The Annals of Applied Probability
2002, Vol. 12, No. 4, 1348–1377

THE STEPPING STONE MODEL: NEW FORMULAS EXPOSE OLD MYTHS

BY J. THEODORE COX AND RICHARD DURRETT

Syracuse University and Cornell University

We study the stepping stone model on the two-dimensional torus. We prove several new hitting time results for random walks from which we derive some simple approximation formulas for the homozygosity in the stepping stone model as a function of the separation of the colonies and for Wright's genetic distance F_{ST}. These results confirm a result of Crow and Aoki (1984) found by simulation: in the usual biological range of parameters F_{ST} grows like the log of the number of colonies. In the other direction, our formulas show that there is significant spatial structure in parts of parameter space where Maruyama and Nei (1971) and Slatkin and Barton (1989) have called the stepping model "effectively panmictic."

1. Introduction. The phrase "isolation by distance" was introduced by Wright (1943) to describe the accumulation of local genetic differences under geographically restricted dispersal. The effects of population subdivision have primarily been studied using two models. The first, introduced by Wright (1943), is the island model, in which a population consists of s colonies of equal size, and migration probabilities are equal for each pair of colonies. The great amount of symmetry of the island model makes it easy to solve the model exactly. An account of the theory can be found in Section 7 of Hudson (1990) or Nei and Takahata (1993).

The second approach, which will be the focus of this investigation, is the stepping stone model of Kimura (1953). This process was studied extensively in the genetics literature for at least two decades before being rediscovered by probabilists Clifford and Sudbury (1973) and Holley and Liggett (1975) under the name the voter model. At that time Kimura and Weiss (1964), Weiss and Kimura (1965), Malécot (1967, 1969 1975), Maruyama (1970, 1971, 1972), Nagylaki (1974) and others had built a detailed theory that could be used to answer many questions of interest to geneticists. Work on the stepping stone model has, of course, continued during the last 25 years. See Crow and Aoki (1984), Strobeck (1987), Slatkin and Barton (1989), Slatkin (1991, 1993), Hey (1991) and Wilkinson-Herbots (1998).

In a parallel endeavor, the voter model has been studied extensively by probabilists. Our investigations here have their roots in the work of Sawyer (1976, 1979) and Cox and Griffeath (1986) who examined the spatial structure of the

---

Received June 2001; revised April 2002.

AMS 2000 subject classifications. 60K35, 92D10.

Key words and phrases. Stepping stone model, voter model, coalescent, fixation indices, identity by descent, heterozygosity.

voter model, and in the work of Cox (1989) and Cox and Greven (1991) on the voter model on a finite set. It is unfortunate for biological applications that most of this work has been carried out under the assumption of nearest neighbor dispersal and with one individual per location. Here we will allow for more general dispersal distributions, and at each spatial location we will have a colony with one or more individuals. The second generalization is needed for modeling. One can imagine that early humans lived in small groups with infrequent exchanges between groups, or one can think of the Napa Valley in California where different *Drosophila* populations live on rotting fruit piles at different vineyards that are separated by several miles.

We represent space as the torus $\Lambda(L) = ((-L/2, L/2] \cap \mathbf{Z})^2$, where $\mathbf{Z}$ is the set of integers. Following the practice in the biology literature, we suppose that at each point $x \in \Lambda(L)$ there is a colony consisting of N diploid or $2N$ haploid individuals, labeled $1, \dots, 2N$. The type of an individual will be a real number in the unit interval (0, 1), new mutant types being chosen randomly from (0, 1). The only thing important about this type space is that each new chosen value will be different from all previous ones. In genetics terms we are using the infinite alleles model.

The population evolves in discrete time, with generation $n+1$ obtained from generation n in the following way. Consider a given individual in colony x. With probability μ, this individual mutates to a new type and, with probability $(1-\mu) \cdot p(x, y)$, assumes the type of an individual chosen at random (in generation n) from the colony at y. All such mutations and choices are assumed to be independent for all individuals at all colonies.

We assume that the transition probability $p(x, y)$ is given by

$$p(x, y) = (1-\nu)I(x, y) + \nu q(y - x), \tag{1.1}$$

where $I(x, y)$ is 1 if $x = y$ and 0 otherwise, and the difference $y - x \in \Lambda(L)$ is computed componentwise and modulo L. We have separated the kernel into two parts since we will be interested in limits as $L \to \infty$ in which the migration rate ν may converge to 0 but $q(z)$ is a fixed displacement kernel. We suppose $q(z)$ is an irreducible probability distribution on $\mathbf{Z}^2$ with $q((0, 0)) = 0$ that has the following properties:

(i) $\mathbf{Z}^2$ *symmetry*. $q((x_1, x_2)) = q((-x_1, -x_2))$ and $q((x_1, x_2)) = q((x_2, x_1))$.
(ii) *Finite range*. $q((x_1, x_2)) = 0$ if $\sup_i |x_i| \geq K$.

