Molecular Ecology (2003) **12**, 1577–1588 doi: 10.1046/j.1365-294X.2003.01835.x

Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers

OLIVIER J. HARDY
Laboratoire de Génétique et Ecologie végétales, Université Libre de Bruxelles, 1850 chaussée de Wavre, B-1160 Bruxelles, Belgium

Abstract

A new estimator of the pairwise relatedness coefficient between individuals adapted to dominant genetic markers is developed. This estimator does not assume genotypes to be in Hardy–Weinberg proportions but requires a knowledge of the departure from these proportions (i.e. the inbreeding coefficient). Simulations show that the estimator provides accurate estimates, except for some particular types of individual pairs such as full-sibs, and performs better than a previously developed estimator. When comparing marker-based relatedness estimates with pedigree expectations, a new approach to account for the change of the reference population is developed and shown to perform satisfactorily. Simulations also illustrate that this new relatedness estimator can be used to characterize isolation by distance within populations, leading to essentially unbiased estimates of the neighbourhood size. In this context, the estimator appears fairly robust to moderate errors made on the assumed inbreeding coefficient. The analysis of real data sets suggests that dominant markers (random amplified polymorphic DNA, amplified fragment length polymorphism) may be as valuable as co-dominant markers (microsatellites) in studying microgeographic isolation-by-distance processes. It is argued that the estimators developed should find major applications, notably for conservation biology.

Keywords: AFLP, dominant markers, isolation by distance, neighbourhood size, RAPD, relatedness

Received 19 November 2002; revision received 30 January 2003; accepted 30 January 2003

Introduction

Several estimators of pairwise relatedness coefficients between individuals using information from co-dominant genetic markers have been developed (e.g. Queller & Goodnight 1989; Loiselle *et al*. 1995; Ritland 1996; Hardy & Vekemans 1999; Lynch & Ritland 1999; Wang 2002). These estimators can be particularly useful to the study of sibship structure, isolation-by-distance in continuous populations, kin selection, inbreeding depression, and for marker-based inferences of quantitative inheritance in natural populations.

To obtain estimates of relatedness coefficients with sufficient precision, the polymorphism available (i.e. the number of polymorphic loci and the number of alleles per locus) is a critical factor (e.g. Ritland 1996; Lynch & Ritland 1999; Wang 2002). In this respect, microsatellites are often recognized as the most efficient genetic markers (Estoup & Angers 1998) because they typically display many alleles per locus (e.g. Streiff *et al*. 1998; Lynch & Ritland 1999) and are co-dominant. Precise relatedness estimates might also be obtained using dominant markers such as amplified fragment length polymorphisms (AFLP) or random amplified polymorphic DNA (RAPD) because these markers usually display a large number of polymorphic (di-allelic) loci (tens to hundreds of polymorphic loci are commonly reported, e.g. Albertson *et al*. 1999; Gerber *et al*. 2000; Degen *et al*. 2001a,b; Wilding *et al*. 2001). Furthermore, in comparison to co-dominant markers, dominant markers can be developed relatively easily even for species for which no prior genetic information is available and at a relatively low cost (Mueller & Wolfenbarger 1999). Consequently, dominant markers may represent excellent alternative tools to co-dominant markers to address questions requiring the estimation of relatedness between individuals. However, to my knowledge, only Lynch &

Correspondence: Olivier Hardy. Fax: + 32 2650 91 70; E-mail: ohardy@ulb.ac.be

Milligan (1994) developed an estimator of pairwise relatedness between individuals adapted to dominant markers. As Lynch & Milligan (1994) assumed Hardy–Weinberg genotypic proportions, their estimator is not adequate to assess relatedness in organisms where selfing occurs (e.g. many plant species), or to study isolation-by-distance processes where biparental inbreeding occurs. Therefore, the development of a new estimator for dominant markers that can account for heterozygote deficiency could be of broad interest for population geneticists.

In this paper, using the framework of quantitative genetics, new estimators of relatedness between individuals are developed, adapted to dominant markers. Contrary to Lynch & Milligan (1994), it will not be assumed that genotypes are in Hardy–Weinberg proportions, but it will be assumed nevertheless that we have an accurate idea of the departure from these proportions (i.e. the inbreeding coefficient is known). Computer simulations are then used to investigate the statistical properties of these new estimators, with the aim of (i) comparing their performance with the estimator proposed by Lynch & Milligan (1994) when inferring relatedness between simple types of relatives in the absence of inbreeding and (ii) assessing their performance to characterize isolation-by-distance processes within a population. Finally, a comparison is presented of the potential power of RAPD/AFLP vs. microsatellite markers to characterize spatial genetic structure.

Theory and analytical developments

Terminology and definitions

Throughout the population genetics literature much confusion occurs in the terminology used to name the different types of relatedness coefficients between individuals. In this paper, the term 'relatedness' coefficient will be used as a generic name for any coefficient describing some feature of the genetic similarity between individuals as a result of common ancestry. Among these relatedness coefficients the closely related 'kinship' (synonymous to 'co-ancestry') and 'relationship' coefficients will be distinguished, which depend on the probabilities of identity of homologous genes from different individuals when single pairs of genes are considered (definitions below). The term 'inbreeding' coefficient will express the similarity between homologous genes occurring within individuals. Additional types of relatedness coefficient exist, for example the 'fraternity' coefficient (Lynch & Walsh 1998), which depends on the probabilities of double identity of homologous genes from different individuals when quadruplets of genes from two diploids are considered. Here, estimators will be derived only for kinship and relationship coefficients.

To make the link with previous publications, the so-called 'relatedness' estimators of Queller & Goodnight (1989) and Lynch & Milligan (1994), as well as the '*r*' coefficients of Lynch & Ritland (1999) and Wang (2002), correspond to the 'relationship' coefficient according to the present terminology [which follows Wright (1922) who first defined this coefficient]. On the contrary, Ritland's (1996) 'relatedness' estimator, also denoted '*r*', as well as the 'ρ' estimator used by Loiselle *et al*. (1995), are estimators of the kinship coefficient following the present terminology.

Problems of consistency across the literature regarding the various relatedness coefficients are not over, as exact definitions of the parameters estimated are also challenging. As a population genetic parameter, the kinship coefficient between two individuals i and j, F_{ij}, is commonly defined as the probability of identity-by-descent (IBD), Θ, between a random gene from i and a random gene from j (e.g. Ritland 1996; Lynch & Ritland 1999). However, as shown by Rousset (2002), the marker-based estimators of F_{ij} coefficients cited above do not estimate a probability of IBD in general, notably when applied on individuals from a population under isolation-by-distance. The parameter assessed by marker-based estimators of F_{ij} is better defined as a ratio of differences of probabilities of identity-in-state (IIS) between homologous genes (Rousset 2002), and is sometimes called the 'conditional kinship' coefficient. I thus define pairwise kinship coefficients as

$$F_{ij} \equiv \frac{Q_{ij} - \bar{Q}}{1 - \bar{Q}} \quad (1)$$

where Q_{ij} is the probability of IIS between random genes from i and j, and $\bar{Q}$ between random genes within a 'reference population' (or a reference sample). Similarly, the relationship coefficient of i relative to j can be defined as

$$r_{ij} \equiv \frac{Q_{ij} - \bar{Q}}{(1 + Q_0)/2 - \bar{Q}} \quad (2)$$

when i and j are assumed to be diploids of identical inbreeding levels, Q_0 being the probability of IIS between genes within individuals (Hardy & Vekemans 1999). This definition is not general in the sense that it does not apply when i and j are not both diploids, situations where r_{ij} can be different from r_{ji}, but such cases will not be considered in this paper. Combining eqn 1 and eqn 2, we see that kinship and relationship coefficients are related in the following way:

$$r_{ij} = \frac{2F_{ij}}{1 + F_I} \quad (3)$$

where F_I is the inbreeding coefficient, defined as

$$F_1 \equiv \frac{Q_0 - \bar{Q}}{1 - \bar{Q}} \quad (4)$$

Hence, under Hardy–Weinberg genotypic proportions with diploids, $r_{ij} = 2F_{ij}$. Note that eqn 4 shows that the inbreeding coefficient is essentially a kinship coefficient between homologous genes within individuals.

