Heredity (2002) 88, 371–380
© 2002 Nature Publishing Group All rights reserved 0018-067X/02 $25.00
www.nature.com/hdy

Inbreeding and relatedness coefficients: what do they measure?

F Rousset
Laboratoire Génétique et Environnement, Institut des Sciences de l'Évolution, Université de Montpellier II, 34095 Montpellier, France

This paper reviews and discusses what is known about the relationship between identity in state, allele frequency, inbreeding coefficients, and identity by descent in various uses of these terms. Generic definitions of inbreeding coefficients are given, as ratios of differences of probabilities of identity in state. Then some of their properties are derived from an assumption in terms of differences between distributions of coalescence times of different genes. These inbreeding coefficients give an approximate measurement of how much higher the probability of recent coalescence is for some pair of genes relative to another pair. Such a measure is in general not equivalent to identity by descent; rather, it approximates a ratio of differences of probabilities of identity by descent. These results are contrasted with some other formulas relating identity, allele frequency, and inbreeding coefficients. Additional assumptions are necessary to obtain most of them, and some of these assumptions are not always correct, for example when there is localized dispersal. Therefore, definitions based on such formulas are not always well-formulated. By contrast, the generic definitions are both well-formulated and more broadly applicable.
Heredity (2002) **88,** 371–380. DOI: 10.1038/sj/hdy/6800065

Keywords: relatedness; population structure; identity-by-descent; coalescence, kin selection

Introduction

Concepts of relatedness, measuring the genetic relationships among individuals, are basic to population genetics. They were initially conceived as measures of genetic likeness due to recent shared ancestry given by pedigree relationships, and as such they are standard tools in quantitative genetics and in kin selection theory. However, there are cases where 'relatedness' measures may be used even though the shared ancestry is not given by a single well-identified pedigree. For example, it was clear since Wright's early work that classical measures of population structure such as '*F*-statistics' (Wright, 1951) may be viewed as measures of relatedness among individuals in spatially subdivided populations. Although 'relatedness' may be defined in an infinite number of ways, not all measures are equally relevant to quantitative models of evolution. It is useful to distinguish parameters that do not depend on mutation (such as 'relatedness' below) and related measures that may depend on mutation (such as 'inbreeding coefficients' below). The most common uses of relatedness measures in spatially subdivided populations are to quantify the relative effects of drift and migration, and to quantify selection in ways more or less analogous to Wright's (1931) initial attempt in this direction. In particular, measures of relatedness may be needed to develop an 'inclusive fitness' framework for measuring selection (eg, Hamilton, 1971; Crow and Aoki, 1984; Taylor, 1988; Rousset and Billiard, 2000).

Some formulas are familiar enough to population geneticists to be taken as basic and even as definitions of relatedness in these different contexts. One such formula expresses the probability that two genes are of a given allelic type, as $rp + (1 - r)p^2$ where p is the 'allele frequency' in a 'reference population' (or 'base population', Falconer and Mackay, 1996) and r is a 'relatedness' measure, or 'probability of identity-by-descent'. Whatever the exact definition of these terms, if r is independent of p, it can be computed by independent methods, such as recursion methods for probabilities of identity, or directly from pedigrees. Such an r can then be used to predict the probability that two genes are of a given allelic type, given p. However, it is assumed that r is independent of p, which raises the question whether it is actually so.

The definition of identity in terms of an ancestral population does lead to some correct computations for a number of basic models, but it may also be questioned *per se*. When confronted with the concept of identity by descent, and to its computation from a real pedigree (eg Hartl and Clark, 1997; Lynch and Walsh, 1998), one may wonder what is the significance of a number that ignores the identity due to common ancestry of members of the 'reference population'.

Some further problems with commonly used definitions of relatedness will be illustrated when the different concepts involved have been defined (see Discussion). However, evidence of difficulties may be found in the claims that there is 'something arbitrary' in the definition of relatedness (Maynard Smith, 1998, p 141;

Correspondence: F Rousset, Laboratoire Génetique et Environnement, Institut des Sciences de l'Évolution, Université de Montpellier II, 34095 Montpellier, France
E-mail: Rousset@isem.univ-montp2.fr
Received 2 August 2001; accepted 15 January 2002

Table 1 Notation for measures of identity and coalescence

p, p_k	frequency of allele k
π_k	expectation of frequency of allele k
Q_j and $\dot{Q}_j$	probabilities of identity in state and identity by descent, respectively, for a pair of genes specified by index j
Q_w, Q_b	w and b are generic notations for identity of pairs of genes compared $w$ithin and $b$etween some structural units
$:k$ as in eg $Q_{j:k}$	allelic type qualifier; eg $Q_{j:k}$ is the probability of identity in state, both genes being of allelic type k
$\|p$ as in eg $Q_{j:k\|p}$	conditioning on allele frequency p; eg $Q_{j:k\|p}$ is $Q_{j:k}$ given allele frequency is p in the population
$\dot{Q}(t^*)$	the probability of coalescence before time t^* (t^* included)
F	generic notation for inbreeding coefficients, $(Q_w - Q_b)/(1 - Q_b)$
$\dot{F}$	same as F but in terms of identity by descent
T, T_w, T_b	average coalescence time; T_w and T_b for genes $w$ithin and $b$etween some structural units, respectively
C	$(T_b - T_w)/T_b$
c_t ($c_{j,t}$)	probability of coalescence, at time t in the past (of a pair of genes specified by index j)

see also Cotterman, 1940, reprinted 1974, quoted below), or that, when computing relatedness, 'we are not attempting to characterize a reality' (Jacquard, 1975, p 342). As emphasized by Grafen (1985), this is certainly not what one should expect from a definition of relatedness suitable for the analysis of biological processes.

There are alternative definitions of relatedness in the literature, but there is little discussion of their relationships to each other. The aim of this paper is to compare some definitions of relatedness parameters and their properties, pointing that these difficulties follow from using some definitions, and not from using some others. Some of the notation used below is summarized in Table 1.

A generic definition of inbreeding coefficients

Several approaches, based either on statistical considerations or on theoretical analysis of evolutionary processes, have led to the following definition of inbreeding coefficients.

Inbreeding coefficients are defined in terms of the probability of identity in state of different pairs of genes. Here the probability of identity in state is simply the probability that two genes are of identical allelic type. For example, for microsatellite loci, allelic type may be characterized by allele size, or it may be characterized by the exact DNA sequence.

Consider a population structured in some way (geography, age structure, etc). The probability of identity will depend on whether one compares genes within subpopulations, between subpopulations, and so on. One may define Q_w, the probability of identity within a 'structural unit', or 'class of genes' (for example among individuals within the same age class, in the same geographical area, etc), and Q_b, the probability of identity between genes in two different structural units, eg two subpopulations. In a generic way one can define a parameter F of the form:

$$F \equiv \frac{Q_w - Q_b}{1 - Q_b}. \quad (1)$$

This definition is 'generic', ie it is not based on the consideration or the properties of a particular model. For example we do not assume a particular mutation model. We only consider that populations follow some unspecified random (stochastic) process. Note that the probability of identity in state is not the frequency of identical pairs of genes in a biological population (which, in many models of interest, will be a random variable, not a parameter). The probability of identity in state is the expectation of the frequency of identical pairs of genes in some sample or population. These expectations are parameters, ie functions of the parameters defining the model, whatever these parameters may be (deme sizes, mutations rates, and so on).

