



Genetic structure of the European beech stands (*Fagus sylvatica* L.): F-statistics and importance of mating system characteristics in their evolution

Joël Cuguen, Dominique Merzeau, Bernard Thiebaut

► To cite this version:

Joël Cuguen, Dominique Merzeau, Bernard Thiebaut. Genetic structure of the European beech stands (*Fagus sylvatica* L.): F-statistics and importance of mating system characteristics in their evolution. *Heredity*, 1988, 60 (1), pp.91-100. 10.1038/hdy.1988.14 . hal-02020253

HAL Id: hal-02020253

<https://hal.science/hal-02020253>

Submitted on 15 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genetic structure of the European beech stands (*Fagus sylvatica* L.): F-statistics and importance of mating system characteristics in their evolution

Joël Cuguen,*§
Dominique Merzeau† and
Bernard Thiebaut*‡

* Unité de Biologie des Populations et des
Peuplements, Centre Louis Emberger, CNRS, Route
de Mende, BP 5051, 34033 Montpellier Cedex,
France.

† Département de Physiologie des Végétaux Ligneux,
Ecologie Génétique, Université de Bordeaux I, Av.
des Facultés, 33045 Talence, France.

‡ Laboratoire de Systématique et d'Ecologie
Méditerranéenne, Université des Sciences et
Techniques du Languedoc, rue Auguste Broussonet,
34000 Montpellier, France.

F-statistics are often used to study allozyme polymorphism in populations and the estimates provide a good basis to better understand selection and mating system effects. In this paper, computer simulation has been developed to determine the effects of the polymorphism level on F_{is} estimates in finite samples. It is shown that finite samples present mostly negative F_{is} estimates when allelic polymorphism is low. In these conditions, the high frequency of these apparent heterozygote excesses is not a consequence of frequency dependent selection acting on rare heterozygotes but only a statistical effect caused by the low probability of encountering rare homozygotes in the samples. The genetic structure of 250 European beech stands was studied using three electrophoretically detectable protein loci. *F*-statistics were estimated using the Weir and Cockerham method. The beech shows the highest interpopulation genetic differentiation estimates among the anemophilous pollinated forest trees. At the population level, an heterozygote deficit can be observed. Selfing estimates are not sufficiently high to explain these deficits. This suggests that limited gene flow exists in these populations and that forests of anemophilous pollinated trees are not panmictic and that these populations will be better understood using neighbourhood concepts.

INTRODUCTION

The knowledge of natural population genetic structure contributes to a better understanding of the role selection and the mating system play in their evolution. Since the theoretical works of Wright (1951, 1965), Cockerham (1969, 1973), Kirby (1975) and Nei (1977), genetic structures are often analysed using *F*-statistics.

The definitions of these parameters are based on the following hypothesis: genetic drift is the main evolutionary force which acts on the evolution of the studied characters. Inter- and intragroup genetic differentiation levels are a function of gene flow intensity in and between groups, and are also

a function of the number of generations since their separation. Three parameters were proposed to describe the properties of hierarchically subdivided natural populations (Wright 1951, 1965). These parameters were defined in terms of the total population (T), subdivisions (S) and individuals (I).

In a diploid individual, the two alleles of a gene may or may not be identical by descent (Malécot 1948). In a group of individuals, association between two identical alleles occurs with a given frequency. And this frequency varies from one group to another, according to the degree of dependence between identical alleles in a group. Therefore,

F_{is} is the average over all subdivisions of the correlation between identical alleles that unite to produce the individuals, relative to the gametes of their own subdivision,

§ Present address: Laboratoire de Génétique Ecologique et de Biologie des Populations Végétales, Université des Sciences et Techniques de Lille, 59655 Villeneuve D'Ascq Cedex, France.

F_{st} is the correlation between randomly chosen identical gametes within subdivisions, relative to gametes of the total population, "coancestry" in the sense of Cockerham (1969, 1973).

F_{it} is the correlation between uniting identical gametes, relative to those of the total population, "inbreeding" (Cockerham *ibidem*).

As pointed out by Wright, the list can be extended if there are further subdivisions. The above three F -statistics are not independent and Wright (1951) demonstrates that:

$$(1 - F_{it}) = (1 - F_{is})(1 - F_{st})$$

For example, when subdivisions are panmictic and isolated for a long time: $F_{is} = 0$ in each group and F_{st} is positive and increases with the number of generations. In this case, $F_{it} = F_{st}$.

However, when subdivisions are constituted of selfed individuals, the equilibrium values of F_{is} , F_{st} and F_{it} are one.

The theoretical points of view of Wright and Cockerham encounter difficulties in the study of natural populations: firstly, without pedigrees it is not possible to recognise alleles identical by descent. Secondly, these authors do not consider direct selection effects which can strongly affect genetic structures.

In natural populations, identity between alleles is generally estimated according to the different allelic forms of a polymorphic gene. But the relation between alleles identical by descent and alike in state ones is not simple. When two alleles are different, it is clear that they have no parental allele in common. But when they are alike, they can be identical by descent or only alike in state. The correctness of an estimate of identity in and between individuals may be increased by considering many polymorphic loci.

In a population and for a diallelic locus, the estimate of F_{is} is often calculated according to the following formula:

$$F_{is} = 1 - H/2p(1 - p)$$

where H is the observed heterozygote frequency in the population and $2p(1 - p)$ is the expected heterozygote frequency according to the Hardy-Weinberg law, calculated from allelic frequencies also estimated from the same sample.