The definitions given above are in terms of probabilities of IIS, which is convenient when using genetic markers because IIS is the sole information available. Alternatively, these coefficients could be defined in terms of ratio of differences of probabilities of IBD, simply replacing all Q by Θ in the preceding definitions, which would be more convenient when dealing with pedigree information. As a ratio of differences of probabilities of IIS approximates the equivalent ratio of differences of probabilities of IBD when mutation can be neglected (Rousset 2002), IIS- and IBD-based definitions are essentially equivalent under the low mutation limit. Thus, F_{ij} as defined above provides an approximation of $(\Theta_{ij} - \bar{\Theta})/(1 - \bar{\Theta})$ rather than of Θ_{ij}.

Conversion between marker-based and pedigree-based relatedness estimates

It is important to keep in mind that relatedness coefficients depend on a 'reference population' (or 'reference sample'), and express a degree of genetic similarity between individuals relative to the average genetic similarity between the individuals found in the reference population. Consequently, negative values of the relatedness coefficients may be obtained, meaning that *i* and *j* are less related on average than random individuals from the 'reference' population. Actually, as any two organisms on earth are assumed to be somehow genetically related by the standard theory of evolution, a relatedness coefficient must always be defined relative to some reference level of relatedness. However, the reference is not always the same according to the way relatedness is computed.

When relatedness is assessed from genetic markers, the 'reference population' is usually a sample of individuals (i.e. $\bar{Q}$ in eqn 1 is the probability of IIS of random genes from the sample). In this case, the average relatedness over all pairs of individuals within the sample is zero by definition. It follows that the relatedness coefficient estimates depend on the sampling scheme [a feature of relatedness coefficients criticised by Rousset (2002) who proposed alternative descriptors]. On the contrary, when relatedness is computed from pedigree information using path analysis (Lynch & Walsh 1998), the 'reference population' is represented by the ancestral individuals at the top of the pedigree (the base generation), and these ancestors are usually assumed to be equally related to each other. Hence, if estimates of relatedness based on genetic markers are to be compared with values expected from pedigree information, it is necessary to take into account that there is a shift of reference population (i.e. shift of $\bar{Q}$). Note that a pedigree provides IBD information, not IIS information, so that the kinship coefficient inferred by path analysis matches the definition given in eqn 1 only when mutations occurring since the ancestors can be neglected. I now show how to recalibrate relatedness coefficients while changing of reference population.

Let F^s represent kinship coefficients relative to a sample of individuals and F^p represents kinship coefficients relative to the base generation of a given pedigree. Using the pedigree information, F^s_{ij} can easily be deduced from F^p_{ij} using eqn 1:

$$F^s_{ij} = \frac{F^p_{ij} - F^p_s}{1 - F^p_s} \tag{5}$$

where F^p_s is the average kinship coefficient between individuals from the sample, relative to the base generation of the pedigree ($F^p_s \geq 0$).

To make the reverse conversion (i.e. estimating F^p_{ij} from the F^s_{ij} estimates obtained using genetic markers) it is necessary to identify pairs of individuals in the sample that are representative of 'unrelated' individuals from the point of view of the pedigree (i.e. individuals that do not share common ancestry since the base generation). Let $\hat{F}^S_{NR}$ be the average estimate of the kinship coefficient relative to the sample for these 'unrelated' individuals ($\hat{F}^S_{NR}$ is expected to be ≤ 0), F^p_{ij} can therefore be estimated as:

$$\hat{F}^p_{ij} = \frac{\hat{F}^s_{ij} - \hat{F}^s_{NR}}{1 - \hat{F}^s_{NR}} \tag{6}$$

Note that eqns 5 and 6 can also be applied on the relationship coefficients defined as in eqn 2, replacing F by r. Moreover, eqn 6 can be applied in a broader context to estimate coefficients relative to the level of relatedness between particular pairs of individuals, in which case $\hat{F}^S_{NR}$ represents the average kinship coefficient relative to the sample for these 'reference' pairs of individuals.

Equation 6 can only be applied when *a priori* knowledge on genealogy that permits identification of pairs of 'unrelated' individuals is available, as for example when maternal sib families are identified so that pairs of individuals from different families can be considered as 'unrelated'. In the absence of such information it might be tempting to use the F^s_{ij} estimates themselves to identify such pairs (e.g. considering the *ij* pairs showing the lowest F^s_{ij} estimates as a reference level), but this is not recommended because it is circular and would cause strongly biased estimates.

The link between the relatedness coefficients presented here and F-statistics merits a few words. F-statistics partition the genetic variance within and among structural entities (e.g. individuals, populations), and are also inbreeding (F_{IS}, F_{IT}) or kinship (F_{ST}) coefficients. The commonly used Weir & Cockerham's (1984) estimator of F_{ST} (Φ), estimates the ratio $(Q_w - Q_a)/(1 - Q_a)$, where subscripts w and a refer

to genes sampled within and among populations, respectively. Hence, F_{ST} is an average kinship coefficient among genes within populations, relative to genes among populations. If eqn 6 is applied on a hierarchically structured sample (e.g. sib families) to estimate the average kinship coefficient among individuals within families, considering individuals from different families as 'unrelated', the mean F_{ij}^{p} obtained is essentially equivalent to an F_{ST} estimate among families.

New estimators of relatedness coefficients using dominant markers

To develop relatedness estimators for a diploid organism using dominant genetic markers, a quantitative genetics approach will be used. The 'reference population' for these estimators will be a given sample of individuals. Firstly, it must be noted that the F_{ij} and r_{ij} definitions given above in terms of probabilities of IIS between genes can also be expressed in terms of correlation coefficients between allelic states of genes, making the link with quantitative genetics more straightforward. Thus, F_{ij} as defined in eqn 1 can equivalently be defined as the expected correlation coefficient between the allelic states of random genes from i and j. Similarly, r_{ij} as defined in eqn 2 can be defined as the expected correlation coefficient between the individual allele frequencies of i and j (individual allele frequencies equal 0, $^1/_2$, or 1; see Hardy & Vekemans 1999). To obtain relatedness estimators for a dominant marker, the correlation between the genotypic values expressed by the marker (the genotypic value of an individual is $X = 1$ if the dominant allele is present, otherwise it is $X = 0$), called the 'genotypic correlation', here denoted ρ_{ij}, is related to the relationship coefficient r_{ij}. The logic is that single locus genotypic values are directly observable by genetic markers (e.g. presence/absence of a band) so that ρ_{ij} values are easily estimated from a sample of genotyped individuals.

