$ jointly maximizing the likelihood $L(1, .x., X)$ are retained as estimates of $\{t_0, \ldots, t_{-z+1}, t_p\}$. In this article, we first estimated diversities $\hat{D}_l$ according to NEI (1978): $\hat{D} = 1 - \sum_{i=1}^{k} \hat{p}_i^2$. Then we numerically determined the values of $\{\hat{t}_0, \ldots, \hat{t}_{-z+1}, \hat{t}_p\}$, maximizing Equation 4 in APPENDIX A.

Confidence intervals: When known, the distribution of the LOD score, the logarithm of the ratio of the maximum likelihood to the likelihood for any value of $\{t_0, \ldots, t_{-z+1}, t_p\}$, LOD $= \text{Log}_{10} (L(\hat{t}_0, \ldots, \hat{t}_{-z+1}, \hat{t}_p)/ L(t_0, \ldots, t_{-z+1}, t_{-z+1}, t_p))$, allows one to build intervals for any given confidence level. The distribution of the LOD for the MIH estimator was empirically determined using large numbers of independent simulations for mating histories with known outcrossing rates. The values of these outcrossing rates are hereafter referred as parametric values (see NUMERICAL VALIDATION). For each simulated mating scenario, we calculated the LOD score based on the maximum-likelihood estimates and the likelihood obtained with the parametric values. To build confidence intervals, LOD_{95} (or LOD_{99}) values were empirically determined as minimum LOD values for which 95% (or 99%) of runs included the parametric values.

Inbreeding equilibrium test: To test whether a population is in inbreeding equilibrium, we calculated the LOD score $\text{Log}_{10} (L(\hat{t}_0, \ldots, \hat{t}_{-z+1}, \hat{t}_p)/L(\hat{t}_f, \ldots, \hat{t}_f, \hat{t}_f))$, where $\{\hat{t}_0, \ldots, \hat{t}_{-z+1}, \hat{t}_p\}$ are the maximum-likelihood (ML) estimates and $\hat{t}_f$ is the rate calculated on the assumption of temporally constant outcrossing rates and inbreeding equilibrium. For LOD score values $>\text{LOD}_{95}$, the hypothesis of inbreeding equilibrium was rejected

TABLE 1

Estimates of successive outcrossing rates in simulated populations

		$t_0 = 0.01$	$t_0 = 0.25$	$t_0 = 0.50$	$t_0 = 0.75$
$t_p = 0.01$	$\hat{t}_0$	0.011 ± 0.009	0.254 ± 0.04	0.50 ± 0.044	0.746 ± 0.039
	$\hat{t}_p$	0.009 ± 0.006	0.01 ± 0.008	0.009 ± 0.012	0.019 ± 0.033
$t_p = 0.25$	$\hat{t}_0$	0.014 ± 0.018	0.247 ± 0.055	0.504 ± 0.053	0.757 ± 0.048
	$\hat{t}_p$	0.253 ± 0.056	0.249 ± 0.045	0.245 ± 0.060	0.228 ± 0.086
$t_p = 0.50$	$\hat{t}_0$	0.013 ± 0.018	0.246 ± 0.055	0.516 ± 0.054	0.743 ± 0.053
	$\hat{t}_p$	0.49 ± 0.056	0.50 ± 0.076	0.489 ± 0.09	0.481 ± 0.141
$t_p = 0.75$	$\hat{t}_0$	0.017 ± 0.023	0.253 ± 0.064	0.495 ± 0.062	0.754 ± 0.059
	$\hat{t}_p$	0.745 ± 0.073	0.756 ± 0.09	0.767 ± 0.13	0.733 ± 0.182

Mean estimates and their standard deviations of two successive outcrossing rates obtained by sampling 150 individuals in 100 simulated independent populations for 16 distinct (t_0, t_p) parametric scenarios: t_0 is the parametric outcrossing rate in the final generation and t_p the one in all previous generations. Individuals were genotyped at four loci of diversity $D = 0.8$.

at the 5% probability level, and we then concluded that $\hat{t}_0$ was significantly different from $\hat{t}_p$. Note that this likelihood ratio is related to a chi-square distribution of 1 d.f. (see Weir 1990, p. 80).

Hereafter, $\hat{t}_f$ is calculated from a multilocus estimate of the $\hat{f}$ inbreeding coefficient (Rousset and Raymond 1995) and then $\hat{t}_f = (1 - \hat{f})/(1 + \hat{f})$.

SIMULATION ALGORITHM

To simulate infinite populations with constant allelic frequencies (no selection) and no genotypic variance in outcrossing rate, the proportions of the $\{S_0, \ldots, S_n\}$ classes were calculated according to the parametric outcrossing values using Equations 1 and 2. Samples of individuals were randomly drawn from this distribution. The multilocus genotype of each individual was then determined by random drawing (Monte Carlo procedure) according to the probabilities of being homozygous or heterozygous given the selfing class, for a given distribution of allelic frequencies. For reasons of simplicity, all simulated populations were in inbreeding equilibrium and outcrossed at rate t_p until the last but one generation and then outcrossed at a rate of t_0 for the final generation.