This estimation shows a great sampling variance. Several authors have tried to estimate it (Rasmussen, 1964; Brown *et al.*, 1975; Vasek and Harding, 1976). Based on different hypotheses, results differ and are therefore difficult to interpret. That is why we chose another method for analysing

empirically the various biases due to sampling and the F_{is} estimation method. In order to do that, we performed sampling simulations on theoretical populations, perfectly known, and estimated F_{is} in each sample.

Furthermore, 250 beech stands have already been sampled in Europe in order to study the alloenzymatic polymorphism of beech (*Fagus sylvatica* L.). So we dispose of an important number of observations in many different natural populations.

F_{is} variations were examined in relation to sample polymorphism, in theoretical and natural populations. Then the organisation of those variations were compared between simulated and natural situations. So, in this paper, we will try to answer the following three questions:

- (a) considering the simulation results, how are F_{is} variations organised in relation to sample polymorphism?
- (b) In European beechstands, does the organisation of F_{is} variations depend on the estimation method or do other mechanisms count, particularly mating system characteristics?
- (c) In this case, what are the mating system characteristics which are responsible for the observed differences between simulations and natural populations: selfing and/or isolation by distance?

MATERIALS AND METHODS

Simulations

We performed sampling simulations on theoretical populations using the Monte-Carlo method.

Creation of theoretical populations

In theoretical populations, the mating system follows the mixed mating model. Populations are assumed to have reached the inbreeding equilibrium and for a diallelic locus, the proportions of the various genotypes are a function of allelic frequencies and the fixation index:

$$AA' = p'^2 + p'q'F'_{is}$$

$$aa' = q'^2 + p'q'F'_{is}$$

$$Aa' = 2p'q'(1 - F'_{is})$$

where AA' and aa' are respectively the homozygote frequencies for alleles A and a , Aa' the frequency of heterozygotes, p' and q' the frequencies of alleles A and a , and F'_{is} the fixation index ($=s/2 - s$,

Haldane, 1924). Ten theoretical populations were created with the following allelic frequencies: $p = 0.50, 0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, 0.10, 0.05$. Two selfing rates were chosen: 0 per cent for total panmixia and 13 per cent according to the selfing estimates in the case of beech (Nielsen and Schaffalitsky de Muckadell, 1954), which means that 20 theoretical populations were created.

Sampling simulation

The sample size is 50, which corresponds to the average sample size for European beech stands. The genotype of each individual is randomly generated using the pseudo-random function of the computer, from an infinite theoretical population. One hundred samples of 50 individuals were constituted from each theoretical population.

European beechstands

Two hundred fifty beech stands were sampled in Europe out of the whole beech area in as different as possible ecological conditions, representing the ecological amplitude of the species. The enzymatic polymorphism has been analysed at three loci, two peroxidases, *Px1* and *Px2*, and one glutamate-oxaloacetate transaminases *GOT1*. For inheritance and the electrophoretic methods see Thiébaud *et al.*, 1982.

F-statistics

Wright's initial formula (1951) are:

$$F_{is} = 1 - H / 2p(1 - p)$$

$$F_{st} = \text{var}(p) / \bar{p}(1 - \bar{p})$$

$$F_{it} = 1 - H / 2\bar{p}(1 - \bar{p})$$

where $2\bar{p}(1 - \bar{p})$ is the average over the samples of expected heterozygote frequencies, H the observed heterozygote frequency in the populations as a whole, $\bar{p}$ the average allelic frequency, $\text{var}(p)$ the variance of the allelic frequencies among the populations. These formula must be corrected according to the number of samples and sample size.

For each sample, whether obtained from simulation or from natural conditions, an estimation of F_{is} is made for each allele at each locus following the corrections of Kirby (1975):

$$F_{is_{ilu}} = 1 - \frac{H_{ilu}}{2p_{ilu}(1 - p_{ilu}) \left(1 + \frac{1}{n_{il}}\right)}$$

where H_{ilu} is the observed heterozygote frequency in population i ($i = 1, 2, \dots, r$), for locus 1 ($1 = 1, 2, \dots, m$) and for allele u ($u = 1, 2, \dots, v$); p_{ilu} is the estimated frequency of this allele and n_{il} the sample size for locus 1 in population i . The average F_{is} for a locus is a weighted average (Wright, 1965; Kirby, 1975; Nei, 1977):

$$F_{is_l} = \frac{\sum_{u=1}^v p_{ilu}(1 - p_{ilu}) F_{is_{ilu}}}{\sum_{u=1}^v p_{ilu}(1 - p_{ilu})}$$

Estimates of the three F -statistics were made for the whole group of beech stands following the method of Weir and Cockerham (1984).

In order to realise an estimation of the variance of these estimates, a jackknife procedure was employed (Miller, 1974; Reynolds *et al.*, 1983; Weir and Cockerham, 1984). The estimates are given $\pm$ the standard error.

RESULTS

Simulations

Panmictic situation

Fig. 1 shows F_{is} variations as a function of sample polymorphism, expressed here as the expected heterozygote frequency in a panmictic situation ($H = 1 - \sum_{u=1}^v p_u^2$). These variations are important, especially when polymorphism is high.