The expression of the correlation between the genotypic values of relatives at a single locus in the presence of inbreeding and dominance effect is complicated, involving six different types of relatedness coefficients and variance components (Cockerham & Weir 1984). However, the expression simplifies considerably in the absence of inbreeding and/or a dominance effect. Without inbreeding (but with dominance),

$$\rho_{ij} = h^2 r_{ij} + (1 - h^2)\,\Delta_{ij} \tag{7}$$

where h^2 is the ratio of the additive variance, Va, over the genotypic variance, Vt, the latter being the sum of the additive and dominance variance components ($Vt = Va + Vd$), and Δ_{ij} is the 'fraternity' coefficient between i and j (Lynch & Walsh 1998). Thus, h^2 can be interpreted as the narrow-sense heritability of the dominant marker. In eqn 7, the relationship coefficient, r_{ij}, accounts for the correlation due to additive effects, whereas the fraternity coefficient, Δ_{ij}, accounts for the correlation due to dominance effects. Thus, when $\Delta_{ij} = 0$ or $h^2 = 1$ (i.e. correlation depends only on additive effects), eqn 7 reduces to $\rho_{ij} = h^2 r_{ij}$, so that the relationship coefficient can be estimated as

$$\hat{r}_{ij} = \hat{\rho}_{ij}/\hat{h}^2 \tag{8}$$

Clearly, eqn 8 provides a biased estimate when $\Delta_{ij} > 0$ and $h^2 < 1$, but Δ_{ij} is positive only when i and j share common ancestry simultaneously through their mothers and fathers ($\Delta_{ij} = 0$, 0 and 0.25 between parent–offspring, half-sibs, and full-sibs, respectively). Hence, without inbreeding, eqn 8 should provide a good approximation of the relationship coefficient between individuals that are not related through both maternal and paternal genes. Moreover, when h^2 tends to 1 (dominance variance negligible), $\rho_{ij} = r_{ij}$, even in the presence of inbreeding (e.g. Lynch & Walsh 1998), so that eqn 8 remains exact. The estimators of relatedness coefficients proposed here will be based on eqn 8. They thus implicitly assume that nonadditive components of the correlation between the genotypic values of the individuals being compared can be neglected. The consequent bias of this simplification will be assessed later by simulation. For now we just need an estimator of h^2.

Standard Fisher decomposition of the variance expressed by a single locus dominant character in a diploid organism permits the expression of h^2 as a function of the frequency of the dominant allele, d, and the inbreeding coefficient (assumed to be constant across individuals), F_I:

$$h^2 = \frac{Va}{Vt} = \frac{2(1 - d(1 - F_I))}{(1 + F_I)(2 - F_I - d(1 - F_I))} \tag{9}$$

The quantity d is unknown but can be estimated from the frequency of the dominant genotype, D, by noting that $(1 - D)$ is the frequency of homozygotes for the recessive allele, so that

$$1 - D = (1 - d)^2(1 - F_I) + (1 - d)F_I \tag{10}$$

Thus, solving eqn 10, h^2 can be expressed as

$$h^2 = \frac{2}{1 + F_I}\,\frac{\sqrt{F_I^2 + 4(1 - F_I)(1 - D)} + F_I}{\sqrt{F_I^2 + 4(1 - F_I)(1 - D)} + 2 - F_I} \tag{11}$$

Equation 11 can be used to estimate h^2, replacing D by the observed frequency of the dominant phenotype in the sample ($\hat{D}$), and F_I by some independent estimate ($\hat{F}_I$). Hereafter, it is assumed that F_I is known. Note that simply replacing D and F_I by unbiased estimates in eqn 11

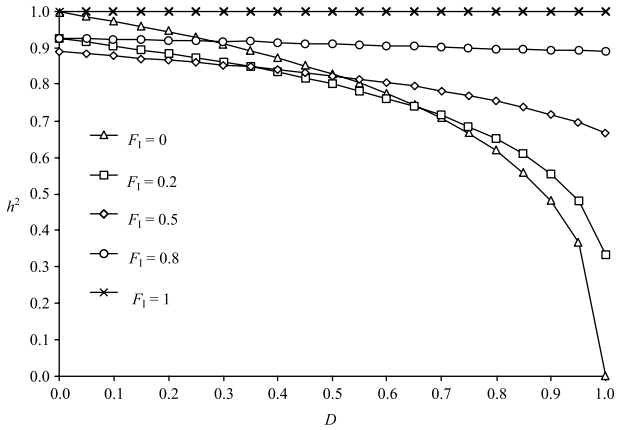

Fig. 1 Narrow sense heritability of the phenotype expressed by a dominant genetic marker (h^2) according to the frequency of the dominant phenotype (D) and the inbreeding coefficient (F_I).

provides a biased h^2 estimator because the equation is all but linear, but problems of bias will be addressed later directly on the relatedness estimators.

Figure 1 illustrates how h^2 varies according to D for different values of F_I, showing that h^2 can become very low for large D and low F_I, but approaches unity when D is low and/or F_I is large, conditions for which eqn 8 is therefore expected to result in a good approximation.

The phenotypic correlation between two individuals can be estimated as

$$\hat{\rho}_{ij} = \frac{Cov(X_i, X_j)}{Var(X)} = \frac{(X_i - \hat{D})(X_j - \hat{D})}{\hat{D}(1-\hat{D})} \quad (12)$$

with $X_i = 0$ for i showing the recessive phenotype, and $X_i = 1$ for the dominant phenotype. Equation 12 gives a somewhat biased estimator because individuals i and j belong to the sample used to estimate D. Most of this sampling bias can be corrected by adding the term $1/(n-1)$ to eqn 12, where n is the sample size. Single locus relationship and kinship coefficient estimates follow naturally as $\hat{r}_{ij} = \hat{\rho}_{ij}/\hat{h}^2$ (eqn 8) and $\hat{F}_{ij} = \hat{r}_{ij}(1 + \hat{F}_I)/2$ (eqn 3). Multilocus estimates can be obtained as weighted averages of single locus estimates. It is suggested that the ratio of the sums of the numerator terms over the denominator terms is taken, so that each locus is weighted approximately by $D(1-D)h^2$, giving more weight to loci with high h^2 values (which are expected to be less biased by dominance effects):

$$\hat{r}_{ij} = \frac{\sum_k (X_{ki} - \hat{D}_k)(X_{kj} - \hat{D}_k)}{\sum_k \hat{D}_k(1-\hat{D}_k)\hat{h}_k^2} \quad (13)$$

or, with the sample size correction,

$$\hat{r}_{ij} = \frac{\sum_k [(X_{ki} - \hat{D}_k)(X_{kj} - \hat{D}_k) + \hat{D}_k(1-\hat{D}_k)\hat{h}_k^2/(n-1)]}{\sum_k \hat{D}_k(1-\hat{D}_k)\hat{h}_k^2} \quad (14)$$

where $\hat{D}_k$ is the observed frequency of the dominant phenotype at locus k in the sample, and $\hat{h}_k^2$ is the estimated heritability for locus k, according to eqn 11. An estimator of the kinship coefficient for dominant markers, Fd_{ij}, can be derived combining eqns 3 and 14:

$$Fd_{ij} = \frac{1+F_I}{2} \frac{\sum_k ((X_{ki} - \hat{D}_k)(X_{kj} - \hat{D}_k) + \hat{D}_k(1-\hat{D}_k)\hat{h}_k^2/(n-1))}{\sum_k \hat{D}_k(1-\hat{D}_k)\hat{h}_k^2} \quad (15)$$

One should note the similarity between these estimators for dominant markers and the estimator of kinship coefficient described in Loiselle *et al*. (1995) and Kalisz *et al*. (2001) for co-dominant markers, here called Fc_{ij}:

$$Fc_{ij} = \frac{\sum_k (Y_{ki} - d_k)(Y_{kj} - d_k)}{\sum_k d_k(1-d_k)} + 1/(2n-2) \quad (16)$$

where Y_{ki} is the kth allele frequency in individual i ($Y_{ki} = 0$, $^1/_2$ or 1), and d_k is the kth allele frequency in the sample (the sums applying over all alleles of all loci). As demonstrated by Hardy & Vekemans (1999), the autocorrelation Moran's I statistic applied on individual allele frequencies is equal to $2/(1 + F_I)$ multiplied by estimator in eqn 16 (neglecting the sampling bias correction), and gives an estimator of the relationship coefficient. Hence, statistical properties that will be assessed for Fc_{ij} are also valid for Moran's I statistic.

