Copyright © 1999 by the Genetics Society of America

Epperson, Bryan K. 1999. Gene genealogies in geographically structured populations. Genetics. 152: 797-806.

Gene Genealogies in Geographically Structured Populations

VMPR
4155
PROB. 1
1.12
AJAIL.

Bryan K. Epperson

Michigan State University, East Lansing, Michigan 48824

Manuscript received July 1, 1998
Accepted for publication March 8, 1999

ABSTRACT

Population genetics theory has dealt only with the spatial or geographic pattern of degrees of relatedness or genetic similarity separately for each point in time. However, a frequent goal of experimental studies is to infer migration patterns that occurred in the past or over extended periods of time. To fully understand how a present geographic pattern of genetic variation reflects one in the past, it is necessary to build genealogy models that directly relate the two. For the first time, space-time probabilities of identity by descent and coalescence probabilities are formulated and characterized in this article. Formulations for general migration processes are developed and applied to specific types of systems. The results can be used to determine the level of certainty that genes found in present populations are descended from ancient genes in the same population or nearby populations *vs.* geographically distant populations. Some parameter combinations result in past populations that are quite distant geographically being essentially as likely to contain ancestors of genes at a given population as the past population located at the same place. This has implications for the geographic point of origin of ancestral, "Eve," genes. The results also form the first model for emerging "space-time" molecular genetic data.

UNDERSTANDING how genetic lineages trace through time and space has long been a central concern of population genetics theory. Stochastic models of the spatial or geographic structure among populations in terms of levels of inbreeding (WRIGHT 1943) or probabilities of identity by descent (IBD; MALÉCOT 1948), and more recently coalescence theory (KINGMAN 1982), underpin such processes. Although much of population genetics theory since Wright and Malécot has focused on measures based on gene frequencies, recent availability of information on the degree of differences among genes, which is obtained with many modern molecular techniques, has stimulated much of the recent modeling efforts. With respect to geographical genetics, the original theoretical works of Wright and Malécot and all works since have focused on describing the geographic or spatial patterns of genetic variation at the present time or some other point in time. MALÉCOT (1946) developed the probability of IBD interpretation and applied it to geographical genetics (MALÉCOT 1948). These are probabilities that genealogical lines of descent occur, with mutations superimposed. Malécot also developed what he termed "les chaînes de parenté gametique" (MALÉCOT 1950, 1973) or probabilities of "gametic kinship chains" (*e.g.*, MALÉCOT 1975), which are the same as the probabilities of coalescences between pairs of genes (*e.g.*, HUDSON 1990). This probability approach can be applied both to measures that include degree of differences among genes (*e.g.*, DNA sequences) under the infinite sites models or to gene frequency data under infinite alleles models, as well as to processes under other mutation models (MALÉCOT 1975). In this article we take a first step in the development of measures that more fully describe spatial-temporal genealogical processes by extending measures of the probability of IBD to time as well as space or geography. That is, we develop space-time probabilities of IBD as measures of the degree of shared ancestry, or alternatively similarity, among genes separated in time as well as space. Additional insights are gained into the nature of genealogical processes in structured populations.

Various methods of estimating probabilities of IBD based on genetic data have been established. For decades various approximate estimators have been developed (*e.g.*, MORTON 1969), although it turns out that approximations involving system mean gene frequencies are not necessary (G. MALÉCOT, personal communication). In any case, it is possible to use the sample (observed) genetic identities to estimate values of probabilities of IBD and hence migration and other parameters (*e.g.*, RANNALA 1996).

The significance of further understanding of space-time processes lies in part in the fact that we often want to understand *historical* migrational processes. There is considerable interest in the geographic origins of genetic polymorphisms and in the origins and spread of populations. Recent examples include the use of molecular data to infer origins of ALU polymorphisms in humans and the geographic origins of "modern" humans themselves (*e.g.*, BATZER *et al.* 1996, 1997; STANLEY 1997). The vast majority of studies utilize present geo-

Address for correspondence: 126 Natural Resources Bldg., Michigan State University, East Lansing, MI 48824.
E-mail: epperson@pilot.msu.edu

graphical patterns of molecular variation. However, to fully understand the significance of present relationships to historical migrational processes we must characterize the direct relationship of present genes to genes in the past located in the same and different populations. It may seem a safe assumption that under most circumstances a given gene at the present time is most closely related to the past or ancient genes in the same population; however, the given gene may also be very closely related to ancient genes in other populations, perhaps in some cases quite distant geographically. Analysis of space-time probabilities allows examination of this. Not only might high probabilities extend for considerable distances, they seemingly could even violate the above assumption if migration rates differ among directions, *i.e.*, anisotropy (EPPERSON 1993a,b). That is, it is conceivable that genes presently found in a given population may be *more* closely related to those ancient genes that were in *other*, perhaps quite distant, geographic regions.

Another emerging feature of modern molecular genetics is the ability to obtain ancient DNA from mummies and even fossils (*e.g.*, KRINGS *et al.* 1997), and considerable genetic data are already available for viruses, such as data in the form of time series of DNA variants in different geographic regions. The methods in this article also model such genes separated in time as well space, which is a first step in developing the probabilistic underpinnings of analyses of space-time molecular genetic data. Moreover, molecular data collected from samples spanning a much shorter time scale (even contemporaneous overlapping generations) also require spatial-temporal measures.

MALÉCOT (*e.g.*, 1950, 1972, 1973, 1975) developed elegant mathematical tools in characterizing the purely spatial probabilities of IBD, that is, how the probabilities that two genes randomly selected from each of two separate (or also the same) populations (at the same time period) change as the distance between populations increases, for a wide range of situations. He further showed exactly how these probabilities of IBD are related to coalescence or gametic kinship chains. Although it is not pursued in great detail in the present article, an analogous but somewhat more complicated relationship exists between space-time probabilities of IBD and space-time coalescence between two genes (separated in time as well as space). We indicate how coalescence analogs of the present general models may be derived from them and hence may be calculated for specific cases.

