How Old Is the Most Recent Ancestor of Two Copies of an Allele?

Nick J. Patterson[1]

Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02139

ABSTRACT

An important clue to the evolutionary history of an allele is the structure of the neighboring region of the genome, which we term the genomic background of the allele. Consider two copies of the allele. How similar we expect their genomic background to be is strongly influenced by the age of their most recent common ancestor (MRCA). We apply diffusion theory, first used by Motoo Kimura as a tool for predicting the changes in allele frequencies over time and developed by him in many articles in this journal, to prove a variety of new results on the age of the MRCA under the simplest demographic assumptions. In particular, we show that the expected age of the MRCA of two copies of an allele with population frequency f is just $2Nf$ generations, where N is the effective population size. Our results are a first step in running exact coalescent simulations, where we also simulate the history of the population frequency of an allele.

CONSIDER two copies of an allele in a panmictic population. What is the age of the most recent common ancestor (MRCA)? This question is interesting for population genetics, but has also taken on importance for medical genetics, because the age controls the expected length of the region over which the genomic backgrounds of the two copies are identical. If the MRCA is old, then, since the time of the ancestor, many nearby mutations or recombinations will usually have occurred. If the MRCA is young, then most often the genomic background of a copy of the allele will be similar to the background of the most recent ancestor. If the genomic background of a disease allele has little variation over a long region, then detection of genomic association becomes much easier than if linkage with the allele exists only over a short distance. Thus, on average, it is easier in a genomic scan to detect young disease-causing mutations than old, although after successful detection precise mapping will be harder (on average) with a young allele.

The association of the age of the MRCA with the extent of genomic diversity is natural to consider and is not a new idea. For example, it is discussed by REICH and GOLDSTEIN (1999), although they are not concerned with the probability distribution of the age. How then can we estimate the age of the MRCA? A quantity that is relatively easy to estimate is the population frequency of an allele. This is correlated with both the age of the mutation and the age of the MRCA. Most rare mutations are young, as has been well understood for a long time and made precise in KIMURA and OHTA (1973), and most young mutations are rare. This is re-

lated to the central idea of a recently proposed powerful test for selection (SABETI *et al.* 2002); suppose we have an allele that is frequent, but the genomic background over a long distance lacks diversity. Under a neutral model, we have a mismatch of information—the lack of diversity suggests a young allele that should not be frequent.

It therefore is interesting to consider the *joint* distribution of population frequency and the age of the MRCA. Related ideas are discussed by SLATKIN and RANNALA (2000) who consider the problem of estimating allele age, using both frequency and linkage disequilibrium information. However, they are chiefly concerned with practical methods of analyzing data, and with the influence of past demographic events, while our aim here is to derive exact probability models and work out the consequences under the simplest possible demography. There are some striking results, including the new result that the expected age of the MRCA of two copies of an allele of population frequency f is proportional to f.

The two most important tools for the analysis of how mutant alleles spread through a population are diffusion theory, largely developed in a genetic context by Kimura over a series of 25 years from 1955 to 1980 (see WATTERSON 1996 for a very clear review), and Kingman's coalescent (KINGMAN 1982; NORDBORG 2001). However, these two essential tools have infrequently been used together.

We work through some ideas motivated by the coalescent, but using diffusion techniques, very much in the spirit of Kimura's work. Indeed we generalize some of his results. There are evident relationships with our work and the theory of the coalescent, which deserve future study. The work of GRIFFITHS and TAVARÉ (1998, 1999, 2003), WIUF and DONNELLY (1999), and GRIFFITHS (2003) could be regarded as applying the theory of the coalescent to the diffusion of alleles. We do the

[1]*Address for correspondence:* Broad Institute, 1 Kendall Square, Cambridge, MA 02139. E-mail: nickp@broad.mit.edu

reverse. It seems that some results are much easier in one approach than in the other.

We briefly mention two other relevant recent articles.

1. BARTON *et al.* (2004): This, like our work, considers coalescence at a single biallelic locus with the allele frequencies varying in time and uses diffusion theory. Comparing their work and ours, although the situations analyzed are similar they are not the same. They consider recurrent mutations while we are interested in alleles where there has been a single mutation event. They also consider the case of selection, and their main technique is to derive a set of coupled ordinary differential equations, which require numerical solution. Our situation is simpler and we can obtain analytic solutions.
2. INNAN and NORDBORG (2003): This article, which does *not* use diffusion theory, considers expressions relating to the length of a polymorphism of known frequency. For instance, they use the theory of the coalescent, conditional on the observed number of alleles in the sample, to estimate the probability of no recombination between two sites of distance *L* apart in the history of the sample.

Using our results we can run coalescent simulations backward in time where we simulate jointly both the ancestral history of a small sample of a population and the population frequency.

Even if (as is the case for humans) the past demography is complex, we believe our results will find application. For instance, if the past population size is not constant, one can "stretch time" so that coalescence occurs at a constant rate, but the rate of mutation is varying. Our techniques should still apply and remain a foundation for simulation.

Our main object of study is the probability distribution of the age of the MRCA of two copies of an allele that has known population frequency *f* today. We ignore the possibility of recurrent mutations, so that the polymorphism arose with a unique mutation event.

Two cases naturally arise. The allele may be *ancestral*, which for the remainder of this article we term *old*, or *nonancestral*, which we term *new*. Suppose we are considering a biallelic polymorphic locus, for instance, an {A, C} SNP. In the absence of other information the prior probability that the A allele is new is $\frac{1}{2}$. If we observe a set *S* of A alleles and study the joint distribution of coalescence times of ancestors of *S* and the population frequency of the A allele, then the setup is *not* symmetric between the cases of A new or old.

For example, it is obvious that for a new allele all coalescent events must occur after the mutation that gave rise to the allele, but for an old allele some coalescent events can occur before the mutation that made the locus polymorphic and some after. This distinction is important for much of our discussion.

We have here an implicit reverse diffusion process where we consider the frequency of the allele back in time from the present. For a new allele the frequency will at some point be 0, for an old allele the frequency will at some point be 1, and the coalescent process will become the ordinary Kingman coalescent.

We assume our organism is diploid and scale time so that unit time is 2*N* generations, where *N* is the effective population size. This is convenient and has become standard in the literature.

Our results, true for a panmictic population of very large constant size *N*, include:

1. The MRCA satisfies a diffusion equation *with killing* (see KARLIN and TAYLOR 1981, for definitions), and the solution can be given as an explicit (infinite) sum of orthogonal polynomials.
2. The expected time to the MRCA of two copies of an allele is *f*, the frequency of the allele. More generally, if *M* is the expected time to the first coalescent event for *k* copies of an allele of frequency *f*, then

$$M = f \bigg/ \binom{k}{2}$$

and the mean times for the whole coalescent subtree of these *k* alleles are the same as for the ordinary coalescent with fixed population size *Nf*. This is true although the probability density of each coalescent time is *not* the same as for the ordinary coalescent.
3. If an allele is known to be new then over a small time interval back from the present, the probability density of the time *t* to the MRCA is approximately the constant 1/*f*.
4. If the allele is not known to be new or old, then the probability density (for small time *t*) is approximately e^{-t}/f. [Corollaries 3 and 4 make these statements precise.]
5. If an allele is known to be new then the expected time *M* to the MRCA is an expression $M_1(y)$ given in Equation 16 below. (This expression also appears in Equation 9.2 of GRIFFITHS and TAVARÉ 2003.)
6. For a new allele, as the frequency *f* tends to 1, the expected time to the MRCA tends to

$$\frac{2(\pi^2 - 9)}{3}.$$

METHODS

Diffusion equations: Suppose $X(t)$ is the population frequency of an allele at time *t*. We begin with the basic diffusion

$$dX(t) = (X(t)(1 - X(t)))^{1/2} dB_t,$$

which is the Wright-Fisher equation for the propagation of the frequency of an allele under no selection, in a

panmictic constant-size population of N diploids, where we scale time so that unit time is $2N$ generations.