Statistical performance of the new relatedness estimators

Marker-based relatedness estimates between individuals are notorious for their extreme associated variance (e.g. Lynch & Ritland 1999), and the problem is even more pronounced with dominant markers (Lynch & Milligan 1994). Hence, a marker-based estimate of relatedness obtained for a single pair of individuals is usually of little help to characterize the genealogical relationship between these individuals, which led Lynch & Milligan (1994) to conclude that the utility of dominant markers (they though RAPD in that time) for relatedness estimation is rather limited. There are however, two ways to overcome this difficulty: (i) using a huge number of polymorphic (and ideally unlinked) loci, or (ii) averaging estimates over many 'analogue' pairwise comparisons between individuals (i.e. pairs of individuals assumed to be equally related according to some independent knowledge). Throughout this paper, the second approach will be used. For example, knowing *a*

priori which pairs of individuals are sibs, an average relatedness estimate between sibs can be computed to infer which type of sibs occurs (half-sibs vs. full-sibs). Similarly, when investigating isolation-by-distance processes under isotropic dispersal, one can assume that the relatedness between individuals separated by a given spatial distance is constant and can be estimated by averaging relatedness estimates for all pairs of individuals separated by this distance. Regarding the first approach, the minimal number of loci required to assess the relatedness between two given individuals will be estimated by investigating the sampling variance associated with pairwise relatedness coefficients.

Performance of the estimators for simple types of relatives under random mating

Relatedness coefficients can be used to distinguish among different types of relatives, such as nonrelatives, parent–offspring, half-sibs, or full-sibs. The statistical properties of the relationship coefficient estimators were assessed on the basis of the average estimates between sibs and nonsibs from a sample of 100 individuals. The latter consisted of 20 families of five sibs (half-sibs or full-sibs) derived from a large random mating population. Such samples were generated by simulating random mating events in a large (infinite) parental population at Hardy–Weinberg equilibrium, assuming independent Mendelian inheritance at 100 diallelic loci (allele frequencies followed a uniform distribution between 0 and 1 across loci). The information at these loci was transformed to represent a dominant marker, and average single locus and multilocus relatedness coefficients for all pairs of sibs and for all pairs of unrelated individuals (i.e. members of different families) were computed on 100 replicates. This was done using the relationship coefficient estimator following eqn 14, hereafter denoted r_{H}, with $F_I = 0$, as well as using the estimator of Lynch & Milligan (1994; equations 17 and 18 in their paper), denoted r_{LM}.

To assess their bias, these r_{ij} estimators were compared with their expectations, which are equal to 0, 0.25 and 0.5 between unrelated individuals, half-sibs and full-sibs, respectively (e.g. Lynch & Walsh 1998), when the 'reference' population is the parental generation. However, it is first necessary to change of shift the reference population, the marker-based r_{ij} estimators being relative to the sample of sib families. To obtain the expected r_{ij} for sibs and unrelated individuals relative to the sample reference, eqn 5 was applied. The mean relatedness between random individuals relative to the parental generation is $r_s^p = 20 * (5 * 4) * 0.5/(100*99) = 0.0202$ in the case of full-sib families, and $r_s^p = 20 * (5 * 4) * 0.25/(100 * 99) = 0.0101$ in the case of half-sib families. Therefore, relative to the sample of sib families, the relatedness between nonrelatives (i.e. nonsibs) is $r_{NR}^s = (0–0.0202)/(1–0.0202) = –0.0206$ and $r_{NR}^s = (0–0.0101)/(1–0.0101) = –0.0102$ in the case of full-sib and half-sib families, respectively, whereas the relatedness between sibs is $r_{ij}^s = (0.5–0.0202)/(1–0.0202) = 0.4897$ and $r_{ij}^s = (0.25–0.0101)/(1–0.0101) = 0.2423$ in the case of full-sib and half-sib families, respectively. Conversely, relatedness between sibs relative to the parental generation can be estimated from the marker-based estimates relative to the sample using nonsibs as reference as in eqn 6:

$$\hat{r}_{ij}^p = (\hat{r}_{ij}^s - \hat{r}_{NR}^s)/(1 - \hat{r}_{NR}^s)$$

Figure 2 shows expectations and mean single locus estimates of r_{ij}^s and r_{ij}^p for nonsibs, half-sibs and full-sibs

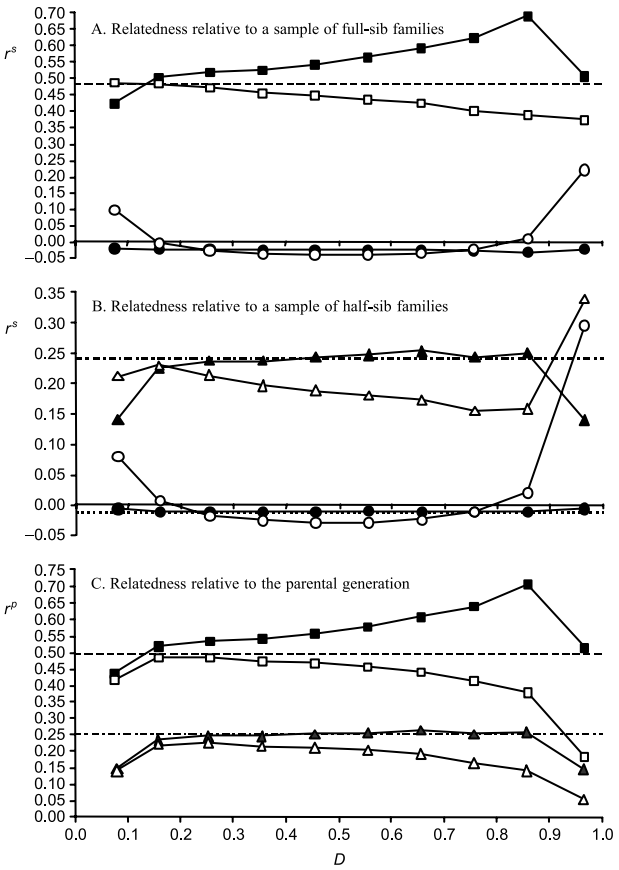

Fig. 2 Dependency of the estimators of relationship coefficients, *r*, on the frequency of the dominant phenotype, *D*. Mean single-locus estimates of the relationship estimator developed in this paper, r_{H} (●, ▲, ■), and Lynch & Milligan (1994) estimator, r_{LM} (○, ○, □), are presented for pairs of unrelated individuals (●, ○), half-sibs (▲, △) and full-sibs (■, □). The estimates are relative to a sample of half-sib families (A), full-sib families (B), or relative to the parental generation using nonsibs as reference (C). Horizontal lines represent theoretical expectations for full-sibs (stippled lines), half-sibs (broken dotted lines), and unrelated individuals, i.e. nonsibs (dotted lines).

according to the frequency class of the dominant phenotype, using either r_H or r_{LM}. The r_H estimator (filled symbols) gives slightly biased estimates for nonrelatives as well as for half-sibs, except at extreme frequencies of the dominant phenotype (downward bias), but it suffers substantial upward bias for full-sibs when the dominant phenotype is frequent. The r_{LM} estimator (open symbols) performs moderately well with nonrelatives, showing a slight downward bias at intermediate frequencies of the dominant phenotype but substantial upward bias at extreme frequencies, and it works fairly badly with half-sibs or full-sibs, generally suffering a downward bias of increasing importance with the frequency of the dominant phenotype.

