Heredity **75** (1995) 382–389 *Received 17 February 1995*

Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci

R. J. PETIT*, N. BAHRMAN & PH. BARADAT

Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, BP 45, F-33611 Gazinet Cedex, France

Thirty genes belonging to three classes of biochemical markers (eight isozyme loci, 16 protein loci revealed by two-dimensional gel electrophoresis, and six terpene loci) were scored in six populations of maritime pine. The purpose of the study was to compare the level of genetic differentiation (G_{ST}) among the populations of this pine in order to test whether differential selective pressures are acting on these markers.

However, each class of loci was found to display different levels of average diversity. Although this should not affect the comparison of the level of differentiation (theoretically a diversity-independent genetic measure) we found here that single-locus values of differentiation significantly depended on the values of diversity. This result was explained analytically by showing that the sampling of a limited number of populations results in G_{ST} taking maximal possible values lower than unity, especially when the level of diversity is low. By removing the less polymorphic loci, measures of differentiation independent of the level of diversity can be obtained. They turned out to be very close for each class of markers indicating the absence (or similar level) of selection acting on the three classes of loci and a high level of differentiation in this pine ($G_{ST} = 0.17$) typical of a species having a highly fragmented range.

Keywords: biochemical markers, genetic diversity, G_{ST}, natural selection, *Pinus pinaster*

Introduction

Genetic differentiation (F_{ST} or G_{ST}), a measure of the variability of allelic frequencies among populations, is a parameter frequently used to describe the organisation of gene diversity among populations of a species. When other evolutionary forces are absent, Wright (1951) has shown that, in an island model of population structure, an equilibrium will arise between drift (which tends to fix different alleles in different populations, as a consequence of their finite size) and migration (which tends to homogenize allelic frequencies). F_{ST} becomes constant and can be directly related to the level of gene flow among populations. Cavalli-Sforza (1966) first stated that, because all genes should be equally affected by drift and migration, they should have similar values of F_{ST}, unless they are differentially selected. This will of course not apply to the comparison of cytoplasmic genes and nuclear genes, because gene flow and effective population sizes are not the same in each case (Birky *et al.*, 1989; Petit *et al.*, 1993a,b).

Here we study the level of differentiation of three classes of biochemical markers encoded by the nucleus (isozyme loci, protein loci revealed using two-dimensional gel electrophoresis and terpene loci) in a conifer, the maritime pine (*Pinus pinaster* Ait.), which has a highly fragmented range in the western Mediterranean region. Because, as we will show later, the level of diversity measured for each of the three groups of loci is different, the question first arises: are levels of diversity and of differentiation correlated? If so, can the level of differentiation observed for each class of marker be nevertheless compared?

If this comparison proves possible, a more fundamental question may then be examined: do these classes of loci exhibit contrasting values of genetic differentiation? Indeed, the selective pressures acting on these markers are potentially heterogeneous. For instance, terpene compounds present in the

*Correspondence.

resin of conifers have sometimes been correlated with resistance against pathogenic insects (e.g. Delorme & Lieutier, 1990) and are known to show fungistatic effects (e.g. Bridges, 1987). In particular, some terpenes of maritime pine (longifolene and limonene) are believed to be involved in the resistance against *Dioryctria splendidella*, a caterpillar which attacks the bark of this tree (Baradat & Marpeau-Bezard, 1988). The genetic basis of several terpene compounds of maritime pine has been extensively studied (see references in Müller-Starck *et al.*, 1992). All the terpene loci identified have two alleles, an allele for high and an allele for low terpene level. The genetics of the many proteins of the haploid megagametophyte of maritime pine has been the topic of several recent investigations (Bahrman & Damerval, 1989; Gerber *et al.*, 1993; Bahrman *et al.*, 1994). The extent to which these anonymous proteins and a set of allozymes developed for this study may be submitted to direct or indirect selective pressures is unknown but, in the case of isozyme loci, overdominance is regularly invoked for such traits as growth rate (e.g. Karl & Avise, 1992; Pogson & Zouros, 1994), though contrary reports exist, also in conifers (Strauss & Libby, 1987).