In this article, we first develop definitions for space-time probabilities of IBD and determine fundamental linkages among them as functions of time and space, and methods of calculating them, for completely general migration processes. We briefly consider the case of a single population and then develop further theoretical results for isotropic, but otherwise general, models. We also develop the Fourier transform of space-time probabilities of IBD in an APPENDIX. Finally, we develop more explicit results for isotropic migration in systems with only one spatial dimension as examples to illustrate some of the paramount features.

RESULTS

General formulation of space-time probabilities of identity by descent and their relationships to spatial probabilities of identity by descent: It is assumed that there is a well-ordered array of populations located in multidimensional space. Each population has N diploid individuals undergoing WRIGHT's (1965) model of life cycle (migration and mutation of gametes, then genetic drift) with, for sake of simplicity, random mating within each population. The latter condition can be relaxed and in some instances does not complicate the situation much (MALÉCOT 1975). Initially, no assumptions need be made about the rates of migration between populations, and thus we may include situations usually referred to as anisotropic. It is assumed only that there are finite numbers of populations that exchange migrants with any given population over one unit of discrete time, hereafter referred to as a generation. The natural definition of space-time probabilities of IBD, denoted $\phi_{n,b}(w, x)$, is the probability that two genes are IBD, where one gene, Γ, is selected at random from a population at generation $n - b$ at location w (where w is a vector of coordinates locating the population), and the other, Γ', is selected at random from a population at the present generation n and located at x. For future reference, the probabilities that the same two genes coalesced s generations prior (time forward) to $n - b$ can be denoted $\pi_{n,b,s}(w, x)$, and sometimes it is convenient to consider "coalescence" events that occur between generations n and $n - b$ (time backward). [Note that this means that, looking backward from generation n, generation $n - b - s$ ($s < 0$) saw the "first" gene that is a direct ancestor of the two genes Γ and Γ'.]

Let $l(w, z)$ be the (time forward) rate of migration from w to z that occurs over one generation. Note that for the purely spatial probabilities of IBD ($b = 0$) $\phi_{n,0}(w, x) = \phi_{n,0}(x, w)$.

Let us first consider the space-time probabilities of IBD for a single generation time lag (*i.e.*, $b = 1$), $\phi_{n,1}(w, x)$. We ignore mutation for the moment, which implies that descendence corresponds to IBD or identity in state (MALÉCOT 1975). To construct $\phi_{n,1}(w, x)$, we may consider three separate probabilities of the ways that Γ and Γ' may be IBD:

1. Probability that Γ' is directly descended from Γ, in which case it is necessarily identical by descent. This is simply the probability that the previous generation ancestor of Γ' came from population w, $l(w, x)$, times

the probability that the gene was in fact Γ, *i.e.*, $1/2N$. Thus this probability is $l(w, x)/2N$.

2. Probability that Γ' is not directly descended from Γ, but is descended from another individual gene in population w, and the latter gene is IBD with Γ. This is the product of two probabilities. The first probability is that Γ' is not directly descended from Γ but is descended from an individual gene in population w, and in the specific case where all populations have the same size this is equal to $l(w, x)(1 - 1/2N)$. Conditional on this, the probability that this gene is IBD with Γ' is simply the within-population probability of IBD at that generation $(n - 1)$, hence $\phi_{n-1,0}(w, w)$, which we may also call $\phi_{n-1,0}(0)$, or more simply $\phi_{n-1}(0)$, because they are the same for all populations.
3. Probability that Γ' is not directly descended from w (*i.e.*, it is descended from some other population, z) but is nonetheless IBD with Γ, $\phi_{n-1,0}(w, z)$. To get this, we must sum over all possible migrations from populations other than w; hence this probability is equal to

$$\sum_{z \neq w} l(z, x)\phi_{n-1,0}(w, z). \tag{1}$$

Thus we have the equation

$$\phi_{n,1}(w, x) = \frac{1}{2N} l(w, x) + \left(1 - \frac{1}{2N}\right) l(w, x)\phi_{n-1,0}(w, w) + \sum_{z \neq w} l(z, x)\phi_{n-1,0}(w, z). \tag{2}$$

The space-time probabilities of IBD with time lag one are linear functions of the spatial probabilities of IBD. Hence the stationarity or equilibrium (as n goes to infinity) conditions are the same as those for spatial IBDs. If we include a recall coefficient, $1.0 > k > 0$, which is most interesting when it represents mutation rates either in the infinite sites (wherein the ϕ's represent probabilities that two nonrecombining haplotypes have no sites that differ) or infinite alleles mutation models, then the spatial probabilities of IBD reach equilibrium (MALÉCOT 1975). At equilibrium, the generation subscripts (n) may be omitted:

$$\phi_1(w, x) = \frac{1}{2N} l(w, x)(1 - k) + \left(1 - \frac{1}{2N}\right) l(w, x)(1 - k)\phi_0(0) + \sum_{z \neq w} l(z, x)(1 - k)\phi_0(w, z). \tag{3}$$

Development of general equations for higher temporal order time lags may be illustrated by first examining the second temporal order probabilities of IBD. In essence, the development follows that for the temporal order one case, except that we must consider all possible paths of migration in the intervening generation(s). There are three components:

1. Probability that Γ' is directly descended from Γ and hence is identical by descent:

$$\frac{1}{2N} \sum_z l(w, z) l(z, x). \tag{4}$$

2. Probability that Γ' is not directly descended from Γ, but is descended from an individual gene in population w, and the latter gene is IBD with Γ':

$$\left(1 - \frac{1}{2N}\right) \sum_z l(w, z)\ l(z, x)\phi_{n-2,0}(w, w). \tag{5}$$

3. Probability that Γ' is not directly descended from w, but is nonetheless IBD with Γ:

$$\sum_{z \neq w} \phi_{n-2,0}(w, z) \sum_y l(z, y) l(y, x). \tag{6}$$

Thus, $\phi_{n,2}(w, x)$ equals

$$\frac{1}{2N} \sum_z l(w, z) l(z, x) + \left(1 - \frac{1}{2N}\right) \sum_z l(w, z) l(z, x)\phi_{n-2,0}(0) + \sum_{z \neq w} \phi_{n-2,0}(w, z) \sum_y l(z, y) l(y, x) \tag{7}$$

and

$$\phi_{n,2}(w, x) = \frac{1}{2N} \sum_z l(w, z) l(z, x)(1 - \phi_{n-2,0}(0)) + \sum_z \phi_{n-2,0}(w, z) \sum_y l(z, y) l(y, x). \tag{8}$$