The upward bias of estimator r_H for full-sibs is in line with our previous expectations, as the genotypic correlation between full-sibs is not purely additive, and h^2 decreases with higher frequencies of the dominant phenotype (Fig. 1). It is worth noting that the effect of extreme frequencies causing lower estimates (Fig. 2) is not confined to the r_H estimator adapted to dominant markers, as such an effect is also observed for estimators using the information from co-dominant markers, such as the one described by eqn 16 (results not shown), a point also made by Rousset (2002).

Average and standard deviations of the multilocus estimates (100 loci) relative to the sample of half-sib families were equal to $r_H = -0.010 \pm 0.002$ and $r_{LM} = 0.000 \pm 0.002$ between nonrelatives (expected value –0.010), and $r_H = 0.240 \pm 0.017$ and $r_{LM} = 0.185 \pm 0.013$ between half-sibs (expected value 0.242). Similarly, with full-sib families, $r_H = -0.024 \pm 0.003$ and $r_{LM} = -0.012 \pm 0.003$ between nonrelatives (expected value –0.021), and $r_H = 0.576 \pm 0.023$ and $r_{LM} = 0.433 \pm 0.017$ between full-sibs (expected value 0.490).

When estimates between sibs were transformed using nonsibs as reference to obtained relatedness relative to the parental generation, average and standard deviations of the multilocus estimates (100 loci) were, for half-sibs, $r_H = 0.248 \pm 0.018$ and $r_{LM} = 0.169 \pm 0.011$ (expected value 0.25), and for full-sibs, $r_H = 0.594 \pm 0.022$ and $r_{LM} = 0.416 \pm 0.015$ (expected value 0.5).

In conclusion, the r_{LM} estimator suffers downward bias for both half and full sibs, whereas the r_H estimator developed in this paper seems essentially unbiased for nonrelatives and half-sibs, but is upwardly biased for full-sibs. The results also suggest that the approach developed for changing of reference population (eqns 5 and 6) is appropriate for comparisons of marker-based relatedness estimates with pedigree-based expectations, as confirmed when this approach was applied on relatedness estimators adapted to co-dominant markers (eqn 16, results not shown).

To assess the sampling variance associated with relatedness estimates between two individuals (single pair), 1000 replicates of the simulation procedure presented above were made and estimates of r_H (relative to the sample) between two half-sibs and between two unrelated individuals were obtained using one to 1000 loci. For n loci, the standard error (SE, the square root of the sampling variance) of the r_H estimates was approximately equal to $\sqrt{3}/n$ (the variance was inversely proportional to the number of loci), although the SE was slightly higher between half-sibs than between nonrelatives (Fig. 3). To find the minimum number of loci required for assessing relatedness between two given individuals, one must define the precision necessary for r_H estimates. For example, to be able to distinguish between half-sibs and unrelated individuals, the SE should be less than about 0.1. To reach such precision, 300 loci would be required (Fig. 3), leading to 92% correct classifications (results not shown). However, such precision is insufficient to distinguish, for example, first cousins from unrelated individuals (r_H differ by about 0.0625), in which case nearly 5000 loci would be necessary to obtain SE ≅ 0.025. In reality, the prospective precision would not be reached even with such a large number of loci because many loci would be linked along the chromosomes, providing nonindependent information (loci were unlinked in the simulations).

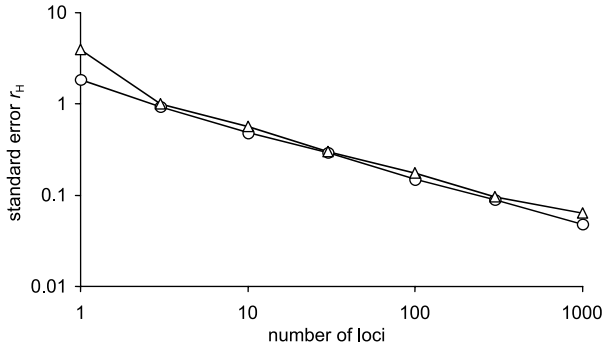

Fig. 3 Standard error of the r_H coefficient computed for a single pair of individuals according to the number of loci (log–log scale) in the case of half-sibs (△) or unrelated individuals (○).

In conclusion, with dominant markers, relatedness coefficients are not efficient tools with which to assess the parentage between two given individuals. For such a purpose, a better alternative would be a likelihood approach, testing whether two individuals fall into a particular parentage class (e.g. Thompson 1975; Mousseau *et al*. 1998), exploiting the interlocus information (which is not exploited by relatedness coefficients).

Performance of the estimators under isolation by distance

Under isolation-by-distance processes, the relationship between F_{ij} (or r_{ij}) and the spatial distance, d_{ij}, between individuals is predicted by analytical models (Rousset 1997, 2000; Hardy & Vekemans 1999; Barton *et al*. 2002). A

convenient theoretical result predicts that, in a two-dimensional space, F_{ij} decreases approximately linearly with the natural logarithm of d_{ij}, $\ln(d_{ij})$, and the rate of decay (slope) is close to $-(1 - F_0)/(4\pi\delta_e\sigma^2)$, where $\pi = 3.14$, δ_e is an effective density of individuals (i.e. accounting for the variance in reproductive success among individuals), σ^2 is the axial variance of gene dispersal distances, and F_0 is the kinship coefficient between adjacent individuals (which is used as an approximation of the kinship between competing gametes before selection; F. Rousset, personal communication). The linearity between F_{ij} and $\ln(d_{ij})$ is actually best observed within a limited distance range, approximately between σ and 20σ (Rousset 1997). The quantity $4\pi\delta_e\sigma^2$ can be interpreted as a neighbourhood size, N_b, which characterizes the extent of local genetic drift (Wright 1943). Hence, characterizing the spatial genetic pattern of populations subject to isolation-by-distance using kinship coefficients can provide indirect estimates of the neighbourhood size.

It was investigated whether the estimator of the kinship coefficient developed herein, Fd_{ij}, permits accurate estimates of the neighbourhood size to be obtained. Because Hardy–Weinberg proportions cannot be assumed under isolation-by-distance, the Lynch and Milligan estimator will not be considered, but results will be compared with those obtained for a diallelic co-dominant marker, using the estimator from eqn 16, Fc_{ij}, so that the loss of accuracy and/or precision due to the dominant nature of a marker will be evaluated.