'1' in the above definition may be viewed the probability of identity of a gene with itself. More generally, inbreeding coefficients may be defined as a ratio of differences in probabilities of identity. A simple conceptual message underlying a ratio of differences is that it compares more and less identical individuals, rather than 'related' *vs* 'unrelated' individuals. Which ratio it is best to consider depends on the biological process considered and, secondarily, may be a matter of convenience. For example, in the analysis of models with localized dispersal (at least), it may be convenient to consider parameters of the form $(Q_w - Q_\mathbf{r})/(1 - Q_w)$, where Q_w is the probability of identity of different genes within a deme, and $Q_\mathbf{r}$ is the probability of identity of genes at some geographical distance $\mathbf{r}$ (Rousset, 1997; Rousset and Billiard, 2000).

The well-known F-'statistics' originally considered by Wright may be defined as above. Such definitions were explicitly considered by, for example, Takahata (1983) and Crow and Aoki (1984) (inspired from Nei's (1973) similar definitions in terms of frequencies of identical pairs of genes) and were further discussed by Cockerham and Weir (1987, 1993) and Nagylaki (1998). For Wright's F_{ST}, Q_w is the probability of identity within a deme and Q_b is the probability of identity between demes. Likewise, Wright's F_{IS}, Q_w is the probability of identity of the two homologous genes in a diploid individual, and Q_b is the probability of identity of two genes in different individuals.

Probabilities of identity in state depend on the mutation process. A measure of relatedness that does not take into account the mutation process may be more appealing. However, in many models of interest, the value of inbreeding coefficients, defined following the above generic expression (1), is only weakly dependent on the mutation model. As emphasized by Crow and Aoki (1984), this is a necessary condition if such measures are to yield information about pedigrees or genealogies, which do not depend on mutation.

The approximate independence from mutation cannot arise *ex nihilo*: it must depend on underlying assumptions

regarding the biological process. We will formulate an assumption in terms of the comparison of distributions of coalescence times of the pairs of genes that define the inbreeding coefficients.

An assumption and its implications

We consider the probability $c_{i,t}$ that two genes have their most recent common ancestor ('coalesce') at time t in the past. The i index corresponds to the type of pair of genes considered (two homologous genes within a diploid individual, two genes in different individuals, and so on) and we will use the w and b indices as in the previous Section.

We assume that the probabilities of coalescence $c_{w,t}$ and $c_{b,t}$ become proportional to each other in the distant past. This simple assumption has a number of consequences, that we first describe graphically, and then more formally.

Graphical argument

To illustrate our argument, we will consider different examples. One example illustrates the computation of relatedness from a pedigree in a panmictic population. To keep mathematics to a minimum, the particular case considered in Figure 1a is relatedness between two genes in a selfed individual in a panmictic population with random mating (including selfing). That is, here identity Q_w is for the two genes borne by a selfed individual, while Q_b is for genes borne by two random individuals in the population. The second example (Figure 1b) is an island model with selfing, detailed in Rousset (1996). The third (Figure 1c) is a stepping stone model.

A property observed in these three examples is that the probabilities of coalescence $c_{w,t}$ and $c_{b,t}$ become proportional to each other in the distant past. Thus we can split the area covered by the probability distribution of coalescence times of more related genes (the area delimited by $c_{w,t}$) into two parts. Take the area below the $c_{b,t}$ curve (the distribution of coalescence times of less related genes) and consider this surface reduced by the value of the ratio $c_{w,t}/c_{b,t}$ for large t. For large t, this reduced area coincides with the area delimited by $c_{w,t}$. The other part is the rest of the area delimited by $c_{w,t}$. These two areas are shown in Figure 1b for the comparison of genes within individuals ($c_{w,t}$) and between individuals within demes ($c_{b,t}$). This second area (lightly shaded in Figure 1b) is restricted to the recent past.

We will see that, as a first approximation, the inbreeding coefficient F, defined as a ratio of differences of probabilities of identity, equals this 'initial area', ie relatedness equals the increased probability of coalescence in the recent past.

Formal argument

The assumption that the probabilities of coalescence $c_{w,t}$ and $c_{b,t}$ become proportional to each other in a distant past may be expressed as follows (Rousset, 2001): for two different pairs of genes, the limit $\lim_{t\to\infty} c_{w,t}/c_{b,t}$ exists and is finite. This limit may be computed in models of population structure, as detailed in the Appendix. For the selfed individual example of Figure 1(a), $c_{w,t}/c_{b,t}$ is constant for any $t > 1$. Actually $c_{w,1} = 1/2$ for genes from the selfed individual, $c_{b,1} = 1/(2N)$ for random individuals, and for both we have $c_{j,t} = (1 - 1/(2N))^{t-2}(1 - c_{j,1})/(2N)$ for $t > 1$. Thus $c_{w,t}/c_{b,t} = N/(2N - 1)$ for $t > 1$.