We call this the *Kimura case.* A key quantity is the *transition function* $K(x, y; t)$, the probability density of $X(t) = y$ conditional on $X(0) = x$. Thus setting

$$P(t, x) = \int_0^1 K(x, y; t)\, dy, \tag{1}$$

$P(t, x)$ is the probability that the allele is polymorphic at time t given that it has frequency x at time 0. The transition function $K(x, y; t)$ satisfies two important partial differential equations:

$$\frac{\partial K}{\partial t} = \frac{\partial^2}{\partial y^2}\left(\frac{y(1-y)K}{2}\right) \tag{2}$$

$$= \frac{x(1-x)}{2}\frac{\partial^2 K}{\partial x^2}. \tag{3}$$

Equation 2 is the *Kolmogorov forward equation* and (3) is the *Kolmogorov backward equation.* Kimura uses these equations creatively in much of his work on diffusions. KIMURA (1955) contains the first appearance of these equations in his work, with an extensive justification and discussion given in KIMURA (1964). A textbook treatment is in KARLIN and TAYLOR (1981), while the original rigorous treatment is by KOLMOGOROFF (1931).

We are interested in the time to the MRCA of two copies of an allele with current population frequency y. We call this the *coalescent case.*

We discuss the diffusion of the population frequency back in time, conditional on coalescence not having occurred. When coalescence occurs we stop (*kill*) the diffusion.

It is natural to consider the function $C(x, y; t)$, which is the joint probability density that $X(0) = y$ and the MRCA has age greater than t conditional on $X(-t) = x$.

Consider the event E that the allele frequency is x at time $-t$ and that the time to the MRCA (TMRCA) occurs earlier than $-t$ so that coalescence has not yet occurred at $-t$. Thus the probability, conditional on the event E, that the MRCA occurs in the interval $(-(t + \delta t), -t)$ is $\delta t / x + o(\delta t)$. This is for the same reason that in the ordinary coalescent the waiting time for the coalescence of two alleles is exponential with mean 1. That is, the instantaneous probability that our diffusion is killed is $1/x$.

Now it follows that the function C satisfies the backward equation:

$$\frac{\partial C}{\partial t} = \frac{x(1-x)}{2}\frac{\partial^2 C}{\partial x^2} - \frac{C}{x}. \tag{4}$$

To motivate Equation 4, note that if we consider a Wright-Fisher diffusion in a population of *constant* frequency x, then the probability $D(t)$ that the MRCA of two alleles is greater than t is just $e^{-t/x}$ and we have

$$\frac{\partial D}{\partial t} = -\frac{D}{x}.$$

Equation 4 also occurs as a special case of Equation 83 in a recent article of GRIFFITHS (2003).

We may consider the forward process where the allele frequency diffuses as in the K process but in addition at time t when the frequency is y the probability in the interval $(t, t + \delta t)$ that the process is killed is $\delta t / y + o(\delta t)$. This then yields the forward equation

$$\frac{\partial C}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial y^2}(y(1-y)C) - \frac{C}{y}. \tag{5}$$

KARLIN and TAYLOR (1981, p. 314) give a simple method of realizing this forward process with killing. Let $X(t)$ be a sample path in the (forward) Wright-Fisher diffusion with $X(0) = x$. As $X(t) \leq 1$ the integral

$$\int_0^\infty \frac{1}{X(t)}\,dt$$

diverges. Let R be a random variable, exponentially distributed with mean 1 and now set ζ (the *killing time*) by

$$\zeta = \inf\left\{s : \int_0^s \frac{1}{X(t)}\,dt > R\right\}.$$

The path X killed at ζ is the realization we seek.

Solution of the transition function: In both the Kimura and coalescent cases, we can write the transition function as an explicit sum of polynomials, orthogonal on the unit interval with respect to an appropriate weight function.

The Kimura case: As we give a detailed argument for the coalescent case and there is now a textbook treatment for the easier Kimura case in KARLIN and TAYLOR (1981, Chap. 16), we just give the result without proof.

Alternative approaches ("lines of descent") are presented in GRIFFITHS (1980) and TAVARÉ (1984).

Here is the result, first obtained by KIMURA (1955),

$$K(x, y; t) = x(1-x)\sum_{i=0}^\infty \frac{J_i^{1,1}(x)J_i^{1,1}(y)}{N_i^{1,1}}e^{-\lambda(i)t} \tag{6}$$

with $\lambda(i)$ given by

$$\lambda(i) = \frac{(i+1)(i+2)}{2}. \tag{7}$$

The $J_i^{1,1}$ are Jacobi polynomials (in this case also called *Gegenbauer* polynomials), and $N_i^{1,1}$ are normalization constants. The J_i and N_i are defined explicitly in the APPENDIX. This result is generalized by Theorem 2 below.

The coalescent case: The diffusion C differs from K by considering the extra chance that coalescence has occurred. As a result $C(x, y; t) < K(x, y; t)$. It follows that $C(x, y; t) \to 0$ as $x \to 0$ or $x \to 1$. Fix y and seek solutions to (4). We seek a solution of the form $\alpha(t)\beta(x)$.

We give details of the analysis in the APPENDIX, but here is a summary of the main results.

THEOREM 1. *Let* $\alpha(t)\beta(x)$ *be a solution to the differential equation* (4). *Then*

1.
$$\alpha(t) = e^{-\lambda t},$$

where $\lambda = (n + 2)(n + 3)/2$ *for some integer* $n \geq 0$.

2. *For a given choice of n of* 1 *above, set*

$$\beta(x) = x^2(1 - x)w_n(x);$$

then $w_n(x)$ *is a polynomial in x.*

3. *Substitute* $x = (1 + z)/2$; *then* $w = w_n$ *satisfies*

$$(1 - z^2)w''(z) + (2 - 6z)w'(z) + n(n + 5)w = 0,$$

which is a special case of the Jacobi differential equation,

$$(1 - z^2)w''(z) + [b - a - (a + b + 2)z]w'(z) + n(n + a + b + 1)w = 0,$$
(8)

with $a = 1$, $b = 3$, *whose solutions are the Jacobi polynomials* $P_n^{a,b}(z)$, *which we define explicitly in the* APPENDIX. (*See, for example,* BIRKHOFF *and* ROTA 1969, *Chap.* 9, *Equation* 32).

4. *The* w_n *are orthogonal with respect to the norm*

$$\|f\|^2 = \int_0^1 x^3(1 - x)f^2(x)\,dx.$$

It now follows from standard results on orthogonal polynomials that we can write $C(x, y; t)$ as

$$C(x, y; t) = \sum_{i=0}^{\infty} c_i(y)x^2(1 - x)J_i^{1,3}(x)e^{-(i+2)(i+3)t/2}, \quad (9)$$

where the coefficient $c_i(y)$ is determined from the orthogonality of the Jacobi polynomials.