On the contrary to what may have been expected in panmictic populations, there are more samples with an excess of heterozygotes ($F_{is} < 0$) than with a deficit ($F_{is} > 0$). The differences are significant (table 1) and the tendency is stronger in the case of low sample polymorphism. Therefore we can define two parts.

The former, when polymorphism is high, with expected heterozygote frequencies lying between 0.25 and 0.50. The shape of the set is rather symmetrical and the number of samples with an excess of heterozygotes does not differ significantly from those with a deficit. The average F_{is} over the samples is equal to zero.

The latter, when polymorphism is low, with expected heterozygote frequencies lower than 0.25. the set is very asymmetric, with many F_{is} negative values. These negative values are aggregated. However, even with the dissymmetry, the F_{is} average remains equal to zero.

Therefore, in a complete panmictic situation, we did not obtain an equal proportion of samples with a deficit or an excess of heterozygotes.

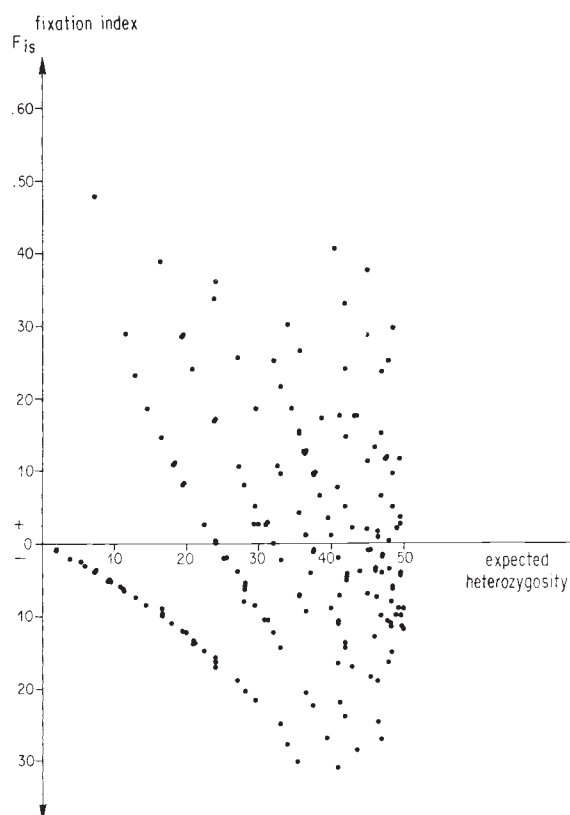


Figure 1 Variations of F_{is} , according to expected heterozygote frequencies in panmixia. Sampling simulations in theoretical panmictic populations, without self fertilisation.

Selfing effect

The selfing rate is 13 per cent. F_{is} variations are important, especially when sample polymorphism is high (fig. 2). The general shape of the set is similar to the precedent, obtained in a complete panmictic situation. But we can observe a displacement to the upper side, in the direction of a heterozygote deficit.

The number of samples with a heterozygote deficit is greater than those with an excess. The differences are highly significant with respect to the hypothesis of an equal probability of excess or deficit. When the sample polymorphism is low, we can still observe an excess of samples with negative F_{is} and an aggregation of these excesses (fig. 2 and tables 1 and 2). However, when polymorphism is high, the number of samples showing heterozygote deficit is significantly greater than the number of samples with excess (tables 1 and 2).

Between the two investigated situations, complete panmixia and mixed mating, F_{is} distributions differ significantly in relation to polymorphism except when the sample polymorphism is low (table 2).

European beech stands

Px1

This set of points shows the same general pattern as in the case of simulations (fig. 3). In particular,

Table 1 Distribution of the number of samples which have negative or positive F_{is} in function of expected heterozygote frequencies in panmixia ($H = 1 - \sum_{u=1}^v pu^2$)

expected H	Simulations				European beech stands					
	selfing = 0		selfing = 13%		Px1		Px2		GOT1	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
0.0-0.1	8	55	13	48	1	10	0	3	1	22
0.1-0.2	36	85	61	66	20	13	7	10	12	17
0.2-0.3	60	76	76	52	32	9	32	9	35	16
0.3-0.4	84	92	117	62	40	15	58	21	32	17
0.4-0.5	263	239	360	143	56	20	58	18	39	22
0.5-0.6	—	—	—	—	—	—	23	3	—	—
total	451	547	627	371	149	67	178	64	119	94
(c) χ^2	9.23		65.67		31.13		53.70		2.93	
$P(\chi^2)$	<0.01		<0.001		<0.001		<0.001		0.09	

(a) = number of samples which have a positive F_{is}

(b) = number of samples which have a negative F_{is}

(c) = comparison with a chi-square test of the distributions of positive and negative F_{is} according to the hypothesis $P(F_{is} > 0) = P(F_{is} < 0)$ $df = 1$

Expected heterozygote frequencies of Px2 can reach 0.67 because there are three alleles at this locus