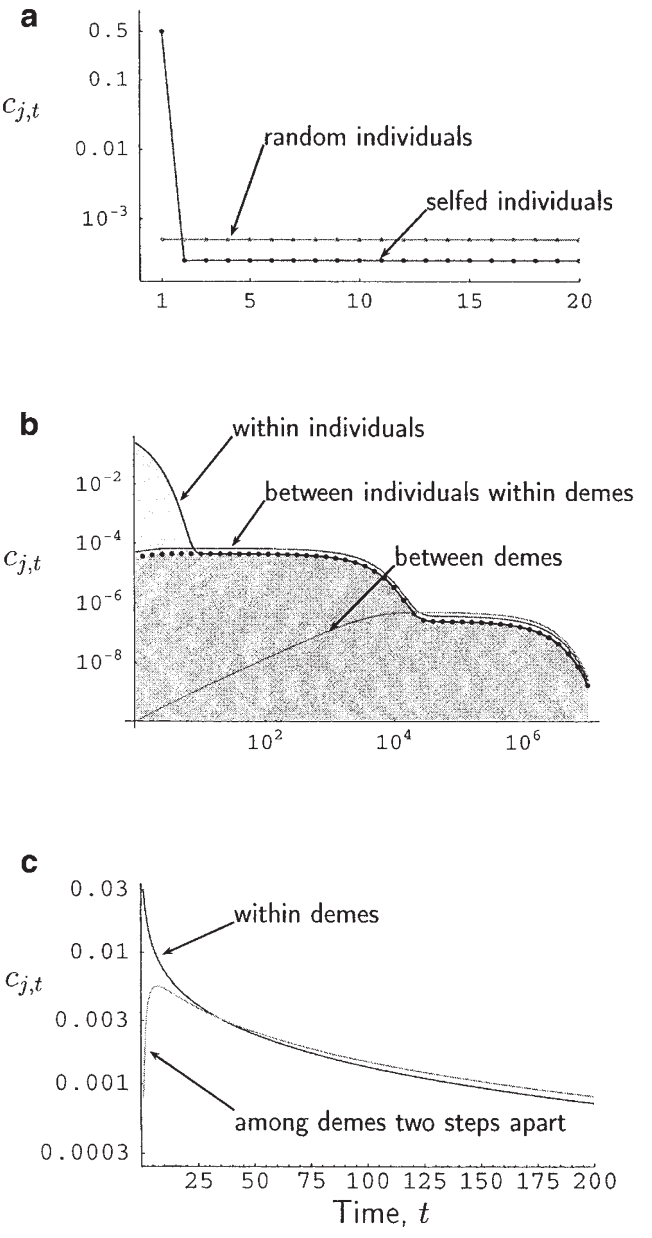

Figure 1 Probabilities $c_{j,t}$ of coalescence at t. This figure compares distributions of coalescence times of different pairs of genes, used to define inbreeding coefficients. Three different cases are considered. (**a**) Selfed individuals in a panmictic, diploid, randomly mating (including selfing) hermaphroditic population of $N = 1000$ individuals. Each offspring may be produced by selfing with probability $1/N$, independently of each other. (**b**) An island model with selfing (see Rousset, 1996, for details), with 100 demes of $2N = 20000$ genes, a dispersal rate $m = 1/N$, and a selfing rate 0.5. $j = 0$: two genes within the same individual; $j = 1$: two genes in different individuals within a deme; $j = 2$: two genes in different demes. Distributions of coalescence times are shown as plain lines. The shaded surface below the dotted line is constructed from the surface covered by the distribution of coalescence times of genes between individuals, reduced as described in the text. The shaded area above the dotted line is the 'initial area' for F_{IS}. Redrawn from Rousset (1996). (**c**) A one-dimensional stepping stone model, 100 demes of $N = 10$ haploid individuals, dispersal rate $m = 1/4$. Redrawn from Rousset (2001).

For models in which $\lim_{t\to\infty} c_{w,t}/c_{b,t}$ exists and is finite, one may then define

$$\omega \equiv 1 - \lim_{t\to\infty} c_{w,t}/c_{b,t}. \quad (2)$$

The height of the 'initial area' at time t is then

$$g(t) \equiv c_{w,t} - (1-\omega)c_{b,t}. \quad (3)$$

Given that the two distributions $c_{w,t}$ and $c_{b,t}$ must each sum to 1 ($\Sigma_{t=1}^{\infty} c_{w,t} = \Sigma_{t=1}^{\infty} c_{b,t} = 1$), if we sum (3) over t, we find that

$$\sum_{t=1}^{\infty} g(t) = \omega. \quad (4)$$

Thus ω is both the 'initial area' and the asymptotic proportional factor between probabilities of coalescence defined by equation 2. These two interpretations of the same quantity have been separately pointed out in different analyses (eg Chesser *et al*, 1993; Rousset, 1996).

For pedigrees in panmictic populations, τ can be defined exactly, such that $g(t) = 0$ for $t > \tau$. In the above example, selfed individuals have $\tau = 1$ ($g(1) = \omega$). More generally, there is no obvious way to define τ accurately: the value of comparing distributions of coalescence times is to provide an intuitive understanding of more exact results. For the example of Figure 1b, a value of τ may be chosen as the time where $c_{w,t} = c_{b,t}$. Thus $\tau \approx 20$ for $c_{0,t}$ *vs* $c_{1,t}$, and $\tau \approx 30000$ for $c_{1,t}$ *vs* $c_{2,t}$. Likewise Figure 1c suggests $\tau \approx 20$.

The validity of the assumption on distributions of coalescence times must itself be proven under any particular model. This is done in the Appendix for the island model, and for local relatedness under isolation by distance. A notable exception concerns average inbreeding coefficients of the form $(Q_w - \bar{Q})/(1-\bar{Q})$, involving the probability of identity within demes, Q_w, and the probability of identity averaged across all possible spatial distances, $\bar{Q}$. In a stepping stone model, a new problem appears: for $Q_b = \bar{Q}$, $\lim_{t\to\infty} c_{w,t}/c_{b,t}$ is approached increasingly slowly as the number of demes increases. Thus the properties and possible uses of such coefficients will be distinct from those reviewed here. Indeed, similar parameters appear in expressions for effective size (eg Wright, 1943; Maruyama, 1972; Whitlock and Barton, 1997), but not as relatedness parameters in some analyses of selection (Rousset and Billiard, 2000).

Definitions in terms of identity by descent

Here we review two definitions of inbreeding coefficients in terms of two concepts of identity by descent. The first definition is related to ω, and the second is a special case of the previous definition of F. Hence, by further showing the relationship between ω and F, we will tie all definitions together.

Identity by descent may be defined as the total probability of coalescence between now and some time t^*. A time-dependent definition of 'F_{ST}' is then obtained by computing a ratio of differences of such identities:

$$F(t^*) \equiv \frac{\sum_{t=1}^{t^*}(c_{w,t} - c_{b,t})}{1 - \sum_{t=1}^{t^*} c_{b,t}} = 1 - \frac{\sum_{t=t^*+1}^{\infty} c_{w,t}}{\sum_{t=t^*+1}^{\infty} c_{b,t}} \quad (5)$$

Similar definitions were considered by Chesser *et al* (1993), Wang (1997), and Whitlock and Barton (1997). The dependence on t^* is removed by considering the asymptotic value of $F(t^*)$ for large t^*. Given $\lim_{t^*\to\infty} c_{w,t^*}/c_{b,t^*} = 1 - \omega$, this asymptotic value is ω. The time scale at which this value is approached is also given by τ since for $t^* \geq \tau$,

$$F(t^*) = \frac{\sum_{t=1}^{t^*}(g(t) + (1-\omega)c_{b,t} - c_{b,t})}{1 - \sum_{t=1}^{t^*} c_{b,t}} \approx \frac{\omega\left(1 - \sum_{t=1}^{t^*} c_{b,t}\right)}{1 - \sum_{t=1}^{t^*} c_{b,t}} = \omega. \quad (6)$$

Identity by descent may also be defined as the probability $\dot{Q}_j$ that there has not been any mutation since the common ancestor, so that

$$\dot{Q}_j = \sum_{t=1}^{\infty} c_{j,t}(1-u)^{2t} \quad (7)$$

(Malécot, 1975, equation 6; Slatkin, 1991). Correspondingly, we can define the identity-by-descent version of F (eg Slatkin, 1991):

$$\dot{F} \equiv \frac{\dot{Q}_w - \dot{Q}_b}{1 - \dot{Q}_b}. \quad (8)$$

Since $\dot{Q}$ is also the identity in state in the 'infinite-allele model', $\dot{F}$ is a special case of F.