The general case: There is a more general diffusion equation for the most recent coalescent event with k copies of an allele. What we have been terming the "coalescent case" is the simplest case where $k = 2$.

If we have k copies, there is a most recent coalescent time, which is the first time in which the k copies have fewer than k distinct ancestors. We call this a *coalescent event.* We now define $C_k(x, y; t)$ as the joint probability that the frequency now (at time 0) is y, and no coalescent event occurred in the time interval $(-t, 0)$, conditional on $X(-t) = x$. If $k = 1$, then coalescence is impossible so we take

$$C_1(x, y; t) = K(x, y; t).$$

The same argument as before shows that C_k satisfies the backward equation,

$$\frac{\partial C_k}{\partial t} = \frac{x(1 - x)}{2}\frac{\partial^2 C_k}{\partial x^2} - S\frac{C_k}{x}, \quad (10)$$

where $S = \binom{k}{2}$.

We can now generalize Theorem 1. Detailed arguments are found in the APPENDIX.

THEOREM 2. *Consider k copies of an allele with population frequency y. Suppose that the diploid population is panmictic with constant size N. Take unit time to be 2N generations. Write X(t) to be the population frequency of the allele at time t, so that X(0) = y. Let $C_k(x, y; t)$ be the probability density, conditional on $X(-t) = x$ of the event that X(0) = y and no coalescence has occurred by time $-t$. (So each of the k ancestors is distinct at time $-t$. If k = 1, coalescence is meaningless and we take $C_1(x, y; t) = K(x, y; t)$). Then under the diffusion approximation*

$$C_k(x, y; t) = y^{k-1}x^k(1 - x)\sum_{i=0}^{\infty}\frac{J_i^{1,2k-1}(x)J_i^{1,2k-1}(y)}{N_i^{1,2k-1}}e^{-\lambda_k(i)t},$$
(11)

where

$$\lambda_k(i) = \frac{(i + k)(i + k + 1)}{2}.$$

The polynomials $J_i^{1,2k-1}$ *and normalizing constants* $N_i^{1,2k-1}$ *are defined in the* APPENDIX.

The age of an allele—a Bayesian interpretation: We now discuss our interpretation of the meaning in our diffusion model of the "age of an allele." We propose first taking a rather natural Bayesian prior and then taking limits. This is an interpretation made in KIMURA and MARUYAMA (1975) and made very explicit in WATTERSON (1976) and SAWYER (1977), although both of these authors preferred a different interpretation involving reversibility. Take the time now to be 0 and assume for the moment that mutations have taken place in a finite time interval $[-T, 0]$ where T is very large. We assume that the prior distribution of the mutation time is uniform in $[-T, 0]$ and that the probability of a double mutation is so low as to be negligible.

We model a mutation as giving birth to an allele with frequency ε that is very small. (Indeed this is biologically as well as mathematically reasonable; the initial frequency is $1/2N$, where N is the effective population size, and N is of course finite.)

We are interested in quantities of the form

$$\int_0^1\int_t^T P(X(-t) = x | X(-s) = \varepsilon)\,ds \cdot f(x)\,dx$$

for various functions $f(x)$.

We always neglect terms of order ε^2. Here is an example. Let $A(t)$ be the probability that the age of a nonancestral allele of frequency y is at least t.

Then it follows from our mutation model that without approximation,

$$A(t) \propto \int_0^1\int_t^T P(X(-t) = x | X(-s) = \varepsilon)\,ds \cdot K(x, y; t)\,dx.$$

Since T is large this can be approximated by the infinite integral

$$\int_0^1 \int_t^\infty P(X(-t) = x | X(-s) = \varepsilon) \, ds \cdot K(x, y; t) \, dx$$

$$= 2 \int_0^\varepsilon \frac{1 - \varepsilon}{1 - x} K(x, y; t) \, dx + 2 \int_\varepsilon^1 \frac{\varepsilon}{x} K(x, y; t) \, dx,$$

where we use Theorem 3 below.

Now $K(x, y; t) = O(x)$ for small x. It follows that the first integral can be neglected and that approximately,

$$A(t) \propto 2\varepsilon \int_0^1 K(x, y; t) / x \cdot dx.$$

Other such integrals in what follows should be interpreted in the same sense.

As $A(0) = 1$, it is easy to determine the constant of proportionality, and we finally obtain

$$A(t) = y \int_0^1 K(x, y; t) / x \cdot dx.$$

Green's functions: We next discuss the *Green's function* of our diffusion, which is a technical tool that has become standard in diffusion theory. KARLIN and TAYLOR (1981) have an extensive treatment. Knowledge of Green's function is the fastest route to deriving a number of the results we will prove. EWENS (1963a, 1964) was the first to make use of Green's function in genetics, and it was used implicitly by Kimura in several of his articles.

The Kimura case: Green's function

$$G_K(x, y) = \int_0^\infty K(x, y; t) \, dt \qquad (12)$$

is well known, and a textbook treatment is in KARLIN and TAYLOR (1981).

The result is as follows.

THEOREM 3 (EWENS 1963a, 1964). *For $x \le y$*

$$G_K(x, y) = 2x/y.$$

For $x > y$

$$G_K(x, y) = 2(1 - x)/(1 - y).$$

Most of the basic results on the Kimura diffusion follow rapidly from knowledge of Green's function. We give two examples.

COROLLARY 1 (KIMURA and OHTA 1973; WATTERSON 1976). *Given an allele has frequency x, the probability the allele is old is x.*

Proof.

$$P(x | \text{allele new}) \propto G_K(\varepsilon, x) = 2\varepsilon/x$$

$$P(x | \text{allele old}) \propto G_K(1 - \varepsilon, x) = 2\varepsilon/(1 - x).$$

Thus

$$P(\text{allele old}|x) = \frac{1/(1 - x)}{1/x + 1/(1 - x)} = x.$$

COROLLARY 2 (WATTERSON 1962; EWENS 1963b). *Let an allele have frequency x; then the expected time to fixation or extinction is*

$$F(x) = -2[x \log x + (1 - x)\log(1 - x)].$$

Proof. If $A(u)$ is a probability density, and $B(u) = \int_u^\infty A(u)$ (the complement of the cumulative density), then it is well known that

$$E_A = \int u A(u) \, du = \int B(u) \, du.$$

Using this,

$$F(x) = \int_0^\infty \int_0^1 K(x, y; t) \, dy \cdot dt$$

$$= \int_0^1 G(x, y) \, dy$$

$$= \int_0^x \frac{2(1 - x)}{1 - y} dy + \int_x^1 \frac{2x}{y} dy$$

$$= -2[x \log x + (1 - x)\log(1 - x)]$$

as asserted.