At equilibrium, when mutation is included,

$$\phi_2(w, x) = (1 - k)^2 \left\{ \frac{1}{2N} \sum_z l(w, z) l(z, x)(1 - \phi_0(0)) + \sum_z \phi_0(w, z) \sum_y l(z, y) l(y, x) \right\}. \tag{9}$$

For temporal lags greater than two time periods, it is possible to construct similar equations by summing over all paths of migrations during intervening time periods. However, the equations become complex and there is a simpler way of expressing space-time probabilities of IBD in terms of those with the next shortest time lag, b. This makes use of the fact that the condition of IBD for $\phi_{n,b}(w, x)$, for genes in a population w at time $n - b$ (b lags in the past), requires that the genes must also be IBD somewhere, z (for all z), at time $n - 1$ (one lag in the past). The value of $\phi_{n,b}(w, x)$ is the sum of the $\phi_{n-1,b-1}(w, z)$ times the probability that such genes in x at time n descended from z in the previous generation, *i.e.*, the migration rate. Hence, for $b > 1$,

$$\phi_{n,b}(w, x) = \sum_z l(z, x)\phi_{n-1,b-1}(w, z). \tag{10}$$

This is also shown by induction in APPENDIX A. For the case of equilibrium with k greater than zero, this reduces

to

$$\phi_b(w, x) = (1 - k) \sum_z l(z, x)\phi_{b-1}(w, z). \qquad (11)$$

Thus all of the space-time probabilities of IBD can be obtained from the spatial probabilities of identity by using Equation 8 (or 9 in the case of equilibrium) and then iterating Equations 10 (or 11).

Special case of a single population: In the case of a single population we may consider the probability, $\phi_{n,b}$, that a gene, Γ', selected at random at generation n, is IBD (barring mutation for the moment), with another gene, Γ, which is selected at random from the population at generation $n - b$. For $b = 1$, the probability that Γ' is directly descended from Γ (and hence is also IBD) is simply $1/2N$. The probability that it is not directly descended from but is nonetheless IBD is $(1 - 1/2N)\phi_{n-1,0}$. Hence $\phi_{n,1} = 1/2N + (1 - 1/2N)\phi_{n-1,0}$. Rather than go through the iterative process, we make use of the fact that because there is a single population, Γ' must be descended from either Γ (with probability $1/2N$) or elsewhere in the population. Hence we have the simple result: $\phi_{n,b} = 1/2N + (1 - 1/2N)\phi_{n-b,0}$. If mutation is included, then

$$\phi_{n,b} = (1 - k)^b\left\{\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\phi_{n-b,0}\right\}. \qquad (12)$$

At equilibrium,

$$\phi_b = (1 - k)^b\left\{\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\phi_0\right\}. \qquad (13)$$

For large b and N and small k,

$$\phi_b = e^{-kb}\phi_0. \qquad (14)$$

Case of multiple populations with arbitrary dimensionality but with isotropic migrations and at equilibrium: In the case of isotropic migration rates (where migration rates are the same for spatial lags for each of two directions within a dimension but may differ between dimensions), it is convenient to use spatial lags rather than absolute locations of populations and to assume that the populations either extend infinitely in all dimensions or are supported by a multidimensional torus (MALÉCOT 1975). In this section, we also assume equilibrium with mutation. Let $l(y)$ be the migration rate from a population at w to a population at $w + y$, and let $\phi_b(y)$ be the probability of IBD between Γ and Γ' at two different populations separated by b generations in time and by spatial lags in the vector y in space. Translation of Equation 3 and factoring $(1 - k)$ gives

$$\phi_1(y) = (1 - k)\left\{\frac{1}{2N}l(y) + \left(1 - \frac{1}{2N}\right) l(y)\phi_0(0) + \sum_{z \neq y} l(z)\phi_0(y - z)\right\} \qquad (15)$$

or equivalently

$$\phi_1(y) = (1 - k)\left\{\frac{1}{2N}l(y) + \left(1 - \frac{1}{2N}\right) l(y)\phi_0(0) + \sum_{z \neq 0} l(y - z)\phi_0(z)\right\} \qquad (16)$$

or

$$\phi_1(y) = (1 - k)\left\{\frac{l(y)}{2N}(1 - \phi_0(0)) + \sum_z l(y - z)\phi_0(z)\right\}. \qquad (17)$$

Similarly,

$$\phi_b(y) = (1 - k) \sum_z l(y - z)\phi_{b-1}(z) \qquad (18)$$

or equivalently

$$\phi_b(y) = (1 - k) \sum_z l(z)\phi_{b-1}(y - z). \qquad (19)$$

Thus,

$$\phi_2(y) = (1 - k)^2 \cdot \sum_z l(y - z)\left\{\frac{l(z)}{2N}(1 - \phi_0(0)) + \sum_x l(x)\phi_0(z - y)\right\} \qquad (20)$$

$$\phi_3(w) = (1 - k)^3 \cdot \sum_y l(w - y)\left\{\sum_z l(y - z)\left\{\frac{l(z)}{2N}(1 - \phi_0(0)) + \sum_x l(x)\phi_0(z - y)\right\}\right\}. \qquad (21)$$

It is easy to see that in general $\phi_b(y_b) =$

$$(1 - k)^b \sum_{y_{b-1}} l(y_b - y_{b-1}) \sum_{y_{b-2}} l(y_{b-1} - y_{b-2}) \ldots \sum_{y_2} l(y_3 - y_2) \cdot \sum_z l(y_2 - z)\left\{\frac{l(z)}{2N}(1 - \phi_0(0)) + \sum_x l(x)\phi_0(z - y)\right\}. \qquad (22)$$

Thus the space-time probabilities of IBD can be determined from the spatial probabilities of IBD using Equation 22.

