

INTRAPOPOPULATION FIXATION INDEX DYNAMICS IN FINITE POPULATIONS WITH VARIABLE OUTCROSSING RATES

Alexandre Siqueira Guedes Coelho¹; Roland Vencovsky^{2*}

¹Depto. de Genética - USP/ESALQ, C.P. 83 - 13400-970 - Piracicaba, SP - Brasil.

²CAPES scholar.

*Corresponding author <rvencovs@esalq.usp.br>

ABSTRACT: The intrapopulation fixation index (f) is inversely related to the outcrossing rate (t). Results obtained from data on molecular markers of natural populations have shown that these values are highly variable, even when measured in the same group of individuals. It is thus suggested that factors besides those described in Wright's genetic equilibrium must be operating. Using simulated data sets this study shows that the finite size condition of a population is sufficient to spread the estimated f values along a range at equilibrium, as opposed to keeping them at the theoretical equilibrium point. The variation in outcrossing rates can amplify this range considerably. Correlation between estimated f values obtained from different loci in this condition showed to be negatively related to the outcrossing rates, and positively related to the variance of these rates along generations. The finite size of populations associated to small fluctuations in t mean values over time may explain the usually reported high variation among estimated f values of different loci.

Key words: population genetics, inbreeding, computer simulation

DINÂMICA DO ÍNDICE DE FIXAÇÃO INTRAPOPOPULACIONAL EM POPULAÇÕES FINITAS COM TAXAS VARIÁVEIS DE FECUNDAÇÃO CRUZADA

RESUMO: O fato de que os valores do índice de fixação intrapopulacional (f) e os das taxas de fecundação cruzada (t) mantêm uma estreita relação entre si é amplamente conhecido. Resultados das análises de marcadores moleculares em populações naturais têm demonstrado uma elevada variação destes valores para locos diferentes avaliados em uma mesma população, sugerindo que outros fatores, além daqueles descritos no modelo de equilíbrio genético de Wright devem estar operando. Pelo uso de simulações, demonstra-se neste trabalho que a condição finita de uma população é suficiente para que os valores estimados de f passem a se estabilizar ao longo de um intervalo de variação e não mais em um único ponto. Demonstra-se ainda que variações nas taxas de fecundação cruzada ao longo das gerações amplificam substancialmente a magnitude deste intervalo. A correlação entre os valores estimados de f obtidos de locos diferentes nestas condições mostrou-se dependente dos valores das taxas de fecundação cruzada e da magnitude da variância destas taxas entre gerações, podendo ser nula sob condições de panmixia e taxas constantes. A condição finita das populações associada a pequenas flutuações na taxa média de fecundação cruzada de diferentes gerações, condições tipicamente encontradas na natureza, podem explicar as discrepâncias entre os valores estimados de f de locos diferentes, comumente reportadas na literatura. Há possibilidade de que a magnitude da variância entre os valores estimados de f de diferentes locos forneça uma estimativa do número efetivo de indivíduos reprodutivamente ativos em uma dada população.

Palavras-chave: genética de populações, endogamia, simulação computacional

INTRODUCTION

In very large populations, reproducing by panmixis, in the absence of selection, migration and mutation, the allele and genotype frequencies are kept in equilibrium, and do not change along generations, keeping a relation described by (Crow & Kimura, 1970):

$$\begin{cases} P_{uu} = p_u^2 \\ P_{uv} = 2p_u p_v \end{cases}$$

in which: p_u is the frequency of allele u ; P_{uu} is the frequency of the homozygous genotype for allele u ; P_{uv} is the frequency of heterozygous genotype for the u and v alleles. These relations are known as the Hardy-Weinberg law and were postulated independently by G. H. Hardy and W. Weinberg in 1908.

Wright (1921; 1922) demonstrated that an equilibrium is also attained if, when keeping the other Hardy-Weinberg conditions, the population reproduces under a fixed outcrossing rate (t) instead of reproducing by

panmixis. Relationships among allele and genotype frequencies in this case are altered and are modified to:

$$\begin{cases} P_{uu} = p_u^2 + p_u(1 - p_u)f \\ P_{uv} = 2p_u p_v(1 - f) \end{cases}$$

in which f is the intrapopulation fixation index (Wright, 1951; 1965). These relations are known as the Wright equilibrium relations and assume great importance in the genetic characterization of natural populations, since they provide not only an indicator parameter of the degree of structure of the genetic variability at the individual level, but also a method to obtain outcrossing rate estimates in natural conditions, given a single generation data set (Fyfe & Bailey, 1951). This is possible because of the fact that under Wright's equilibrium conditions the f and t values keep a relation that may be expressed by (Bennet & Binet, 1956):

$$f = \frac{1-t}{1+t}$$

The extensive use of this knowledge, however, was delayed until the development of genotype evaluation techniques at the molecular level, carried out along the last years. The development of molecular marker techniques along the last 30 years, and notably during the last 20 years, resulted in an actual revolution in the capacity of genotype evaluation under natural conditions. These techniques allowed the use of a much broader variety of species in genetic studies in contrast to the use of a few model species in the past.

Papers reporting the genetic characterization of natural populations are now abundant, many of which providing highly variable estimates for the intrapopulation fixation index obtained from information on different loci. In this respect, one must recognize that variation among different populations/species reflect different population structures and reproductive strategies that evolved under different ecological pressures (Loveless & Hamrick, 1984). On the other hand, variation in the same population along generations or over different loci, must reflect different aspects of its dynamic, being then a consequence of the demographic stochastic process and genetic drift (Gillespie, 1998).