The effects of mutation

Given there is some τ such that $\Sigma_{t=1}^{\tau} g(t) \approx \omega$ and that mutation can be neglected in the first τ generations, we may intuitively expect that the inbreeding coefficient F will be weakly dependent on mutation and will be approximately ω. Slatkin (1991) noticed a relationship between $\dot{F}$ and the average coalescence times of pairs of genes, which can be extended to the identity in state parameter F as follows. In a finite population and for different mutation models, $Q_j = 1 - 2uT_j + O(u^2)$ where T_j is the average coalescence time of a pair of genes of type j, and $O(u^2)$ is a residual term which scales as the square of the mutation rate. It follows that the limit value of F is a ratio of coalescence times, T_w and T_b:

$$\lim_{u\to 0}\frac{Q_w - Q_b}{1 - Q_b} = \lim_{u\to 0}\frac{1 - 2uT_w - (1 - 2uT_b)}{1 - (1 - 2uT_b)} = \frac{T_b - T_w}{T_b} \equiv C. \quad (9)$$

Thus, in the low mutation limit, the identity in state and identity-by-descent parameters measure the same 'relatedness' measure C (Slatkin, 1995; Rousset, 1996).

The effects of mutation rate may be understood as follows. Let q_t be the probability of identity in state of a pair of genes which coalesce t generations in the past. If q_t were a linear function of the coalescence time of these pairs of genes ($q_t = 1 - 2ut$, for example), one would have $F = C$. More generally, writing $q_t = 1 - 2ut + R(t)$ where $R(t) = O(u^2)$ is the deviation from linearity, the difference between F and C is proportional to $\Sigma_1^{\infty} R(t)g(t)$. Hence the difference between F and C is more

important when the relationship between divergence $1 - q_t$ and coalescence time t is more strongly nonlinear and when $g(t)$ remains large in the distant past. The latter condition occurs in island models with low migration rates, or over large distances under models of isolation by distance (Slatkin, 1995; Rousset, 1996, 1997).

There are simple mathematical analogies between the $1 - Q$ terms and measures of divergence between pairs of genes based on sequence divergence (eg Hudson, 1990), additive genetic variance (eg Lande, 1992), or variance in allele size (eg Slatkin, 1995). In each case these measures of divergence between pairs of genes are assumed to be linearly related to their realized coalescence time, hence the value of the F_{ST} analogues, defined from these measures of divergence, is C.

Exact relationships between ω and inbreeding coefficients

When does $\dot{F} = \omega$? From equations 2 and 7, it follows that

$$\dot{Q}_w - \dot{Q}_b = \sum_{t=1}^{\infty} ((1-\omega)c_{b,t} + g(t) - c_{b,t})(1-u)^{2t}$$

$$= \sum_{t=1}^{\infty} (-\omega c_{b,t} + g(t))\,(1-u)^{2t} \quad (10)$$

$$= \omega\left(1 - \sum_{t=1}^{\infty} c_{b,t}\,(1-u)^{2t}\right) - \sum_{1}^{\infty} g(t)\,(1-(1-u)^{2t}) \quad (11)$$

(where we have inserted $\omega - \Sigma_t g(t)$ which is null by equation (4))

$$= \omega\,(1-\dot{Q}_b) - \sum_{1}^{\infty} g(t)\,(1-(1-u)^{2t}). \quad (12)$$

Hence

$$\dot{F} \equiv \frac{\dot{Q}_w - \dot{Q}_b}{1-\dot{Q}_b} = \omega - \frac{\sum_{1}^{\infty} g(t)\,(1-(1-u)^{2t})}{1-\dot{Q}_b}. \quad (13)$$

Hence in general $\dot{F} \neq \omega$. The low mutation limit value of $\dot{F}$ may be written

$$C = \omega - \frac{\sum_{1}^{\infty} tg(t)}{T_b}. \quad (14)$$

Hence in general, $\lim_{u\to 0} F = C \neq \omega$. But there is an important exception, that of migration models with an infinite number of demes, such as the infinite island model or more generally models of isolation by distance on an infinite lattice. In the latter case it is shown in the Appendix that

$$\lim_{u\to 0} \dot{F} = C = \omega \quad (15)$$

ie $\Sigma_1^{\infty} tg(t)/T_b \to 0$ as the number of demes $n \to \infty$.

It may be checked from the algebra of island or isolation by distance models that $\dot{F}$ is weakly dependent on the number of demes, as noted for related quantities by Crow and Aoki (1984) or Rousset (1997). Further, ω for the finite population model is itself close to ω for the infinite population model, so $\dot{F}$ for the finite population model is close to ω for the infinite population model. Since $F(t^*)$ is asymptotically equivalent to ω (equations 5 and 6), $F(t^*)$ is asymptotically equivalent to the low-mutation value of $\dot{F}$ when this value is ω, ie for large number of subpopulations. These results tie together the different definitions of relatedness or inbreeding coefficients for low mutation and large number of subpopulations.

Concepts of reference population

We have seen that F approximates a ratio of differences in probabilities of identity by descent ($\dot{F}$), rather than a probability of identity by descent. Such conclusions may seem to conflict with usual arguments according to which 'inbreeding coefficients' measure 'identity by descent' (eg Hartl and Clark, 1997; Lynch and Walsh, 1998). Here we discuss such an argument, based on the concept of a 'reference population', and show that when it is correctly interpreted, it leads to the same ratios of differences of identities as considered above. Under some conditions, this reduces to an identity by descent.

Definitions of relatedness in terms of a 'reference population' were introduced by Cotterman (1940, reprinted 1974):

> "[A definition of identity] should also be, if possible, a mathematically exact one, but so far the author has been unable to fulfill this requirement. We may say that [identical] genes shall be taken to mean two or more genes derived recently, in terms of generations of adults, from some common gene or one from the other. But precisely how recently? Again, in the absence of a definite criterion we may say 5 or 6 generations for the human population. Though this is quite arbitrary, it is nevertheless serviceable for several reasons.
>
> First, the chance that mutation should have occurred during this time is in most cases quite negligible, whereas it would not be so for some longer period. Hence in the solution of many statistico-genetic problems we may choose to assume that mutation is absent and that all derivative genes must be identical with but little loss of accuracy. Secondly, inbreeding which comes about through the occurrence of a common ancestor more distantly removed than 5 or 6 generations will have entirely negligible genetic effects…"

This is often interpreted as follows.

Relatedness relative to a reference biological population

One defines relatedness as the total probability of coalescence between now and t^*, $\dot{Q}(t^*) \equiv \Sigma_{t=1}^{t^*} c_{w,t}$ (this is the first definition of identity by descent previously considered). Let p be the frequency of allele k in a 'reference' biological population at time t^*. Consider at t^* the probability $Q_{.k}(t^*)$ that two genes are identical in state, and both of type k. If we suppose that there is no mutation between now and t^*, then given p, the probability of identity is

$$Q_{.k}(t^*) = \dot{Q}(t^*)p + (1-\dot{Q}(t^*))p^2 \quad (16)$$

(eg Crow and Kimura, 1970, section 3.2). This is of the form $rp + (1-r)p^2$ *for* $r = Q_{.k}(t^*)$.

To obtain (16), one assumes first that the ancestral allele frequency at time t^* is identical to the present allele frequency. That is, we neglect drift in allele frequency p (and mutation) over time span t^*. This results from considering 'infinite' populations, for t^* bounded (equation 16 is of interest only for t^* bounded, since as $t^* \to \infty$, $\dot{Q}(t^*) \to 1$ so that one would have $Q_{.k} = p$, a result that contains no information about relatedness). The fraction r of pairs of genes that have coalesced by time t^* then accounts for the term rp. Second the argument assumes

that genes that have not coalesced by time t^* are effectively independent. This accounts for the term $(1 - r)p^2$.