Suppose all we know about a biallelic marker is that it is polymorphic. If x is the population frequency of one of the two alleles, what is the natural prior probability distribution $P(x)$ for x? From Green's function and our Bayesian interpretation, we see that the appropriate prior is the *improper* prior,

$$P(x) \propto \frac{2\varepsilon}{x} + \frac{2\varepsilon}{1 - x}$$

$$\propto \frac{1}{x(1 - x)}.$$

This prior is often used in the remainder of the article. An interesting consequence, given in KRUGLYAK and NICKERSON (2001, Box 1), though with a different argument, is that if a SNP is detected from examination of two chromosomes, then the distribution of the allele frequency of the discovered SNP is uniform. This follows because if a SNP has frequency x, the probability that two copies of the marker do not agree is proportional to $x(1 - x)$ and multiplication by the prior gives a constant, independent of x.

The coalescent case: Our process now is subject to killing at rate $1/x$.

Define Green's function:

$$G_C(x, y) = \int_0^\infty C(x, y; t) \, dt. \qquad (13)$$

Standard techniques (and some heavy algebra using Maple) yield the following.

THEOREM 4. *For $x \le y$*

$$G_C(x, y) = \frac{-2(x^2 + 2x \log(1 - x) - 2x - 2 \log(1 - x))}{y^2 x}.$$

For $x > y$

$$G_C(x, y) = \frac{2(1 - x)(2y - y^2 + 2 \log(1 - y) - 2y \log(1 - y))}{(1 - y)y^2 x}.$$

THEOREM 5. *For an allele of unknown ancestry status and frequency y:*

1. *Let $A(t)$ be the probability that the MRCA of two copies is at least t. Then*

$$A(t) = y(1 - y)\left[\begin{array}{l} \int_0^1 \dfrac{C(x, y; t)}{x(1 - x)}dx \\ + \lim_{\varepsilon \to 0} \dfrac{1}{2\varepsilon}\int_0^t e^{-(t-\tau)}C(1 - \varepsilon, y; \tau)\,d\tau \end{array}\right]. \tag{14}$$

2. *The expected age of the MRCA is y. More generally the expected age of the first coalescent event, for k copies of the allele is y/S, where $S = \binom{k}{2}$.*

3. *The expected age of each coalescent event, in the coalescent tree of k copies of an allele with frequency y is the same as in the ordinary coalescent in a fixed size population of size y.*

We give details in the APPENDIX, but briefly sketch the central, rather simple idea to prove 2 and 3. An easy calculation shows that if τ is the time to the coalescent event, $M = E(\tau)$, and $\bar{x}$ is the expected population frequency at time τ, then $\bar{x} = SM$. But from general principles $\bar{x} = y$. This gives statement 2. Statement 3 now follows by induction.

Although the *expectations* are the same as for the ordinary coalescent, the *distributions* of the coalescent times are different, because the population size of the allele varies with time.

THEOREM 6. *For a new allele, of frequency y:*

1. *Let $B(t)$ be the probability that the MRCA of two copies has age at least t. Then*

$$B(t) = y\left[\int_0^1 \frac{C(x, y; t)}{x}dx\right]. \tag{15}$$

2. *The expected age of the MRCA is $M_1(y)$, where*

$$M_1(y) = -4 + 2 \log y + \frac{2\pi^2 - 12 \text{ dilog } y}{3y} + \frac{2 \log y}{1 - y} \tag{16}$$

and

$$\text{dilog } x = \int_1^x \frac{\log t}{1 - t}dt.$$

We again give details in the APPENDIX.

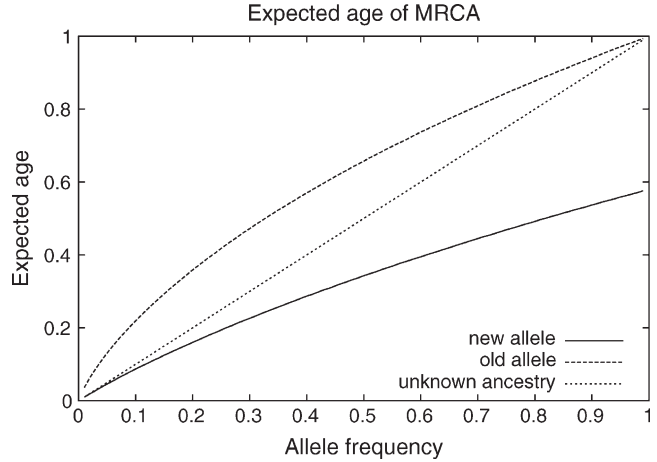If we consider an old allele, rather than a new one,

FIGURE 1.—Expected age of MRCA (old and new allele). We show the expected age of the MRCA of two copies of an allele conditioned both on the population frequency of the allele and on its ancestral status. If the ancestral status of an allele is unknown, then the expected age of the MRCA is simply the population frequency, which is the appropriately weighted sum of the top and bottom curves.

as in Theorem 6, then the expected age of the MRCA of two copies, $M_2(y)$ say, is

$$M_2(y) = 1 - \frac{(1 - y)M_1(y)}{y}. \tag{17}$$

This follows immediately, since by Theorem 5 and Corollary 1:

$$(1 - y)M_1(y) + yM_2(y) = y.$$

In Figure 1 we show plots of M_1 and M_2.

Our formulas have implications for coalescent theory. As an illustration, and a check on our algebra, suppose we observe n alleles and of these i carry a mutation, known to be nonancestral. WIUF and DONNELLY (1999, Equation 28) give a formula for the expected age of the MRCA of the i alleles, using coalescent arguments.

Take as an example $i = 2$, $n = 4$. Then with the same time normalization as we have been using, the Wiuf-Donnelly formula gives the expected MRCA age as 5/18. On the other hand, if we observe two copies of a new mutation out of four, the posterior distribution of the population frequency y is readily seen to be $12y(1 - y)^2$ and so the expected age of the MRCA is

$$\int_0^1 12y(1 - y)^2 M_1(y)\,dy,$$

which Maple shows is indeed 5/18. If the mutation is not known to be new or old, we get an easier integral for the expected age of the MRCA,

$$\int_0^1 6y^2(1 - y)\,dy = \frac{1}{2},$$

which we checked by coalescent simulations.

Time to coalescence
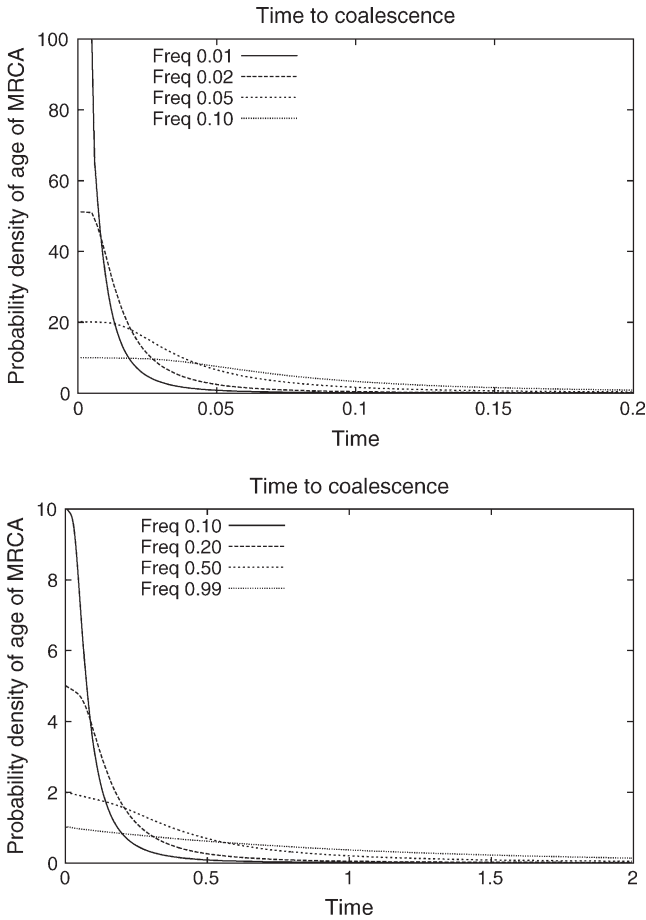

Time to coalescence

FIGURE 2.—The probability distribution of the coalescence time. We display the probability density of the coalescent time for two copies of an allele of a known population frequency. The ancestral status of the allele is assumed to be unknown. Note the approximate invariance up to scaling.