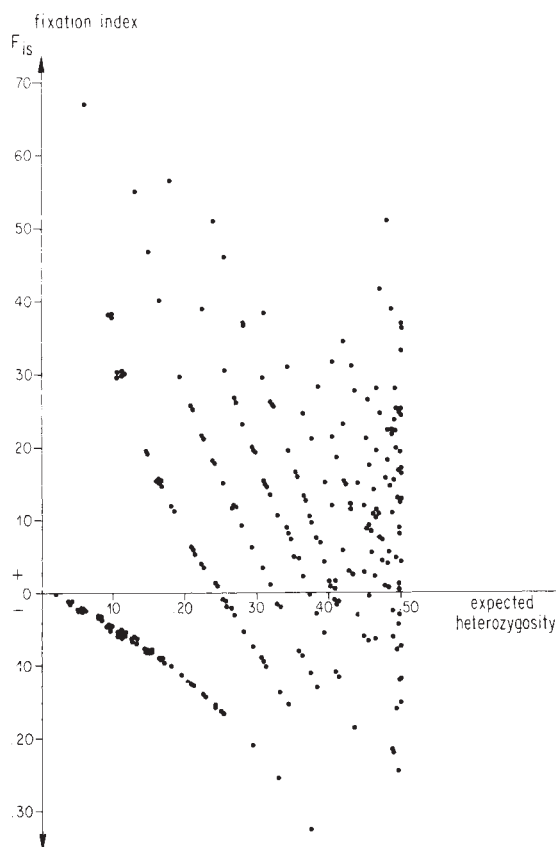


Figure 2 Variations of F_{is} , according to expected heterozygote frequencies in panmixia. Sampling simulations in theoretical panmictic populations, with 13 per cent of self fertilisation.

the asymmetry becomes very important when polymorphism decreases. However, we can observe a displacement to the upper side. Therefore, heterozygote deficit is greater in natural populations than in simulated populations even with 13 per cent selfing.

Among the beech stands, the number of samples showing heterozygote deficit is significantly greater than the number of samples with excess (table 1). F_{is} distributions significantly differ between natural populations and panmictic simulations, except when the expected heterozygote frequencies are lower than 0.10 (tables 1 and 2). However, the distributions do not differ between the beech stands and the 13 per cent selfing simulations.

Considering all of the beech stands, the F -statistics estimates were: $F_{is} = 0.151 \pm 0.016$, $F_{st} = 0.078 \pm 0.008$ and $F_{it} = 0.217 \pm 0.016$.

Px2

Here too, we can observe more heterozygote deficits than excesses (fig. 4 and table 1). However, when polymorphism is low, the aggregation of the excesses is less pronounced than for Px1. A first explanation is that there are three alleles at this locus instead of two for the preceding.

The simulations are made for a two allele locus and therefore we cannot statistically compare the results for this locus with those from simulations. However, in natural populations the aggregation

Table 2 Comparison with a chi-square test of the distributions of samples which have positive or negative F_{is} , in function of expected heterozygote frequencies in panmixia ($H = 1 - \sum_{u=1}^v p_u^2$), $df = 1$

		Expected heterozygote frequencies				
		0.0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5
selfing = 0% /	χ^2	31.06	19.56	1.88	0.36	1.15
$P(F_{is} > 0) = P(F_{is} < 0)$	$P(\chi^2)$	<0.001	<0.001	0.17	0.55	0.28
selfing = 13% /	χ^2	19.51	0.20	4.50	16.90	93.62
$P(F_{is} > 0) = P(F_{is} < 0)$	$P(\chi^2)$	<0.001	0.66	0.03	<0.001	<0.001
selfing = 0% /	χ^2	1.63	8.69	6.15	11.24	39.23
selfing = 13%	$P(\chi^2)$	0.20	<0.01	0.01	<0.001	<0.001
Px1 /	χ^2	(a)	10.67	14.53	10.53	12.10
selfing = 0%	$P(\chi^2)$	—	<0.01	<0.001	<0.01	<0.001
Px1 /	χ^2	(a)	1.66	5.59	1.03	0.15
selfing = 13%	$P(\chi^2)$	—	0.20	<0.02	0.31	0.70
GOT1 /	χ^2	(a)	1.45	8.91	4.74	2.92
selfing = 0%	$P(\chi^2)$	—	0.23	<0.01	0.03	0.09
GOT1 /	χ^2	(a)	0.42	1.33	0.00	1.53
selfing = 13%	$P(\chi^2)$	—	0.52	0.25	0.99	0.22
Px1 /	χ^2	(a)	2.28	1.02	0.67	1.51
GOT1	$P(\chi^2)$	—	0.13	0.31	0.41	0.22

(a) Insufficient sample size to compute a chi-square test

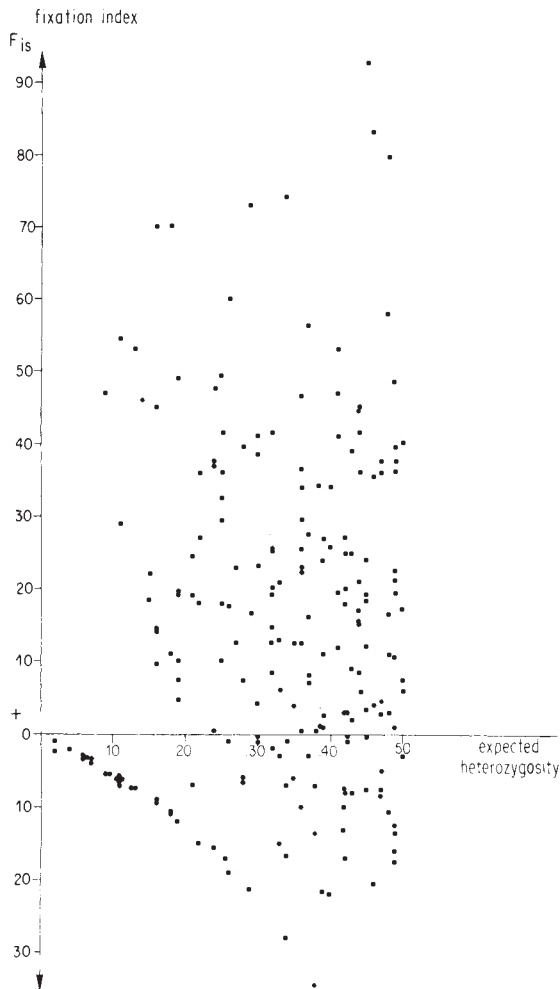


Figure 3 Variations of F_{is} , according to expected heterozygote frequencies in panmixia: peroxidases 1.