The present paper evaluates, through computer simulation, the effects resulting from the finite condition of populations and those resulting from variation in the outcrossing rates, along generations and among individuals of a given generation, over the estimated values of the intrapopulation fixation index. Information herein produced might be valuable for the development of methods that would allow a better characterization of the dynamics of the genetic properties of populations, and of their implications in the recognition of better conservation and management strategies.

METHODOLOGY

In a way to allow the simulation and evaluation of a diverse set of conditions, a computer program (available from the authors) based on the Delphi/Pascal language was developed. Computer simulations were carried out using the following algorithm: 1. attribute to each of the L loci of the N individuals a given genotype; 2. identify an individual at random (i); 3. attribute to the outcrossing rate a given value (t); 4. generate a random number from a $[0, 1]$ uniform distribution (x); 5. if $x < t$, obtain a second random individual (j), if not, take $j=i$; 6. for each of the L loci, take one allele at random from i and combine it with a randomly taken allele from j ; 7. repeat 2 to 6 steps N times, providing the N individuals that will constitute the next generation; 8. repeat step 7 for the number of desired generations.

Initial genotypes were obtained from expected genotype frequencies under Hardy-Weinberg equilibrium, considering two equally frequent alleles per locus. Data were obtained after 20 generations, when Wright's equilibrium conditions prevail. The effects of the following factors were evaluated: a) population size; b) variation in the generation mean outcrossing rate; c) variation among the outcrossing rates of different individuals from a given generation.

The effects of the variation of the outcrossing rates, both among individuals from the same generation and among different generations, were evaluated admitting that the outcrossing rate is a Beta distributed random variable (t), restricted to the $[0, 1]$ interval, displaying a Beta density function given by:

$$\phi(t) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1} (1-t)^{\beta-1}$$

in which: $\Gamma(\cdot)$ is the gamma function and α and β are distribution parameters, being $\alpha > 0$ and $\beta > 0$.

Mean and variance of t are then given, respectively, by:

$$\bar{t} = \frac{\alpha}{\alpha + \beta} \quad Var(t) = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

Parameters α and β used in simulation were defined in a way to characterize species showing different mating systems. Simulations were carried out under conditions typical for autogamous species ($\bar{t} = 0.05$), mixed mating species ($\bar{t} = 0.50$) and allogamous species ($\bar{t} = 0.95$). Beta distributions with variances equal to 0.000025 and 0.0025 were established for each case (Figure 1). Simulations with variance equal to 0.025 were carried out specifically for $\bar{t} = 0.50$, because of the fact that for mean outcrossing rates equal to 0.05 and 0.95, the correspond-

ing density functions would represent a behavior hardly found in the nature, in which the most likely values of t would be the extreme ones, simultaneously. For each simulated case, the α and β values were obtained from the mean and variance of t values using the relations:

$$\alpha = \frac{t^2[1-t]}{Var(t)} - \bar{t} \quad \beta = \frac{\alpha}{t} - \alpha$$

Simulated values for the Beta distribution were generated following a procedure adapted from Dagpunar (1988).

Considering the magnitude of the evaluated populations, the intrapopulation fixation index values were estimated using the traditional expression:

$$\hat{f} = \frac{\tilde{h} - \tilde{H}}{\tilde{h}}$$

in which: \tilde{h} is the expected heterozygosity under Hardy-Weinberg equilibrium obtained directly from sample observed allele frequencies and \tilde{H} is the sample observed heterozygosity.

RESULTS AND DISCUSSION

Considering the dynamics of the fixation index under the different simulated conditions, the estimated values of f (\hat{f}) reach the equilibrium condition rather quickly, this condition being represented in certain cases by an interval of variation and not anymore by a single value (Figures 2, 3 and 4). The \hat{f} values obtained showed an increasing magnitude of variation for decreasing population sizes, exception being made to the simulation carried out with $t=0.00$. Variation in mean outcrossing rates along different generations also noticeably influenced the dynamic of f values obtained in simulations. The magnitude of variation of \hat{f} showed to be positively associated to variation in t values. On the other hand, situations in which the outcrossing rates were variable only among individuals of the same generation, the mean values kept constant along generations and with a population size of 1,000,000 individuals, did not result in a detectable variation of \hat{f} values along generations.

Variation of outcrossing rates (t) along generations resulted in a stronger effect on the behavior of \hat{f} values than that derived from variation of t among different individuals (Tables 1, 2 and 3). The weaker influ-