We will suppose that $L \geq 2K$ so that we do not get confused when we try to define the corresponding random walk transition probability on the torus. The first assumption implies that a single step taken according to q has zero mean and covariance $\sigma^2 I$, where $\sigma^2 = \sum_{z \in \mathbf{Z}^2} z_1^2 q(z) = \sum_{z \in \mathbf{Z}^2} z_2^2 q(z)$. The finite range condition implies $\sigma^2 < \infty$.

To study the behavior of the stepping stone model, we work backward in time to define a coalescing random walk with killing. Individuals whose state is the result of a new mutation are killed, since we no longer have to work backward to determine their state. Other particles make a jump from colony x to colony y with probability $p(x, y)$ and land at a randomly chosen site within the colony. Suppose, for the moment, that the mutation rate $\mu = 0$ and consider the genealogy of a sample of size 2 chosen at random from the population. As we work backward, let T_0 be the amount of time required until the two lineages first reside in the same colony and let t_0 be the total amount of time needed for the two lineages to coalesce to one. Let X_k be the difference in the locations of the two particles (computed modulo L). Since the two particles were chosen randomly from the torus, the distribution of X_0 is the uniform distribution on $\Lambda(L)$, which we denote by π. Let P_π denote the distribution of the difference of two random walks starting from a pair of points chosen randomly on the torus. Using the methods of Section 2 of Cox (1989), we can prove the following, which generalizes results in Flatto, Odlyzko and Wales (1985).

THEOREM 1. *For any $t > 0$, as $L \to \infty$, uniformly for $\nu \in (0, 1]$,*

$$P_\pi\left(T_0 > \frac{L^2 \log L}{2\pi \nu\sigma^2} t\right) \to e^{-t}. \tag{1.2}$$

SKETCH OF PROOF. To explain the size of the normalizer, note that $P_\pi(X_n = 0) = 1/L^2$, and the local central limit theorem implies $P_0(X_n = 0) \sim (2\pi(2\nu\sigma^2)n)^{-1}$, so

$$\begin{aligned} 1 = \sum_{n=0}^{L^2-1} P_\pi(X_n = 0) &= \sum_{m=0}^{L^2-1} P_\pi(T_0 = m) \sum_{k=0}^{L^2-m} P_0(X_k = 0) \\ &\approx P_\pi(T_0 \le L^2) \frac{\log(L^2)}{2\pi(2\nu\sigma^2)}. \end{aligned}$$

Rearranging gives $P_\pi(T_0 \le L^2) \approx 2\pi\sigma^2\nu/\log L$. To see why the limit is exponential, one can show that X_n comes to equilibrium in time $o(L^2 \log L/\nu)$, so the limit distribution of $T_0/(L^2 \log L/\nu)$ must have the lack of memory property. □

NOTE. We will use the notation $a \approx b$ to mean that a is approximately equal to b, but we will not be precise about the conditions or details of the approximation. We will also use the notation $a_n \sim b_n$ as $n \to \infty$ to mean that $\lim_{n\to\infty} a_n/b_n = 1$.

The typical distance between two points compared in Theorem 1 is of order L. When we look at closer distances the result changes. Let P_x denote the law of the difference of two walks when one starts in colony 0 and one in colony x. (If $x = 0$ we pick two distinct individuals from colony 0.) Letting $|x|$ denote the usual Euclidean norm, we can now state the following result.

THEOREM 2. *Suppose* $x = x_L$ *satisfies* $\lim_{L\to\infty}(\log^+|x|)/\log L = \beta \in [0,1]$. *Then, for any* $t > 0$, *as* $L \to \infty$, *uniformly for* $\nu \in (0,1]$,

$$P_x\left(T_0 > \frac{L^2 \log L}{2\pi\nu\sigma^2}t\right) \to \beta e^{-t}. \tag{1.3}$$

SKETCH OF PROOF. This generalizes a result in Cox and Greven (1991). The main idea behind (1.3) is that if $(\log^+|x|)/\log L \to \beta$, then

$$P_x(T_0 \le L^2/\nu) \to 1 - \beta.$$

When $T_0 > L^2/\nu$, it is likely that $T_0 > L^2\sqrt{\log L}/\nu$ at which time the distribution of X_n has become uniform over the torus, and the longer time behavior is as in Theorem 1. □

Getting the two lineages to the same colony at time T_0 is only the first part of the coalescence time t_0. For the second part after T_0, we need only be concerned with the distribution of t_0 under P_0. We begin with a surprising formula of Strobeck (1987), which shows that $E_0 t_0$ is independent of the value of the migration rate $\nu > 0$ and of the dispersal kernel $q(x)$. The proof we give here is new and is a simple application of the cycle trick for Markov chains [see Theorem 5.4.3 in Durrett (1996)].