NUMERICAL VALIDATION

To cover a breadth of situations, both t_0 and t_p took values in $\{0.01, 0.25, 0.50, 0.75\}$. For each of the 16 ($t_0$, t_p) pairs, we simulated 100 independent populations of 150 individuals genotyped for four loci with diversity value $D = 0.8$. Actual $\hat{D}$ values used in MIH estimators were estimated for each locus from the simulated data sets according to Nei (1978).

Table 1 gives means and standard deviations of the estimated values of $\hat{t}_0$ and $\hat{t}_p$ over 100 independent simulations for the 16 parametric scenarios. The $\hat{t}_0$ and $\hat{t}_p$ values obtained were reliable even with a low number of loci (four), and no significant bias was observed in

their means. The parametric value of t_p slightly affected the standard deviation of $\hat{t}_0$, whereas the standard deviation of $\hat{t}_p$ increased with a high value of t_0.

No significant effect of (t_0, t_p) values on LOD score threshold values at 95 and 99% was apparent from the distribution of LOD scores for all parametric values of (t_0, t_p) (ANOVA, results not shown). Mean LOD score threshold values were 2.99 and 4.76 for 95th and 99th percentiles, respectively. By default, these empirical score values can be used to calculate confidence intervals. We also computed $\hat{t}_f$ and their LOD score values. The proportion of estimates for which inbreeding equilibrium is rejected (which can be interpreted as the proportion of significant differences between $\hat{t}_0$ and $\hat{t}_p$) is reported in Table 2 for each simulated (t_0, t_p) pair. When parametric t_0 was equal to parametric t_p, the hypothesis of inbreeding equilibrium (H_0) was accepted 98 times out of 100 on average. For low to intermediate parametric outcrossing rates, H_0 was always rejected when the parametric value of t_0 effectively differed from that of t_p. But with increasing parametric outcrossing rates, and particularly t_0, H_0 could be wrongly accepted. The worst such situations were those for which t_0 was high and the difference with t_p low (H_0 being accepted 81 times out of 100 when $t_0 = 0.75$ and $t_p = 0.50$).

This numerical validation shows that MIH estimates of t one generation (or more) before that observed are sensitive to the "resetting effect" of a high level of outcrossing in the final generation. In such cases, most inbred individuals outcross to give S_0 individuals and the proportions of selfing classes S_n with $n > 0$ are therefore dramatically reduced. The next section analyzes the variance of these ML estimators under different experimental designs (sample size, and number and diversity of loci).

RELIABILITY OF ESTIMATES BASED ON MIH

The reliability of MIH estimators of successive outcrossing rates was studied as a function of (i) the number

TABLE 2

Test of the inbreeding equilibrium hypothesis

		$t_0 = 0.01$	$t_0 = 0.25$	$t_0 = 0.5$	$t_0 = 0.75$
$t_p = 0.01$	$\hat{t}_f$	0.01	0.15	0.34	0.61
	H_0	98/100	0/100	0/100	0/100
$t_p = 0.25$	$\hat{t}_f$	0.12	0.25	0.43	0.67
	H_0	0/100	97/100	36/100	6/100
$t_p = 0.50$	$\hat{t}_f$	0.20	0.33	0.51	0.71
	H_0	0/100	40/100	99/100	81/100
$t_p = 0.75$	$\hat{t}_f$	0.28	0.40	0.55	0.75
	H_0	0/100	3/100	70/100	98/100

Mean over 100 simulation runs of outcrossing rate t_f estimated on the assumption of inbreeding equilibrium according to varying mating scenario (t_0, outcrossing rate in the final generation; t_p, outcrossing rate in all previous generations). First line, $\hat{t}_f$ multilocus estimate using Wright's inbreeding coefficient; second line, number of simulations out of 100 for which the hypothesis of inbreeding equilibrium could be accepted when using maximum-likelihood estimators of successive outcrossing rates.

of sampled individuals, which determines the sampling error of the estimated proportions in selfing classes, and (ii) the numbers and diversity of genotyped loci, which determine the accuracy of individual assignment to selfing classes. The theoretical variance of the maximum-likelihood estimator can be calculated by the second partial derivatives of the likelihood expression. Because of the complexity of these derivatives, we chose to analyze the different sources of variation involved in the likelihood separately.

Sampling error in selfing class distribution: Sample size has an obvious effect on the accuracy of ML estimates through sampling errors in the representation of the different selfing classes. As the proportions of selfing classes $\{Q_0, \ldots, Q_n\}$ depend on the outcrossing rates experienced, the optimal sampling size will vary with the mating history and the number of past generation outcrossing estimates desired. The higher the outcrossing rate in the most recent generation, the lower the number of individuals distributed in $\{S_1, \ldots, S_n\}$, and the more individuals will be needed.

When sampling within selfing classes is the only considered effect, the variance of the t_{-n} estimate in a population having a steady outcrossing rate $t = t_0 = t_{-1} = t_{-n}$ is

$$V(\hat{t}_{-n}) \approx \frac{t}{N(1-t)^{n-1}} \qquad (8)$$

(APPENDIX C).