Case of multiple populations with one spatial dimension but with isotropic migrations and at equilibrium: The case of populations located along a single dimension illustrates several aspects of the space-time probabilities of IBD. In this case each spatial index is an integer, not a vector. Using Fourier transforms, MALÉCOT (1972) showed that in the infinite system case (or effectively so for a large torus), the spatial probabilities of IBD are approximately related by

$$\phi_0(x) = \phi_0(0)\left(1 - \frac{(2k)^{1/2}}{\sigma}\right), \qquad (23)$$

where σ^2 is the variance in the distance of migration. A continuous approximation for large distances of separation is

$$\phi_0(x) = \phi_0(0)e^{-x(\sqrt{2k}/\sigma)}. \tag{24}$$

MALÉCOT (1972) also showed that

$$\phi_0(0) = \frac{1}{4N\sigma\sqrt{2k} + 1}. \tag{25}$$

Development of the Fourier transforms of the space-time probabilities of IBD for the one-dimension case is presented in APPENDIX B and can be used for obtaining analytic results on the space-time probabilities of IBD.

For the special case of the strict stepping-stone model, $\sigma^2 = 2m$, where m is the migration rate between adjacent populations, so that

$$\phi_0(x) = \phi_0(0)\ (1 - \sqrt{k/m})^x \tag{26}$$

$$\phi_0(0) = \frac{1}{8N\sqrt{km} + 1}, \tag{27}$$

letting $\phi_0(0) = a$ and

$$g = (1 - \sqrt{k/m}) \tag{28}$$

we have $\phi_0(y) = ag^y$. For $y \neq -1$, 0, or 1,

$$\phi_1(y) = (1 - k)\ \{m(\phi_0(y + 1) + \phi_0(y - 1)) + (1 - 2m)\phi_0(y)\} \tag{29}$$

$$\phi_1(y) = (1 - k)\ \{m(ag^{y+1} + ag^{y-1}) + (1 - 2m)ag^y\} \tag{30}$$

$$= (1 - k)ag^{y-1}\ \{mg^2 + m + (1 - 2m)g\}. \tag{31}$$

Letting $c = mg^2 + m + (1 - 2m)g$, we have $\phi_1(y) = (1 - k)acg^{y-1}$. Using this and Equation 11, and for $-2 > y$, or $y > 2$, we have $\phi_2(y) = (1 - k)^2\ ac^2\ g^{y-2}$. Iterating Equation 11, we have for $-b > y$ or $y > b$, the simple relationship $\phi_b(y) = (1 - k)^b\ ac^b g^{y-b}$. This indicates that for populations where direct descendence is not possible among the two populations (separated by b in time and y in space and $-b > y$, or $y > b$) in the strict stepping-stone model, the space-time probabilities of IBD exponentially decrease with spatial distance.

For $y = -1$ or 1,

$$\phi_1(y) = (1 - k)\left\{\frac{m}{2N}(1 - \phi_0(0)) + m(\phi_0(y + 1) + \phi_0(y - 1)) + (1 - 2m)\phi_0(y)\right\}. \tag{32}$$

For $y = 0$,

$$\phi_1(y) = (1 - k)\left\{\frac{1 - 2m}{2N}(1 - \phi_0(0)) + m(\phi_0(y + 1) + \phi_0(y - 1)) + (1 - 2m)\phi_0(y)\right\}. \tag{33}$$

Examples that capture many of the salient features of space-time probabilities of IBD for isotropic migration processes are shown for strict stepping-stone equilibrium models with one spatial dimension in Tables 1–4. These were calculated using Equations 31, 32, and 33 in conjunction with Equation 11. Calculations were first done assuming 40,000 populations to avoid any possible edge effects, but the same numbers occurred for calculations using 400 populations. Computations for 400 populations for 10,000 generations used ~5 sec of CPU on a Sun Microsystems Sparcstation 20. The key features are: (1) the degree to which the probabilities do not decrease smoothly in time or space; (2) the degree to which the function over space may become more flat as time lag increases; and (3) the general effects of the parameters. Population size, N, affects all space-time probabilities of IBD in exactly the same way. Precisely, as is clear in the Fourier transform in APPENDIX B (Equation B19), the probability of IBD within a population, $\phi_0(0)$, decreases linearly with N, and all of the space-time probabilities decrease by $(1 - \phi_0(0))/2N$, so that the *relative* values of the $\phi_b(y)$ (for b and y not equal to zero) are unaffected by N. For the models shown in Tables 1–4, an arbitrary but small population size ($N = 100$) was used to better show the effects of the other parameters, the rates of migration and mutation. Of course, for larger migration rates the purely spatial probabilities of IBD are smaller for short distances, but they also decrease more slowly as distance increases (MALÉCOT 1975). It is worth pointing out that higher migration rates, m, imply less "information" in terms of purely spatial patterns, because these cause more distant populations to be less different relative to nearby populations. Also obvious is that as k (*e.g.*, mutation rate) increases the spatial correlations are generally smaller, but it is noteworthy that the decrease with distance becomes sharper. That is, local sharing of stochastic effects of drift among nearby populations is relatively stronger when stronger force is acting in the direction of homogenizing the entire system of populations.

For short time lags, particularly for time lag 1 the probabilities of IBD for spatial lags 0 or ±1 may actually be greater than $\phi_0(0)$, especially if the mutation rate is not too large. The effect increases as the migration rate increases. Similarly, the probabilities of the type $\phi_b(0)$ tend to decrease rapidly as migration rate increases. Naturally the larger the value of k, the faster the decreases with time lag, generally. However, for some combinations of migration rate and k, small increases in probabilities of IBD can occur from b to $b + 1$ even at large spatial and temporal lags.