Materials and methods

Plant materials and experimental methods

Six populations of maritime pine (*Pinus pinaster* Ait.) were studied: Landes (south-west France), Leiria (Portugal), Vivario (Corsica), Cuenca (Aragon, Spain), Fontanin (Liguria, Italy) and Monte Pino (Sardinia). Their locations are given in Table 1 and Fig. 1 of Bahrman *et al.* (1994). The trees were selected in homogeneous stands within larger maritime pine forests (of at least several hectares). Minimal distance between sampled trees was 10 m, maximum distance was 500 m. Terpene compounds were extracted from the oleoresin obtained from cortical tissues of a shoot of the upper part of the crown from at least 60 adult trees (i.e. more than 10 years old) of each population and the analysis of samples was carried out by means of gas–liquid

Table 1 Allele frequencies, number of sampled alleles, mean diversity (h_T) and differentiation (G_{ST}) for protein, isozyme and terpene loci of *Pinus pinaster*

	Origin									
Proteins	Corsica	Portugal	Spain	Italy	Sardinia	France	Mean	No. alleles	h_T	G_{ST}
*Pprs**								*192*	*0.455*	*0.141*
Allele 1	0.000	0.031	0.000	0.062	0.000	0.000	0.016		0.031	0.033
Allele 2	0.031	0.343	0.406	0.281	0.218	0.625	0.317		0.434	0.147
Alelle 3	0.968	0.625	0.593	0.656	0.781	0.375	0.666		0.445	0.143
*Mprk**	*0.437*	*0.843*	*0.937*	*0.562*	*0.750*	*0.500*	*0.672*	*192*	*0.441*	*0.149*
*Fprj**	*0.625*	*0.750*	*0.125*	*0.687*	*0.906*	*0.218*	*0.552*	*192*	*0.495*	*0.322*
*Gprl**	*0.687*	*0.875*	*0.375*	*0.781*	*0.906*	*0.312*	*0.656*	*192*	*0.452*	*0.236*
*Mprg**	0.969	*0.625*	*0.594*	*0.625*	*0.782*	*0.407*	*0.667*	*192*	*0.446*	*0.132*
*Kprl**	*0.844*	*0.469*	*0.750*	*0.719*	*0.813*	*0.782*	*0.730*	*192*	*0.395*	*0.073*
*Xprl**	*0.782*	*0.344*	*0.625*	*0.625*	*0.719*	*0.532*	*0.605*	*192*	*0.479*	*0.079*
Vprd	*0.875*	*1.000*	*1.000*	*0.750*	*1.000*	*1.000*	*0.938*	*192*	*0.117*	*0.152*
Wprh	*0.938*	*0.938*	*0.625*	*1.000*	*0.906*	*0.781*	*0.865*	*192*	*0.234*	*0.132*
*Upri**	*0.562*	*0.875*	*0.719*	*0.250*	*0.906*	*0.781*	*0.682*	*192*	*0.434*	*0.227*
*Hprw**	*0.906*	*0.343*	*0.469*	*0.906*	*0.688*	*0.594*	*0.651*	*192*	*0.455*	*0.189*
*Tprv**	*0.281*	*0.813*	*0.781*	*0.281*	*0.438*	*0.719*	*0.552*	*192*	*0.495*	*0.204*
*Uprk**	*0.437*	*0.875*	*0.844*	*1.000*	*0.688*	*0.906*	*0.792*	*192*	*0.330*	*0.201*
*Zpra**								*192*	*0.505*	*0.218*
Allele 1	0.250	0.000	0.125	0.000	0.250	0.094	0.120		0.211	0.096
Allele 2	0.469	0.937	0.719	0.312	0.594	0.906	0.656		0.452	0.220
Allele 3	0.281	0.063	0.156	0.688	0.156	0.000	0.224		0.348	0.288
*Vprb**	*0.844*	*0.562*	*0.469*	*0.875*	*0.812*	*0.468*	*0.672*	*192*	*0.441*	*0.136*
*Fprd**	*0.469*	*0.938*	*0.782*	*0.625*	*0.532*	*0.844*	*0.698*	*192*	*0.422*	*0.132*
All proteins									**0.412**	**0.170**
14 proteins*									**0.446**	**0.174**