The performances of the F_{ij} estimators were assessed on data sets obtained by simulating an individual-based isolation-by-distance model, similar to that described in Hardy & Vekemans (1999). Basically, a population of 1600 hermaphrodite diploid individuals filling a 40×40 squared grid was simulated with discrete (i.e. nonoverlapping) generations. Each individual was characterized at up to 200 diallelic loci. The initial generation was generated by drawing alleles at random for each locus, where initial allele frequencies followed a uniform distribution between 0.05 and 0.95 across loci. In subsequent generations, new individuals were produced by drawing, independently for each locus, an allele from each of two parents from the previous generation. Parents were randomly chosen assuming that gametes disperse according to a centred isotropic bivariate normal distribution of variance $\sigma^2 = 4$ lattice units^2. The self fertilization rate (i.e. the probability that the two gametes came from the same parent) depends on the dispersal law, but it could be forced, in proportion s, to control the inbreeding level. Mutation and immigration were not implemented, hence allele frequencies fluctuated by genetic drift alone. Actually, a substantial level of immigration would have affected the spatial genetic structure, resulting in biased estimates of local gene dispersal (Hardy & Vekemans 1999). Simulations were stopped after 200 generations, a time sufficiently long for the spatial genetic structure to reach a quasi-equilibrium state (Hardy & Vekemans 1999), and sufficiently short to avoid a substantial reduction of genetic variability by drift. As $\delta = 1$ for the lattice, the neighbourhood size was $N_b = 4\pi[4 + 4(1 - s)]/2$, giving $N_b = 50.3$ when $s = 0$ (no forced autogamy).

According to runs, all individuals along one to 10 parallel transects were sampled, excluding individuals at five or fewer positions from the population edge to avoid border effects. Hence, sample size varied between 30 and 300. On each sample, the pairwise F_{ij} were estimated first by treating genotypes as co-dominant markers (eqn 16), Fc_{ij}, and second by transforming genotypes into phenotypes of dominant markers (eqn 15), Fd_{ij}. To calculate Fd_{ij}, the F_I considered is the mean inbreeding coefficient calculated over all loci.

Single locus and multilocus Fc_{ij} and Fd_{ij} values were regressed on the natural logarithm of i–j spatial distances, providing regression slopes, b. N_b can be estimated as $-(1 - F_0)/b$, where F_0 is the average F_{ij} estimate for adjacent i, j individuals, but because such an estimate of N_b can reach extreme values when b approaches zero, or meaningless negative values when b is positive, the reciprocal estimate, $b/(F_0 - 1)$ was reported, and compared to the expected $1/N_b$ value. It is worth noting that the $b/(F_0 - 1)$ ratio is independent of the reference population used to computed kinship coefficients [the $(1 - \bar{Q})$ terms in the numerator and denominator cancel], so that changing the reference population is needless.

The average and standard deviations of $b/(F_0 - 1)$ when kinship coefficients are estimated using the information available from a dominant or a co-dominant marker are illustrated on Fig. 4 for different sampling schemes, where the sample size or the number of loci scored varied (there was no forced autogamy). These averages are all similar to each other and close to the expected value ($1/N_b = 0.0199$). Some downward bias is however, observed when the sample size is small (30 individuals), this effect being observed both for dominant and co-dominant markers. The standard deviations reduce when the sample size or the number of loci increase. As a rough approximation, the variance is reduced by two (the standard deviation by $\sqrt{2}$) when the number of loci or the number of sampled individuals is doubled. On average, the variance is 65% larger (the standard deviation 28% larger) for the estimates based on the information from a dominant marker relative to a co-dominant marker.

When autogamy is forced to some level, results remain very similar, Fd_{ij} and Fc_{ij} leading both to accurate estimates, but the difference of precision (standard deviations) between estimates based on dominant and co-dominant information is reduced (results not shown). At the extreme, when $F_I = 1$ (complete homozygosity), Fd_{ij} values are almost identical to Fc_{ij} values.

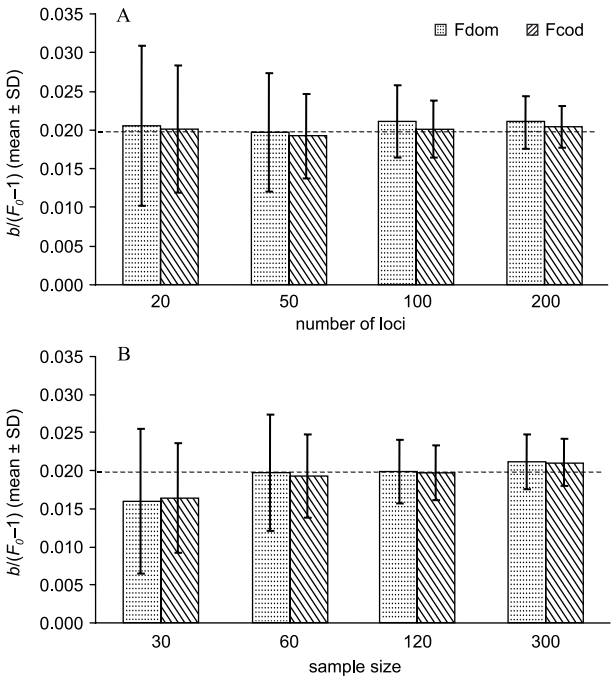

Fig. 4 Mean and standard deviation of the ratio $b/(F_0 - 1)$, estimating the reciprocal of the neighbourhood size, for a dominant (stippled bars) or a co-dominant (hatched bars) marker according to (A) the number of loci assayed (with a sample size = 60), (B) the sample size (with 50 loci assayed). Stippled lines show the expected value according to the simulated gene dispersal parameters ($1/N_b = 0.0199$).

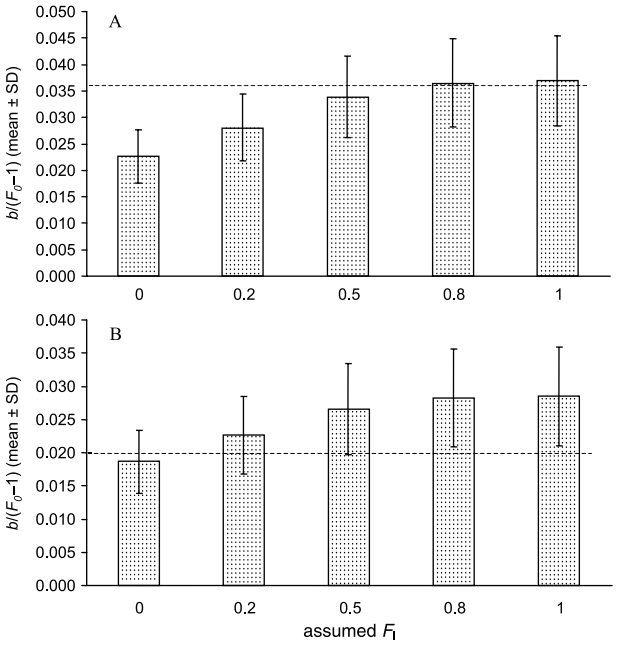

Fig. 5 Mean and standard deviation of the ratio $b/(F_0 - 1)$, estimating the reciprocal of the neighbourhood size, for a dominant marker according to the assumed inbreeding coefficient when (A) the actual inbreeding coefficient is high ($F_I = 0.82$), (B) the actual inbreeding coefficient is low ($F_I = 0.04$). Stippled lines show the expected values according to the simulated gene dispersal parameters: (A) 90% forced autogamy, $1/N_b = 0.0362$; (B) no forced autogamy, $1/N_b = 0.0199$. Estimates are based on a sample of 120 individuals assayed at 50 loci.