THEOREM 3. $E_0 t_0 = 2NL^2$.

PROOF. When both lineages are in the same colony, they have probability $1/2N$ per generation to hit, so $E_0 t_0 = 2NE_0 T_0'$, where T_0' is the first time $t \ge 1$ that the two lines are in the same colony. (Recall that E_0 refers to picking two distinct individuals from the same colony.) The stationary distribution for the migration process is uniform, so it follows from the cycle trick representation of the stationary distribution that $E_0 T_0' = L^2$. □

Comparing Theorems 1 and 3, we see that there are two extreme possibilities

$$ET_0 = O(L^2 \log L/\nu) \ll O(2NL^2) = Et_0 \quad \text{or} \quad ET_0 \gg Et_0. \tag{1.4}$$

In the first case the two lineages will come to the same colony in $o(NL^2)$ so the actual starting positions of the particles does not matter, and the limit distribution will have the lack of memory property.

THEOREM 4. *If* $\lim_{L\to\infty} N\nu/\log L = \infty$, *then, for any* $t > 0$ *as* $L \to \infty$,

$$\sup_{x\in\Lambda(L)} |P_x(t_0 > 2NL^2 t) - e^{-t}| \to 0. \tag{1.5}$$

Conventional wisdom [see pages 125–126 of Kimura and Maruyama (1971)] says that "marked local differentiation of gene frequencies can occur if $N\nu < 1$ where N is the effective size of each colony and ν is the rate at which each colony exchanges individuals with the four surrounding colonies." In the other direction "if $N\nu > 1$ local differentiation is less pronounced and especially if $N\nu \ge 4$, the whole population tends to behave as a panmictic unit." As Theorem 4 and the next result show, $N\nu$ must be larger than $\log L$ in order for the system to behave as if it were homogeneously mixing.

THEOREM 5. *If* $\lim_{L\to\infty} 2N\pi\sigma^2\nu/\log L = \alpha \in [0,\infty)$, *then, for any* $t > 0$ *as* $L \to \infty$,

$$P_\pi\left(t_0 > (1+\alpha)\frac{L^2 \log L}{2\pi\nu\sigma^2}t\right) \to e^{-t}. \tag{1.6}$$

If $x = x_L$ *satisfies* $\lim_{L\to\infty}(\log^+|x|)/\log L = \beta \in [0,1]$, *then, as* $L \to \infty$,

$$P_x\left(t_0 > (1+\alpha)\frac{L^2 \log L}{2\pi\nu\sigma^2}t\right) \to \left(\beta + (1-\beta)\frac{\alpha}{1+\alpha}\right)e^{-t}. \tag{1.7}$$

SKETCH OF PROOF. It suffices to prove the second conclusion. By the reasoning for Theorem 2, the probability the two lineages will enter the same colony before time L^2/ν is about $1-\beta$. As the proof will show, when they do they will be in the same colony a geometrically distributed number of times with mean $\log L/\pi\nu\sigma^2$ before time L^2/ν. Therefore the probability of coalescence before time L^2/ν is approximately

$$\frac{1/(2N)}{1/(2N) + \pi\nu\sigma^2/\log L} \to \frac{1}{1+\alpha} \qquad \text{as } L \to \infty.$$

If a large multiple of L^2/ν units of time elapses between times that the two lineages are in the same colony, then the relative positions of the particles randomize over the torus and we are back to waiting for T_0 to happen under P_π. After a geometrically distributed number of attempts, each with success probability $1/(1+\alpha)$, the cycle ends with successful coalescence. Since the sum of a geometric number of exponentials is exponential, the result follows. □

In Section 2 we use Theorems 1–5 to compute various quantities of interest in genetics. Our main aim there is to dispel the myth that the stepping stone model is too complicated for practical applications. The remainder of the paper is devoted to proofs. Theorems 1 and 2 are proved in Section 3, Theorem 4 in Section 4 and Theorem 5 in Section 5.