Table 1 *Continued*

Isozymes	Origin						Mean	No. alleles	h_T	G_{ST}
	Corsica	Portugal	Spain	Italy	Sardinia	France				
Got	*1.000*	*1.000*	*1.000*	*0.776*	*1.000*	*1.000*	*0.963*	*1856*	*0.071*	*0.191*
6-Pgd	*0.983*	*1.000*	*0.984*	*1.000*	*1.000*	*0.984*	*0.992*	*1980*	*0.013*	*0.011*
Idh	*1.000*	*1.000*	*0.992*	*1.000*	*1.000*	*0.972*	*0.994*	*1967*	*0.023*	*0.023*
Mdh-2								*1806*	*0.112*	*0.064*
Allele 1	0.000	0.000	0.017	0.000	0.000	0.000	0.003		0.004	0.011
Allele 2	0.928	0.938	0.786	1.000	0.994	0.966	0.935		0.111	0.068
Allele 3	0.072	0.062	0.197	0.000	0.006	0.034	0.062		0.108	0.061
*Mdh-3**								*1316*	*0.611*	*0.222*
Allele 1	0.568	0.186	0.333	0.935	0.697	0.186	0.484		0.501	0.283
Allele 2	0.261	0.497	0.556	0.000	0.298	0.678	0.382		0.473	0.217
Allele 3	0.171	0.317	0.111	0.065	0.006	0.136	0.134		0.247	0.105
*Pgi**								*2018*	*0.273*	*0.251*
Allele 1	0.018	0.039	0.079	0.000	0.011	0.040	0.031		0.016	0.009
Allele 2	0.930	0.837	0.882	1.000	0.421	0.867	0.823		0.271	0.257
Allele 3	0.053	0.124	0.039	0.000	0.568	0.093	0.146		0.260	0.259
Pgm	*1.000*	*1.000*	*0.944*	*1.000*	*1.000*	*1.000*	*0.991*	*1865*	*0.021*	*0.040*
*Skdh**								*1704*	*0.567*	*0.081*
Allele 1	0.609	0.579	0.521	0.592	0.879	0.346	0.588		0.488	0.075
Allele 2	0.046	0.298	0.165	0.000	0.064	0.269	0.140		0.248	0.084
Allele 3	0.345	0.124	0.314	0.408	0.057	0.385	0.272		0.398	0.085
All isozymes									**0.211**	**0.161**
3 isozymes*									**0.484**	**0.172**

Terpenes	Origin						Mean	No. alleles	h_T	G_{ST}
	Corsica	Portugal	Spain	Italy	Sardinia	France				
*β-Pinene-0**	*0.817*	*0.433*	*0.867*	*0.645*	*0.734*	*0.254*	*0.625*	*840*	*0.474*	*0.198*
*3-Carene-0**	*0.667*	*0.800*	*0.883*	*0.952*	*0.938*	*0.632*	*0.812*	*840*	*0.310*	*0.101*
*Myrcene-1**	*0.550*	*0.700*	*0.750*	*0.371*	*0.391*	*0.474*	*0.539*	*840*	*0.504*	*0.144*
*Limonene-0**	*0.833*	*0.533*	*0.783*	*0.677*	*0.734*	*0.763*	*0.721*	*840*	*0.408*	*0.045*
*Longifolene-0**	*0.783*	*0.917*	*0.550*	*0.839*	*0.750*	*0.754*	*0.766*	*840*	*0.364*	*0.069*
*Caryophyllene-1**	*0.533*	*0.933*	*0.850*	*0.307*	*0.453*	*1.000*	*0.679*	*840*	*0.440*	*0.309*
All terpenes									**0.417**	**0.139**
All 30 loci									**0.365**	**0.167**
24 loci									**0.442**	**0.169**

For diallelic loci, only the frequency of the most frequent allele is given.
*Loci which have a level of diversity higher than 0.27 and were used to compute the second mean estimate of differentiation.

chromatography according to Bernard-Dagan *et al.* (1971). The genetic interpretation (including dominance relationships) of six terpene loci β-pinene, 3-carene, myrcene, limonene, longifolene and caryophyllene) has been studied previously (Baradat *et al.*, 1972, 1975; Marpeau *et al.*, 1975; Marpeau-Bezard *et al.*, 1983; Baradat & Marpeau-Bezard, 1988). Seeds were collected from the same 60 adult trees and bulked in a single seed-lot from which random samples were taken for the other analyses. Two-dimensional electrophoresis was performed using proteins extracted from 32 haploid megagametophytes of each population, according to Bahrman & Damerval (1989). In general using this technique allowed more than 100 heterozygous loci to be detected in a parental tree. Only 20 per cent of this variation corresponds to isoelectric point variants (allelic variation of structural genes) and 80 per cent