For long temporal lags, there can be remarkable "flattening" of the probabilities of IBD function on distance, especially when mutation and migration are both strong. Nonetheless, it is also remarkable that the curves are relatively flat only up to 10 to 100 distance units in

TABLE 1

Space-time probabilities of identity by descent for the isotropic strict stepping-stone model, with migration rate $m = 0.1$ and mutation rate $k = 10^{-6}$ and population size arbitrarily set at $N = 100$

	Distance lag y											
Time lag b	0	1	2	3	4	5	10	40	70	100	1,000	10,000
0	0.7981	0.7956	0.7931	0.7905	0.7880	0.7856	0.7732	0.7031	0.6394	0.5814	0.0336	0.0000
1	0.7984	0.7957	0.7931	0.7905	0.7880	0.7856	0.7732	0.7031	0.6394	0.5814	0.0336	0.0000
2	0.7979	0.7957	0.7931	0.7905	0.7880	0.7856	0.7732	0.7031	0.6394	0.5814	0.0336	0.0000
3	0.7974	0.7956	0.7931	0.7905	0.7880	0.7856	0.7732	0.7031	0.6394	0.5814	0.0336	0.0000
4	0.7971	0.7956	0.7931	0.7906	0.7880	0.7856	0.7732	0.7031	0.6394	0.5814	0.0336	0.0000
5	0.7968	0.7955	0.7931	0.7906	0.7880	0.7856	0.7732	0.7031	0.6394	0.5814	0.0336	0.0000
10	0.7957	0.7949	0.7929	0.7905	0.7880	0.7856	0.7732	0.7031	0.6394	0.5814	0.0336	0.0000
100	0.7892	0.7890	0.7883	0.7873	0.7858	0.7841	0.7731	0.7031	0.6394	0.5814	0.0336	0.0000
1,000	0.7696	0.7696	0.7694	0.7690	0.7686	0.7680	0.7632	0.7031	0.6394	0.5814	0.0336	0.0000
10,000	0.7082	0.7082	0.7082	0.7081	0.7079	0.7078	0.7064	0.6804	0.6338	0.5805	0.0336	0.0000

most realistic scenarios. Still this may be a substantial distance.

For the space-time coalescence probabilities, the n is not necessary so long as we are careful not to exceed n generations going backward in time and equilibrium is sufficient but not necessary in this regard. We derived equations analogous to Equations 3, 9, and 11 for ϕ, for example, the following (for $s > 1$)

$$\pi_{b,s}(y) = \sum_z \sum_x l(z)\, l(x)\, \pi_{b,s-1}(y + z - x) - \sum_z l(z)\, l(y + z)\, \pi_{b,s-1}(0)/2N \tag{34}$$

and (for $s = 1$)

$$\pi_{b,1}(y) = \sum_z l(z)\, l(y + z)/2N. \tag{35}$$

The exact same equations are found for the coalescence probabilities for two sampled genes separated in space but not time (MALÉCOT 1975), so that the equation holds for all $b > 0$. However, we note that for $s = 0$ the value is not 0 as it is by definition in the purely spatial case. Because we can choose any b, we can iterate these equations from the initial time lag of 0. Thus it is possible to determine the theoretical spatial and space-time coalescence times for general systems and to conduct coalescence analysis of data collected from different time periods. It can also be shown that

$$\phi_b(n, y) = \sum_{s=0}^{n-b} (1 - k)^{2s} \pi_{b,s}(y). \tag{36}$$

Note the summation includes $s = 0$ (probability that Γ' is a direct descendent of Γ).

DISCUSSION

The mathematical relationships developed in this article demonstrate that the probabilities ($\phi_b(y)$) of IBD between genes separated by time lags (b) as well as distance lags (y) in space are (usually complex) linear functions of the spatial probabilities of IBD for general

TABLE 2

Space-time probabilities of identity by descent for the isotropic strict stepping-stone model, with migration rate $m = 0.01$ and mutation rate $k = 10^{-6}$ and population size arbitrarily set at $N = 100$

	Distance lag y											
Time lag b	0	1	2	3	4	5	10	40	70	100	1,000	10,000
0	0.9259	0.9167	0.9075	0.8984	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
1	0.9261	0.9167	0.9075	0.8984	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
2	0.9259	0.9167	0.9075	0.8984	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
3	0.9257	0.9167	0.9075	0.8984	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
4	0.9255	0.9167	0.9075	0.8984	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
5	0.9254	0.9167	0.9075	0.8984	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
10	0.9245	0.9166	0.9075	0.8984	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
100	0.9163	0.9134	0.9067	0.8982	0.8894	0.8805	0.8374	0.6194	0.4582	0.3389	0.0000	0.0000
1,000	0.8930	0.8922	0.8899	0.8861	0.8811	0.8751	0.8370	0.6194	0.4582	0.3389	0.0000	0.0000
10,000	0.8214	0.8212	0.8206	0.8195	0.8180	0.8160	0.8006	0.6193	0.4582	0.3390	0.0000	0.0000

TABLE 3

Space-time probabilities of identity by descent for the isotropic strict stepping-stone model, with migration rate $m = 0.1$ and mutation rate $k = 10^{-4}$ and population size arbitrarily set at $N = 100$

	Distance lag y											
Time lag b	0	1	2	3	4	5	10	40	70	100	1,000	10,000
0	0.2833	0.2743	0.2657	0.2573	0.2491	0.2413	0.2054	0.0783	0.0299	0.0114	0.0000	0.0000
1	0.2843	0.2747	0.2657	0.2573	0.2491	0.2413	0.2054	0.0783	0.0299	0.0114	0.0000	0.0000
2	0.2824	0.2747	0.2657	0.2573	0.2491	0.2413	0.2054	0.0783	0.0299	0.0114	0.0000	0.0000
3	0.2808	0.2746	0.2657	0.2573	0.2491	0.2413	0.2054	0.0783	0.0299	0.0114	0.0000	0.0000
4	0.2796	0.2743	0.2657	0.2573	0.2491	0.2413	0.2054	0.0783	0.0299	0.0114	0.0000	0.0000
5	0.2785	0.2739	0.2657	0.2573	0.2491	0.2413	0.2054	0.0783	0.0299	0.0114	0.0000	0.0000
10	0.2747	0.2719	0.2652	0.2572	0.2491	0.2413	0.2055	0.0784	0.0299	0.0114	0.0000	0.0000
100	0.2515	0.2508	0.2488	0.2456	0.2413	0.2362	0.2052	0.0784	0.0299	0.0114	0.0000	0.0000
1,000	0.1847	0.1846	0.1842	0.1835	0.1826	0.1814	0.1718	0.0784	0.0300	0.0114	0.0000	0.0000
10,000	0.0441	0.0441	0.0441	0.0441	0.0440	0.0439	0.0434	0.0342	0.0207	0.0101	0.0000	0.0000