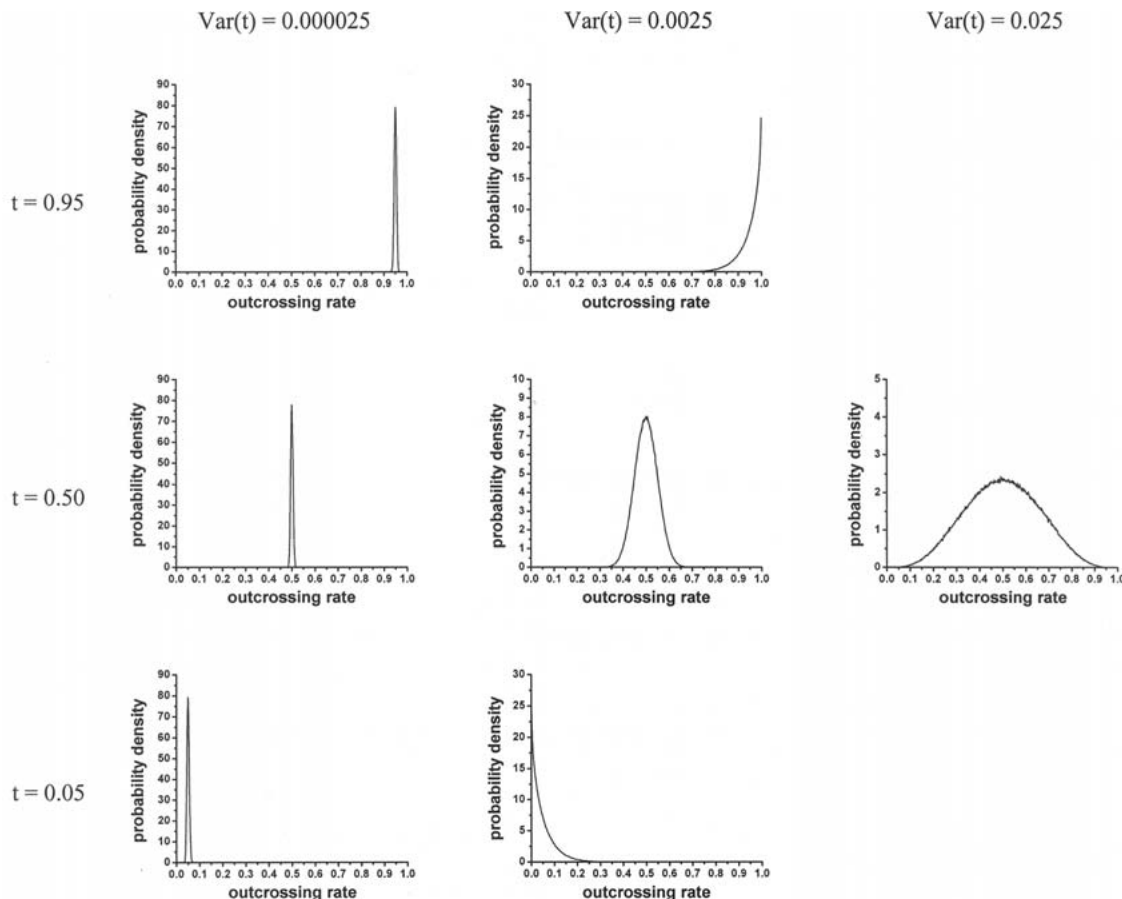


Figure 1 - Simulated probability distributions for different values attributed to the mean (\bar{t}) and to the variance [$Var(t)$] of the outcrossing rates.

Table 1 - Mean (\bar{f}) and variance ($s_{\hat{f}}^2$) of estimated values for the intrapopulation fixation index along 500 generations in finite populations with varying sizes (N) and different values for the outcrossing rates

\bar{t}	$N=100$		$N=10,000$		$N=1,000,000$	
	\bar{f}	$s_{\hat{f}}^2$	\bar{f}	$s_{\hat{f}}^2$	\bar{f}	$s_{\hat{f}}^2$
0.00	1.000	0.000×10^{-2}	1.000	0.000×10^{-5}	1.000	0.000×10^{-7}
0.25	0.596	1.284×10^{-2}	0.600	8.262×10^{-5}	0.600	7.555×10^{-7}
0.50	0.330	1.124×10^{-2}	0.334	10.883×10^{-5}	0.333	9.723×10^{-7}
0.75	0.139	1.136×10^{-2}	0.143	9.498×10^{-5}	0.143	9.425×10^{-7}
1.00	-0.006	1.049×10^{-2}	-0.001	10.326×10^{-5}	0.000	10.250×10^{-7}

Table 2 - Mean (\bar{f}) and variance ($s_{\hat{f}}^2$) of estimated values for the intrapopulation fixation index along 500 generations in populations with varying mean outcrossing rates along generations ($N=1,000,000$).

\bar{t}	$Var(t)=0.000025$		$Var(t)=0.0025$		$Var(t)=0.025$	
	\bar{f}	$s_{\hat{f}}^2$	\bar{f}	$s_{\hat{f}}^2$	\bar{f}	$s_{\hat{f}}^2$
0.05	0.904	29.864×10^{-6}	0.889	32.314×10^{-4}	-	-
0.50	0.333	12.773×10^{-6}	0.334	11.726×10^{-4}	0.336	12.773×10^{-2}
0.95	0.026	7.492×10^{-6}	0.024	5.724×10^{-4}	-	-

Table 3 - Mean (\bar{f}) and variance ($s_{\hat{f}}^2$) of estimated values for the intrapopulation fixation index along 500 generations in populations with varying outcrossing rates among individuals ($N=1,000,000$).

\bar{t}	$Var(t)=0.000025$		$Var(t)=0.0025$		$Var(t)=0.025$	
	\bar{f}	$s_{\hat{f}}^2$	\bar{f}	$s_{\hat{f}}^2$	\bar{f}	$s_{\hat{f}}^2$
0.05	0.905	2.035×10^{-7}	0.905	2.295×10^{-7}	-	-
0.50	0.333	9.387×10^{-7}	0.333	9.839×10^{-7}	0.333	9.663×10^{-7}
0.95	0.026	9.849×10^{-7}	0.026	8.999×10^{-7}	-	-

ence of the variation of outcrossing rates among individuals from the same generation over the dynamic of \hat{f} values along generations should, however, be considered carefully. Once the effect of the variation of mean outcrossing rates along generations over the behavior of \hat{f} values is recognized, one should consider that for small population sizes, variation of t values at individual level might considerably contribute to the increment in t value oscillation, and in this case, it would result in an indirect effect on the variance of \hat{f} values.

The detected variation in \hat{f} values does not include that derived from statistical sampling: resulted from the fact that only a portion of the actual population was evaluated. In this sense, even in situations in which the evaluation of every individual of a given population is possible, an oscillation in the \hat{f} values from different loci along generations must be expected, resulting from the instability of the outcrossing rates and from the events of genetic sampling of gametes.