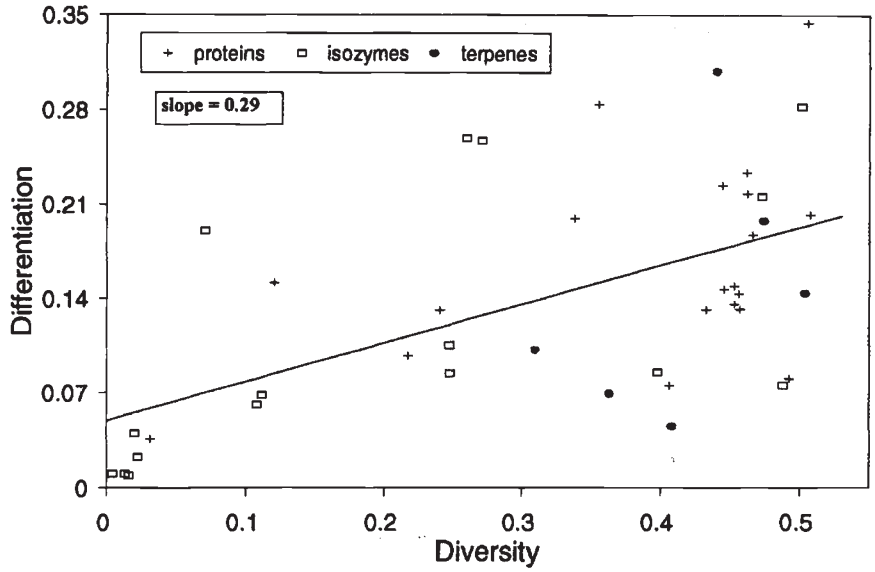

Fig. 1 Relationship between differentiation and diversity. 'Allelic' values of diversity and differentiation were obtained by considering that each triallelic locus can be split into three diallelic loci as explained in the text. A significant positive effect of diversity on differentiation is apparent.

of variants express presence/absence or quantitative variations (Bahrman & Petit, 1995). Here we scored only 16 loci for which we had studied the segregation using megagametophytes from a single tree (Bahrman & Damerval, 1989). For the isozymes, a selection of eight polymorphic loci was studied in megagametophytes and embryos of over 120 seeds from each population using starch gel electrophoresis according to Conkle *et al.* (1982). The loci surveyed were phosphoglucoisomerase (*Pgi*, EC 5.3.1.9), phosphoglucomutase (*Pgm*, EC 2.7.5.1), shikimate dehydrogenase (*Skdh*, EC 1.1.1.25), isocitrate dehydrogenase (*Idh*, EC 1.1.1.42), malate dehydrogenase (*Mdh-2* and *Mdh-3*, EC 1.1.1.37), 6-phosphogluconate dehydrogenase (*6-Pgd*, EC 1.1.1.43) and glutamate oxaloacetate transaminase (*Got*, EC 2.6.1.1).

Statistical analyses

The analysis of diversity was performed using Nei's (1987) single-locus estimators: mean intrapopulation diversity (h_S), total diversity (h_T) and differentiation (G_{ST}). The estimators of Weir & Cockerham (1984) were also tested but the results were very similar and are not presented. For the test of the dependence of single-locus G_{ST} estimates on the diversity estimates h_T, we examined whether the slope of the linear regression differed from zero. For this purpose, each of the six triallelic loci was transformed into three diallelic loci by pooling in turn two of the three alleles. However, we conservatively assumed that there were only 28 degrees of freedom (i.e. the number of loci minus two) because the alleles belonging to the same triallelic locus are not completely independent. Mean estimators were calculated according to Nei (1977), where the mean intrapopulation and total diversities $\bar{h}_S$ and $\bar{h}_T$ are obtained by averaging over loci, whereas the mean differentiation is given by:

$$\overline{G}_{ST} = \frac{\bar{h}_T - \bar{h}_S}{\bar{h}_T} \quad (1)$$

(and not by the average of G_{ST} over loci).

Results

The sample sizes (number of alleles studied) and allelic frequencies at the 30 loci are given in Table 1. There are four isozymes and two protein loci which are triallelic, the remaining loci being diallelic. The mean total diversity is $\bar{h}_T = 0.365$ and the mean differentiation is $\overline{G}_{ST} = 0.167$. This high level of differentiation is further illustrated by the existence of three private alleles (i.e. alleles unique to individual populations) at the loci *Got*, *Mdh-2* and *Pgm*. For the three classes of markers, the total diversity varies from 0.211 for isozymes to 0.412 for proteins, and the differentiation from 0.139 for terpenes to 0.170 for proteins. Hence, despite an important difference in the level of diversity revealed by each class of markers, the levels of differentiation observed are quite similar. The values of differentiation of the 42 diallelic 'loci' are plotted against their values of diversity in Fig. 1. The coefficient of regression of G_{ST} on h_T (0.287) is highly significant ($P<0.001$, 28 d.f.). To study further this relationship, we selected

four loci showing contrasted values of diversity: *Mdh-3* ($h_T = 0.611$), *Pgi* ($h_T = 0.273$), *Mdh-2* ($h_T = 0.112$) and *6-Pgd* ($h_T = 0.013$). For each locus, we resampled the same number of individuals in the six populations to generate 1000 bootstrap samples. Total diversity and differentiation were then recomputed and differentiation was plotted against diversity for the 1000 bootstrap samples. The four examples are illustrated in Fig. 2(a–d). With decreasing values of diversity, there is an increasing correlation between diversity and differentiation: $r = 0.029$ for *Mdh-3*, $r = 0.194$ for *Pgi*, $r = 0.478$ for *Mdh-2* and $r = 0.600$ for *6-Pgd*. All these values are highly significant except the first one.