These results demonstrate that the kinship estimator developed here is well suited to characterize isolation-by-distance. The inherent bias of the estimator, as observed previously for full-sibs, appears to be negligible (even under inbreeding), meaning that essentially all the phenotypic correlation among individuals at a dominant marker is additive.

Up to now, the true inbreeding coefficient was assumed to be known when computing Fd_{ij}, but this information is not necessarily available in practice. It is thus necessary to assess the impact of an error made on the assumed inbreeding coefficient on the accuracy of the estimates. Figure 5 shows that when there is little inbreeding ($F_I = 0.04$), assuming erroneously a large value of the inbreeding coefficient causes an overestimation of the degree of spatial structure, hence an underestimation of the neighbourhood size. Reciprocally, when there is much inbreeding (here caused by forcing 90% of self-fertilization, so that $F_I = 0.82$), assuming erroneously a low value of the inbreeding coefficient causes an underestimation of the degree of spatial structure, hence an overestimation of the neighbourhood size. It is noteworthy that a moderate error on the assumed F_I value, for example of no more than 0.2, causes a bias in the ratio $b/(F_0 - 1)$) of less than 15%. A large error (e.g. assuming $F_I = 1$ when the actual $F_I = 0$, or reciprocally) causes a bias of no more than 40%. Thus, the accuracy of an estimate of the neighbourhood size using dominant markers appears fairly robust to the error made on the assumed level of inbreeding.

Relative performances of RAPD or AFLP vs. microsatellite markers to characterize spatial genetic structure

Microsatellite markers are generally considered the best tool to address questions relative to microgeographic genetic structure because they usually display high numbers of alleles per locus compared to other co-dominant markers (Estoup & Angers 1998), the level of polymorphism available being a critical factor to get precise inferences. The dominant markers provided by RAPD or AFLP techniques may also be efficient for such studies because they usually display a large number of loci (e.g. Albertson *et al*. 1999; Mueller 1999; Gerber *et al*. 2000; Degen *et al*. 2001a,b; Wilding 2001). To compare the precision that can be obtained with RAPD or AFLP markers and microsatellites when characterizing

Table 1 Comparison of the precision [SD of $b/(1-F(1))$ ratios] provided by dominant and co-dominant markers with realistic levels of polymorphism when characterizing spatial genetic structure in simulated isolation-by-distance processes

Marker/species/sample size used to set initial allele frequencies in simulations	No. of polymorphic loci*	$b/(1-F(1))$	
		Mean	SD
Dominant marker			
RAPD/*Vouacapoua*/59	40	0.0206	0.0049
AFLP/*Quercus*/43	147	0.0210	0.0024
Co-dominant marker			
Microsatellite/*Vouacapoua*/187	8 (45)	0.0201	0.0061
Microsatellite/*Quercus*/43	6 (96)	0.0200	0.0038

SD is the standard deviation computed over 100 independent simulation runs.
*Total no. of alleles.

microgeographic genetic structure, a population was simulated under isolation-by-distance, as described before, and dominant or co-dominant markers with realistic levels of polymorphism were considered. Examples used were the levels of polymorphism observed within populations for two allogamous tree species: *Vouacapoua americana* (Aublet) for RAPD (Degen *et al*. 2001a) and microsatellites (Dutech *et al*. 2002), and *Quercus petraea* for AFLP (Gerber *et al*. 2000) and microsatellites (Streiff *et al*. 1998). The polymorphism available from microsatellite markers (total number of alleles) was moderate in *Vouacapoua americana* and higher in *Quercus petraea* (Table 1). Similarly, for the dominant markers, the available polymorphism (number of polymorphic loci) was higher for AFLP markers in *Quercus petraea* than for RAPD markers in *Vouacapoua americana* (Table 1; note that this does not mean that gene diversity is higher in *Quercus*, just that there was more polymorphism available from the markers developed on *Quercus*). The initial allele frequencies used for the simulations were thus (i) those observed at microsatellite loci in either of the two tree species, or (ii) the square root of the frequency of the null phenotype at each polymorphic RAPD or AFLP band, which should closely approximate the null allele frequencies as each species showed genotypic proportions close to Hardy–Weinberg expectations. After 200 generations, regression slopes, b, of Fc_{ij} (for microsatellites) or Fd_{ij} (for RAPD or AFLP, with $F_I = 0$) values on the logarithm of the spatial distance were computed on a sample of 180 individuals, and 100 replicates were run. The mean and standard deviations of the $b/(F_0 - 1)$ ratios are reported in Table 1.

For all parameter sets, the mean $b/(F_0 - 1)$ values were close to their theoretical expectation (0.0199), although dominant markers gave slightly upwardly biased estimates. As expected, within dominant or within co-dominant markers, higher precisions (lower SD) were obtained when the number of polymorphic loci or the total number of alleles was higher (Table 1). The precision provided by RAPD and AFLP markers was of the same order of magnitude as that provided by microsatellites, and within each species the dominant marker actually provided somewhat more precision than microsatellites (Table 1). Hence, RAPD and AFLP techniques can be as valuable as microsatellites in the characterization of spatial genetic structure, at least when the inbreeding coefficient can be estimated independently (and RAPD bands are reliable).

The same sample of 43 mapped individuals of *Quercus petraea* was scored at AFLP and microsatellite loci, so that the actual spatial genetic structures assessed by each marker type could be compared. Figure 6 shows that both types of markers provided congruent pictures of the genetic structure. Consistency between markers is furthermore demonstrated by the $b/(F_0 - 1)$ ratios, which were equal to 0.0101 (SE = 0.0062) for microsatellites and 0.0135 (SE = 0.0049) for AFLP (approximate standard errors were obtained by jackknifing over loci). When 10 000 randomizations of individual spatial locations were performed to test for the spatial structure, the observed $b/(F_0 - 1)$ ratio exceeded the value obtained after randomization in 9855 cases for microsatellites and 9957 cases for AFLP, showing that the power to detect spatial structure was somewhat higher with AFLP markers than microsatellites for this particular example.

In the example given, the method used to infer gene dispersal from spatial genetic structure may seem to be not very efficient as the estimated standard errors are close to

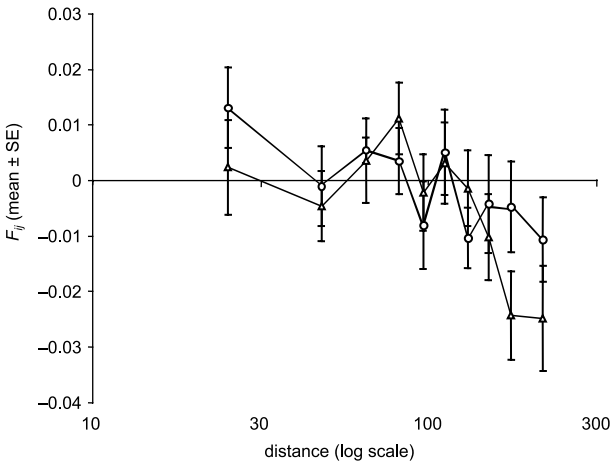

Fig. 6 Mean kinship coefficients between individuals in a population of *Quercus petraea* as assessed using six microsatellite markers (○) and 147 AFLP markers (△). The sample consisted of 43 genotyped individuals; each of the 10 distance classes involves 89–92 pairs of individuals. Error bars represent mean ± SE, the latter being assessed by a jackknifing procedure over loci.

the average $b/(F_0 - 1)$ values. The reason is the small sample size (microsatellites were originally scored on 166 trees but the data set had to be reduced to allow legitimate comparison with AFLP markers which were scored on only 43 trees) and the efficiency of pollen dispersal in oaks causing a weak level of structuring in *Q. petraea*. When the same analysis was performed on 46 mapped individuals of the co-occurring species, *Q. robur*, also scored for the same set of AFLP (Gerber *et al*. 2000) and microsatellite (Streiff *et al*. 1999) markers, no spatial genetic structure could be detected with any of the markers in this species (results not shown), which is consistent with the better seed dispersal abilities of *Q. robur* compared to *Q. petraea* [Streiff *et al*. (1999) detected genetic structuring with microsatellites in *Q. robur* but using a sample of 183 trees]. It is thus clear that gene dispersal inference in species with extended dispersal abilities requires fairly large sample sizes to obtain reasonably precise estimates using microsatellites or AFLP/RAPD markers.