of the heterozygote excesses when polymorphism is low, and the heterozygote deficit when polymorphism is high are similar to those of the simulations.

The F -statistics estimates are: $F_{is} = 0.122 \pm 0.010$, $F_{st} = 0.035 \pm 0.003$ and $F_{it} = 0.153 \pm 0.010$.

GOT1

For this locus, the F_{is} distribution pattern is the same as for the two first loci, but the upward tendency is less pronounced (fig. 5).

The number of samples with negative F_{is} values does not statistically differ from the number with positive F_{is} values, although there is a tendency for a heterozygote deficit (table 1).

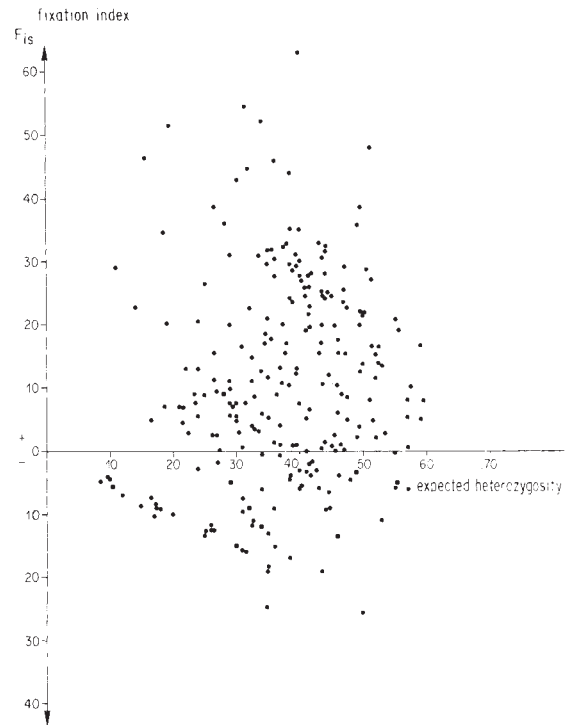


Figure 4 Variations of F_{is} , according to expected heterozygote frequencies in panmixia, peroxidases 2.

The observed distribution differs from that of the panmictic simulations, except when polymorphism is low (tables 1 and 2). However there are no differences with results obtained with 13 per cent selfing.

The three estimated F -statistics are: $F_{is} = 0.082 \pm 0.013$, $F_{st} = 0.084 \pm 0.008$ and $F_{it} = 0.159 \pm 0.013$.

DISCUSSION AND CONCLUSIONS

Simulations

Whether for a simulation or for a beech stand, the theoretical population and the natural population are both objects which we study through images: samples.

When F_{is} is estimated in a small sample, the estimation shows great variations. Theoretically, F_{is} is an interesting parameter for the study of genetic population structures in relation to selection and mating system. But it would be dangerous to interpret the estimates from such limited samples. The corrections given by Weir and Cockerham (1984) make it possible to estimate this parameter and other F -statistics with less biases

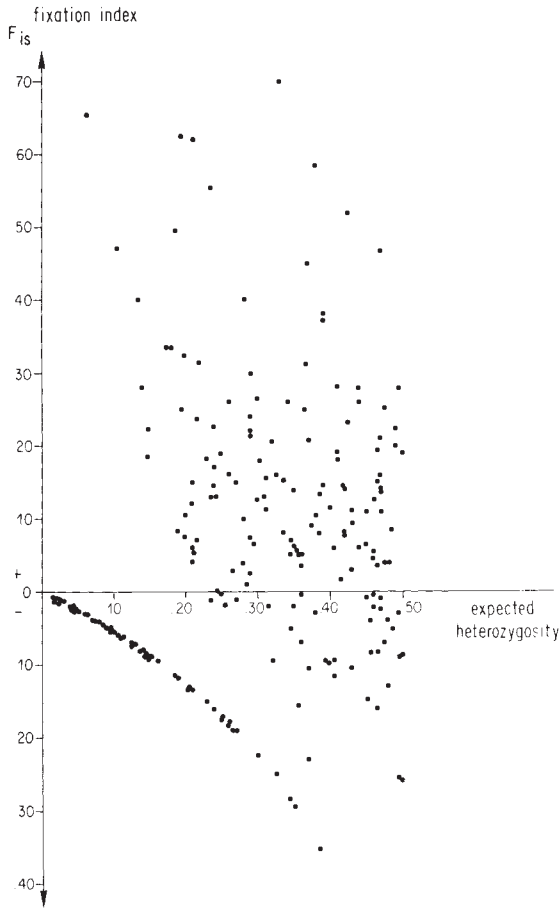


Figure 5 Variations of F_{is} , according to expected heterozygote frequencies in panmixia, glutamate oxaloacetate transaminases.

for a group of populations. But under these conditions, F_{is} becomes an average and can mask inter-population variations. And these variations are precisely the most interesting ones, because they reflect the diversity of the evolutionary forces in natural populations.