migration models with arbitrary numbers of spatial dimensions, with isotropic or anisotropic migration, at equilibrium or not. Equations were generated that can be iterated so that the space-time probabilities of IBD can be calculated from the spatial probabilities of IBD, again for the same range of general models. In all of these systems the effects of number of individuals within populations, N, are simple. The probabilities of IBD within the same population at the same time and the purely spatial probabilities of IBD decrease linearly with N (*e.g.*, MALÉCOT 1975), and APPENDIX Equation B19 shows that the space-time probabilities also decrease with N. Moreover, we would expect, although did not show, that variation of N among populations would have little effect, as is the case for spatial correlations of allele frequencies (BODMER and CAVALLI-SFORZA 1968). Theory for purely spatial probabilities of IBD has already been developed for equilibrium and nonequilibrium general models by MALÉCOT (1948, 1972, 1973, 1975) and others, and these can also be related to coalescence events between pairs of genes or gametic kinship chains (MALÉCOT 1975). Malécot developed analytical expressions for spatial probabilities of IBD for isotropic migration models but not for anisotropic models, which complicates the Fourier transform, but in principle this should be possible. We focused on using isotropic equilibrium models as examples to illustrate some key features of the space-time probabilities of IBD and as an example of their Fourier transforms (which can also be found for isotropic nonequilibrium systems for space-time as well as spatial probabilities of IBD).

Several fundamental features of space-time probabilities of IBD were illustrated using the equilibrium one-dimensional strict stepping-stone migration process, for which the purely spatial probabilities have the simple form of an exponential decrease with distance of spatial separation. First, for relatively short time lags, the probabilities of IBD for relatively small distances can exhibit complex behavior, which would not necessarily be expected from consideration of purely spatial patterns. Probabilities of IBD for two genes existing at different generations but within the same population or between

TABLE 4

Space-time probabilities of identity by descent for the isotropic strict stepping-stone model, with migration rate $m = 0.01$ and mutation rate $k = 10^{-4}$ and population size arbitrarily set at $N = 100$

	Distance lag y											
Time lag b	0	1	2	3	4	5	10	40	70	100	1,000	10,000
0	0.5556	0.5000	0.4500	0.4050	0.3645	0.3281	0.1937	0.0082	0.0003	0.0000	0.0000	0.0000
1	0.5566	0.5000	0.4500	0.4050	0.3645	0.3281	0.1937	0.0082	0.0003	0.0000	0.0000	0.0000
2	0.5554	0.5000	0.4500	0.4050	0.3645	0.3281	0.1937	0.0082	0.0003	0.0000	0.0000	0.0000
3	0.5542	0.5000	0.4500	0.4050	0.3645	0.3281	0.1937	0.0082	0.0003	0.0000	0.0000	0.0000
4	0.5531	0.5000	0.4500	0.4050	0.3645	0.3281	0.1937	0.0082	0.0003	0.0000	0.0000	0.0000
5	0.5520	0.5000	0.4500	0.4050	0.3645	0.3281	0.1937	0.0082	0.0003	0.0000	0.0000	0.0000
10	0.5467	0.4998	0.4500	0.4050	0.3645	0.3281	0.1937	0.0082	0.0003	0.0000	0.0000	0.0000
100	0.4955	0.4802	0.4452	0.4043	0.3647	0.3284	0.1939	0.0082	0.0003	0.0000	0.0000	0.0000
1,000	0.3594	0.3566	0.3485	0.3357	0.3190	0.2995	0.1937	0.0083	0.0004	0.0000	0.0000	0.0000
10,000	0.0846	0.0845	0.0841	0.0834	0.0824	0.0812	0.0719	0.0087	0.0004	0.0000	0.0000	0.0000

two nearby populations may actually increase as the time lag increases. Such effects are greater when mutation rates are higher and are slightly increased (for very short time periods) when migration rates are higher. Naturally, mutation tends to decrease overall probabilities of IBD and low migration rates tend to increase short distance probabilities of IBD. For the purely spatial probabilities, they drop off more sharply with distance when there are higher mutation rates or when there are lower migration rates.

For longer time lags, b, even when there are sharp declines in purely spatial probabilities of IBD (as distance increases), which occur particularly when mutation rates are high and migration rates are low, there can be a remarkable degree of "flatness" when b increases. That is, looking at ancient generations, IBD decreases much more slowly with geographic or spatial distance. Such effects are greatest when migration rates are high, as would be expected. The same also occurs in space-time correlations in allele frequencies (Epperson 1993b). This means that if we consider a gene at the present generation at a particular geographic location, its ancestors are essentially just as likely to have been from populations within such a range of distances as from that location itself. For example, with a migration rate, m, of 10% and a per sequence mutation rate, k, of 10^{-6} (Table 1) a population located 100 times more distant than the average distance between adjacent populations is still ~82% as likely to share identity as is an ancient gene in the same population (location) itself, for genes from 10,000 generations ago. For another example, with $m = 0.10$ and $k = 10^{-4}$ (Table 3), at the same generation the value at spatial lag 100, $\phi_0(100)$ (0.0114) is only 4% as large as at the origin, $\phi_0(0)$ (0.2833), whereas for 10,000 generations ago, $\phi_{10,000}(100)$ (0.0101) is 23% as large as $\phi_{10,000}(0)$ (0.0441). It may be expected that a system of populations existing in two dimensions would show even greater flatness, as is the case for space-time correlations of gene frequencies (Epperson 1993b). Moreover, anisotropic migration could cause distant populations to be *more* likely than the same (or very nearby) populations to be the ancestral source of present variants (Epperson 1993b). Theoretical analyses of space-time probabilities of IBD thus can also provide a means for determining the degree of certainty that may be placed on the historic geographic origins of molecular genetic polymorphisms.

Finally, this article developed coalescence probabilities for two genes in samples separated in time as well space. Thus the coalescent can be extended to ancient DNA, and, for example, an ancient DNA sample could be placed in a gene genealogy reconstruction using the coalescent.

I thank two anonymous reviewers for helpful comments on an earlier version of the manuscript. This research was supported in part by grants from McIntire-Stennis and the Michigan Agricultural Experiment Station.