The assumption that such genes are 'effectively independent', given they have not coalesced by time t^*, is the weak part of this argument. The comparison of distributions of coalescence times is helpful in understanding why the underlying logic is not generally correct, but is still correct in some classical models.

Consider again Figure 1. In the finite island model, the more demes, the lower the probability that ancestral lineages meet in the same deme at time t. More precisely, if we let the number of demes $n \to \infty$, for all t the probability of identity $c_{b,t}$ of genes in different demes $\to 0$ (it is $O(1/n)$). This means that the probability distribution of coalescence times of genes in different demes flattens down on the x-axis, for all t. Thus, either genes coalesce in the 'recent' past within the same deme where they are both located, or the ancestral lineages separate in different demes, and in the latter case, these lineages may be considered 'independent' (eg Hudson, 1998). A technical assumption underlies this reasoning. Genes in different demes are independent if mutations occurs faster that the coalescence of genes from different demes. For low mutation ($u \to 0$), this is obtained by assuming that the number of demes $n \to \infty$ and that $nu \to \infty$.

The argument for the computation of relatedness coefficients from pedigrees follows exactly the same logic. In an 'infinite' panmictic population, genes in randomly chosen individuals have an 'infinitely small' probability of coalescing in a recent past. Relatedness measures the probability of coalescence before ancestral lineages 'leave' the pedigree considered. In addition the time span t^* may be identified by an exact argument ($t^* = \tau$, the base of the pedigree), and thus relatedness may be computed from an examination of pedigrees.

By contrast, in the stepping stone case, when the number of demes $n \to \infty$, there is still a positive probability that nearby genes coalesce in a recent past ($c_{b,t}$ does not decrease to 0 for all t). Thus genes in different demes cannot be considered independent.

One remaining question is whether equation 16 is correct in cases where its previous 'proof' fails. More generally, we may ask whether the expected frequency $Q_{:k|p}$ of pairs of genes both of type k, given allele frequency p, is of the form

$$Q_{:k|p} = rp + (1 - r)p^2 \tag{17}$$

for some r independent of allele frequency. It cannot be true at extreme allele frequencies in finite populations, as seen in the trivial case of only one copy of the allele. Then $Q_{:k|p} = 0$, so $r < 0$ according to the above formula. More importantly, simulations (Figure 2) suggest notable discrepancies from equation 17, which seem to persist when the number of demes increases, for the stepping stone model. On the other hand, discrepancies are weak in the island model, and decrease with an increasing number of demes (details not shown). This contrast could be expected from the distinction we have drawn between island and stepping stone models.

In some formulations, one can consider a local relatedness statistic, where p is an allele frequency in some local sample rather than in the total population (eg Ritland, 1996; Lynch and Ritland, 1999; Weir, 2001). A discrepancy from equation 17 may also be observed

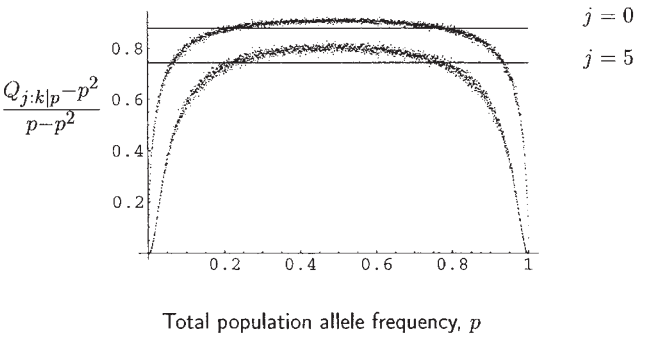

Figure 2 The relationship between identity and frequency in the total population. If equation 16 is valid then $(Q_{j:k|p} - p^2)/(p(1-p)) = (Q_{j:k} - \mathrm{E}[p^2])/(\mathrm{E}[p(1-p)])$. Therefore, discrepancies with equation 16 are shown by plotting estimates of $(Q_{j:k|p} - p^2)/(p(1-p))$ (dots) *vs* $(Q_{j:k} - \mathrm{E}[p^2])/(\mathrm{E}[p(1-p)])$ (straight lines), for two values of j (0 and 5), in a one-dimensional stepping stone model with $n = 200$ demes of 10 haploid individuals. The dispersal rate was $m = 0.2$, and a two allele model with mutation rate $u = 10^{-5}$ was considered.

when a local allele frequency is considered, as shown in Figure 3 for $\tilde{p} < 0.1$ or $\tilde{p} > 0.9$.

Another interpretation of the reference population

There is an alternative, much less common, interpretation of the reference population and of allele frequency in this population. Here the concept of 'population' refers to an infinite number of replicates of the mutation-drift process considered.

The values of probabilities of identity Q or $\dot{Q}$, previously considered in equations 1 and 7, refer to such a concept of population, in the same way that the expectation of a Normal random variable is the average value in an infinite number of samples from a Normal distribution. Likewise, allele frequency in this 'population' is the expected frequency π_k of allele k in the process considered. For example, in a two-allele model with symmetrical mutation rate between the two alleles, the expected frequency π_k of each allele is $1/2$, while the real-

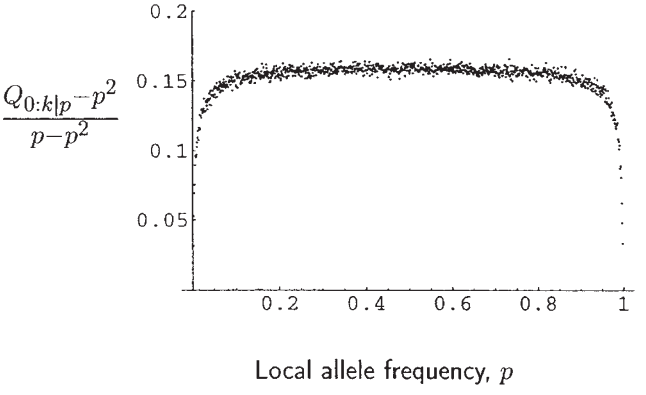

Figure 3 The relationship between identity and frequency in a local sample. In contrast to Figure 2, p is here the allele frequency in a sample of 1000 genes. A two-dimensional stepping-stone population of 100×100 demes of 10 haploid individuals was considered, and an exact coalescent algorithm (R Leblois and FR, unpublished results) was used to generate more than 400 000 samples of 1000 genes on a square of 10×10 demes. The one-dimensional dispersal rate was $m = 0.2$, and a two allele model with mutation rate $u = 5\ 10^{-6}$ was considered.

ized frequency p_k in any particular biological population is a random variable with expectation 1/2 (Cockerham and Weir (1987) used the notation p for what is π_k here). In considering replicates of the process, the probability that two independent genes are both of type k is π_k^2, not the expectation $E[p_k^2]$. $E[p_k^2]$ would arise when considering random sampling of two genes from one biological population, hence such genes are not independent in that they both depend on events that led to a given allele frequency p_k in the biological population.