The distribution of the time to the MRCA: *Some numerical plots:* Consider an allele of frequency y at $t = 0$. We are again interested in coalescence of two copies of the allele. We let $V(t) = V(t|y)$ be the probability density of the event that the coalescence time is $-t$, which we can compute numerically using Theorem 2 and Equations A22 and A23.

In Figure 2 we show plots of $V(t)$. We first show a plot for $y = 0.01, 0.02, 0.05, 0.10$. Then we plot $y = 0.10, 0.20, 0.50, 0.99$.

The frequency in the top graph is lower than that in the bottom graph (except we chose to plot frequencies 0.1 and 0.99) by a factor of 10 and the graphs are similar except for the scale. As a rough approximation the probability density $P(y, t)$ that the MRCA of two alleles of population frequency y has age t is a function of y/t.

A simulation scheme: We simulate the coalescent with side information—the population frequency of an allele.

In the ordinary coalescent, if we begin with k haploids, the next coalescent time t is exponentially distributed with mean $1/\binom{k}{2}$ and at coalescence we have $k - 1$ distinct ancestors. Thus a simulation procedure can be iterated until $k = 1$.

In our case, if we begin with k haploids, each with an allele of frequency y, we can proceed analogously by sampling (t, x), where x is the population frequency at coalescence. Just as in an ordinary coalescent simulation we can iterate, until $k = 1$.

The required probability densities are given by Equations A22 and A23 in the APPENDIX. To use these we will need to evaluate $C_k(x, y; t)$ numerically. The series expansion given by the orthogonal polynomial expansion of Theorem 2 converges very fast unless t is small, but in that case the series has poor numerical behavior. We therefore give alternative formulas for the coalescent times valid for small t. Taken together these give satisfactory results for all t.

Taylor series for the coalescent time: Our orthogonal polynomial expansions are not numerically very stable for small time t. The instability is related to the well-known Gibbs phenomenon. (See GOTTLIEB and SHU 1997, for a modern review and some sophisticated techniques for alleviating the problem.) To develop random sampling procedures valid for all regions of time, we wanted to have high-accuracy solutions to the density of the coalescent time valid for small time.

VORONKA and KELLER (1975) and TIER and KELLER (1978) develop, for the Kimura case, highly accurate asymptotic series when time t is small. Their methods use techniques from the theory of wave propagation. We instead use methods that are more elementary and seem sufficient for our needs.

Fix the frequency y and let $C(x, y; t)$ be defined as before. We term a function $f(x)$ on the unit interval *well-behaved* if for all derivatives $f^{(k)}(x)$ of f and all partial derivatives

$$\frac{\partial^l C(x, y; t)}{\partial x^l}$$

$$\lim_{x \to 0} \lim_{t \to 0} f^{(k)}(x) \frac{\partial^l C(x, y; t)}{\partial x^l} = 0$$

$$\lim_{x \to 1} \lim_{t \to 0} f^{(k)}(x) \frac{\partial^l C(x, y; t)}{\partial x^l} = 0.$$

If $f(x)$ is a rational function, it is well behaved. Let

$$F(y, f; t) = \int_0^1 f(x) C(x, y; t) \, dx.$$

There is an interesting recursion for the Taylor series of F as a power series in t around $t = 0$.

THEOREM 7. *Let $f(x)$ be a well-behaved function on the unit interval. Let*

$$F(y, f; t) = \int_0^1 f(x) C(x, y; t) \, dx.$$

Define $f_0 = f$ and iteratively for $n = 1, 2, \ldots,$ let

$$f_n(x) = \frac{\partial^2(x(1-x)f_{n-1}(x)/2)}{\partial x^2} - f_{n-1}(x)/x.$$

Then $f_n(y)$ is the nth partial derivative of F with respect to t, evaluated at $t = 0$.

Let an allele have frequency y, and let $Y(t)$ be the probability distribution of the coalescence time of two copies of the allele.

COROLLARY 3. *If the allele is new, then the Taylor series $T(t)$ for Y at $t = 0$ is*

$$Y(t) \approx T(t) = 1/y.$$

COROLLARY 4. *If the allele is not known to be new or old, then the Taylor series $T(t)$ for Y at $t = 0$ is*

$$Y(t) \approx T(t) = e^{-t}/y.$$

Proof. For Corollary 3, using Green's function, we see that we evaluate the Taylor series by setting $f_0 = f(x) = y/x^2$ in Theorem 7. But then $f_1(x) = 0$ and the result follows.

For Corollary 4 set $f_0 = f(x) = y(1-y)/x^2(1-x)$. Now $f_1 = -f_0$ and iterating, $f_n = (-1)^n f_0$. Again the result follows.

DISCUSSION

Kingman's coalescent has for 20 years been the dominant paradigm for the analysis of ancestral trees. Our work provides convincing evidence that diffusion theory still has much to offer. Suppose, for example, the frequency of some allele in the population is known. One wishes to sample from the probability distribution of the coalescent, for k copies of the allele. Our theory gives the joint distribution of the next coalescent time t and the population frequency $X(t)$, at that time. Sample (t, x) according to this distribution. The procedure may be iterated, just as in an ordinary coalescent simulation. Equations A22 and A23 of the APPENDIX give the required probability densities.

We found it interesting that the theory is much simpler if the ancestral state is unknown than if it is known. The underlying reason is suggested by the proof of our Theorem 5. If the ancestral state is unknown then the expected population frequency in the past is the same as it is now, and this causes substantial simplification.

LITERATURE CITED

ABRAMOWITZ, M., and I. A. STEGUN, 1964 *Handbook of Mathematical Functions.* Dover Publications, New York.

BARTON, N., A. ETHERIDGE and A. STURM, 2004 Coalescence in a random background. Ann. Appl. Probab. **14:** 754–785.

BIRKHOFF, G., and G.-C. ROTA, 1969 *Ordinary Differential Equations,* Ed. 2. Wiley, New York.

EWENS, W., 1963a The diffusion equation and a pseudo-distribution in genetics. J. R. Stat. Soc. Ser. B **25:** 405–412.

EWENS, W., 1963b The mean time for absorption in a process of genetic type. J. Austr. Math. Soc. **3:** 375–383.

EWENS, W., 1964 The pseudo-transient distribution and its uses in genetics. J. Appl. Probab. **1:** 141–156.

GOTTLIEB, D., and C.-W. SHU, 1997 On the Gibbs phenomenon and its resolution. SIAM Rev. **39:** 644–668.