2. Applications to the stepping stone model. Our results in Theorems 4 and 5 about the coalescence time t_0 allow us to compute the distribution of any quantity that involves only pairwise comparison of DNA sequences. In this section we will consider three of these: (i) the probability two randomly chosen individuals are identical by descent, (ii) the decay of genetic correlation with distance and (iii) Wright's (1951) measure of population subdivision, F_{ST}.

(i) *Identity by descent.* Two individuals will be identical by descent if (and only if) no mutation has occurred before t_0. If these individuals are picked at random from the population and we let h denote the probability they are identical by descent, then

$$h = E_\pi (1-\mu)^{2t_0}. \tag{2.1}$$

If t_0/c_L equals (approximately) 0 with probability $1-\rho$ and a mean one exponential random variable with probability ρ, as is the case in all of the conclusions of Theorems 4 and 5, then

$$h \approx (1-\rho) + \rho \int_0^\infty e^{2c_L \ln(1-\mu)t} e^{-t}\, dt = (1-\rho) + \frac{\rho}{1-2c_L \ln(1-\mu)}. \tag{2.2}$$

In the case of Theorem 4, $\rho = 1$ and $c_L = 2NL^2$. Using the approximation $\ln(1-x) \approx -x$ for small x and letting $N_T = NL^2$ denote the total number of individuals in the system,

$$h \approx \frac{1}{1-4NL^2 \ln(1-\mu)} \approx (1+4N_T\mu)^{-1} \tag{2.3}$$

for small μ, the classic result for a homogeneously mixing population with N_T individuals. [See, e.g., page 10 of Hudson's (1991) survey.] In contrast, if we let $L \to \infty$ with constant colony size N, then we end up in the $\alpha = 0$ case of the first result of Theorem 5. Again $\rho = 1$, but this time the normalizing constant is

$$c_L = \frac{L^2 \log L}{2\pi\sigma^2\nu},$$

so using (2.2) we end up with

$$h \approx \left(1 + \frac{L^2 \log L)}{\pi\sigma^2\nu}\mu\right)^{-1}. \tag{2.4}$$

Turning (2.3) around, we see that the value of h given in (2.4) is that of a homogeneously mixing population with effective population size

$$N_e = \frac{L^2 \log L}{4\pi\sigma^2\nu}. \tag{2.5}$$

For a numerical example consider a 50 × 50 grid of colonies of size 20 for a total of 50,000 individuals ($L = 50$, $N = 20$). Suppose there is migration with equal probability to each of the other 24 points in a 5×5 square centered at the point ($\sigma^2 = 50/24$), and let $\nu = 0.1$ have the number of migrants per generation $N\nu = 2$. In this case the effective population size is

$$N_e \approx \frac{2500(3.91202)}{4(3.14159)(50/24)0.1} \approx 3736$$

versus the actual population size $N_T = 50{,}000$.

It is interesting to contrast the last calculation with the island model in which there are s subpopulations with N individuals and transition probability

$$p(x, y) = \begin{cases} 1 - \nu, & x = y, \\ \nu/(s-1), & x \neq y. \end{cases}$$

Nei and Takahata (1993) found that the island model behaves like a homogeneously mixing population of size

$$N_e = Ns\left(1 + \frac{(s-1)^2}{4N\nu s^2}\right). \tag{2.6}$$

Note that the first factor Ns is the actual population size, while the factor in parentheses is larger than 1, and is close to $1 + (1/4N\nu)$ when the number of subpopulations s is large. Thus, in contrast to the stepping stone model, in the island model the effective population size N_e is always larger than the actual population size.

(ii) *Decay of correlation with distance*. Pick one individual from the colony at 0 and one from the colony at x (if $x = 0$ pick two distinct individuals from the colony at 0). Let $\phi(x)$ be the probability the two individuals are identical by descent. As Slatkin (1991) has observed, when the mutation rate is small, $1 - \phi(x) \approx 2\mu E_x t_0$. Noting that $E_x t_0 = E_x T_0 + E_0 t_0$, we have

$$\phi(0) - \phi(x) \approx 2\mu E_x T_0. \tag{2.7}$$

By Theorem 2, if $L \to \infty$ and $\log^+ |x| / \log L \to \beta$, then we should have

$$E_x T_0 \approx \beta \frac{L^2 \log L}{2\pi\sigma^2\nu}.$$

Replacing β by $\log^+ |x| / \log L$ on the right-hand side and using (2.7), we see that, for small μ,

$$\phi(0) - \phi(x) \approx 2\mu \frac{\log^+ |x|}{\log L} \cdot \frac{L^2 \log L}{2\pi\sigma^2\nu} = \mu \cdot \log^+ |x| \cdot \frac{L^2}{\pi\sigma^2\nu}. \tag{2.8}$$

That is, the difference $\phi(0) - \phi(x)$ is proportional to $\log^+ |x|$.