In this way, the relationship between identity and allele 'frequency' may be intuitively understood as follows. Either the genes are identical by descent as defined by equation 7 (with probability $\dot{Q}_j$ for some specific class j of pair of genes, as above) or they are not (with probability $1 - \dot{Q}_j$) and then they are considered 'independent', that is, both of type k with probability π_k^2. One may then write

$$Q_{j:k} = \dot{Q}_j\pi_k + (1 - \dot{Q}_j)\pi_k^2 \Rightarrow \dot{Q}_j = \frac{Q_{j:k} - \pi_k^2}{\pi_k - \pi_k^2}, \qquad (18)$$

where $Q_{j:k}$ is the probability that two genes from some specific class j are both of type k. This result is not strictly correct, but almost so. For example in a symmetrical two-allele model the exact value of $(Q_{j:k} - \pi_k^2)/(\pi_k - \pi_k^2)$ is the value of identity by descent in a model with a two-fold mutation rate, whatever the model of population structure (eg Tachida, 1985).

Equation 18 is of the form $r\pi_k + (1 - r)\pi_k^2$ for $r = \dot{Q}_j$, which suggests that $\dot{Q}_j$ is a relatedness measure. In the infinite island model, this result may be obtained for low mutation ($u \to 0$), by assuming that the number of demes $n \to \infty$ and that $nu \to \infty$. As previously noted, the latter assumption means that mutations occur faster that the coalescence of genes from different demes. It ensures that $\dot{Q}_w < 1$ and that $\dot{Q}_b = 0$ in the limit, so that $\dot{F} = \dot{Q}_w$. This supports the computation of relatedness, r, as identity by descent, $\dot{Q}_w$. A similar argument can be made for pedigree relatedness in panmictic populations. However, these are the exceptions. More generally, the low mutation limit of $\dot{Q}_j$ is 1, which bears no information about the genealogical relationships of different individuals.

Rather, we may recover the interpretation of inbreeding coefficients in terms of ω, as follows. With probability $1 - \omega$ (which corresponds to the area below the dotted line in Figure 1b), the probability of identity of pairs of genes 'within' is the same as the probability of identity of genes 'between', and with probability ω (the 'initial area') the coalescence event has occurred recently in a common ancestor, which was of allelic type k with probability π_k. Then

$$Q_{w:k} \approx \omega\pi_k + (1 - \omega)Q_{b:k} \Rightarrow \frac{Q_{w:k} - Q_{b:k}}{\pi_k - Q_{b:k}} \approx \omega \qquad (19)$$

where $Q_{w:k}$ and $Q_{b:k}$ are probabilities of identity, both genes being of allelic type k, 'within' and 'between' classes of genes as above. Summing this expression over alleles, one has

$$Q_w \approx (1 - \omega)Q_b + \omega \Rightarrow \frac{Q_w - Q_b}{1 - Q_b} \approx \omega, \qquad (20)$$

Equation 19 may simply be viewed as a generalization of equation 18 where almost any probability of identity Q_b may be considered, instead of the probability of identity π_k^2 of 'independent' genes.

Discussion

Coherent definitions

Relatedness and identity by descent are often identified to each other. This identification seems supported by a number of efficient computation techniques based on them. On the other hand, it leads to inconsistencies which are easily resolved by using alternative definitions. For example, inconsistencies arise whenever relatedness is defined as a probability of identity by descent, and an (unbiased) estimator of it is defined, such that the average estimated relatedness among all sampled individuals is null (as for example some estimators of Queller and Goodnight, 1989; Ritland, 1996; Lynch and Ritland, 1999). This would imply that the average relatedness parameter among all sampled individuals is null, and therefore that the 'probability of identity by descent' is negative for some pairs of individuals.

Actually, these estimators may be understood as follows. The estimated 'relatedness' between individuals x and y may be written $(\tilde{Q}_{xy} - \Sigma_k\tilde{p}_k^2)/(1 - \Sigma_k\tilde{p}_k^2)$ (eg Ritland, 1996), where $\tilde{Q}_{xy}$ is the observed frequency of identical alleles between the two individuals, and $\tilde{p}_k$ is the frequency of allele k in the sample. $\Sigma_k\tilde{p}_k^2$ is identical to the frequency of pairs (sampled with replacement) of genes in the sample, which we may interpret as an estimator of the average probability of identity in state among pairs of genes, $\bar{Q}$, given the sampling design. Hence these estimators may be understood as estimators of a ratio of probabilities of identity in state, $(Q_{xy} - \bar{Q})/(1 - \bar{Q})$, which approximate the equivalent ratio of probabilities of identity by descent, $(\dot{Q}_{xy} - \bar{Q})/(1 - \bar{Q})$. Such coefficients measure how much higher (or lower) the probability of recent coalescence is for the pair x, y relative to the average probability for all pairs considered.

Other, sometimes trivial, inconsistencies abound. For example, definitions of relatedness as 'identity by descent' are also not general enough to include negative correlations between genes, such as heterozygote excesses (negative F_{IS}). This problem also arises when defining inbreeding coefficients as ratios of expected mean squares in an analysis of variance (eg Weir and Cockerham, 1984; Cockerham and Weir, 1987). Actually, inbreeding coefficients of the form F bear a more complex relationship with expected mean squares (Rousset, 2001). It is also well-recognized that in various models, F-'statistics' approach their equilibrium values, after temporal variations in demographic parameters, faster than gene diversities (Takahata, 1983; Slatkin, 1994; Pannell and Charlesworth, 1999). This again shows a difference between F-'statistics' and probabilities of identity.

These distinctions are blurred in the infinite island model (and for pedigree relationships in infinite panmictic populations), where the identity by descent in different demes may be considered null in a limit case (given the implicit technical assumption $nu \to \infty$, detailed above). In this case the ratio of differences of probabilities reduces to a single probability of identity by descent, which is also the probability that genes lineages coalesce before a dispersal event occurs. This is very helpful in obtaining approximations based on such models, but this

does not logically establish the approximation used (eg identity by descent) as a coherent definition of the quantity approximated (eg relatedness in a finite population).

Beyond the logical consistency of definitions, we may also question the claim that the probability that two genes are of a given allelic type can be written as $rp + (1 - r)p^2$, where p is the allele frequency in a 'reference population' and r is a 'relatedness' measure independent of p. As we have seen, there may not be any reference biological population such that this relationship is satisfied. Hence, interpreting p as frequency in an 'ancestral reference population' (equation 16) is not generally valid. It is again essentially correct in infinite panmictic (for pedigree analyses) and infinite island populations, but not in other cases, particularly with localized dispersal.

Such conclusions emphasizes the relevance of a statistical framework in which none of these conceptual ambiguities arise. We have simply distinguished between random variables (allele frequencies in a biological population) and their expectations (their expected value under the effects of drift and mutation). The distinction between frequencies in biological populations and their expectations is not the one between sample values and values in a biological population. Allele frequencies in a population are often random variables in theoretical models of interest (such as the neutral models of population structure). Thus, in a classical statistical perspective, they should not appear in the definition of parameters; only their expectations should. The motivation for this statistical framework is simply that, if we are to make inference about the parameters of a process characterized by (say) subpopulations of size N and a dispersal rate m among them, the statistical inferences must deal with functions of N and m but not of a random variable such as p or a relatedness 'parameter' that would be a function of p (Nagylaki, 1998).