LITERATURE CITED

Batzer, M. A., S. S. Arcot, J. W. Phinney, M. Alegria-Hartman, D. H. Kass *et al.*, 1996 Genetic variation of recent Alu insertions in human populations J. Mol. Evol. **42:** 22–29.

Batzer, M. A., S. T. Sherry, P. L. Deininger and M. Stoneking, 1997 Alu repeats and human evolution-response. J. Mol. Evol. **45:** 7–8.

Bodmer, W. F., and L. L. Cavalli-Sforza, 1968 A migration matrix model for the study of random genetic drift. Genetics **59:** 565–592.

Epperson, B. K., 1993a Recent advances in correlation studies of spatial patterns of genetic variation. Evol. Biol. **27:** 95–155.

Epperson, B. K., 1993b Spatial and space-time correlations in systems of subpopulations with genetic drift and migration. Genetics **133:** 711–727.

Hudson, R. R., 1990 Gene geneologies and the coalescent process, pp. 1–44 in *Oxford Surveys in Evolutionary Biology*, edited by D. J. Futuyma and J. Antonovics. Oxford University Press, Oxford.

Kingman, J. F. C., 1982 The coalescent. Stochast. Proc. Appl. **13:** 235–248.

Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking *et al.*, 1997 Neanderthal DNA-sequences and the origin of modern humans. Cell **90:** 9–30.

Malécot, G., 1946 La consanguinité dans une population limitée. C. R. Acad. Sci. **222:** 841–843.

Malécot, G., 1948 *Les Mathématiques de l'Hérédité.* Masson, Paris.

Malécot, G., 1950 Quelques schémas probabilistes sur la variabilité des populations naturelles. Ann. Univ. Lyon Sci. Sect. A **13:** 37–60.

Malécot, G., 1972 Génétique des populations naturelles dans le cas d'un seul locus. II. Étude du coefficient de parenté. Ann. Génét. Sél. Anim. **4:** 385–409.

Malécot, G., 1973 Génétique des populations diploïdes naturelles dans le cas d'un seul locus. III. Parenté, mutations et migration. Ann. Génét. Sél. Anim. **5:** 333–361.

Malécot, G., 1975 Heterozygosity and relationship in regularly subdivided populations. Theor. Popul. Biol. **8:** 212–241.

Morton, N. E., 1969 Human population structure. Annu. Rev. Genet. **3:** 53–74.

Rannala, B., 1996 The sampling theory of neutral alleles in an island population of fluctuating size. Theor. Popul. Biol. **50:** 91–104.

Stanley, S. E., 1997 Alu repeats and human-evolution. J. Mol. Evol. **45:** 6–7.

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

Wright, S., 1965 The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19:** 395–420.

Communicating editor: M. W. Feldman

APPENDIX A

To demonstrate the validity of Equation 10 in the text consider the equation for $\phi_{n,2}(w, x)$,

$$\phi_{n,2}(w, x) = \sum_k l(k, x)\phi_{n-1,1}(w, k), \tag{A1}$$

and text Equation 2,

$$\phi_{n-1,1}(w, k) = \frac{1}{2N} l(w, k) + \left(1 - \frac{1}{2N}\right) l(w, k)\phi_{n-2,0}(0) + \sum_{z \neq w} l(z, k)\phi_{n-2,0}(w, z). \tag{A2}$$

Substituting the right-hand side of Equation A2 for $\phi_{n-1,1}(w, k)$ in Equation A1 produces

$$\phi_{n,2}(w, x) = \sum_k l(k, x)\left\{\frac{1}{2N} l(w, k) + \left(1 - \frac{1}{2N}\right) l(w, k) \cdot \phi_{n-2,0}(0) + \sum_{z \neq w} l(z, k)\phi_{n-2,0}(w, z)\right\}. \tag{A3}$$

The simplest way to show the correspondence of this equation to text Equations 7 or 8 is to multiply through the brackets and compare term by term. The first term in Equation A3 is, substituting the dummy variable z for k,

$$\sum_z l(z, x)\frac{1}{2N} l(w, z), \tag{A4}$$

which is a rearrangement of the first term in text Equation 7. The same rearrangement of the second term in (A3) is

$$\sum_z l(z, x)\left(1 - \frac{1}{2N}\right) l(w, z)\phi_{n-2,0}(0) \tag{A5}$$

and this clearly equals the second term of Equation 7. For the third term we first interchange the summation signs, giving

$$\sum_{z \neq w} \phi_{n-2,0}(w, z)\sum_k l(z, k)\, l(k, x). \tag{A6}$$

Substituting y for k yields the third term of Equation 7. Thus Equations A3 and 7 are equivalent. Because the higher-order lags (>2) involve the same types of terms, it follows that Equation 10 is true by induction. Of course, the analog of Equation 11 is true also when there is mutation and the equilibrium is obtained.

APPENDIX B

As an example, we develop the Fourier transform for the case of isotropic migration, equilibrium, and one spatial dimension. The Fourier transform is $F = \Sigma_y \alpha^y f(y)$, where $\alpha = e^{-i\theta}$, and let us define $K(\alpha)$ as

$$K_b(\alpha) = F(\phi_b(y)) = \sum_y \alpha^y \phi_b(y). \tag{B1}$$

Applying the transform to both sides of text Equation 17, we have

$$K_1(\alpha) = (1 - k)\left\{\sum_y \alpha^y \frac{l(y)}{2N}(1 - \phi_0(0)) + \sum_y \sum_z \alpha^{y-z}\alpha^z l(y - z)\phi_0(z)\right\}. \tag{B2}$$

Defining $L(\alpha) = \Sigma_y \alpha^y\, l(y)$ and recognizing the relationship of the products of Fourier transforms of two functions to the convolution of two functions and that we can interchange the order of the summations in the second term lead to

$$K_1(\alpha) = (1 - k)\left\{\frac{L(\alpha)(1 - \phi_0(0))}{2N} + L(\alpha)K_0(\alpha)\right\}. \tag{B3}$$

MALÉCOT (1972) showed that

$$K_0(\alpha) = \frac{(1 - \phi_0(0))}{2N}\,\frac{(1 - k)^2 L(\alpha)L(1/\alpha)}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)}. \tag{B4}$$