Figure 5 shows the observed probability distributions for the intrapopulation fixation index values

obtained along generations in populations of 10,000 individuals, with different outcrossing rates. These distributions, obtained with constant outcrossing rates among individuals and generations, exhibit strong similarity with Beta probability density functions defined for the interval $(0, N)$, with means equal to the simulated t values and variances given by the same expression of sampling variance of \hat{t} , reported by Vencovsky (1994):

$$Var(\hat{t}) = \frac{4}{(1+f)^4} Var(\hat{f})$$

with

$$Var(\hat{f}) = \frac{(1-f)^2(1-2f)}{N} + \frac{f(1-f)(2-f)}{2Np_u(1-p_u)}$$

The probability density function of interest, expressed in terms of f is given by:

$$\phi(\hat{f}) = \frac{\Gamma(\alpha + \beta) \left(\frac{1-\hat{f}}{1+\hat{f}}\right)^{\alpha-1} \left(N - \frac{1-\hat{f}}{1+\hat{f}}\right)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta) N^{\alpha+\beta-1}} \frac{2}{(1+\hat{f})^2}$$

$$\beta = \frac{N\alpha(1+f)}{1-f} - \alpha$$

with parameters:

$$\alpha = \frac{1}{N} \left[\frac{(1-f)^2(1+f)^2 \left(N - \frac{1-f}{1+f}\right)}{4 \cdot \text{Var}(\hat{f})} - \frac{1-f}{1+f} \right]$$

The usefulness of the $(0, N)$ interval derives from the fact that the \hat{f} values are allowed to be negative, implying in \hat{t} values that may be larger than 1.0 (apparent outcrossing rate).

The similarity between these functions suggests that, in cases in which the outcrossing rates are kept relatively constant along generations, the behavior of \hat{f} val-

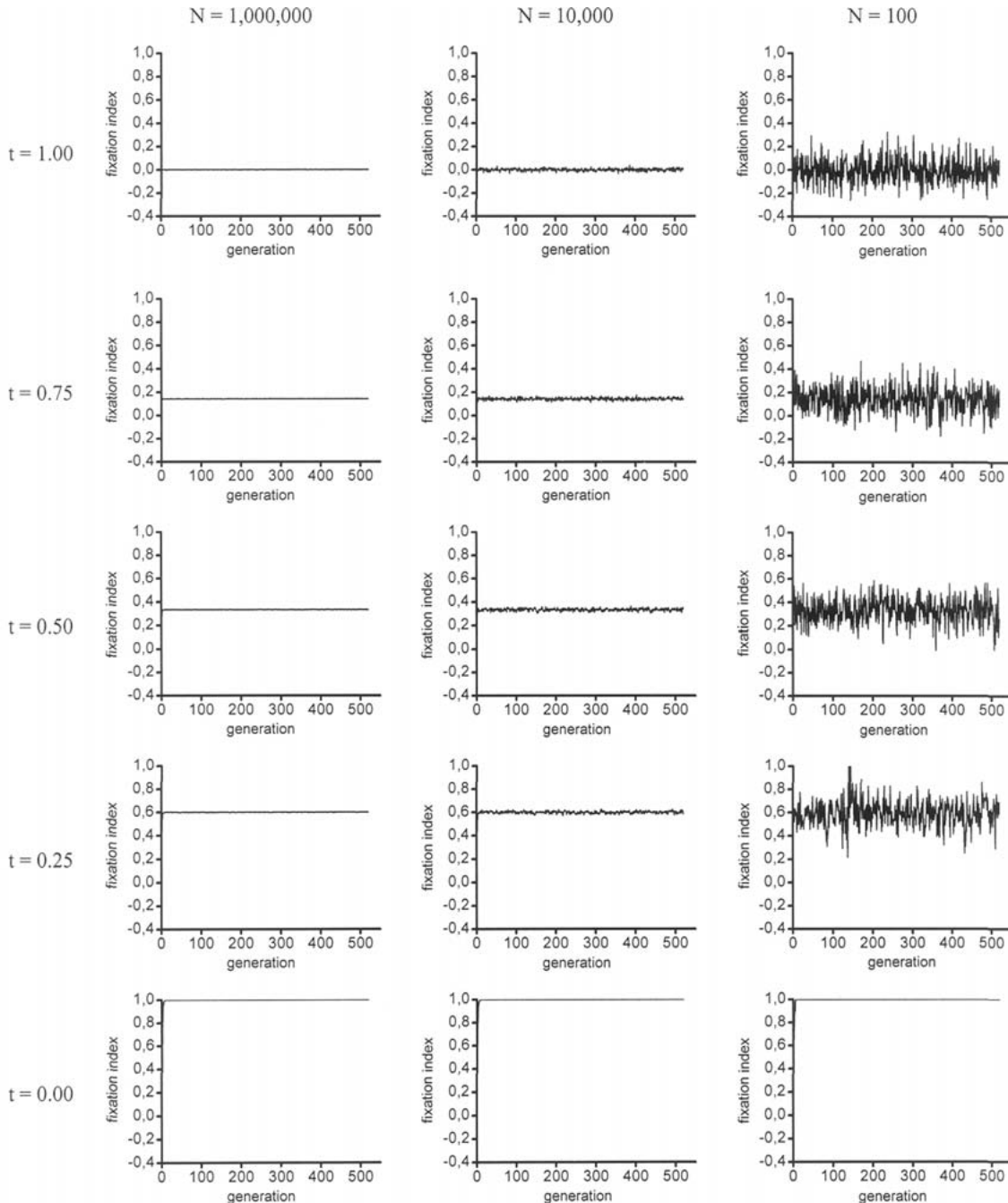


Figure 2 - Intrapopulation fixation index in populations of different sizes, with different outcrossing rates, along 520 generations.