Figure 1 plots the standard deviation over $\hat{t}_{-n}$, when sample size is $N = 100$, and illustrates that the lower the successive outcrossing rates, the deeper the possible insight into the past mating behavior of the population. For low or intermediate outcrossing rates, the accuracy of the recent outcrossing rate estimates t_0, t_{-1}, and t_{-2} is thus expected to be good with a small sample size. Using Equation 8, sample sizes can be determined to obtain sufficiently accurate estimates of successive outcrossing rates.

Discrimination between selfing classes: Accuracy in the determination of outcrossing rates relies greatly on the ability to correctly assign an individual to its specific selfing class on the basis of its heterozygosity. To illustrate the importance of the number and diversity of loci used to discriminate selfing classes, we plot on Figure 2 the distribution of the number of heterozygous loci expected for S_0, S_1, S_2, and S_3 individuals in the case of $L = 20$ independent loci with equal Nei's diversity (a, $D = 0.8$; b, $D = 0.3$). In this simple case, the distribution of the MIH is binomial,

$$P(K = k/\text{Self} = n) = \binom{k}{L}\left(\frac{D}{2^n}\right)^k\left(1 - \frac{D}{2^n}\right)^{L-k}, \qquad (9)$$

where K is the number of heterozygous loci.

With few loci, MIH distribution of S_0 individuals differentiates from the other distributions, allowing us to estimate t_0. Oppositely, S_1, S_2, and S_3 distributions are strongly overlapping, and more loci of high polymorphism are to be used to separate them and thus allow
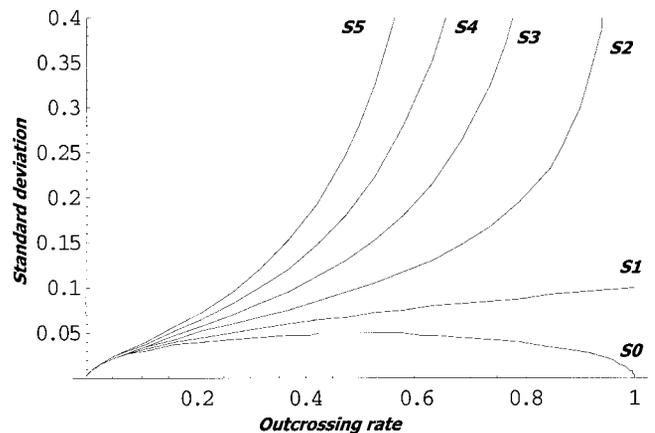

FIGURE 1.—Change in standard deviation of estimates of outcrossing rates for the five last generations according to the outcrossing rate of a population in inbreeding equilibrium in a sample of 100 individuals (see text for details).
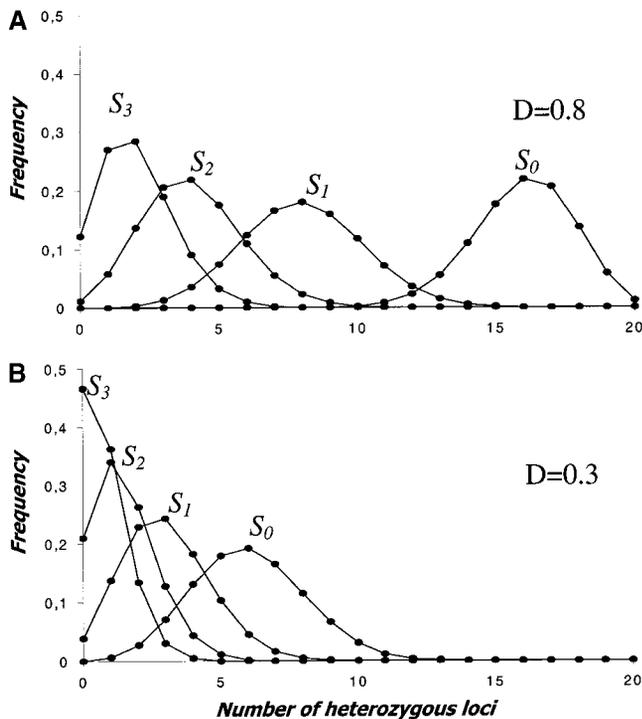
FIGURE 2.—MIH distributions: distribution of the number of heterozygous loci expected in individuals belonging to S_0, S_1, S_2, and S_3 selfing classes when observing 20 independent loci with equal heterozygosities of (A) $D = 0.8$ and (B) $D = 0.3$.
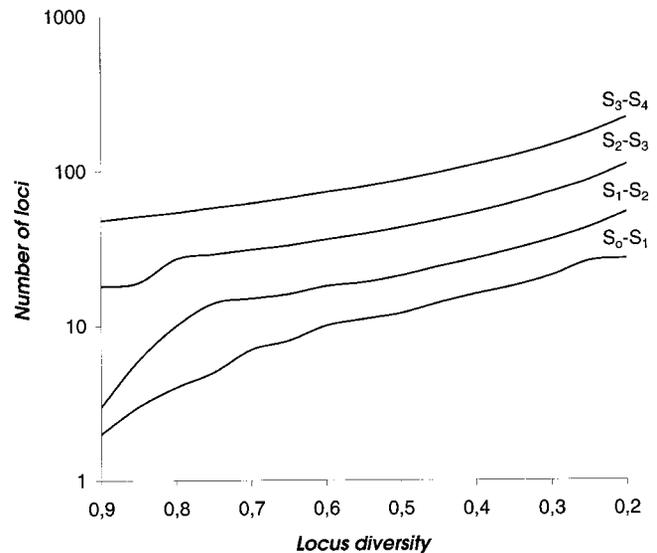

FIGURE 3.—Number of loci required to separate MIH distributions of two successive selfing classes with <20% of overlapping area, according to their Nei diversity.

access to the corresponding outcrossing rate. To compare sets of loci of different levels of diversity, Equation 9 was arbitrarily used to calculate the number of loci required to have <20% of the area of a given distribution overlapping with the others (Figure 3).