To get a check on the quality of this approximation, we may use the exact results of Maruyama (1970) [see formula (5.12) on page 512] to compute $\phi(m, 0)$ for the case $L = 50$, $N = 20$, with nearest neighbor migration ($\sigma^2 = 0.5$) at rate $\nu = 0.05$,

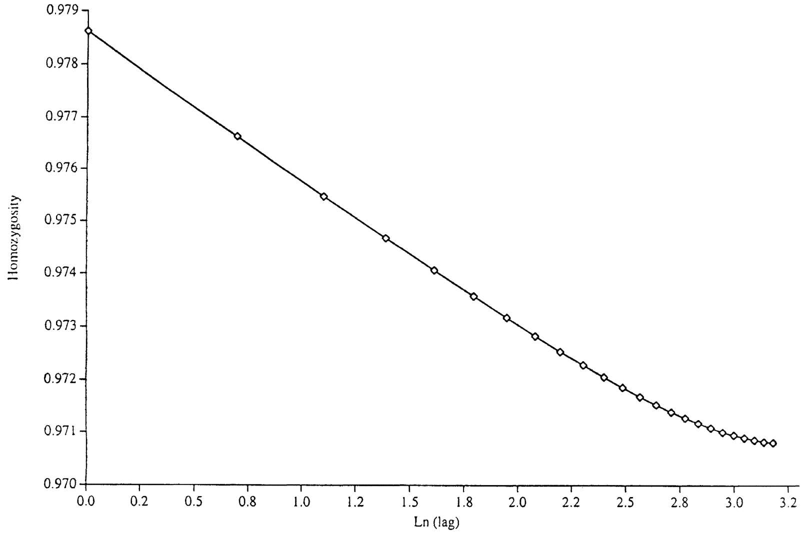

FIG. 1. *Homozygosity as a function of the log of the distance.*

and with a mutation rate $\mu = 10^{-7}$. Figure 1 shows $\phi(m, 0)$ plotted against $\log m$. Note that $\phi(m, 0)$ does decrease roughly linearly with the logarithm of the distance between the colonies until the effect of wraparound on the torus kicks in.

The heterozygosities $1 - \phi(m, 0)$ are quite small here, but they do range from 0.021 to 0.029, an increase of about 40% as we move across the system. The fact that there is spatial structure in this case contradicts Slatkin and Barton (1989) who predict that the population should be effectively panmictic. Their heuristic argument is simple. The amount of time until a lineage encounters a mutation is of order $1/\mu$. The amount of variance in the random walk at this time is $\sigma^2\nu/\mu$, so if this is much smaller than L^2 two lineages from opposite points of the torus will have little chance to meet before a mutation and there will be significant spatial structure. The last conclusion is clearly correct. However, Slatkin and Barton argue that conversely if $\sigma^2\nu/\mu$ is much larger than L^2 then the genealogies will have wrapped around the torus many times and spatial structure is lost. The preceding example shows that there is a flaw in this reasoning, since in that case

$$\frac{\sigma^2\nu}{\mu} = \frac{(0.5)(0.05)}{10^{-7}} = 250{,}000 \gg 2500 = L^2,$$

but there is significant spatial structure.

To further investigate the values of μ for which a system is effectively panmictic, we have considered the last example for various values of the mutation rate. Figure 2 plots $\phi(0)$ as determined by Maruyama's (1970) exact formula, a "large mutation approximation" due to Nagylaki (1974) and the mean field