In Fig. 1, the six loci showing values of diversity smaller than 0.05 are all characterized by particularly low values of differentiation. To explain this result, it must be considered that, for a diallelic locus when the number of populations studied is limited, the maximum value which can be observed for G_{ST} (assuming that both alleles are restricted to as few populations as possible) is $G_{ST(max)}$ (<1), where $G_{ST(max)}$ is given by the equation:

$$G_{ST(max)} = \mathrm{Var}_{(max)}(p)/[p(1-p)]. \quad (2)$$

The maximum variance of p (the mean allelic frequency) is obtained when as many populations as possible are fixed for this allele. If L is the number of populations sampled in the survey, and int(p) indicates the integer of p and rad(p) its radical, then we have: (a) int(Lp) populations fixed for this allele; (b) one population characterized by a frequency of rad(Lp); and (c) the other ($L - \mathrm{int}(Lp) - 1$) populations fixed for the alternative allele. This supposes

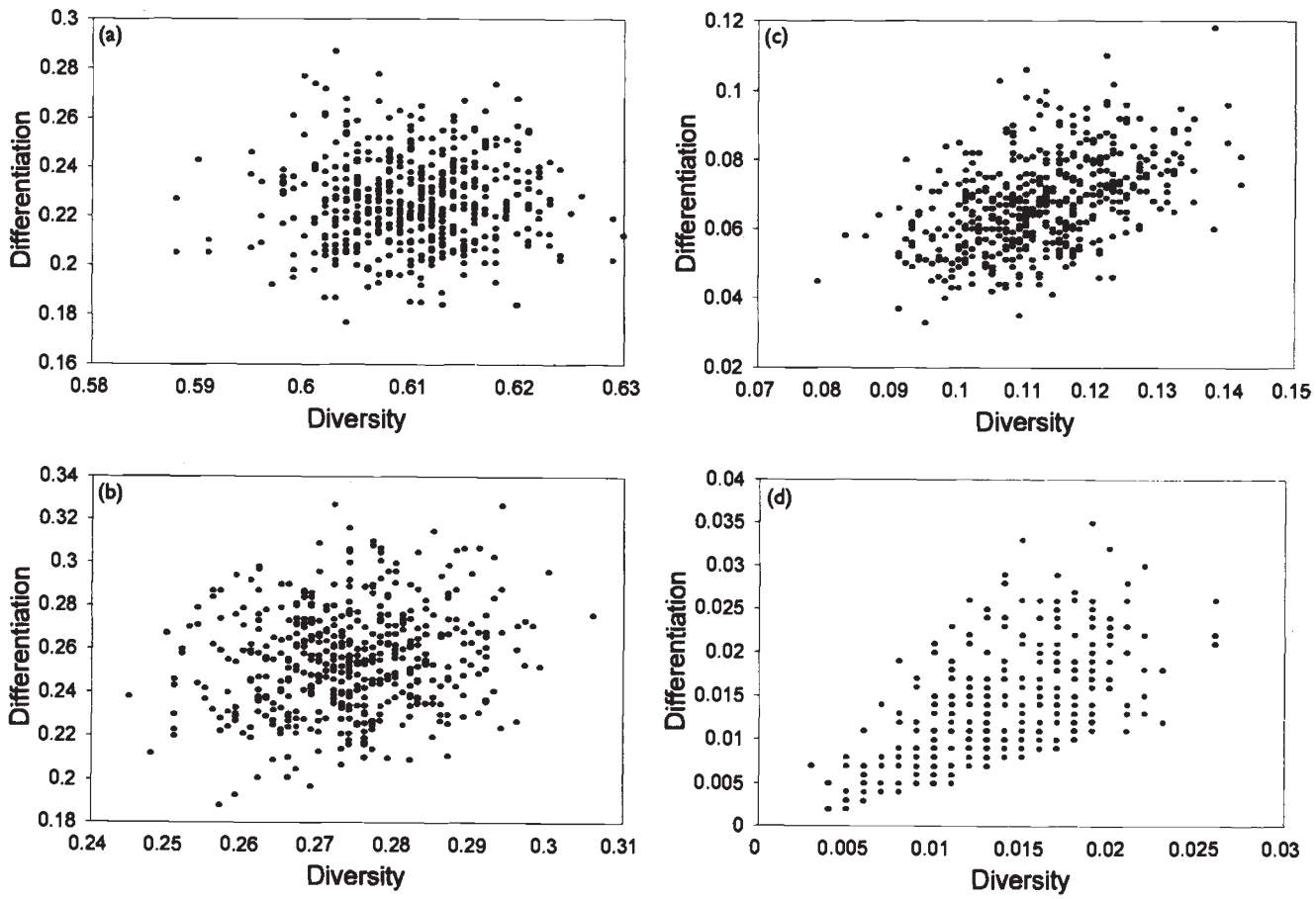

Fig. 2 Relationship between bootstrap values of differentiation and diversity at four loci. For each locus, individuals were sampled 1000 times with replacement in each of the six populations, and mean diversity h_T and differentiation G_{ST} were computed each time and plotted on the graph to examine their relationship. In Fig. 2(a) there is no relationship at the locus *Mdh-3* characterized by a high level of diversity ($h_T = 0.611$, $r = 0.029$). In Fig. 2(b) (locus *Pgi*, $h_T = 0.273$), the correlation is slight ($r = 0.194$). The relationship between differentiation and diversity becomes apparent (Fig. 2c) at the locus *Mdh-2* ($r = 0.478$) characterized by an intermediate level of diversity ($h_T = 0.112$). The locus *6-Pgd* (Fig. 2d) illustrates the case of a nearly fixed locus ($h_T = 0.013$) where the relationship between differentiation and diversity is striking ($r = 0.600$).

that the sample size is the same for each population. The maximum variance of p is then:

$$\mathrm{Var}_{(\max)}(p) = \frac{\sum_{i=1}^{L}(p_i - p)^2}{L} \quad (3)$$

which gives:

$$\mathrm{Var}_{(\max)}(p) = \{\mathrm{int}(Lp)(1-p)^2 + [\mathrm{rad}(Lp) - p]^2 \times [L - \mathrm{int}(Lp) - 1](0-p)^2\}/L \quad (4)$$