GRIFFITHS, R., 1980 Lines of descent in the diffusion approximation of neutral Wright-Fisher models. Theor. Popul. Biol. **17:** 37–50.

GRIFFITHS, R., 2003 The frequency spectrum of a mutation, and its age, in a general diffusion model. Theor. Popul. Biol. **64:** 241–251.

GRIFFITHS, R., and S. TAVARÉ, 1998 The age of a mutation in a general coalescent tree. Stoch. Models **14:** 273–295.

GRIFFITHS, R., and S. TAVARÉ, 1999 The ages of mutations in gene trees. Ann. Appl. Probab. **9:** 567–590.

GRIFFITHS, R., and S. TAVARÉ, 2003 The genealogy of a neutral mutation, pp. 393–413 in *Highly Structured Stochastic Systems,* edited by P. GREEN, N. HJORT and S. RICHARDSON. Oxford University Press, London/New York/Oxford.

INNAN, H., and M. NORDBORG, 2003 The extent of linkage disequilibrium and haplotype sharing around a polymorphic site. Genetics **165:** 437–444.

KARLIN, S., and H. TAYLOR, 1975 *A First Course in Stochastic Processes.* Academic Press, San Diego.

KARLIN, S., and H. TAYLOR, 1981 *A Second Course in Stochastic Processes.* Academic Press, San Diego.

KIMURA, M., 1955 Solution of a process of random genetic drift with a continuous model. Proc. Natl. Acad. Sci. USA **41:** 144–150.

KIMURA, M., 1964 Diffusion models in population genetics. J. Appl. Probab. **1:** 177–232.

KIMURA, M., and T. MARUYAMA, 1975 Moments for sum of an arbitrary function of gene frequency along a stochastic path of gene frequency change. Proc. Natl. Acad. Sci. USA **72:** 1602–1604.

KIMURA, M., and T. OHTA, 1973 The age of a neutral mutant persisting in a finite population. Genetics **75:** 199–212.

KINGMAN, J., 1982 The coalescent. Stoch. Proc. Appl. **13:** 235–248.

KOLMOGOROFF, A., 1931 Über die analytischen methoden in der wahrscheinlichkeitrechnung. Math. Ann. **104:** 415–458.

KRUGLYAK, L., and D. NICKERSON, 2001 Variation is the spice of life. Nat. Genet. **27:** 234–236.

NORDBORG, M., 2001 Coalescent theory, pp. 179–212 in *Handbook of Statistical Genetics,* edited by D. BALDING, M. BISHOP and C. CANNINGS. John Wiley & Sons, New York.

REICH, D., and D. GOLDSTEIN, 1999 Estimating the age of mutations using variation at linked markers, pp. 129–138 in *Microsatellites: Evolution and Applications,* edited by D. GOLDSTEIN and C. SCHLÖTTERER. Oxford University Press, London/New York/Oxford.

SABETI, P., D. REICH, J. HIGGINS, H. LEVINE, D. RICHTER *et al.,* 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419:** 832–837.

SAWYER, S., 1977 On the past history of an allele now known to have frequency p. J. Appl. Probab. **14:** 439–450.

SLATKIN, M., and B. RANNALA, 2000 Estimating allele age. Annu. Rev. Genomics Hum. Genet. **1:** 225–249.

TAVARÉ, S., 1984 Line-of-descent and genealogical processes, and their application in population genetic models. Theor. Popul. Biol. **26:** 119–164.

TIER, C., and J. KELLER, 1978 Asymptotic analysis of diffusion equations in population genetics. SIAM J. Appl. Math. **34:** 549–576.

VORONKA, R., and J. KELLER, 1975 Asymptotic analysis of stochastic models in population genetics. Math. Biosci. **25:** 331–362.

WATTERSON, G., 1962 Some theoretical aspects of diffusion theory in population genetics. Ann. Math. Stat. **33:** 939–957 (correction: **34:** 352).

WATTERSON, G., 1976 Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. Theor. Popul. Biol. **10:** 239–253.

WATTERSON, G., 1996 Motoo Kimura's use of diffusion theory in population genetics. Theor. Popul. Biol. **49:** 154–188.

WIUF, C., and P. DONNELLY, 1999 Conditional genealogies and the age of a neutral mutant. Theor. Popul. Biol. **56:** 183–201.

Communicating editor: S. TAVARÉ

APPENDIX

Proof of Theorem 1: As we have already argued, $C(x, y; t) < K(x, y; t)$, and so $C(x, y; t) \to 0$ as $x \to 0$ or $x \to 1$. Fix y. We seek solutions to (4) of the form $\alpha(t)\beta(x)$. From the equation, $\alpha'(t)/\alpha(t)$ is independent of t. It follows that $\alpha(t) = e^{-\lambda t}$ for some λ and

$$(1 - \lambda x)\beta(x) = \frac{x^2(1 - x)}{2}\beta''(x). \quad (A1)$$

We want solutions with $\beta(0) = \beta(1) = 0$. We first show $\lambda \neq 0$. Consider the homogeneous differential equation

$$\frac{x^2(1 - x)}{2}\frac{d^2w}{dx^2} - w = 0. \quad (A2)$$

This equation has two independent solutions, as can be verified directly by differentiation:

$$a_1(x) = 2 - x + \frac{2(1 - x)\log(1 - x)}{x} \quad (A3)$$

$$a_2(x) = \frac{1 - x}{x}. \quad (A4)$$

We find that

$$a_1(0) = 0 \quad (A5)$$

$$a_1(1) = 1 \quad (A6)$$

$$a_2(x) \to \infty \quad \text{as } x \to 0 \quad (A7)$$

$$a_2(1) = 0. \quad (A8)$$

And so no nonzero linear combination gives 0 at both ends of the unit interval. Thus $\lambda \neq 0$.

Next, in contrast to the Kimura case, we show $\lambda \neq 1$. Indeed if $\lambda = 1$ Equation A1 is

$$\beta(x) = \frac{x^2}{2}\beta''(x),$$

which has the two independent solutions x^2 and $1/x$ and again no linear combination is feasible.

We now seek solutions with a convergent power series in the open interval $(0, 1)$,

$$\beta(x) = \sum_{n=0}^{\infty} b_n x^n$$

so that

$$\beta''(x) = \sum_{n=0}^{\infty} c_n x^n,$$

where

$$b_{n+2} = \frac{c_n}{(n + 1)(n + 2)}.$$

Equation A1 now gives

$$\frac{2c_{n+1}}{(n + 2)(n + 3)} - \frac{2\lambda c_n}{(n + 1)(n + 2)} = c_{n+1} - c_n \quad (A9)$$

and so

$$c_{n+1} = c_n\left(1 - \frac{2\lambda}{(n + 1)(n + 2)}\right)\Big/\left(1 - \frac{2}{(n + 2)(n + 3)}\right). \quad (A10)$$