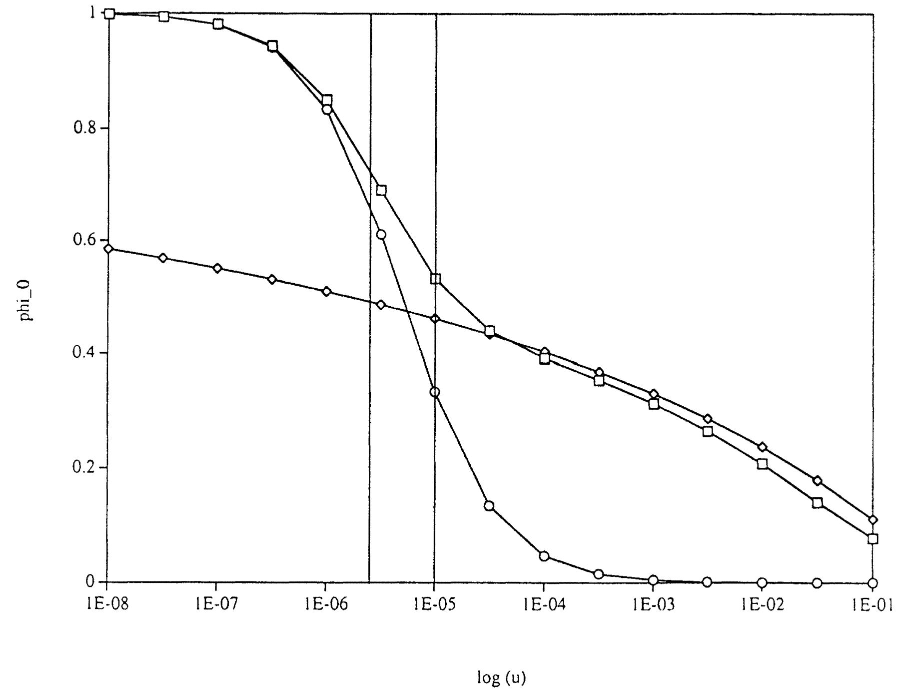

FIG. 2. *Comparison of the probability that two individuals in the same colony are identical by descent as computed by Maruyama's exact solution (squares), Nagylaki's large mutation approximation (diamonds) and the result for a homogeneously mixing population (circles).*

value given in (2.3). Let $\mu_0 = \nu\sigma^2/L^2 = 10^{-5}$ and $\mu_1 = \nu\sigma^2/(L^2 \log L) \approx (10^{-5})/3.91$. Note that, as our theory predicts, the large mutation approximation works well for $\mu \gg \mu_0$, while mean field theory works well for $\mu \ll \mu_1 \approx 2.5 \times 10^{-6}$.

(iii) *Wright's* (1951) *statistic* F_{ST}. This statistic was invented to quantify the amount of genetic differentiation in a spatially distributed population. Following Nei (1975), we define it as

$$F_{ST} = \frac{\phi(0) - \bar{\phi}}{1 - \bar{\phi}}, \tag{2.9}$$

where, as before, $\phi(0)$ is the probability of identity by descent for two individuals sampled from the same colony and $\bar{\phi}$ is the probability for two individuals sampled at random from the entire population. As before, when the mutation rate is small, $1 - \bar{\phi} \approx 2\mu E_\pi t_0$ and $1 - \phi(0) \approx 2\mu E_0 t_0$, so

$$F_{ST} \approx \frac{E_\pi t_0 - E_0 t_0}{E_\pi t_0} = \frac{E_\pi T_0}{E_\pi T_0 + E_0 t_0}. \tag{2.10}$$

Using Theorems 1 and 3, it follows that if $2N\pi\sigma^2\nu/\log L \to \alpha \in (0,\infty)$, then

$$F_{ST} \approx \frac{L^2 \log L/(2\pi\nu\sigma^2)}{L^2 \log L/(2\pi\nu\sigma^2) + 2NL^2} \approx \frac{1}{1+2\alpha}. \tag{2.11}$$

This says that F_{ST} is close to 0 if and only if $N\nu \gg \log L$. Crow and Aoki (1984) did a numerical study of F_{ST} for the nearest neighbor stepping stone model and found (see page 6075) that F_{ST} is roughly proportional to $\log n$, where $n = L^2$ is the number of colonies. In the cases they considered, $\log L/(2\pi\nu\sigma^2) \ll 2N$, so the first term in the denominator of (2.11) can be ignored, and we have

$$F_{ST} \approx \frac{\log L}{4N\pi\nu\sigma^2} = \frac{1}{8N\pi\nu\sigma^2}\log(L^2),$$

confirming their prediction.

It is interesting to compare (2.11) with the corresponding formula for the island model. In this case Nei and Takahata (1993) have shown

$$F_{ST} = \frac{1}{1 + 4N\nu s^2/(s-1)^2}. \tag{2.12}$$

Most authors assume s is large to suppress the factor $s^2/(s-1)^2$, so that $F_{ST} \approx (1+4N\nu)^{-1}$, and one can estimate the scaled migration rate $N\nu$ by

$$\widehat{M} = \frac{1}{4}\left(\frac{1}{F_{ST}} - 1\right) \tag{2.13}$$

[see, e.g., page 265 of Slatkin (1993)]. Suppose the population being sampled has a stepping stone structure, and one uses the island model formula (2.13). Using (2.11) in (2.13) and the assumption $2N\pi\sigma^2\nu/\log L \approx \alpha$, we find that

$$\widehat{M} \approx \frac{\alpha}{2} \approx N\nu\sigma^2 \cdot \frac{\pi}{\log L}, \tag{2.14}$$

so there is a bias which depends on the size of the system. Here, we have kept the σ^2 with $N\nu$, since in the stepping stone model the scaled migration rate $N\nu\sigma^2$ is what one can most simply estimate from observations.