A very rough way to compare the potential precision offered by different markers is to compute, for dominant markers, the number of polymorphic bands, and for co-dominant markers, the total number of alleles minus the number of loci. The logic behind this is that for a locus with equi-frequent alleles, the variance of kinship or relationship coefficients estimates is approximately proportional to the number of alleles minus one (Lynch & Ritland 1999). Actually, the exact variance depends on the allele frequencies, the degree of relatedness, and the statistic considered but one can check that this simple guideline ranks correctly the level of precision obtained by simulations (Table 1). It should be noted that in all the simulations presented here, independence among the loci was assumed. Clearly, when tens or hundreds of loci are assayed in some species, such as in RAPD or AFLP studies, the occurrence of pairs of linked loci is highly probable. This should not affect the accuracy of the estimates, but should somewhat lower their precision because there is some redundancy in the information. Therefore, in the presence of linked loci, methods of numerical resampling of the loci (e.g. jackknife, bootstrap) are likely to provide somewhat underestimated standard errors.

Discussion

Because they suffer high sampling variance, marker-based relatedness coefficients are not efficient to identify precisely how two given individuals are genealogically related, especially using dominant markers. However, when prior information allows the identification of pairs of individuals that should be equally related, average relatedness coefficients can be very useful. It is shown herein that the estimators of relatedness coefficients developed here are well suited to characterize spatial genetic structure, leading to essentially unbiased estimates. They may be less suited to characterize sibship structure because they overestimate relatedness between full-sibs, but would still be useful to distinguish half-sib families from full-sib families, because the bias for full-sibs amplifies the difference in relatedness estimates between half- and full-sibs.

When relatedness coefficients are used to assess parentage relationships between individuals by comparing them to theoretical expectations, it is essential to account for the references to which these coefficients are relative. This is true irrespective of the dominant or co-dominant nature of the markers. Thus, the expected marker-based relationship coefficient between, say, two half-sibs is not necessarily 0.25, as it also depends on the relatedness among all pairs of individuals present in the sample used as reference. An approach to change of reference was presented and proved efficient. But such a method is applicable only if some prior knowledge about the genealogical structure of the sample is available (i.e. the origin of each sampled individual with respect to family, nest, or another structural unit is known). On a sample of undefined individuals, marker-based relatedness coefficients are generally useless, in part because the associated error on an estimate between two individuals is much too high, even when many loci are available.

It has also been shown that RAPD and AFLP markers can be as efficient as microsatellites in characterizing spatial genetic structure. Nevertheless, dominant markers require prior knowledge of the inbreeding coefficient. The latter can be obtained (i) using a co-dominant marker, (ii) using the dominant markers if an outbred offspring generation can be obtained from the parental generation (Lynch & Milligan 1994), or (iii) from knowledge of the mating system. However, for some species, none of these approaches may be applicable. Although the accuracy of the estimates is rather robust to a moderate error on the assumed inbreeding coefficient, the method might be improved to avoid the need of an independent assessment of the inbreeding level. To this end, a potential option would be to develop an estimation procedure in the framework of Bayesian inference, where the inbreeding coefficient would be given uniform prior probabilities between 0 and 1 (Holsinger *et al*. 2002).

For population geneticists, the possibility of using RAPD or AFLP markers to assess relatedness between individuals and to study microgeographic isolation-by-distance processes is promising because, compared to microsatellite markers, many polymorphic loci can be obtained fairly easily, in a relatively short time, and at a relatively low cost (Mueller 1999). Hence, the methods developed in this paper should find major population genetics applications, notably in the field of conservation genetics, where molecular markers need to be developed at reasonable cost.

The estimators of relatedness coefficients described in this paper will be available in the software SPAGEDI (Hardy

& Vekemans 2002), which can be downloaded from the following website: http://www.ulb.ac.be/sciences/lagev/spagedi.html.

Acknowledgements

I thank Cyril Dutech, Henri Caron and Stephanie Mariette for providing me with data sets. I am also grateful to Xavier Vekemans, Antoine Kremer, Sophie Gerber as well as three anonymous referees for their comments on a previous version of the manuscript. I thank the Belgian National Fund for Scientific Research (FNRS) where I am a Postdoctoral Researcher.

References

Albertson RC, Markert JA, Danley PD, Kocher TD (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the USA*, **96**, 5107–5110.

Barton NH, Depaulis F, Etheridge AM (2002) Neutral evolution in spatially continuous populations. *Theoretical Population Biology*, **61**, 31–48.

Cockerham CC, Weir BS (1984) Covariances of relatives stemming from a population undergoing mixed self and random mating. *Biometrics*, **40**, 157–164.

Degen B, Caron H, Bandou E *et al*. (2001a) Small scale spatial genetic structure of six tropical tree species in French Guiana. In: *Genetic Response of Forest Systems to Changing Environmental Conditions* (eds Müller-Starck G, Schubert R), pp. 75–92. Kluwer Academic Publishers, Dordrecht.

Degen B, Caron H, Bandou E *et al*. (2001b) Fine-scale spatial genetic structure of eight tropical species as analysed by RAPDs. *Heredity*, **87**, 497–507.

Dutech C, Seiter J, Petronelli P, Joly HI, Jarne P (2002) Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana. *Molecular Ecology*, **11**, 725–738.

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: *Advances in Molecular Ecology* (ed. Carvalho GR), pp. 55–86. IOS Press, Amsterdam.

Gerber S, Mariette S, Streiff R, Bodénès C, Kremer A (2000) Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology*, **9**, 1037–1048.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Holsinger KE, Lewis PO, Dey DK (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, **11**, 1157–1164.

Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: The roles of dispersal, mating, history and selection. *Evolution*, **55**, 1560–1568.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, **3**, 91–99.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753–1766.

Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland MA.

Mousseau TA, Ritland K, Heath DD (1998) A novel method for estimating heritability using molecular markers. *Heredity*, **80**, 218–224.

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution*, **14**, 389–394.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetics Research, Cambridge*, **67**, 175–185.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Rousset F (2002) Inbreeding and relatedness coefficients: what do they measure? *Heredity*, **88**, 371–380.

Streiff R, Labbe T, Bacilieri R *et al*. (1998) Within-population genetic structure in *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*, **7**, 317–328.

Thompson EA (1975) The estimation of pairwise relationships. *Annals of Human Genetics*, **39**, 173–188.

Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**, 1203–1215.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611–619.

Wright S (1922) Coefficients of inbreeding and relationship. *American Naturalist*, **56**, 330–338.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

O. J. Hardy is a postdoctoral researcher developing approaches to infer gene dispersal parameters and adaptive responses at quantitative traits from the analysis of spatial genetic structure. He is also developing computer programs for data analyses.