The organisation of F_{is} variations is not the same whatever the sample polymorphism. When this polymorphism is low, with or without selfing, the F_{is} negative values are more frequent than the positive ones. When the polymorphism is high, in a complete panmictic situation, the negative and positive values are equifrequent, while with 13 per cent selfing, positive values are always more frequent than negative values.

When the polymorphism decreases in the sample, the asymmetry and the aggregation of the samples have a statistical origin caused by the allelic frequency estimation method used in a finite

sample. Indeed, for a diallelic locus, allelic frequencies are estimated from observed genotypic frequencies:

$$p = Aa/2 + AA1 \quad (1)$$

$$q = 1 - p = 1 - H/2 - H1$$

and

$$2pq = 2(Aa/2 + AA1)(1 - Aa/2 - AA1)$$

where H and $H1$ are respectively the observed frequencies of Aa heterozygotes and AA homozygotes. Substituting $2pq$ by the estimate, F_{is} becomes:

$$F_{is} = 1 - \frac{Aa}{2(Aa/2 + AA1)(1 - Aa/2 - AA1)}$$

When p tends toward zero, the frequencies of AA homozygotes and Aa heterozygotes also tend toward zero because:

$$AA1 = p^2 + pqF_{is}$$

$$Aa = 2pq(1 - F_{is})$$

But, as AA decreases faster than Aa , there is a progressive estimation dependence between p and Aa . According to (1), when AA is near zero, p is mostly estimated from Aa . And when AA is equal to zero, $p = Aa/2$. Under these conditions, the fixation index is estimated according the following simplified formula:

$$F_{is} = 1 - 1/(1 - Aa/2)$$

Here we can see the total dependence between F_{is} and the observed heterozygote frequency. Whatever the observed Aa , $1 - Aa/2$ is still lower than one, and F_{is} shows a systematic heterozygote excess in the sample.

The F_{is} values are positioned on a graph whose equation is given by the latter formula. When an allele is rare, the probability of finding it as a homozygote in the sample is low. Most often, it is only found as a heterozygote. Under these conditions, there are only two genotypes in the sample instead of three. In this case, the heterozygote sampling still respects the binomial distribution whereas the rare homozygote sampling already respects the Poisson distribution. The probability of rare homozygote absence in the sample can be estimated according to the Poisson formula:

$$P(0) = e^{-\mu}$$

with $\mu = N(p^2 + pqF_{is})$. When p increases, this probability decreases and tends toward zero. Thus the three genotypic forms reappear in the sample,

and when the sample size N increases, $P(0)$ decreases. The proportion of samples with only two genotypes decreases as the polymorphism and the sample size increase.

Therefore the interpretation of the sign and the absolute value of F_{is} , estimated from a finite sample, is different according to the polymorphism. In samples with a low polymorphism, especially when only two genotypes are present, it is normal to find more heterozygote excesses than deficits. The frequency of these excesses is not a consequence of overdominance or frequency dependent selection but only a statistical effect caused by the low probability of encountering rare homozygotes in the samples.

Natural populations

In beech stands, the positive F_{is} values are more frequent than the negative ones. If we consider all of the sampled populations, they present the same tendency for a heterozygote deficit at the three loci. The average F_{is} at each locus is estimated at: 0.151 ± 0.016 ($Px1$), 0.122 ± 0.010 ($Px2$) and 0.082 ± 0.013 ($GOT1$). The average F_{is} over the three loci is 0.120. These estimates are all positive and this trend suggests an influence due to the mating system.

The differentiation coefficient, F_{st} , varies among the loci: 0.078 ± 0.008 ($Px1$), 0.035 ± 0.003 ($Px2$) and 0.084 ± 0.008 ($GOT1$). The average F_{st} over the three loci is 0.064.

The estimated F -statistics of beech can be compared to those of other species, also estimated on the same scale, over the whole distribution area (table 3). The beech shows the highest estimates

among the anemophilous forest trees, only *Pinus banksiana* has similar estimates. However herbaceous species can present higher estimates, in relation to their less allogamous mating system but also to their low longevity and therefore to a higher average population age. The estimates currently realised were obtained from more than 200 beech-stands and are very accurate for each of the three loci. But the number of sampled loci is low and we must be careful about generalising our interpretations.

As pointed out by Brown (1979), allogamous species often exhibit a heterozygote deficit relative to panmixia while autogamous species show the opposite phenomenon, a heterozygote excess. That is what he called the "heterozygosity paradox". And among the different origins he listed, two can be applied to the beech.