Discrimination between S_0 and S_1 can be achieved with a reasonable number of loci (<10) provided their diversity is at least ∼0.5. Discriminating S_1 from S_2 is still possible using markers with high diversity values, but discrimination between higher inbreeding levels will require a huge number of loci (∼30 highly diverse loci to discriminate between S_2 and S_3). Thus the MIH approach in practice seems limited to three successive outcrossing rate estimates (t_0, t_{-1}, t_p).

Errors in diversity estimates: According to WEIR (1990), variance in $\hat{D}$ at a locus is expressed as $V(\hat{D}) \approx 2(1 + f)/I[\Sigma_u p_u^3 - (\Sigma_v p_v^2)^2]$, where f is the inbreeding level, I the sample size, and p_u the frequency of allele u. From this expression, it follows that asymmetrical allelic frequencies at a locus lead to higher variances in $\hat{D}$ than equal allelic frequencies and could thus reduce the accuracy of outcrossing rate estimates. We explored the sensitivity of MIH outcrossing estimators to uneven allelic frequencies by simulating four series of populations with varying allelic frequencies but constant diversities and found a negligible effect on the resulting estimates and their variances (data not shown).

Designing an experimental protocol: As discussed pre-

viously, the optimal experimental design depends on the mating history of the population and also on the polymorphism of the marker used. Thus, to complete a specific experimental design, it could be of interest to first obtain a rough value for (t_0, t_p) from a subsample of individuals analyzed with few markers. Then the genotyping could be completed by increasing the number of loci or the number of individuals analyzed. This optimal experimental design could be determined by simulating populations with parameters varying around the rough estimates.

OUTCROSSING RATE ESTIMATES IN WHEAT POPULATIONS

Plant material: Using restriction fragment length polymorphism (RFLP) markers, MIH maximum-likelihood estimation of outcrossing rates was applied to nine wheat populations derived from two composite crosses (hereafter called PA and PB). The composites stem from successive crosses of two distinct sets of 16 inbred lines (DAVID et al. 1997; ENJALBERT et al. 1999a). A pyramidal crossing design was used to create these composites. It required 4 yr of manual crosses: eight one-way hybrids were created the first year, four two-way hybrids the second year, and so on. Beginning in 1984, six populations were then grown in different field locations. The individuals used in this study were derived from the 1994 harvest, after at least 12 generations of open pollination, predominantly selfing. Thus, the heterozygosity due to the manual hybridization of parental lines will have practically vanished and should not interfere with the estimation of the recent outcrossing rates in these populations.

An additional three populations were derived from

TABLE 3

Successive outcrossing rates in nine wheat populations

Population	No. of plants	$\hat{t}_0$ (C.I.)	$\hat{t}_p$ (C.I.)	$\hat{t}_f$ (LOD($\hat{t}_f$, $\hat{t}_f$))
PA Moulon	78	0.04 (0.005–0.13)	0.06 (0.03–0.10)	0.06 (0.19) NS
PA Rennes	77	0.01 (0.0–0.10)	0.13 (0.08–0.20)	0.06 (5.69)*
PA Toulouse	78	0.06 (0.01–0.16)	0.15 (0.09–0.21)	0.09 (3.37)*
PB Moulon	78	0.0 (0.0–0.04)	0.09 (0.05–0.14)	0.02 (14.7)*
PB Rennes	77	0.0 (0.0–0.05)	0.26 (0.16–0.33)	0.09 (19)*
PB Venours	77	0.015 (0.0–0.08)	0.08 (0.05–0.14)	0.04 (5.26)*
6SSD PA0	77	0.004 (0.0–0.06)	0.16 (0.11–0.22)	0.05 (19)*
6SSD PB0	78	0.00 (0.0–0.04)	0.12 (0.08–0.18)	0.03 (22)*
4SSD R.S.	77	0.00 (0.0–0.04)	0.21 (0.14–0.28)	0.06 (30)*

Maximum-likelihood estimates of outcrossing rates based on multilocus individual heterozygosity (MIH) in nine populations of wheat (see text). C.I., confidence interval based on minimum and maximum values observed in the confidence area defined by LOD = $\text{Log}_{10} L(\hat{t}_0, \hat{t}_p)/L(t_0, t_p) \le 2.8$. t_f is the multilocus estimate using Wright inbreeding coefficient when inbreeding equilibrium is assumed. (LOD($\hat{t}_f$, $\hat{t}_f$)), LOD score of the hypothesis of inbreeding equilibrium H_0. H_0 is accepted if LOD($\hat{t}_f$, $\hat{t}_f$) ≤ 2.8 and rejected otherwise and marked with an asterisk (2.8 is the LOD_{95} value obtained by simulating populations with the same characteristics as studied wheat populations). NS, not significant.

the same initial populations by four (SR) and six (PA0 and PB0) generations of single seed descent. As selfing was performed by bagging the spikes, these populations are expected to have been purely selfing since the controlled crosses.