To simplify the exposition, let $a = (1 - \phi_0(0))/2N$, and thus

$$K_1(\alpha) = (1 - k)L(\alpha)\left\{a + \frac{a(1 - k)^2 L(\alpha)L(1/\alpha)}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)}\right\} \tag{B5}$$

$$= (1 - k)L(\alpha)a\left\{1 + \frac{(1 - k)^2 L(\alpha)L(1/\alpha)}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)}\right\} \tag{B6}$$

$$= (1 - k)L(\alpha)a \cdot \left\{\frac{1 - (1 - k)^2 L(\alpha)L(1/\alpha) + (1 - k)^2 L(\alpha)L(1/\alpha)}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)}\right\}. \tag{B7}$$

Thus,

$$K_1(\alpha) = \frac{(1 - k)L(\alpha)a}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)}. \tag{B8}$$

Similarly, transforming text Equation 18 for $b > 1$ gives

$$K_b(\alpha) = (1 - k)L(\alpha)K_{b-1}(\alpha). \tag{B9}$$

Thus, for $b = 2$,

$$K_2(\alpha) = (1 - k)L(\alpha)\frac{(1 - k)L(\alpha)a}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)} \tag{B10}$$

or

$$K_2(\alpha) = \frac{(1 - k)^2 L^2(\alpha)a}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)}. \tag{B11}$$

Repeating this process, we see that for $b \geq 1$,

$$K_b(\alpha) = \frac{(1 - k)^b L^b(\alpha)a}{1 - (1 - k)^2 L(\alpha)L(1/\alpha)}. \tag{B12}$$

It is possible to obtain some analytical solutions for the value of $\phi_b(y)$ as a function of y and b by taking the inverse of the Fourier transform. Indeed, the inversion involves the same roots as in the purely spatial case because the denominator is the same. That is, we need only consider the singularities as the denominator goes to zero. We use an approach similar to the residue theorem. We let $H(\alpha) = L(\alpha)L(1/\alpha)$ and recognize that we need only consider the singularity where $H(\alpha) = 1/(1 - k)^2$ (because we have assumed that $k > 0$), which we set equal to $1 + k_1$. We need consider only the poles α_1 and $\alpha_2 = 1/\alpha_1$, which are very close to 1.0. Using the inversion formula of MALÉCOT (1972),

$$K_b(\alpha) = \sum_{j=1,2} \frac{A_j}{\alpha - \alpha_j} \tag{B13}$$

and

$$\phi_b(y) = \frac{1}{2\pi i}\sum_j \int_c \frac{\alpha_j^{r-1}A_j}{\alpha - \alpha_j} d\alpha \sim \alpha_1^{r-1}A_1 , \tag{B14}$$

where

$$A_1 = \lim_{\alpha\to\alpha_1} (\alpha - \alpha_1) K_b(\alpha) \tag{B15}$$

$$A_1 = a(1 - k)^b \lim_{\alpha\to\alpha_1} \frac{(\alpha - \alpha_1) L^b(\alpha)}{1 - (1 - k)^2 L(\alpha) L(1/\alpha)}. \tag{B16}$$

Because both the numerator and the denominator go to zero in the limit, we can use l'Hopital's rule,

$$A_1 = a(1 - k)^b \lim_{\alpha\to\alpha_1} \frac{(d/d\alpha)(\alpha - \alpha_1) L^b(\alpha)}{(d/d\alpha)\{1 - (1 - k)^2 L(\alpha) L(1/\alpha)\}} \tag{B17}$$

or

$$A_1 = a(1 - k)^b \lim_{\alpha\to\alpha_1} \frac{L^b(\alpha) + (\alpha - \alpha_1) bL'(\alpha) L^{b-1}(\alpha)}{-(1 - k)^2 H'(\alpha)}. \tag{B18}$$

Malécot (1972) showed that $H'(\alpha_1) = 2\sigma^2(\alpha_1 - 1) + o(\alpha_1 - 1)^2 = -2\sigma\sqrt{k_1} + o(4k/\sigma^2)$, where σ^2 is the variance in the distance of migration. When k is small, $H'(\alpha_1) = -2\sigma\sqrt{2k}$. Taking the limit, and substituting back for a and putting this altogether, we have

$$\phi_b(y) = \frac{(1 - k)^b(1 - \phi_0(0)) L^b(\alpha_1)}{4N\sigma\sqrt{2k}} \left(1 - \frac{\sqrt{2k}}{\sigma}\right). \tag{B19}$$

It is of interest to take the Taylor series expansion of $L^b(\alpha_1)$ about 1.0:

$$L^b(\alpha_1) = L^b(1) + \frac{dL^b(1)}{d\alpha_1}(\alpha_1 - 1) + \frac{d^2L^b(1)}{d^2\alpha_1} \cdot \frac{(\alpha_1 - 1)^2}{2!} + o(\alpha_1 - 1)^3. \tag{B20}$$

It is easy to show that $L^b(1) = 1.0$, and that $dL^b(1)/d\alpha_1 = bl$ (for $b > 1$), where l is the average movement of gene migration. In the isotropic case, l is zero. Also in the isotropic case the value of $d^2L^b(1)/d^2\alpha_1 = b\sigma^2$. Thus,

$$L^b(\alpha_1) = 1 + \frac{b\sigma^2(\alpha_1 - 1)^2}{2!} + o(\alpha_1 - 1)^3 = 1 + \frac{b\sigma^2(\sqrt{2k/\sigma})^2}{2} + o\left(\frac{\sqrt{2k}}{\sigma}\right)^3. \tag{B21}$$

The first two terms of Equation B21 provide a good approximation if k is very small compared to σ and $1/b$ (the latter is required because higher terms of the Taylor series may involve higher powers of b). Under these conditions the following approximation is good: $L^b(\alpha_1) = 1 - b\sigma^2((\sqrt{2k})/2\sigma)^2 = 1 + bk$. Note also that when b is large the probability of IBD decreases essentially exponentially with the time lag as well as with distance of separation.

Fourier transform methods can be developed for space-time probabilities of IBD for systems with two or more spatial dimensions, although the notation becomes more complicated.