In (1998) Seielstad, Minch and Cavalli-Sforza used F_{ST} values with the island model formula to make the following estimates of human migration rates from three different types of genetic material: mitochondrial DNA (mtDNA), the Y chromosome and the non-sex chromosomes called autosomes.

Genetic system	F_{ST}	$N_e\nu$	N_e
mtDNA	0.186	4.38	$N/2$
autosomes	0.144	1.49	$2N$
Y chromosome	0.645	0.55	$N/2$

To explain the last column, note that males have one Y chromosome and females zero; mtDNA is haploid and inherited from an individual's mother, so in mtDNA evolution males can be ignored.

If we imagine the world to be a 50 × 50 grid ($L = 50$), then the correction factor $\pi/\log L$ is about 0.4, and there is a significant amount of bias from using an island model formula on a stepping stone population. The point is moot, however. Seielstad, Minch and Cavalli-Sforza (1998) used the fact that the ratio of estimates from mtDNA and Y chromosome data is 8 to conclude that the migration rate for females is 8 times that for males. When this is done the extra factor cancels out.

3. Proofs of Theorems 1 and 2. In this section we prove the indicated results by following the basic approach of Cox (1989), where simple random walk in continuous time is treated. Here, we work with discrete time, use a more general random walk and need uniformity in the parameter ν. We will begin with Theorem 2, which implies Theorem 1, since, under P_π, the probability that the distance between the two chosen particles is larger than $L/(\log L)$ tends to 1 as $L \to \infty$.

Let W_n^1 and W_n^2 be independent random walks on $\mathbf{Z}^2$ that take jumps according to q with probability ν and do not move with probability $1 - \nu$. Let X_n be the difference random walk on the torus

$$X_n = W_n^2 - W_n^1 \mod L,$$

so that $T_0 = \inf\{n \geq 1 : X_n = 0\}$. To get rid of the dependence on ν, it is useful to scale time by the factor ν. Let $Y_s = X_{[s/\nu]}$ for $s \geq 0$, where $[t]$ is the largest integer less than or equal to t, and let $\rho_s^L(x, y) = P_x(Y_s = y)$. Note that $\rho_s^L(x, y)$ is defined for all $s \geq 0$ and is constant between integer multiples of $1/\nu$. Let $\hat{T}_0 = \min\{s \geq 0 : Y_s = 0\}$ and define the transforms

$$F_L(x, \lambda) = E_x \exp(-\lambda \hat{T}_0),$$
$$G_L(x, \lambda) = \int_0^\infty e^{-\lambda t} \rho_t^L(x, 0)\, dt.$$

Breaking things down according to the value of $\hat{T}_0$, and using the Markov property of Y_s (which holds at times $s = m\nu$), shows that $G_L(x, \lambda) = F_L(x, \lambda) G_L(0, \lambda)$ and hence that

$$F_L(x, \lambda) = \frac{G_L(x, \lambda)}{G_L(0, \lambda)}. \tag{3.1}$$

We will prove Theorem 2 by determining the limiting behavior of $G_L(x, \lambda)$ for appropriate λ and applying (3.1). To do this, we need several preliminary facts which are consequences of the local central limit theorem.

LEMMA 3.1. (a) *Let* $\varepsilon_L = 1/\sqrt{\log L}$. *There is a finite constant* C *such that*, *if* $L \geq 2$,

$$\sup_{u \geq \varepsilon_L L^2} \sup_{x \in \Lambda(L)} \varepsilon_L L^2 \cdot \rho_u^L(x, 0) \leq C. \tag{3.2}$$

(b) *If* $t_L \to \infty$ *as* $L \to \infty$, *then*, *uniformly for* $\nu \in (0, 1]$,

$$\sup_{u \geq t_L L^2} \sup_{x \in \Lambda(L)} L^2 |\rho_u^L(x, 0) - L^{-2}| \to 0. \tag{3.3}$$