Selfing

Although the beech is essentially allogamous, there can be some self-fertilisation. Nielsen and Shaffalitsky-de-Muckadell (1954) did auto-pollinations on 23 trees. Under these conditions, the estimated average selfing rate approaches 13 per cent. However, the selfing rate may widely vary on several levels: intraindividual, interindividual and interpopulation (Shen *et al.*, 1981; Shaw and Allard, 1982). The estimate proposed for beech may be imprecise because of the few traces studied. Plus the fact that autopollination without competition with allopollen may have made the authors overestimate the selfing rate in regard to natural conditions. Nevertheless, 13 per cent selfing is similar to the estimates realized on open pollina-

Table 3 F -statistics of some plant species (in part from Guries and Ledig, 1982)

Species	Number of of loci	F_{is}	F_{st}	F_{it}	authors
Herbaceous					
<i>Polygonum pensylvanicum</i>	3	0.67	0.12	0.71	Kubetin and Schaal, 1979
<i>Liatris cylindracea</i>	15	0.41	0.07	0.43	Schaal, 1975
<i>Sarracenia purpurea</i>	4-5	-0.10	0.23	—	Schwaegerle and Schaal, 1979
<i>Phlox cuspidata</i>	5	0.67	0.41	0.80	Levin, 1978
<i>Phlox roemariana</i>	4	0.04	0.21	0.54	Levin, 1978
<i>Oenothera laciniata</i>	5	0.13	0.24	0.37	Ellstrand and Levin, 1980
<i>Oenothera grandis</i>	6	0.04	0.09	0.19	Ellstrand and Levin, 1980
<i>Desmodium nudiflorum</i>	5	0.04	0.17	0.20	Schaal and Smith, 1980
Forest trees					
<i>Pinus rigida</i>	21	0.01	0.02	0.03	Guries and Ledig, 1982
<i>Pinus contorta</i>	21	0.02	0.02	0.04	Dancik and Yeh, 1983
<i>Pinus banksiana</i>	21	0.10	0.02	0.12	Dancik and Yeh, 1983
<i>Fagus sylvatica</i>	3	0.12	0.06	0.18	Present study

tions which vary between 0 and 15 per cent in anemophilous forest trees (Squillace, 1974). Although low, this selfing rate can produce a heterozygote deficit at the inbreeding equilibrium. This deficit can be estimated throughout the fixation index with the Haldane formula (1924). On one hand, if we admit to 13 per cent selfing, we should expect F_{is} to be equal to 0.069. Compared to the observed F_{is} in natural conditions, there should be a residual F_{is} equal to 0.051 which is not trivial. On the other hand, if only self-fertilisation is responsible for the observed deficit, the selfing rate should, according to the Haldane formula, be equal to 21 per cent in order to obtain $F_{is} = 0.120$. This last average selfing rate is too high for an anemophilous forest tree. Consequently, self-fertilisation can explain only a part of the observed heterozygote deficit and another origin must be invoked.

Limited gene flow

The beech stands present very different physiologies according to where they are located in the area of the species. In the optimal zone, in latitude as well as in altitude, beech stands are continuous, monospecific, with millions of individuals. In contradiction with the general view of the anemophilous forest trees mating system, it is likely that reproduction occurs between closely spaced individuals and that neighbourhoods exist in these forests. On the contrary, on the outskirts of the distribution area, the forests are often discontinuous with a small population effective size and little interpopulation gene flow because of distance and phenological differences. In the first case, genetic structures could best be described by an isolation by distance model (Wright, 1943, 1946) and in the second case, by an island model (Wright, 1931) or a stepping stone model (Kimura, 1953; Kimura and Weiss, 1964). These two kinds of genetic models lead to a relatedness between individuals in the populations, therefore to a heterozygote deficit and an inter- and intrapopulation genetic differentiation.

In the first case, intrapopulation pollen flow is limited to the nearest neighbours. Consequently, there is an increasing relatedness between individuals in the population and a decrease of heterozygote frequency in relation to neighbourhood size and the number of generations since the population foundation. Simulations of reproduction, following the "Isolation by distance" model, with different pollen effective sizes, make it possible to establish a relation between neighborhood

size and heterozygote deficit (Cuguen, 1986; Cuguen and Thiebaut, in preparation). Under these conditions, the $F_{is} = 0.051$ not explained by selfing may correspond to a neighbourhood size varying between 25 and 50. This indirect estimate, although very imprecise, would allow an estimate of the average neighbourhood size of beech, difficult to estimate with inductive methods. According to Wright (1946), a neighbourhood size lower than 200 can produce an intrapopulation genetic differentiation.

In the second case, where the populations are quite discontinuous, the limited effective size does not allow the constitution of distinct neighbourhoods. Under these conditions, reproduction is mostly panmictic in such small and isolated populations. However, here too the finite effective size in these marginal beech stands leads to an increasing relatedness between individuals and a diminution of heterozygosity.

Therefore limited gene flow between individuals and self fertilisation are likely to be at the origin of the genotypic structures of European beech populations.

REFERENCES

- BROWN, A. H. D. 1979. Enzyme polymorphism in plant populations. *Theor. Pop. Biol.*, 15, 1-42.
- BROWN, A. H. D., MATHESON, A. C. AND ELDRIDGE K. G. 1975. Estimation of the mating system of *Eucalyptus obliqua* L'Hérit. by using allozyme polymorphisms. *Aust. J. Bot.*, 23, 931-949.
- COCKERHAM, C. C. 1969. Variance of gene frequencies. *Evolution*, 23, 72-84.
- COCKERHAM, C. C. 1973. Analyses of gene frequencies. *Genetics*, 74, 679-700.
- CUGUEN, J. 1986. Différenciation génétique inter- et intrapopulations d'un arbre forestier anémophile: le cas du Hêtre (*Fagus sylvatica* L.). Diplôme de Doctorat de l'Université des Sciences et Techniques du Languedoc, 96 pp. et annexes 75 pp.
- CUGUEN, J., THIEBAUT, B., N'TSIBA, F. AND BARRIERE, G. 1985. Enzymatic variability of beech stands (*Fagus sylvatica* L.) on three scales in Europe: Evolutionary mechanisms. In Jacquart, P., Heim, G. and Antonovics, J. (eds.) *Genetic Differentiation and Dispersal in Plants*, NATO ASI Series, Vol 1. G5, pp.17-39.
- CUGUEN, J. AND THIEBAUT, B. 1987. Genetic differentiation over microgeographical space in two beech stands (*Fagus sylvatica* L.): Simulations versus experimental results. (in prep.).
- DANCIK, B. P. AND YEH, F. C. 1983. Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. *Can. J. Genet. Cytol.*, 25, 57-64.
- ELLSTRAND, N. C. AND LEVIN, D. A. 1980. Recombination system and population structure in *Oenothera*. *Evolution*, 34, 923-933.