Relatedness in kin selection theory

The distinctions made here are relevant to assess the validity of uses of 'relatedness' in some other contexts. The reference population framework underlies Hamilton's (1964, 1970) development of kin selection theory. The 'regression definition of relatedness' (eg Grafen, 1985) is a reformulation of this framework. It defines relatedness r from an assumed relationship between the frequency q of allele k in some individual related to a 'focal' individual, the allele frequency X in this 'focal' individual, and the allele frequency p in the biological population. This relationship is:

$$\mathrm{E}(q|p) = rX + (1 - r)p. \qquad (21)$$

Here $\mathrm{E}(q|p)$ is the expectation of q conditional on allele frequency p in the population, and r is assumed independent of p. Consider for example a subdivided haploid population. We can compute the probability $Q_{:k|p}$ of identity in state (both genes being of the allelic type k) between a focal individual and its neighbors in the same deme, conditional on an allele frequency p in the population. $Q_{:k|p}$ is the product of the probability that a gene from a neighbor is of type k when a focal individual is of type k (which is $r + (1 - r)p$ from the above expression), times the probability that a focal individual bears allele k (which is the allele frequency in the population, p). That is, $Q_{:k|p} = (r + (1 - r)p)p$, which is equation 16 if $r = Q(t^*)$. Thus the domain of validity of the 'regression definition' is the same as the domain of validity of equation 16. This formulation was appropriate for Hamilton's original model, but recognizing its shortcomings motivates a more general approach to modelling selection in subdivided populations (Rousset and Billiard, 2000), where generalized relatedness measures take the form of ratios such as F, considered in the low mutation limit.

Estimation of relatedness

The implications for estimators of inbreeding coefficients are less clear. The infinite island model is not at issue here. In this model, relatedness may be interpreted as the probability of coalescence before migration of any ancestral lineage. This is useful for constructing likelihood methods under island models (eg Wakeley and Aliacar, 2001), and can be generalized to other models where the genes within units (demes or individuals) coalesce at a faster time scale than genes in different units (Nordborg, 1997; Nordborg and Donnelly, 1997).

Some well-known estimators of F-statistics, such as 'Weir and Cockerham's (1984) estimator, are not based on equation 16. However, estimators that weight alleles according to their frequencies differently from Weir and Cockerham's one, might in principle be affected. Equation 17 is also used for computing the likelihood of matches of genotypes of different individuals (eg in forensic applications, Weir, 2001). The computer simulations (Figure 3) suggest that these computations would be affected under localized dispersal, when using highly polymorphic markers with several 'rare' alleles. The resulting bias may be small, and more realistic simulations would be required to evaluate it.

Conclusion

We have compared different definitions of inbreeding coefficients and of relatedness, and emphasized that definitions of inbreeding coefficients as ratios of differences of probabilities of identity in state are always well-formulated and broadly applicable. They do not constrain one to think in terms of the models to which less general definitions may apply, such as the infinite island model. These alternative definitions relieve us from the ambiguities of the concepts of 'reference population' and 'unrelated individuals'. They do not approximate a probability of identity by descent but rather a ratio of differences of probabilities of identity by descent. Nevertheless, we can recover from such definitions the classical rules for computing relatedness as 'identity by descent', either from a pedigree in a panmictic population, of in infinite island models.

Acknowledgements

I thank R Leblois for help with simulations, an anonymous reviewer for several useful comments on this paper, and C Chevillon, M Lascoux, Y Michalakis, M Raymond, S Otto and O Ronce for comments on various versions. This is paper ISEM 02-014.

References

Chesser RK, Rhodes OE, Sugg DW, Schnabel A (1993). Effective sizes for subdivided populations. *Genetics* **135**: 1221–1232.

Cockerham CC, Weir BS (1987). Correlations, descent measures: drift with migration and mutation. *Proc Natl Acad Sci USA* **84**: 8512–8514.
Cockerham CC, Weir BS (1993). Estimation of gene flow from *F*-statistics. *Evolution* **47**: 855–863.
Cotterman CW (1940). Reprinted 1974. A calculus for statistico-genetics. Ph.D. thesis, Ohio State University, Columbus. In: Ballonoff P (ed) *Genetics and Social Structure*, Dowden: Hutchinson & Ross, Stroudsburg, Pennsylvania, pp 157–272.
Crow JF, Aoki K (1984). Group selection for a polygenic behavioural trait: estimating the degree of population subdivision. *Proc Natl Acad Sci USA* **81**: 6073–6077.
Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*. Harper & Row: New York.
Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*. 4th edn. Longman: Harlow, UK.
Grafen A (1985). A geometric view of relatedness. *Oxford Surv Evol Biol* **2**: 28–89.
Hamilton WD (1964). The genetical evolution of social behavior. I. *J Theor Biol* **7**: 1–16.
Hamitlon WD (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**: 1218–1220.
Hamilton WD (1971). Selection of selfish and altruistic behavior in some extreme models. In: Eisenberg JF, Dillon WS (eds) *Man and Beast: comparative social behaviour*, Smithsonian Institution Press: Washington, pp 58–91.
Hartl DL, Clark AG (1997). *Principles of Population Genetics*. 3rd edn. Sinauer: Sunderland, Mass.
Hill WG (1972). Effective size of populations with overlapping generations. *Theor Popul Biol* **3**: 278–289.
Horn RA, Johnson CR (1985). *Matrix Analysis*. Cambridge University Press: Cambridge.
Hudson RR (1990). Gene genealogies and the coalescent process. *Oxford Surv Evol Biol* **7**: 1–44.
Hudson RR (1998). Island models and the coalescent process. *Mol Ecol* **7**: 413–418.
Jacquard A (1975). Inbreeding: one word, several meanings. *Theor Popul Biol* **7**: 338–363.
Lande R (1992). Neutral model of quantitative genetic variance in an island model with local extinction and recolonization. *Evolution* **46**: 381–389.
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: Sunderland.
Malécot G (1975). Heterozygosity and relationship in regularly subdivided populations. *Theor Popul Biol* **8**: 212–241.
Maruyama K, Tachida H (1992). Genetic variability and geographical structure in partially selfing populations. *Jap J Genet* **67**: 39–51.
Maruyama T (1972). Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. *Genetics* **70**: 639–651.
Maynard Smith J (1998). *Evolutionary Genetics*. 2nd edn. Oxford University Press: Oxford.
Nagylaki T (1998). Fixation indices in subdivided populations. *Genetics* **148**: 1325–1332.
Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321–3323.
Nordborg M (1997). Structured coalescent processes on different time scales. *Genetics* **146**: 1501–1514.
Nordborg M, Donnelly P (1997). The coalescent process with selfing. *Genetics* **146**: 1185–1195.
Pannell JR, Charlesworth B (1999). Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* **53**: 664–676.
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. *Genet Res* **67**: 175–185.
Rousset F (1996). Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* **142**: 1357–1362.
Rousset F (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
Rousset F (1999). Genetic differentiation in populations with different classes of individuals. *Theor Popul Biol* **55**: 297–308.
Rousset F (2001). Inferences from spatial population genetics. In: Balding DJ, Bishop M, Cannings C (eds) *Handbook of Statistical Genetics*, John Wiley & Sons: Chichester, UK, pp 239–269.
Rousset F, Billiard S (2000). A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J Evol Biol* **13**: 814–825.
Sawyer S (1976). Results for the stepping stone model for migration in population genetics. *Ann Prob* **4**: 699–728.
Slatkin M (1991). Inbreeding coefficients and coalescence times. *Genet Res* **58**: 167–175.
Slatkin M (1994). Gene flow and population structure. In: Real LA (ed) *Ecological Genetics*, Princeton University Press: Princeton, New Jersey. pp 3–17.
Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
Tachida H (1985). Joint frequencies of alleles determined by separate formulations for the mating and mutation systems. *Genetics* **111**: 963–974.
Takahata N (1983). Gene identity and genetic differentiation of populations in the finite island model. *Genetics* **104**: 497–512.
Taylor PD (1988). An inclusive fitness model for dispersal of offspring. *J Theor Biol* **130**: 363–378.
Wakeley J, Aliacar N (2001). Gene genealogies in a metapopulation. *Genetics* **159**: 893–905.
Wang J (1997). Effective size and *F*-statistics of subdivided populations. II. Dioecious species. *Genetics* **146**: 1465–1474.
Weir BS (2001). Forensics. In: Balding DJ, Bishop M, Cannings C (eds) *Handbook of Statistical Genetics*, Wiley: Chichester, UK, pp 721–739.
Weir BS, Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
Whitlock MC, Barton NH (1997). The effective size of a subdivided population. *Genetics* **146**: 427–441.
Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.
Wright S (1943). Isolation by distance. *Genetics* **28**: 114–138.
Wright S (1951). The genetical structure of populations. *Ann Eugenics* **15**: 323–354.