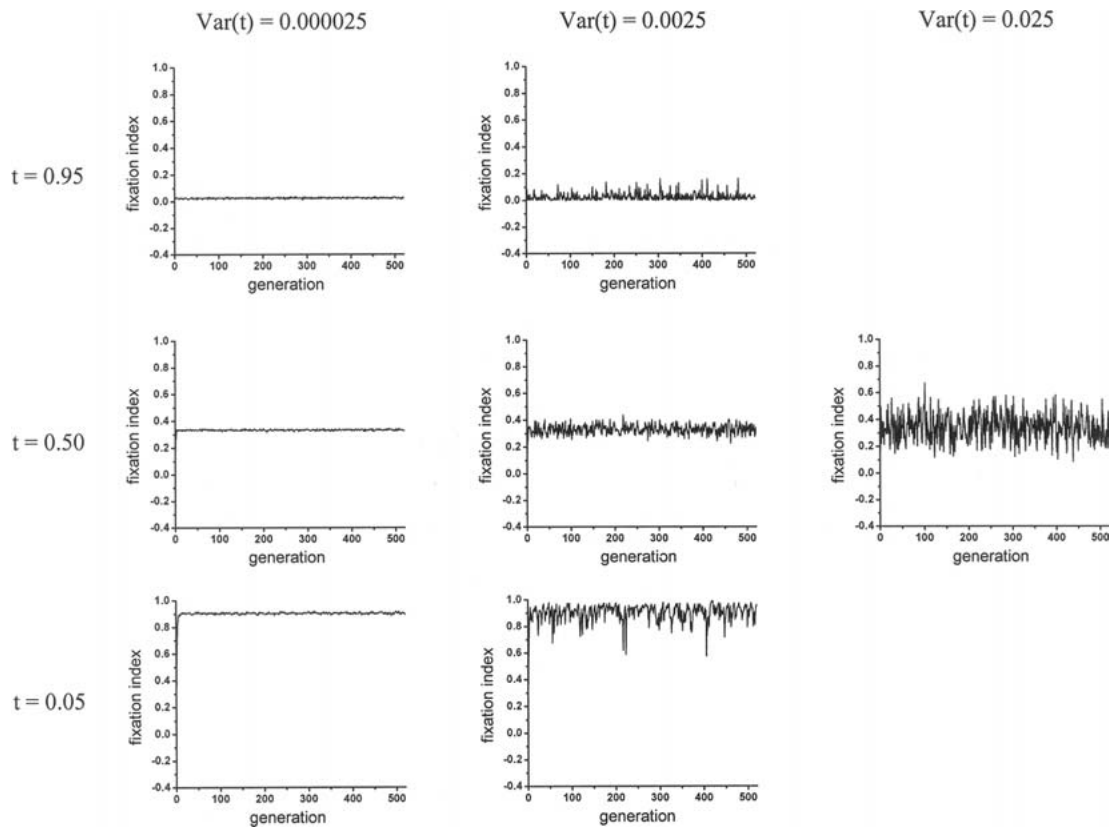


Figure 3 - Intrapopulation fixation index in populations exhibiting variation in outcrossing rates along generations, with mean given by t and variance $Var(t)$, along 520 generations ($N=1,000,000$).

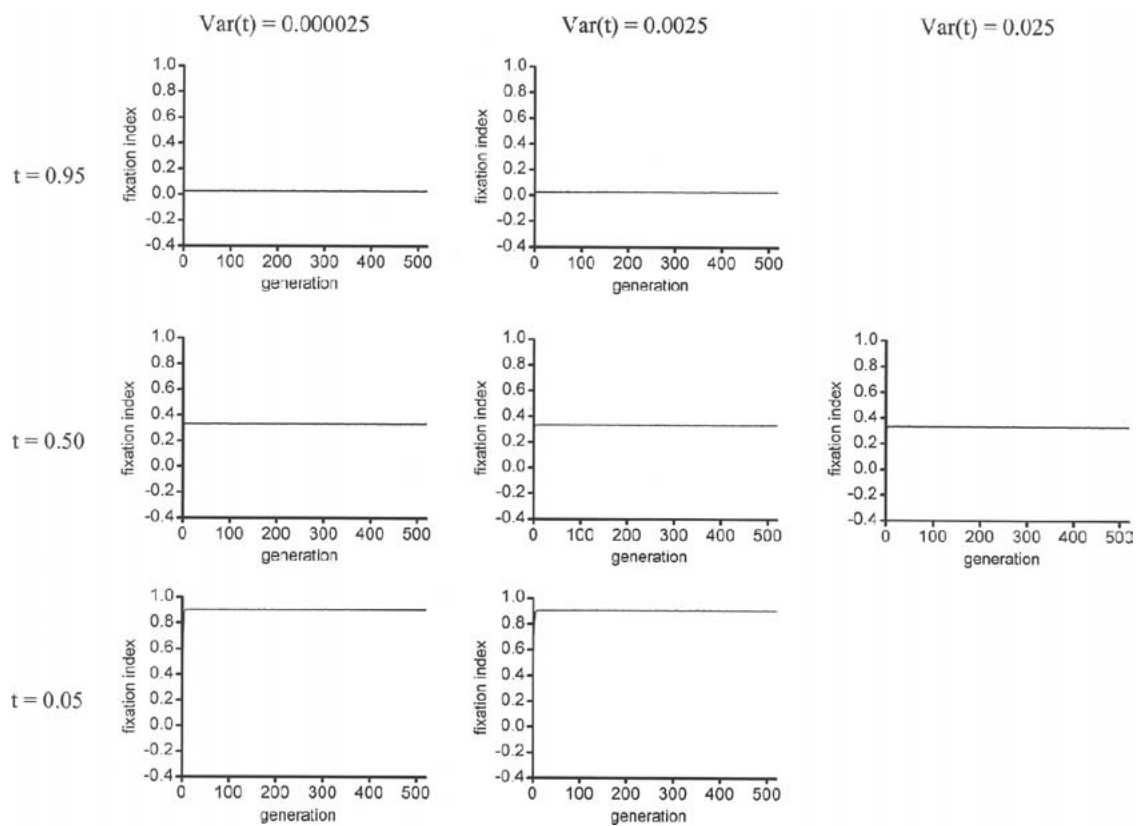


Figure 4 - Intrapopulation fixation index in populations exhibiting variation in outcrossing rates among individuals, with mean given by t and variance $Var(t)$, along 520 generations ($N=1,000,000$).