About 78 individuals per population were genotyped using 14 restriction enzyme and probe combinations, providing 25 independent codominant RFLP loci (ENJAL-BERT *et al.* 1999b). To verify whether our data were able to provide proper estimates of outcrossing rates and to calculate the LOD_{95} value, we simulated populations with identical sample size, loci number, and diversity as observed in wheat populations.

Results: The mean number of alleles per locus was 2.6 and the mean Nei gene diversity was 0.34. Most markers were located on distinct chromosome arms. After the four manual crosses, which occurred during the building of the initial population, linkage disequilibrium was low in PA0, PB0, and in all derived populations (ENJALBERT *et al.* 1999a). Our data should thus make it possible to discriminate satisfactorily $S_0 - S_1$ (0.8) and $S_1 - S_2$ (0.63) and to estimate two outcrossing rates t_0 and t_p. Note that for the three single seed descent (SSD) populations, t_p was not an equilibrium value, but was a mean of the manual crosses and selfing that occurred in the previous generations. The LOD_{95} value obtained

by simulation was 2.8, close to values obtained in previous simulations. Table 3 gives the estimates ($\hat{t}_0$, $\hat{t}_p$) and corresponding confidence intervals, together with $\hat{t}_f$ and its LOD score. Globally, outcrossing is low in the nine populations. The greatest departures from inbreeding equilibrium were observed for the three SSD populations for which $\hat{t}_0$ was estimated to be zero, as expected, while high $\hat{t}_p$ values (for a selfing species) suggested considerable outcrossing in the past (LOD score >19). As also expected, the four-generation-old SSD population (SR) has a higher $\hat{t}_p$ than the six-generation-old SSD populations (PA0 and PB0). More surprisingly, all populations but one (PA Moulon) grown under natural selection and open pollination were not in inbreeding equilibrium and sometimes indicated highly contrasting outcrossing values. Another point to note is that $\hat{t}_0$ was always lower than $\hat{t}_p$. This could be for various reasons, discussed below.

DISCUSSION

We developed a maximum-likelihood estimator of successive outcrossing rates based on MIH. This method provides estimates and confidence intervals for the two or three most recent outcrossing rates, under the assumptions of no selection, no allelic frequency changes,

unlinked loci at linkage equilibrium, and random mating for outcrossing gametes. To build confidence intervals, an empirical LOD_{95} value of ~3.0 should be used to cover most situations, except those with few loci. We verified that this LOD_{95} value did not differ greatly for other simulated populations with varying sampling size and loci number or diversity (results not shown).

The accuracy and number of estimates that can be made depend greatly on the outcrossing level of the population: the MIH estimator is most informative for populations with intermediate or low outcrossing rates, because of the resetting effect of outcrossing on mating history. The accuracy of the estimates is related to the size of the sample analyzed and to the number and heterozygosities of the markers used. Increasing the diversity of loci rather than their number allows for better discrimination between selfing classes and thus a deeper insight into populations' mating history, whereas increasing the sample size improves the accuracy of estimates of the sizes of distinct inbreeding classes. To improve the statistical power of a data set and to test inbreeding equilibrium, we suggest simulating data within a range of expected t values and sequentially accumulating the molecular data, adding more individuals or more loci to obtain the desired accuracy.

This is apparently the first method that leads to a test for inbreeding equilibrium using a single generation analysis. It is of great importance to the analysis of heterozygosity, as using mean heterozygosity to estimate the outcrossing rate t_f will lead to misleading conclusions in a population with varying outcrossing rates. As pointed out by BROWN and ALBRECHT (1980), the estimates of inbreeding coefficient will be biased to higher values by varying outcrossing rates. Considering a population that has a constant outcrossing rate t_p for a long time and shifting the final generation from t_p to t_0, then the value of t_f assuming inbreeding equilibrium is $t_f = (t_0 + t_p)/(2 + t_p - t_0)$ (APPENDIX B). Let us imagine two populations, the first with ($t_p = 0.3$, $t_0 = 0.1$) and the second with ($t_p = 0.1$, $t_0 = 0.3$). The first population that had a high level of outcrossing for more generations than the second one will have the lower $\hat{t}_f$ value (0.18 vs. 0.22). As experimentally demonstrated for the SSD populations analyzed (Table 3), our method thus avoids this pitfall that could lead to wrong decisions in population management.

In wheat bulk populations, MIH estimators detected significant temporal variation in outcrossing rates in eight populations out of nine. In the controlled populations (SR and SSD), the MIH estimators were close to the theoretical expected values. In all experimental wheat populations, the most recent outcrossing rate was estimated to be lower than the previous ones. This may indicate that a considerable year-to-year effect exists for outcrossing in these populations, as already shown for commercial cultivars sensitive, for example, to low light

levels (for example, the Moulin variety in DEMOTES-MAINARD et al. 1996). However, analyses of climatic conditions do not clearly explain why the 1994 season involved more selfing than previous years. Heterotic selection (positive correlation between heterozygosity and fitness; see DAVID 1998 for example) could explain this situation. Since it decreases the reduction in heterozygosity by favoring the most heterozygous genotypes, heterotic selection leads to overestimation of t_p, whereas t_0 estimates remain unaffected (since genotypes were sampled prior to juvenile selection). This hypothesis is consistent with other experiments that will be presented elsewhere.