It follows that if there exists n with $(n + 1)(n + 2) > 2\lambda$, and $c_n \neq 0$ that the sequence c_m converges to a nonzero value, L say, as $m \to \infty$. But then as $x \to 1$, $x^2(1 - x)\beta''(x) \to L$, and $\beta(x) \to L/(2(1 - \lambda))$. This is a contradiction. Therefore $\beta''(x)$ [and $\beta(x)$] is a polynomial, and $\lambda = (n + 1)(n + 2)/2$ for some integer $n \geq 0$. As $\lambda \neq 1$ we in fact obtain

$$\lambda = \frac{(n + 2)(n + 3)}{2}, \quad n = 0, 1, \ldots.$$

From (A1) and noting that $\lambda > 1$, $\beta(x)$ is divisible by $x^2(1 - x)$. Set

$$w(x) = \frac{\beta(x)}{x^2(1 - x)}.$$

Substituting in (A1) we get after a little algebra

$$x(1 - x)w'' + (4 - 6x)w' + n(n + 5)w = 0, \quad (A11)$$

a hypergeometric equation,

$$x(1 - x)w''(x) + (\gamma - (\alpha + \beta + 1)x)w'(x) - \alpha\beta w = 0. \quad (A12)$$

(See, for example, BIRKHOFF and ROTA 1969, Chap. 9, Equation 18.) Here,

$$\gamma = 4$$

$$\alpha + \beta = 5$$

$$\alpha\beta = -2(\lambda - 3) = -n(n + 5) \quad (n > 0).$$

Substituting $x = (1 + z)/2$ we obtain Equation A13.

$$(1 - z^2)w''(z) + (2 - 6z)w'(z) + n(n + 5)w = 0, \quad (A13)$$

which is a special case of the Jacobi equation (8). Theorem 1 now follows from standard facts about the Jacobi polynomials.

The general case—Proof of Theorem 2: We now generalize the analysis to the solution of Equation 10, which is the corresponding equation for considering the co-

alescence of k copies of an allele. As before set $S = \binom{k}{2}$. Again we seek solutions of the form $\alpha(t)\beta(x)$, and

$$\alpha(t) = e^{-\lambda t}.$$

The corresponding equation to (A1) for general k is

$$(S - \lambda x)\beta(x) = \frac{x^2(1 - x)}{2}\beta''(x). \qquad (A14)$$

Substitute

$$w(x) = \frac{\beta(x)}{x^k(1 - x)}.$$

Making the same transformation as before, by setting $x = (1 + z)/2$ we obtain the differential equation, generalizing (A13),

$$(1 - z^2)w''(z) + ((2k - 2) - (2k + 2)z)w'(z) + n(n + 2k + 1)w = 0, \qquad (A15)$$

a Jacobi equation with $a = 1$, $b = 2k - 1$.

The eigenvalues are

$$\lambda_n = (n + k)(n + k + 1)/2 \qquad (A16)$$

and the solution to Equation A15 is the Jacobi polynomial $P_n^{1,2k-1}(z)$.

Thus the equation generalizing (9) is

$$C_k(x, y; t) = \sum_{i=0}^{\infty} c_i(y) x^k (1 - x) J_i^{1,2k-1}(x) e^{-\lambda_i t},$$

where λ_i is given by (A16). Multiply by $x^{k-1} J_i^{1,2k-1}(x)$ and integrate. We define

$$N_i^{a,b} = \int_0^1 x^b (1 - x)^a (J_i^{a,b}(x))^2 dx$$

and get, using $C_k(x, y, 0) = \delta(x - y)$,

$$y^{k-1} J_i^{1,2k-1}(y) = c_i(y) N_i^{1,2k-1}. \qquad (A17)$$

This is enough to prove Theorem 2.

In Table A1 we give the first few values of J_n and N_n, which are useful for checking a software implementation that uses Theorem 2.

Proof of Theorem 5: Let $A(t)$ be the probability that the MRCA of two copies of an allele with frequency y is at least t, where the ancestral status of the allele is unknown. Then

$$A(t) \propto \int_0^\infty \int_0^1 K(\epsilon, x; \tau) C(x, y; t)\, dx \cdot d\tau$$

$$+ \int_0^\infty \int_0^1 K(1 - \epsilon, x; \tau) C(x, y; t)\, dx \cdot d\tau$$

$$+ \int_0^t e^{-(t-\tau)} C(1 - \epsilon, y; \tau)\, d\tau$$

$$\propto 2\epsilon \int_0^1 \frac{C(x, y; t)}{x(1 - x)} dx$$

$$+ \int_0^t e^{-(t-\tau)} C(1 - \epsilon, y; \tau)\, d\tau,$$

where the last term corresponds to the case that the allele is old and that the coalescent event occurs before the mutation event, so that the allele was fixed in the population at the time of coalescence.

As $A(0) = 1$ we get

$$A(t) = y(1 - y)\left[\begin{array}{l} \int_0^1 \dfrac{C(x, y; t)}{x(1 - x)} dx \\[2mm] + \lim_{\epsilon \to 0} \dfrac{1}{2\epsilon} \int_0^t e^{-(t-\tau)} C(1 - \epsilon, y; \tau)\, d\tau \end{array}\right], \qquad (A18)$$

which is what we required. For the second part we give a Martingale argument that has its own interest and avoids any heavy computation.

We first give a "reversibility" argument that is now well known (see WATTERSON 1976, for a detailed description and applications). Write

$$R(x, y; t) = P(X(-t) = x | X(0) = y),$$

calculated for the Kimura diffusion. First assume that $0 < x < 1$. It follows from the Kolmogorov equations or directly from Theorem 2 that

$$\frac{K(x, y; t)}{x(1 - x)} = \frac{K(y, x; t)}{y(1 - y)}.$$

Under our Bayesian interpretation,

$$R(x, y; t) = P(X(-t) = x | X(0) = y)$$

$$= \frac{y(1 - y) K(x, y; t)}{x(1 - x)}$$

$$= K(y, x; t)$$

$$= P(X(t) = x | X(0) = y).$$

At the boundary we can also check that with our Bayesian interpretation

$$P(X(-t) = 1 | X(0) = y) = P(X(t) = 1 | X(0) = y)$$

$$P(X(-t) = 0 | X(0) = y) = P(X(t) = 0 | X(0) = y).$$

It follows that if we fix y, then $R(x, y; t)$ and $K(y, x; t)$ are equivalent stochastic processes. However, $K(y, x; t)$ is evidently a *Martingale* (KARLIN and TAYLOR 1975) and therefore $R(x, y; t)$ is also.

Next we give a result, of independent interest, relating the expected coalescent time to the expected allele frequency at coalescence. The result holds in wide generality.

LEMMA 1. *Consider k copies of an allele with population frequency f. Do not assume details of the process such as whether the allele is old or has unknown status, selection . . . . Normalize time so that unit time is $2N$ generations. Let the mean time to the first coalescent event be $M(f)$. Set $S = \binom{k}{2}$. Let $X(f)$ be the expected frequency of the allele at first coalescence. Then*

$$SM(f) = X(f).$$

TABLE A1

The first few values of J_n and N_n

Degree (n)	Kimura		Coalescent	
	$J_n^{1,1}$	$N_n^{1,1}$	$J_n^{1,3}$	$N_n^{1,3}$
0	1	1/6	1	1/20
1	$4x - 2$	2/15	$6x - 4$	2/35
2	$15x^2 - 15x + 3$	3/28	$28x^2 - 35x + 10$	1/18
3	$56x^3 - 84x^2 + 36x - 4$	4/45	$120x^3 - 216x^2 + 120x - 20$	4/77

We give the first few values of J_n and N_n required for applications of Theorem 2 in either the Kimura case or the coalescent case with two copies of an allele.