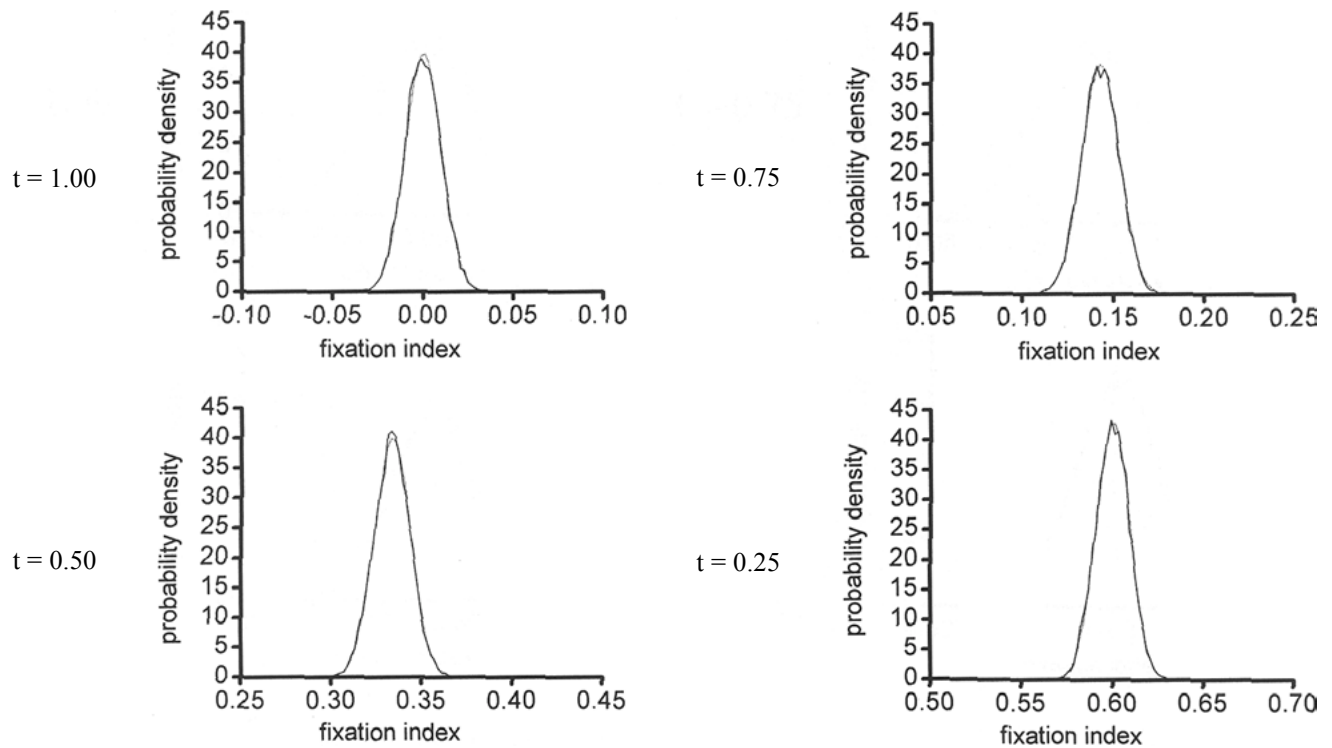


Figure 5 - Probability distributions of intrapopulation fixation index values in populations of 10,000 individuals, with different outcrossing rates, along 10,000 generations. Corresponding Beta probability density functions are also shown (dotted lines).

ues may be described using a Beta probability density function, defined by parameters that are function of the population size, of the f value correspondent to the actual outcrossing rate, and of the allele frequencies.

Figure 6 shows the observed probability distributions for the \hat{f} values obtained from 10,000 loci evaluated in a single generation, in populations of 1,000 individuals, with different outcrossing rates. The figure also shows the probability density functions defined in a similar manner to that used to describe the behavior of \hat{f} values in different generations. The fact that the dynamic behavior of \hat{f} values in both cases, a given locus along generations and different loci in a given generation, presented a similar pattern suggests that data from loci in a single generation might provide information on the single locus dynamics along generations. Although data from different generations are rare in the literature, information from different loci is not and could thus provide an estimate of the effective number of individuals reproductively active in a given population (N).

Figure 7 shows the predicted and observed values for the variance of \hat{f} among loci in different situations. In practical terms, the variance of \hat{f} values being considered does not include the effects resulting from statistical sampling, and would be better interpreted as a variance component of \hat{f} values associated to different loci, than as a mean square of the sample estimated values for \hat{f} .

Estimates of the correlation coefficient between \hat{f} values obtained from two different loci along generations are represented in Figure 8. The magnitude of the correlation between \hat{f} values in different loci along generations decreases with the increase in the outcrossing rate, reaching values close to zero under panmixis. Variation in the mean outcrossing rate along generations results in a detectable positive effect over this correlation (Figure 9). In a simulation of a population with 10,000 individuals, with a constant outcrossing rate of 0.95, the correlation of \hat{f} values of different loci along generations were close to zero. Inclusion of variance in mean outcrossing rates along generations increased the variance of \hat{f} values 9 fold, at the same time that it increased the correlation of \hat{f} from different loci to approximately 0.88. Although the correlation between \hat{f} values estimated for different loci increases with the variance of outcrossing rates along generations the heterogeneity among \hat{f} values from different loci in a given generation does not change, since it results from the genetic sampling resulting from the finite condition of populations.