For the short period of time relevant here (two to four generations), low temporal variation of allelic frequencies is expected in populations of reasonable size and should thus only slightly affect the outcrossing estimates, especially since mean diversity indices seem to be only slightly sensitive to allelic frequency variation, as found in *B. truncatus* (VIARD et al. 1997). Nevertheless, attention should be paid to populations of low effective size, rapidly evolving under strong natural selection or submitted to migration. A misleading situation would also be the pooling into a single sample of strongly spatially structured populations. In this case, heterozygous deficiency could lead to the underestimation of outcrossing (as would homogamy or crosses between relatives, as shown for *Secale cereale*; PEREZ DE LA VEGA and ALLARD 1984).

CONCLUSION

Multilocus individual heterozygosity has been used here to analyze the historical outcrossing rates of populations, yielding a better understanding of the dynamics of plant populations. The availability of highly polymorphic codominant markers, such as microsatellites (TAUZ et al. 1986), now makes it possible to envisage powerful MIH studies within realistic experiments. As progeny array analysis is not always possible for practical reasons, our procedure offers an alternative to estimation based on inbreeding level. Further theoretical developments are still needed for formal descriptions of the estimator's properties, particularly in the case of using loci in linkage disequilibrium and heterotic selection.

Decomposing a population into selfing classes could also be used to study heterosis *in situ*. For example, if individual fitness can be measured *in situ* and individuals can be genotyped to estimate their probabilities of belonging to different selfing classes, weighted mean fitnesses of successive selfing classes can provide information about inbreeding depression. Additionally, the study of two successive generations makes it possible to measure correlations between frequencies of S_n classes and S_{n-1} classes in the previous generation. Any correlation not fitted with the value expected from the final

generation outcrossing rate will measure a selection effect.

LITERATURE CITED

Barrett, S. C. H., B. C. Husband and W. W. Cole, 1993 Variation in outcrossing rates in *Eichhornia paniculata*: Temporal changes in populations of contrasting style morph structure. Plant Species Biol. **8:** 141–148.

Bennett, J. H., and F. E. Binet, 1956 Association between mendelian factors with mixed selfing and random mating. Heredity **10:** 51–55.

Brown, A. H. D., and L. Albrecht, 1980 Variable outcrossing and the genetic structure of predominantly self-pollinated species. J. Theor. Biol. **82:** 591–606.

Brown, A. H. D., and R. W. Allard, 1970 Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. Genetics **66:** 133–145.

David, J. L., M. Zivy, M. L. Cardin and P. Brabant, 1997 Protein evolution in dynamically managed populations of wheat: adaptative responses to macro-environmental conditions. Theor. Appl. Genet. **95:** 932–941.

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

Demotes-Mainard, S., G. Doussinault and J. M. Meynard, 1995 Effect of low radiation and low temperature at meiosis on pollen viability and grain set in wheat. Agronomie **15:** 357–365.

Demotes-Mainard, S., G. Doussinault and J. M. Meynard, 1996 Abnormalities in the male developmental programme of winter wheat induced by climatic stress at meiosis. Agronomie **16:** 505–515.

Enjalbert, J., C. Bœuf, H. Belcram and P. Leroy, 1999a Use of multiparental inbred populations to determine allelic relationships of molecular markers. Plant Breed. **118:** 88–90.

Enjalbert, J., I. Goldringer, S. Paillard and P. Brabant, 1999b Molecular markers to study genetic drift and selection in wheat populations. J. Exp. Bot. **332:** 283–290.

Godt, M. J. W., and J. L. Hamrick, 1991 Estimates of outcrossing rates in *Lathyrus latifolius* populations. Genome **34:** 988–992.

Hayes, H. K., 1918 Natural crossing in wheat. A cause of impurities in breeding plots. Belief of some agronomists that hybrids frequently revert to the parental type. J. Hered. **9:** 326–330.

Krueger, S. K., and S. J. Knapp, 1991 Mating systems of *Cuphea laminuligera* and *Cuphea lutea*. Theor. Appl. Genet. **82:** 221–226.

Li, H. B., Q. Zhang, A. M. Liu, J. S. Zou and Z. M. Chen, 1996 A genetic analysis of low-temperature-sensitive sterility in indica-japonica rice hybrids. Plant Breed. **115:** 305–309.

Mandal, A. K., and R. A. Ennos, 1995 Mating system analysis in a natural population of *Acacia nilotica* subspecies *kraussiana*. Forest Ecol. Manage. **79:** 235–240.

Moran, G. F., J. C. Bell and A. R. Griffin, 1989 Reduction in levels of inbreeding in a seed orchard of *Eucalyptus regnans* F. Muell. compared with natural populations. Silvae Genet. **38:** 32–36.