Appendix

Migration matrix and other models of population structure

In many models of population structure without demographic fluctuations, identity by descent obeys expressions of the form

$$\dot{\mathbf{Q}} = \gamma(\mathbf{A}\dot{\mathbf{Q}} + \bar{\mathbf{A}}\mathbf{c}) \quad \text{(A.1)}$$

where $\dot{\mathbf{Q}}$ is a vector of stationary probabilities of identity by descent, $\gamma \equiv (1 - u)^2$, $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are two matrices ($\mathbf{A}$ is further irreducible), and $\mathbf{c}$ is a vector expressing the gain in identity due to coalescence events – typically it contains elements c_i either null or of the form $(1 - \dot{Q}_i)/N_i$. In the island and isolation by distance models, $\mathbf{A} = \tilde{\mathbf{A}}$. See Maruyama and Tachida (1992) for a detailed example. See Rousset (1999) for models with $\mathbf{A} \neq \bar{\mathbf{A}}$ (eg spatially- and age-structured populations).

Equation A.1 can also be written in terms of a matrix $\mathbf{G}$, previously considered by Hill (1972), as

$$\dot{\mathbf{Q}} = \gamma\,(\mathbf{G}\dot{\mathbf{Q}} + \bar{\mathbf{A}}\delta) \quad \text{(A.2)}$$

where all elements g_{ij} of $\mathbf{G}$ are the sum of the factors of

$\dot{Q}_j$ in the ith elements of $\mathbf{A}\dot{\mathbf{Q}}$ and of $\bar{\mathbf{A}}\mathbf{c}$, and $\tilde{\mathbf{A}}\boldsymbol{\delta}$ is the remaining term of $\tilde{\mathbf{A}}\mathbf{c}$ where $\boldsymbol{\delta}$ is a vector of elements either null, or of the form $\delta_i = 1/N_i$ if c_i was of the form $(1 - \dot{Q}_i/N_i$. It follows that

$$\dot{\mathbf{Q}} = (\mathbf{I} - \gamma\mathbf{G})^{-1}\,\gamma\bar{\mathbf{A}}\boldsymbol{\delta}. \tag{A.3}$$

Let $\mathbf{e}_1, \ldots, \mathbf{e}_k$ be the right eigenvectors of $\mathbf{G}$, each being the column vector $\mathbf{e}_j$ $(e_{j1}, \ldots, e_{jk})$. The eigenvalues λ_i associated with each $\mathbf{e}_i$ obey $1 > \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_k$ (from the Perron-Frobenius theorem for irreducible non-negative matrices; see Horn and Johnson, 1985, section 8.4.4). The vector $\tilde{\mathbf{A}}\boldsymbol{\delta}$ may be written as $\Sigma_j a_j \mathbf{e}_j$ for some a_j's so that

$$\dot{\mathbf{Q}} = \sum_j \frac{\gamma a_j \mathbf{e}_j}{1 - \gamma\lambda_j} = \sum_{t=1}^{\infty}\sum_{j=1}^{k} \gamma^t \lambda_j^{t-1}\, a_j \mathbf{e}_j \tag{A.4}$$

which shows that the probability of coalescence is $c_{i,t} = \Sigma_j \lambda_j^{t-1}\, a_j e_{ji}$. Note that all e_{1i}'s are nonzero (this also follows from the Perron-Frobenius theorem) and that

$$\lim_{t\to\infty} c_{w,t}/c_{b,t} = e_{1w}/e_{1b} \tag{A.5}$$

where the indices w and b are used as in the main text. The function $g(t)$ of the main text may then be written

$$g(t) = c_{w,t} - (1 - \omega)c_{b,t} = \sum_{j=2}^{k} \lambda_j^{t-1}\, a_j(e_{jw} - (1 - \omega)e_{jb}). \tag{A.6}$$

When the dimension of the matrix increases indefinitely with the number of demes, as for models of isolation by distance, it is not obvious that $\lim_{t\to\infty} c_{w,t}/c_{b,t}$ is defined (the fact that it is for each model with a finite number of demes is not sufficient when the limit is approached more and more slowly as the number of demes increases). However for lattice models of isolation by distance, $\lim_{t\to\infty} c_{w,t}/c_{b,t}$ follows from an expression given by Sawyer (1976) for $c_{\mathbf{r},t}$ in these models, where the index $\mathbf{r}$ is here used from genes at distance $\mathbf{r}$ on the lattice. In Sawyer's notation, $c_{\mathbf{r},t}$ is $\Pr[M = t|Z_0 = \mathbf{r}]$ and is given by his equation 4.29. From this equation, one has eg $\omega = \lim_{t\to\infty} c_{\mathbf{0},t}/c_{\mathbf{r},t} = N/[N + b(\mathbf{r})]$ where N is the number of haploid adults per deme and $b(\mathbf{r})$ is the 'recurrent potential' whose definition is given by Sawyer, equation 4.3. For example in a one-dimensional lattice

$$b(r) \equiv \frac{1}{\pi}\int_0^{\pi} \frac{(1 - \cos rx)\psi^2(e^{ix})}{1 - \psi^2(e^{ix})}\,dx \tag{A.7}$$

where ψ is the characteristic function of dispersal distance. With N haploid adults per deme, one has $\lim_{u\to 0} \dot{F}/(1 - \dot{F}) = b(\mathbf{r})/N$ (Rousset, 1997, equation A10), hence $\lim_{u\to 0} \dot{F} = N/[N + b(\mathbf{r})] = \omega$.