- FELBER, F. AND THIEBAUT, B. 1984. Etude préliminaire sur le polymorphisme enzymatique du Hêtre, *Fagus sylvatica* L.: variabilité génétique de deux systèmes de peroxydases en relation avec les conditions écologiques. *Oecol. Plant.* 5, 133-150.
- GURIES, R. P. AND LEDIG, F. T. 1982. Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). *Evolution*, 36, 382-402.
- HALDANE, J. B. S. 1924. A mathematical theory of natural and artificial selection. part 11. The influence of partial self-fertilization, inbreeding, assortative mating and selective fertilization on the composition of mendelian populations and on natural selection. *Proc. Camb. Phil. Soc. Biol. Sci.*, 158-163.
- KIMURA, M. 1953. "Stepping stone" model of population. *Ann. Rept. Nat. Inst. Genet. of Japan*, 3, 62-63.
- KIMURA, M. AND WEISS, G. H. 1964. The stepping stone model of populations structure, and the decrease of genetic correlation with distance. *Genetics*, 49, 561-576.
- KIRBY, G. C. 1975. Heterozygote frequencies in small subpopulations. *Theor. Pop. biol.*, 8, 31-48.
- KUBETIN, W. R. AND SCHAAL, B. A. 1979. Apportionment of isozyme variability in *Polygonum pensylvanicum* (Polygonaceae). *Syst. Bot.*, 4, 148-156.
- LEVIN, D. A. 1978. Genetic variation in annual phlox: self-compatible versus self-incompatible species. *Evolution*, 32, 245-263.
- MALECOT, G. 1948. *Les mathématiques de l'hérédité*, Masson et Cie, Paris, 63pp.
- MILLER, R. G. 1974. The jackknife—a review. *Biometrika*, 61, 1-15.
- NEI, M. 1977. *F*-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.*, 41, 225-233.
- NIELSEN, P. C. AND SCHAFFLATISKY DE MUCKADELL, M. 1954. Flower observations and controlled pollinations in *Fagus*. *Z. Forstgenet.*, 3, 6-17.
- RASMUSSEN, D. I. 1964. Blood group polymorphism and inbreeding in natural populations of the deer mouse *Peromyscus maniculatus*. *Evolution*, 18, 219-229.
- REYNOLDS, J., WEIR, B. S. AND COCKERHAM, C. C. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, 105, 767-779.
- SCHAAL, B. A. 1975. Population structure and local differentiation in *Liatris cylindracea*. *Amer. Natur.*, 109, 511-528.
- SCHAAL, B. A. AND SMITH, W. G. 1980. The apportionment of genetic variation within and among populations of *Desmodium nudiflorum*. *Evolution*, 34, 214-221.
- SCHWAEGERLE, K. E. AND SCHAAL, B. A. 1979. Genetic variability and founder effect in the pitcher plant, *Sarracenia purpurea* L. *Evolution*, 33, 1210-1218.
- SHAW, D. V. AND ALLARD, R. W. 1982. Estimation of outcrossing rates in Douglas-fir using isozyme markers. *Theor. Appl. Genet.*, 62, 113-120.
- SHEN, H. H., RUDIN, D. AND LINDGREN, D. 1981. Study of the pollination pattern in a Scots pine seed orchard by means of isozyme analysis. *Silvae Genet.*, 30, 7-15.
- SQUILLACE, A. E. 1974. Average genetic correlations among offsprings from open-pollinated forest trees. *Silvae Genet.*, 23, 149-156.
- THIEBAUT, B., LUMARET, R. AND VERNET, P. 1982. The bud enzymes of beech (*Fagus sylvatica* L.). Genetic distinction and analysis of polymorphism in several French populations. *Silvae Genet.*, 31, 51-60.
- VASEK, F. C., AND HARDING, J. 1976. Outcrossing in natural populations. V. Outcrossing, inbreeding, and selection in *Clarkia exilis* and *Clarkia tembloriensis*. *Evolution*, 30, 403-411.
- WEIR, B. S. AND COCKERHAM, C. C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, 36, 1358-1370.
- WRIGHT, S. 1931. Evolution in mendelian populations. *Genetics*, 16, 97-159.
- WRIGHT, S. 1943. Isolation by distance. *Genetics*, 28, 114-138.
- WRIGHT, S. 1946. Isolation by distance under diverse systems of mating. *Genetics*, 31, 39-59.
- WRIGHT, S. 1951. The genetical structure of populations. *Ann. Eugenics*, 15, 323-354.
- WRIGHT, S. 1965. The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, 19, 395-420.