From a practical point of view, the finite condition of populations might result in a pronounced variation in the intrapopulation fixation index values estimated from different loci in a given generation. Small populations, with dozens of individuals, and variation in outcrossing rates are conditions ubiquitous in nature

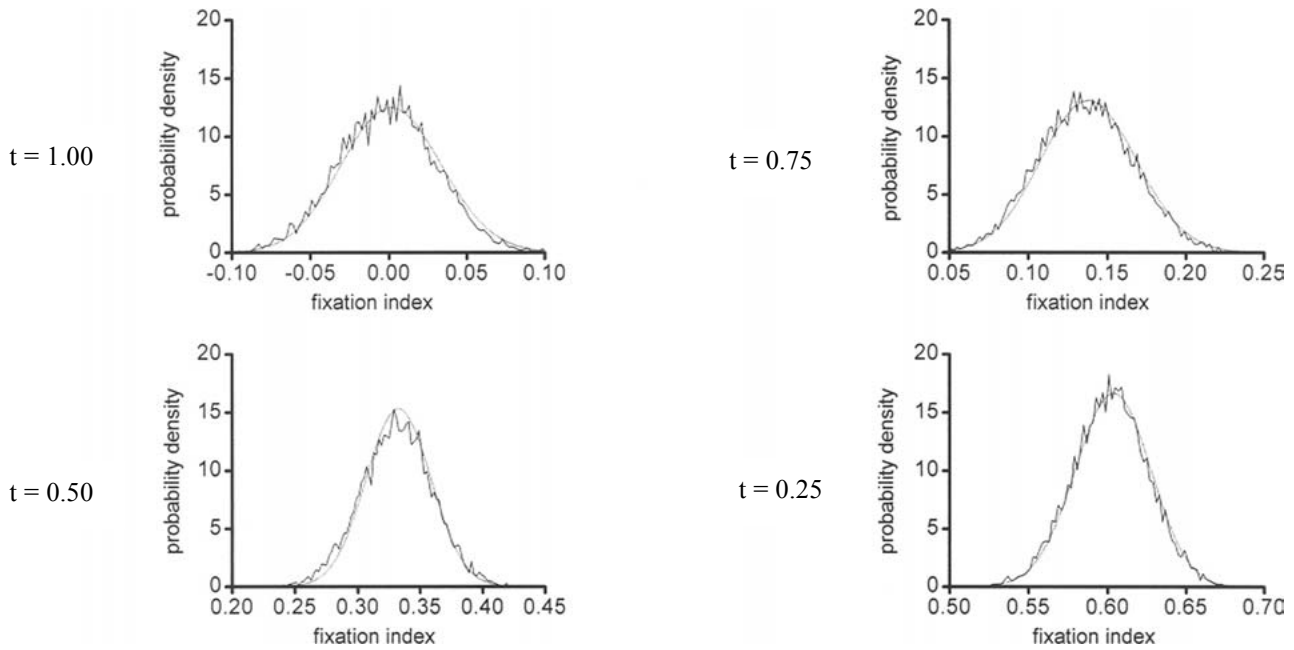


Figure 6 - Probability distributions of intrapopulation fixation index values from 10,000 loci in a population of 1,000 individuals, with different values of outcrossing rates. Corresponding Beta probability density functions are also shown (dotted lines).

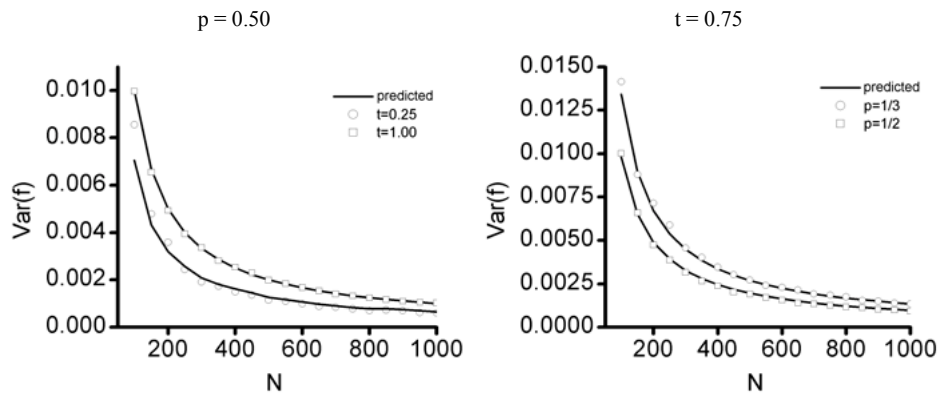


Figure 7 - Predicted and observed values for the variance of intrapopulation fixation index values [$Var(\hat{f})$] among loci, evaluated in a single generation of populations with different outcrossing rate values (t), allele frequencies (p) and population sizes (N). Values obtained from 10,000 loci.