By substituting formula (4) in formula (2) and simplifying we get:

$$G_{ST(\max)} = \{\mathrm{int}(Lp)(1-2p) + \mathrm{rad}(Lp)^2 - 2\mathrm{rad}(Lp)p + Lp^2\}/Lp(1-p). \quad (5)$$

The maximum possible values of G_{ST} as a function of the allelic frequency are plotted in Fig. 3 where the number of populations studied (L) takes different values between 5 and 100. Because only six populations were examined in the present study, the less polymorphic loci were constrained to take low values of differentiation.

This suggests that, in order to obtain estimates of differentiation independent of the level of diversity, the loci displaying little variability should be eliminated in studies involving only a few populations. For example, for diallelic loci we could set $\min(p, 1-p) \geq 1/L$, which is equivalent to $h_T \geq 2(L-1)/L^2$). We have therefore eliminated here all loci with a diversity lower than 0.27. Seven loci were therefore eliminated for subsequent comparisons (five isozyme and two protein loci). After deleting the appropriate loci, the mean level of differentiation was recomputed for these two classes of markers and for all markers simultaneously. The new measures of diversity (Table 1) are appreciably higher, as might be expected, but the new values of differentiation are only slightly higher even for the isozymes where five of the eight loci had been eliminated.

Because new mean estimates of differentiation independent of the mean total diversities are available for each class of marker, they can be compared to each other. They appear to be extremely similar, in contrast to the large variability observed for single-locus estimates.

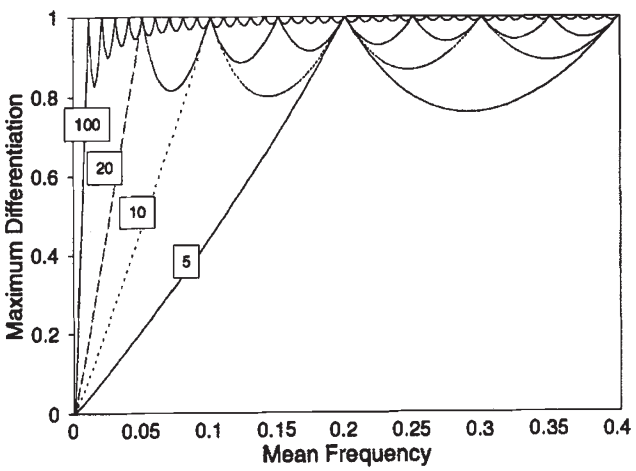

Fig. 3 Maximum possible values of differentiation at a diallelic locus when the number of populations sampled is limited. For a fixed number of populations (5, 10, 20 or 100) the upper limit of G_{ST} is provided as a function of the mean frequency of the rarer allele using eqn 5. For instance, when only five populations are sampled, regardless of the sample size per population, a locus with allele frequencies of 0.95 and 0.05 cannot have a G_{ST} value higher than 0.2. It will reach this maximal value if the less frequent allele is found only in a single population.