Proof. Let $A(t)$ be the probability of no coalescence by time t back from the present. Then by a standard argument

$$M = \int_0^\infty A(t)\, dt. \tag{A19}$$

Let $P(t, x)$ be the joint probability that coalescence has not occurred at time $-t$ and that the frequency at time $-t$ is x. To make the notation clear we assume $0 < x < 1$ and define $Q(t)$ to be the joint probability that coalescence has not occurred at time $-t$ and that the frequency at time $-t$ is 1. [In general, $\lim_{x\to 1} P(t, x) \neq Q(t)$.] Then

$$A(t) = \int_0^1 P(t, x)\, dx + Q(t). \tag{A20}$$

Next define $B(t, x)$ as the joint probability density that coalescence occurs at time $-t$ and that the frequency is x ($0 < x < 1$). Then

$$B(t, x) = SP(t, x)/x.$$

Thus

$$
\begin{aligned}
X &= \int_0^\infty \left(\int_0^1 B(t, x)\, x \cdot dx \right) + SQ(t)\, dt \\
&= S\int_0^\infty \int_0^1 P(t, x)\, dx + Q(t)\, dt \\
&= S\int_0^\infty A(t)\, dt \\
&= SM,
\end{aligned}
$$

where we use Equation A19. This proves the lemma.

Robert Griffiths (personal communication) pointed out that this lemma is a result about any stochastic process $X(t)$ on the positive real line that is killed at a rate proportional to $1/X(t)$.

Continuing with the proof of Theorem 5, we have constant population size, no selection, and the population frequency is f. We suppose that the allele status (new or old) is unknown.

Let τ be the first coalescence time given k copies of an allele. Now τ is a *stopping time* with finite mean, and it follows from basic Martingale results that

$$E(X(-\tau)|(X(0) = f)) = f. \tag{A21}$$

Set, as before, $S = \binom{k}{2}$. By Lemma 1 we have

$$E(X(-\tau)|X(0)) = SE(\tau|X(0))$$

and it follows that

$$E(\tau|X(0) = f) = \frac{f}{S}.$$

This completes the proof of part 2 of Theorem 5. The last part follows easily by induction.

For numerical work we need expressions for the probability density

$$P(t, x) = P(\tau = t, X(-t) = x|X(0) = y)$$

[so we could write, more explicitly, $P(t, x; y)$ for $P(t, x)$] and for

$$Q(t) = P(\tau = t, X(-t) = 1|X(0) = y).$$

Then by an argument we have already given, for $0 < x < 1$,

$$P(t, x) = y(1 - y) S\frac{C_k(x, y; t)}{x^2(1 - x)}, \tag{A22}$$

while

$$Q(t) = y(1 - y)\lim_{\varepsilon \to 0} 1/2\varepsilon \int_0^t Se^{-S(t-u)} C_k(1 - \varepsilon, y; u)\, du. \tag{A23}$$

Here C_k is given explicitly by the series expansion of Equation 11.

Proof of Theorem 6: The same argument as the proof of the first part of Theorem 5 gives our expression for $B(t)$. For the second part, we have

$$M_1(y) = \int_0^\infty B(t)\, dt = y\int_0^1 G_C(x, y)/x \cdot dx.$$

Now using Theorem 4 and Maple we can integrate this in closed form and obtain Equation 16.

Proof of Theorem 7: We see that $F(y, f; 0) = f(y)$. Define

$$F^{(k)} = \frac{\partial^k F(y, f; t)}{\partial t^k} \quad (t = 0). \qquad (A24)$$

We just prove the result for $k = 1$. The general case is essentially the same. For the first derivative

$$\frac{\partial F}{\partial t} = \int_0^1 f(x) \frac{\partial C(x, y; t)}{\partial t} dx$$

$$= \int_0^1 f(x) \left[\frac{x(1 - x)}{2} \frac{\partial^2 C}{\partial x^2} - \frac{C}{x} \right] dx$$

$$= \int_0^1 \frac{\partial^2 (x(1 - x) f(x)/2)}{\partial x^2} C - f(x) C/x \cdot dx$$

and so

$$F^{(1)} = \lim_{t \to 0} F\left(y, \frac{\partial^2 (x(1 - x) f(x)/2)}{\partial x^2} - f(x)/x; t \right)$$

$$= \left(\frac{\partial^2 (x(1 - x) f(x)/2)}{\partial x^2} - f(x)/x \right)_{x=y}.$$

This proves our result.

Jacobi polynomials—some formulas: We give here some formulas involving Jacobi polynomials, necessary for calculations using the orthogonal polynomial expansion of Theorem 1. Our main source is ABRAMOWITZ and STEGUN (1964, Chap. 22).

The Jacobi polynomials $P_n^{a,b}(z)$ are orthogonal on the interval $[-1, 1]$ with respect to the weight function $W_{a,b}(z) = (1 - z)^a (1 + z)^b$. We normalize so that

$$P_n^{a,b}(1) = \binom{n + a}{n}. \qquad (A25)$$

With this normalization

$$\int_{-1}^1 W_{a,b}(z) P_n^{a,b}(z) P_m^{a,b}(z) \, dz = \delta_{mn} A_n^{a,b}, \qquad (A26)$$

where $A_n^{a,b}$ is given by

$$A_n^{a,b} = \frac{2^{(a+b+1)} \Gamma(n + a + 1) \Gamma(n + b + 1)}{(2n + a + b + 1) n! \Gamma(n + a + b + 1)} \qquad (A27)$$

and $P_n^{a,b}(z)$ is defined by

$$P_n^{a,b}(z) = \binom{n + a}{n} a_{0,n}(z) \qquad (A28)$$

(see ABRAMOWITZ and STEGUN 1964, Equation 22.18), where $a_{0,n}(z)$ is defined recursively by

$$a_{n,n}(z) = 1 \qquad (A29)$$

$$a_{m-1,n}(z) = 1 - \frac{(1 - z) t_m a_{m,n}(z)}{2m(a + m)} \quad \text{for } m = n, n - 1, \ldots, 1 \qquad (A30)$$

and

$$t_m = (n - m + 1)(a + b + n + m). \qquad (A31)$$

The above formulas are recommended for numerical work if one wishes to evaluate $P_n^{a,b}(z)$.

It is more convenient for us to work with polynomials $J_n^{a,b}(x)$ defined on the unit interval by

$$J_n^{a,b}(x) = P_n^{a,b}(2x - 1),$$

which are orthogonal on the unit interval with weight function $w_{a,b}(x) = x^b (1 - x)^a$ and

$$\int_0^1 W_{a,b}(x) J_n^{a,b}(x) J_m^{a,b}(x) \, dx = \delta_{mn} N_n^{a,b}, \qquad (A32)$$

where

$$N_n^{a,b} = \frac{A_n^{a,b}}{2^{a+b+1}}$$

$$= \frac{\Gamma(n + a + 1) \Gamma(n + b + 1)}{(2n + a + b + 1) n! \Gamma(n + a + b + 1)} \qquad (A33)$$

and $A_n^{a,b}$ is given by Equation A27.

We have $a = b = 1$ for the Kimura case and $a = 1$, $b = 3$ for the coalescent case. With $a = 1$ our normalization is $P_n^{a,b}(1) = J_n^{a,b}(1) = n + 1$.