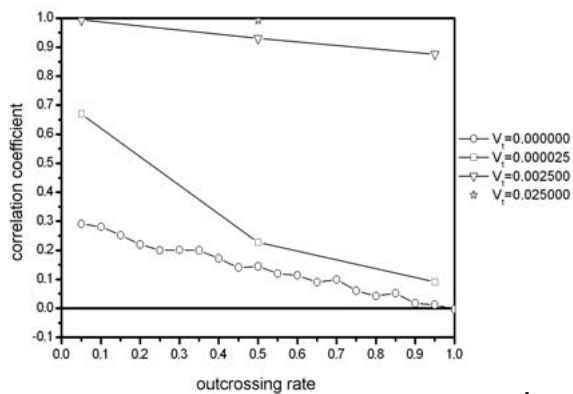


Figure 8 - Observed correlation coefficient values among \hat{f} values from different loci along 1,000 generations of populations with different values for the mean and variance of the outcrossing rates along generations ($N=10,000$).

and the resulting effect of these factors might be contributing to the high level of variation of \hat{f} values among loci evaluated in the same population, commonly reported in the literature.

As a final comment, the difficulty in performing a good characterization of intrapopulation fixation index values by the use of point estimates should be stressed. The dynamic behavior of this parameter at equilibrium would certainly be better characterized by a probabilistic approach in which probabilities of occurrence of different \hat{f} values could be evaluated. Without any loss of information, an approach of such a nature would allow not only to evaluate the \hat{f} mean value, but also its instability, that seems to be ubiquitous in natural conditions.

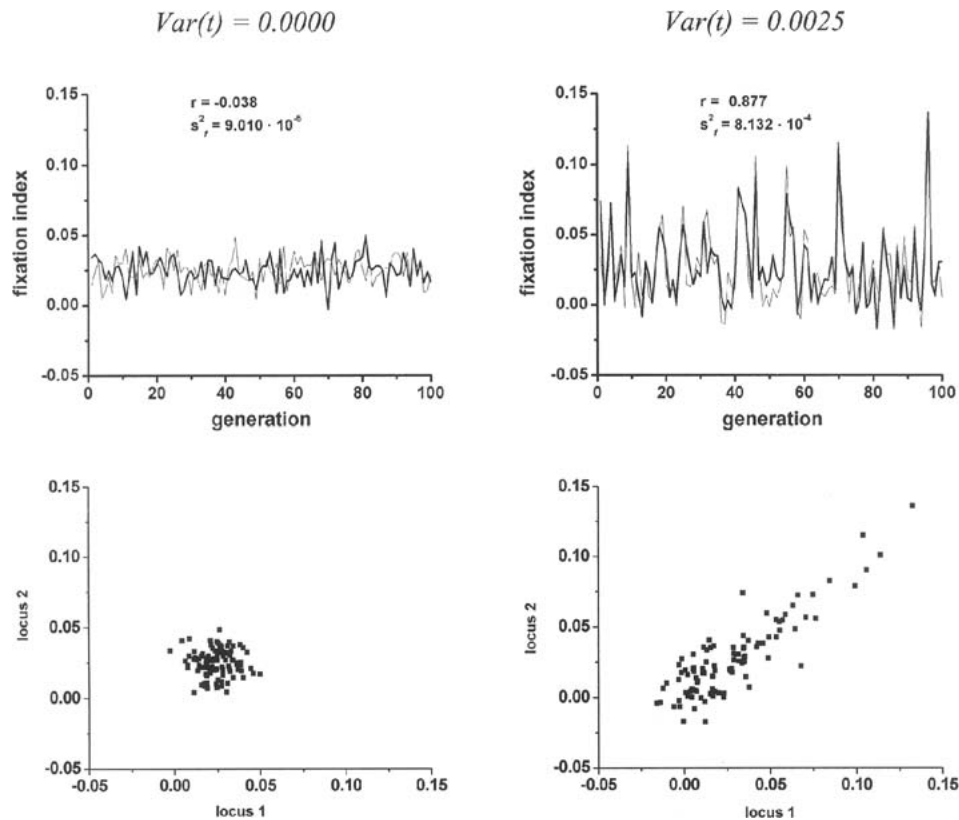


Figure 9 - Intrapopulation fixation index dynamics in two loci in two populations of the same size ($N=10,000$), same outcrossing rate ($\hat{t}=0.95$) but different values for the variance of t along generations.

CONCLUSIONS

The estimated values of the intrapopulation fixation index exhibit a dynamic behavior under equilibrium, that is strongly influenced by population size and variation in mean outcrossing rates along generations. The instability of \hat{f} values from different loci evaluated in the same population, commonly reported in the literature, might be explained by the finite condition of natural populations.

REFERENCES

- BENNET, J.H.; BINET, F.E. Association between Mendelian factors with mixed selfing and random mating. *Heredity*, v.10, p.51-55, 1956.
- CROW, J.F.; KIMURA, M. *An introduction to population genetics theory*. New York: Harper & Row, 1970. 591p.
- DAGPUNAR, J. *Principles of random variate generation*. Oxford: Clarendon Press, 1988. 248p.
- FYFE, J.L.; BAILEY, N.T.J. Plant breeding studies in leguminous forage crops: I. Natural cross-breeding in winter beans. *Journal of Agricultural Science*, v.41, p.371-378, 1951.
- GILLESPIE, J.H. *Population genetics: a concise guide*. Baltimore: Johns Hopkins University Press, 1998. 184p.
- LOVELESS, M.D.; HAMRICK, J.L. Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, v.15, p.65-95, 1984.
- VENCOVSKY, R. Variance of an estimate of outcrossing rate. *Brazilian Journal of Genetics*, v.17, p.349-351, 1994.
- WRIGHT, S. Systems of mating. *Genetics*, v.6, p.111-178, 1921.
- WRIGHT, S. Coefficients of inbreeding and relationship. *American Naturalist*, v.56, p.330-338, 1922.
- WRIGHT, S. The genetical structure of populations. *Annals of Eugenics*, v.15, p.323-354, 1951.
- WRIGHT, S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, v.19, p.395-420, 1965.

Received July 22, 2002