Discussion

To our knowledge, this is the first population genetics report where protein loci revealed using two-dimensional gel electrophoresis have been studied. Moreover, the protein loci are compared to other classes of biochemical markers (isozymes and terpenes). The fact that strikingly different values of diversity were observed among the three classes of markers is probably not of great biological significance, or, if so, is difficult to interpret. Note, indeed, that at loci characterized by the presence/absence of a product (the terpene loci in particular), only two alleles can be distinguished and that these loci have therefore a limited maximum value of diversity. In addition, in the case of the abundant proteins studied by two-dimensional electrophoresis, a bias was made in the selection of the polymorphic spot because we studied here only those polypeptides which were shown to be inherited in a Mendelian way, i.e. for which the maternal tree studied by Bahrman & Damerval (1989) was heterozygous. This certainly eliminated many less variable loci. For isozymes also, only loci which were known from preliminary studies to be at least slightly polymorphic were selected. More generally, different genes are likely to show extremely different values of diversity, because of different functional constraints acting on their products (Kimura, 1983). One of the main advantages of the coefficient of differentiation stems from the fact that, in theory, it should be independent of the diversity of the particular locus studied,

because it is precisely normalized by the diversity. Hence, comparisons of the level of differentiation of noncoding fractions of the genome with coding ones, or of conserved with more variable protein genes, should be valid. However, as we have shown here, differentiation is independent of diversity only as long as the number of populations studied is large enough. Note that, in a compilation of 655 studies of gene diversity in plants, the average number of populations studied was only 12.3 (Hamrick *et al.*, 1992). Hence, comparisons of levels of differentiation will not be rigorously valid in many cases and will lead to underestimates. However, because the mean differentiation is actually a weighted average which tends to correct for this problem, this limits to a large extent the importance of the bias for multilocus data. When comparisons must be made at the single-locus level, however, the problem may become quite acute unless many populations are sampled to estimate the differentiation.

It may be argued that as diversity approaches zero then so must differentiation, and that the results that we have obtained are quite obvious. However, when total diversity is zero, differentiation is undefined, and it is not clear whether cases of complete fixation should be given a value of G_{ST} of one or of zero (see discussion in Weir & Cockerham, 1984, p. 1360). Moreover, similar dependence of differentiation on diversity does not seem to have been reported to date. For instance, Nei (1987, p. 190) points out that G_{ST} is "highly dependent on the value of h_T. When this is small, G_{ST} may be large even if the absolute gene differentiation is small". Actually, as we have shown, small values of diversity are more likely to be associated with small values of differentiation.

Rare alleles should be on average of more recent origin than frequent ones, because at least some of them may have just appeared through mutation. If this is so, then we may expect that they are also more restricted geographically (i.e. more differentiated), because they have not yet reached equilibrium conditions between drift and migration and may have dispersed only to neighbouring populations. Clearly, as we have shown here, G_{ST} is not an appropriate measure to test these kinds of interesting predictions because of the artefacts that occur when computing this measure for rare alleles in limited surveys of genetic variation.

By removing those loci characterized by lower values of diversity, we obtained diversity-independent measures of differentiation that can be compared with each other. No differences were apparent among the various classes of loci compared. Hence, the results are consistent with the hypothesis that all these markers behave as if they were neutral and principally reflect factors such as drift which act at the whole genome level. In a recent study, Karl & Avise (1992) have found much lower levels of differentiation for isozyme loci than for noncoding nuclear regions, which led these authors to postulate that isozymes were subject to homogenizing selection as a consequence of heterozygote superiority. In the present study, this appears unlikely given the very high level of differentiation measured with most markers. This differentiation more probably reflects the very dissected circum-Mediterranean range of maritime pine and the history of the species.

Acknowledgements

We thank Dr A. Kremer, Dr R. Bacilieri and two anonymous reviewers for helpful comments on the manuscript and Th. Labbé for skilled computing assistance with the bootstrap. This research was supported in part by the contract BRG of MRT and by the contract EEC 'Forest' MA2B-CT91-0040.

References

BAHRMAN, N. AND DAMERVAL, C. 1989. Linkage relationships of loci controlling protein amounts in maritime pine (*Pinus pinaster* Ait.). *Heredity*, **63**, 267–274.

BAHRMAN, N. AND PETIT, R. J. 1995. Genetic polymorphism in maritime pine (*Pinus pinaster* Ait.) assessed by two-dimensional gel electrophoresis of needle, bud and pollen proteins. *J. Mol. Evol.*, **41**, (in press).

BAHRMAN, N., ZIVY, M., DAMERVAL, C. AND BARADAT, PH. 1994. Organisation of the variability of abundant proteins in seven geographical origins of maritime pine (*Pinus pinaster* Ait.). *Theor. Appl. Genet.*, **88**, 407–411.