Murawski, D. A., J. L. Hamrick, S. P. Hubbell and R. B. Foster, 1990 Mating systems of two Bombacaceous trees of a neotropical moist forest. Oecologia **82:** 501–506.

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

Penteado, M. I. D. O., P. Garcia and M. Perez De La Vega, 1996 Genetic variability and mating system in three species of the genus Centrosema. J. Hered. **87:** 124–130.

Perez de la Vega, M., and R. W. Allard, 1984 Mating system and genetic polymorphism in populations of *Secale cereale* and *S. vavilovii*. Can. J. Genet. Cytol. **26:** 308–317.

Ritland, K., 1983 Estimation of mating systems, pp. 289–302 in *Isozymes in Plant Genetics and Breeding*, edited by S. D. Tanksley and T. J. Orston. Elsevier, Amsterdam.

Ritland, K., 1984 The effective proportion of self fertilisation with consanguineous matings in inbred populations. Genetics **106:** 139–152.

Ritland, K., and S. K. Jain, 1981 A model for the estimation of outcrossing rate and gene frequencies using *n* independent loci. Heredity **47:** 35–52.

Rousset, F., and M. Raymond, 1995 Testing heterozygote excess and deficiency. Genetics **140:** 1413–1419.

Schoen, D. J., and A. H. D. Brown, 1991 Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proc. Natl. Acad. Sci. USA **88:** 4494–4497.

Shaw, D. V., and R. W. Allard, 1982 Estimation of outcrossing rates in Douglas-fir using isozymes markers. Theor. Appl. Genet. **62:** 113–120.

Shaw, D. V., A. L. Khaler and R. W. Allard, 1981 A multilocus estimation of mating system parameters in plant populations. Proc. Natl. Acad. Sci. USA **78:** 1298–1302.

Sproule, A. T., and B. P. Dancik, 1996 The mating system of black spruce in North-Central Alberta, Canada. Silvae Genet. **45:** 159–164.

Tauz, D., M. Trick and G. A. Dover, 1986 Cryptic simplicity in DNA is a major source of genetic variation. Nature **322:** 652–656.

Viard, F., F. Justy and P. Jarne, 1997 Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncatus*. Genetics **146:** 973–982.

Weir, B. S., 1990 *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.

Wright, S., 1969 *Evolution and the Genetics of Populations, Vol. II: The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

Communicating editor: A. H. D. Brown

APPENDIX A

Likelihood of individual x of genotype G_x could be written as

$$L(x) = t_0 \prod_{l=1}^{L} D_l^{a_l}(1 - D_l)^{1-a_l}$$
$$+ \sum_{n=1}^{N}\left(t_{-n}\prod_{i=1}^{n-1}(1 - t_{-i})\prod_{l=1}^{L}\left(\frac{D_l}{2^n}\right)^{a_l}\left(1 - \frac{D_l}{2^n}\right)^{1-a_l}\right) + R.$$

The mating behavior is estimated for z previous generations and outcrossing rates of all generations before generation z are assumed to be constant and equal to t_p. Then $z + 1$ outcrossing rates $(t_0, \ldots, t_{-z+1}, t_p)$ must be estimated jointly by the maximum-likelihood technique. Hence, $L(x)$ can be developed as

$$L(x) = t_0 \prod_{l=1}^{L} D_l^{a_l}(1 - D_l)^{1-a_l}$$
$$+ \sum_{n=1}^{z-1}\left(t_{-n}\prod_{i=0}^{n-1}(1 - t_{-i})\prod_{l=1}^{L}\left(\left(\frac{D_l}{2^n}\right)^{a_l}\left(1 - \frac{D_l}{2^n}\right)^{1-a_l}\right)\right).$$
$$+ \sum_{m=z}^{N}\left(t_p(1 - t_p)^{m-z}\prod_{i=0}^{z-1}(1 - t_{-i})\prod_{l=1}^{L}\left(\left(\frac{D_l}{2^m}\right)^{a_l}\left(1 - \frac{D_l}{2^m}\right)^{1-a_l}\right)\right) + R,$$

with $R = \displaystyle\sum_{m=N}^{\infty}\left(t_p(1 - t_p)^{m-z}\prod_{i=0}^{z-1}(1 - t_{-i})\prod_{l=1}^{L}\left(\left(\frac{D_l}{2^m}\right)^{a_l}\left(1 - \frac{D_l}{2^m}\right)^{1-a_l}\right)\right).$ (A1)

For highly selfing classes $(m > N \approx 20)$,

$$1 - \frac{D_l}{2^m} \approx 1, \quad \text{thus } R = \left(t_p\prod_{i=0}^{z-1}(1 - t_{-i})\right)\prod_{l=1}^{L}(D_l)^{a_l}\left(\sum_{m=N}^{\infty}(1 - t_p)^{m-z}\prod_{l=1}^{L}\left(\frac{1}{2^m}\right)^{a_l}\right).$$