BARADAT, PH., BERNARD-DAGAN, C., FILLON, C., MARPEAU, A. AND PAULY, G. 1972. Les terpènes du Pin maritime: aspects biologiques et génétiques. II. Hérédité de la teneur en monoterpènes. *Ann. Sci. For.*, **29**, 307–334.

BARADAT, PH., BERNARD-DAGAN, C., PAULY, G. AND ZIMMERMANN-FILLON, C. 1975. Les terpènes du Pin maritime. Aspects biologiques et génétiques. III. Hérédité de la teneur en myrcène. *Ann. Sci. For.*, **32**, 29–54.

BARADAT, PH. AND MARPEUA-BEZARD, A. 1988. *Le pin maritime. Biologie et génétique des terpènes pour la connaissance et l'amélioration de l'espèce.* Thèse doctorat d'Etat, Université Bordeaux-I.

BERNARD-DAGAN, C., FILLON, C., PAULY, G., BARADAT, PH. AND ILLY, G. 1971. Les terpènes du pin maritime: aspects biologiques et génétiques. I. Variabilité de la composition monoterpénique dans un individu et entre provenances. *Ann. Sci. For.*, **28**, 223–258.

BIRKY, C. W., JR, FUERST, P. AND MARUYAMA, T. 1989. Organelle gene diversity under migration, mutation,

and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, **121**, 613–627.

BRIDGES, J. R. 1987. Effects of terpenoid compounds on growth of symbiotic fungi associated with the Southern Pine Beetle. *Phytopathology*, **77**, 83–85.

CAVALLI-SFORZA, L. L. 1966. Population structure and human evolution. *Proc. R. Soc. B*, **164**, 362–379.

CONKLE, M. T., HODGSKISS, P. D., NUNNALLY, L. B. AND HUNTER, S. C. 1982. Starch gel electrophoresis of Conifer seeds: a laboratory manual. USDA-Forest Service, Pacific Southwest Forest and Range Experiment Station. Gen. Techn. Rep. PSW-64, 1–18.

DELORME, L. AND LIEUTIER, F. 1990. Monoterpene composition of the preformed and induced resins of Scots pine, and their effect on bark beetles and associated fungi. *Eur. J. Forest Pathol.*, **20**, 304–316.

GERBER, S., RODOLPHE, F., BAHRMAN N. AND BARADAT, PH. 1993. Seed-protein variation in maritime pine (*Pinus pinaster* Ait.) revealed by two-dimensional electrophoresis: genetic determinism and construction of a linkage map. *Theor. Appl. Genet.*, **85**, 521–528.

HAMRICK, J. L., GODT, M. J. AND SHERMAN-BROYLES, S. L. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

KARL, S. A. AND AVISE, J. C. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, **256**, 100–102.

KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

MARPEAU, A., BARADAT, PH. AND BERNARD-DAGAN, C. 1975. Les terpènes du Pin maritime: aspects biologiques et génétiques. IV. Hérédité de la teneur en deux sesquiterpènes: le longifolène et le caryophyllène. *Ann. Sci. For.*, **32**, 185–203.

MARPEAU-BEZARD, A., BARADAT, PH. AND BERNARD-DAGAN, C. 1983. Les terpènes du Pin maritime: aspects biologiques et génétiques. V. Hérédité de la teneur en limonène. *Ann. Sci. For.*, **32**, 185–203.

MÜLLER-STARCK, G., BARADAT, PH. AND BERGMANN, F. 1992. Genetic variation within European tree species. *New Forests*, **6**, 23–47.

NEI, M. 1977. *F*-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.*, **41**, 225–233.

NEI, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

PETIT, R. J., KREMER, A. AND WAGNER, D. B. 1993a. A finite island model for organelle and nuclear genes in plants. *Heredity*, **71**, 630–641.

PETIT, R. J., KREMER, A. AND WAGNER, D. B. 1993b. Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theor. Appl. Genet.*, **87**, 122–128.

POGSON, G. H. AND ZOUROS, E. 1994. Allozyme and RFLP heterozygosities as correlates of growth rate in the scallop *Placopecten magellanicus*: a test of the associative overdominance hypothesis. *Genetics*, **137**, 221–231.

STRAUSS, S. H. AND LIBBY, W. J. 1987. Allozyme heterosis in radiata pine is poorly explained by overdominance. *Am. Nat.*, **130**, 879–890.

WEIR, B. S. AND COCKERHAM, C. C. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

WRIGHT, S. 1951. The genetical structure of populations. *Ann. Eugen.*, **15**, 323–354.