If h is the number of heterozygous loci in the genotype, that is, $h = \Sigma_{l=0}^{L} a_l$, then

$$\sum_{m=x}^{\infty} (1 - t_p)^x \prod_{l=1}^{L} \left(\frac{1}{2^m}\right)^{a_l} = \sum_{m=x}^{\infty} \left(\frac{1 - t_p}{2^h}\right)^m = \frac{1 - ((1 - t_p)/2^h)^{x+1}}{1 - (1 - t_p)/2^h},$$

leading to

$$R \approx t_p \prod_{n=0}^{z-1} (1 - t_{-n}) \frac{((1 - t_p)/2^h)^{N-z+1}}{1 - (1 - t_p)/2^h} \prod_{l=1}^{L} D_l^{a_l}. \quad \text{(A2)}$$

Extracting the constant terms $\Pi_{l=1}^{L} D_l^{a_l}$ from Equations A1 and A2,

$$L'(x) = t_0 \prod_{l=1}^{L} (1 - D_l)^{1-a_l} + \sum_{n=1}^{N} \left(\frac{1}{2^{nh}} t_{-n} \prod_{i=0}^{n-1} (1 - t_{-i}) \prod_{l=1}^{L} \left(1 - \frac{D_l}{2^n}\right)^{1-a_l}\right)$$
$$+ t_p \prod_{n=0}^{z-1} (1 - t_{-n}) \frac{((1 - t_p)/2^h)^{N-z+1}}{1 - (1 - t_p)/2^h}. \quad \text{(A3)}$$

For computational ease, the logarithm of Equation A3 was used in our program. This C program calculating t estimates could be downloaded at ftp://moulon.in-ra.fr/pub/moulon/enj.

APPENDIX B

At inbreeding equilibrium,

$$f = 1 - \frac{H_o}{H_i} = \frac{1 - t_f}{1 + t_f},$$

where H_o is the observed frequency of heterozygotes, and H_i their expected frequency under random mating. Considering a population outcrossing at a constant rate t_p excepted in the last generation (t_0), then

$$f' = 1 - \frac{H_o'}{H_i} = 1 - \frac{H_o(1 - t_0)/2 + t_0 H_i}{H_i} = \frac{1 - t_0}{1 + t_p}$$

$$t_f = \frac{t_0 + t_p}{2 + t_p - t_0}.$$

APPENDIX C

Sampling variance over each selfing class could be derived from Equation 2:

$$Q_0 = t_0 \quad \text{and} \quad \hat{t}_0 = \hat{Q}_0$$

$$Q_1 = t_{-1}(1 - t_0) \quad \text{then} \quad \hat{t}_{-1} = \frac{\hat{Q}_1}{1 - \hat{Q}_0}$$

$$Q_2 = t_{-2}(1 - t_{-1})(1 - t_0) = t_{-2}\left(1 - \frac{Q_1}{1 - Q_0}\right)(1 - Q_0)$$

$$\text{then} \quad \hat{t}_{-2} = \frac{Q_2}{1 - \hat{Q}_0 - \hat{Q}_1}.$$

By recurrence,

$$\hat{t}_{-n} = \frac{\hat{Q}_n}{1 - \sum_{k=0}^{n-1} \hat{Q}_k} = \frac{\hat{Q}_n}{1 - \hat{Q}_{<n}}.$$

Mean and variance of t_{-n} could be approximated using Taylor's second-order expansion (*i.e.*, the Delta method; WEIR 1990, p. 44). All moments of $\hat{Q}_n$ derive from moments of binomial or multinomial distributions,

$$E(\hat{t}_{-n}) = \frac{E(\hat{Q}_n)}{E(\hat{Q}_{\geq n})} + \frac{E(\hat{Q}_n)\text{Var}(Q_{\geq n})}{E^3(\hat{Q}_{\geq b})} - \frac{\text{Cov}(\hat{Q}_n, \hat{Q}_{\geq n})}{E^2(Q_{\geq n})}$$
$$= \frac{Q_n}{Q_{\geq n}} + \frac{Q_n Q_{\geq n}(1 - Q_{\geq n})}{N Q_{\geq n}^3} - \frac{Q_n(1 - Q_{\geq n})}{N Q_{\geq n}^2},$$

where N is the number of sampled individuals. Thus $E(\hat{t}_{-n}) = Q_n/Q_{\geq n} = t_{-n}$ as expected:

$$V(\hat{t}_{-n}) \approx \frac{V(\hat{Q}_n)}{E^2(\hat{Q}_{\geq n})} + \frac{V(\hat{Q}_{\geq n})E^2(\hat{Q}_n)}{E^4(\hat{Q}_{\geq n})} - 2\frac{E(\hat{Q}_n)\text{Cov}(\hat{Q}_n, \hat{Q}_{\geq n})}{E^3(\hat{Q}_{\geq n})}$$

$$V(\hat{t}_{-n}) \approx \frac{Q_n(1 - Q_n)}{N Q_{\geq n}^2} + \frac{Q_n^2 Q_{\geq n}(1 - Q_{\geq n})}{N Q_{\geq n}^4} - 2\frac{Q_n^2(1 - Q_{\geq n})}{N Q_{\geq n}^3} \approx \frac{Q_n Q_{\geq n+1}}{N Q_{\geq n}^3}.$$

In an infinite population in inbreeding equilibrium with a constant outcrossing rate t, it could be shown that $Q_n = t(1 - t)^n$. In this simple case, variance of $\hat{t}_{-n}$ is

$$V(\hat{t}_{-n}) \approx \frac{t}{N(1 - t)^{n-1}}.$$