(c) *If* $u_L \to \infty$ *as* $L \to \infty$, *then*

$$\sup_{x \in \Lambda(L)} \sup_{u_L(1+|x|)^2 \leq t \leq \varepsilon_L L^2} \bigl|\bigl(2\pi(2\sigma^2)t\bigr)\rho_t^L(x, 0) - 1\bigr| \to 0. \tag{3.4}$$

(d) *There is a finite constant* C *such that*

$$\sup_{u \geq 0,\, x \in \Lambda(L)} (1 + |x|^2)\rho_u^L(x, 0) \leq C. \tag{3.5}$$

To encourage the reader to skip the proof of Lemma 3.1, we have hidden it away in the Appendix. In each case one first uses the local central limit theorem to prove the result for the random walk with jump kernel q. Then one uses the fact that ρ_t^L corresponds to a binomial$(2[t/\nu], \nu)$ number of steps according to q.

The next result is:

LEMMA 3.2. *Suppose* $\log^+ |x|/\log L \to \beta \in [0, 1]$ *as* $L \to \infty$. *Then*, *uniformly for* $\nu \in (0, 1]$,

$$\frac{G_L(x, \lambda/L^2 \log L)}{\log L} \to \lambda^{-1} + (1 - \beta)\frac{1}{2\pi\sigma^2}. \tag{3.6}$$

It follows from (3.1) and Lemma 3.2 that if $\log^+ |x|/\log L \to \beta \in [0, 1]$ as $L \to \infty$, then

$$F_L\biggl(x, \frac{2\pi\sigma^2\lambda}{L^2 \log L}\biggr) \to \frac{\lambda^{-1} + (1 - \beta)}{\lambda^{-1} + 1}.$$

That is, as $L \to \infty$, uniformly in $\nu \in (0, 1]$,

$$E_x \exp\biggl(-\frac{2\pi\sigma^2\lambda}{L^2 \log L}\hat{T}_0\biggr) \to (1 - \beta) + \beta\frac{1}{1 + \lambda}.$$

The right-hand side above is the Laplace transform of the distribution that is $1 - \beta$ times a point mass at 0 plus β times a mean one exponential. By a standard argument, Theorem 2 follows.

PROOF OF LEMMA 3.2. We will write $o(1)$ for a quantity that tends to 0, uniformly in $\nu \in (0,1]$, as $L \to \infty$. Let $\varepsilon_L = 1/(\log L)^{1/2}$ and assume that $t_L \to \infty$ and $u_L \to \infty$ such that $t_L/(\log L)^{1/2} \to 0$ and $u_L/\log L \to 0$ as $L \to \infty$. We compute the left-hand side of (3.6) as follows. By (3.3),

$$\begin{aligned}
&\frac{1}{\log L}\int_{t_L L^2}^{\infty} \exp\left(\frac{-\lambda t}{L^2 \log L}\right)\rho_t^L(x,0)\,dt \\
&\qquad = \frac{1}{\log L}\int_{t_L L^2}^{\infty} \exp\left(\frac{-\lambda t}{L^2\log L}\right)\frac{1+o(1)}{L^2}\,dt \qquad (3.7)\\
&\qquad = \frac{1+o(1)}{\lambda}\exp\left(\frac{-\lambda t_L}{\log L}\right) \to \lambda^{-1}
\end{aligned}$$

as $L \to \infty$. By (3.2),

$$\begin{aligned}
&\frac{1}{\log L}\int_{\varepsilon_L L^2}^{t_L L^2} \exp\left(\frac{-\lambda t}{L^2\log L}\right)\rho_t^L(x,0)\,dt \\
&\qquad \le \frac{1}{\log L}\int_{\varepsilon_L L^2}^{t_L L^2}\frac{C}{\varepsilon_L L^2}\,dt \le \frac{C t_L}{\varepsilon_L \log L} \to 0 \qquad (3.8)
\end{aligned}$$

as $L \to \infty$. Next, using (3.4), $\log \varepsilon_L/\log L \to 0$ and $\log u_L/\log L \to 0$, it follows that

$$\begin{aligned}
&\frac{1}{\log L}\int_{u_L(1+|x|^2)}^{\varepsilon_L L^2}\exp\left(\frac{-\lambda t}{L^2\log L}\right)\rho_t^L(x,0)\,dt \\
&\quad = \frac{1}{\log L}\int_{u_L(1+|x|)^2}^{\varepsilon_L L^2}\frac{1+o(1)}{2\pi(2\sigma^2)t}\,dt \qquad (3.9)\\
&\quad = \frac{1+o(1)}{2\pi(2\sigma^2)\log L}\bigl(2\log L - 2\log(1+|x|) + \log\varepsilon_L - \log u_L\bigr) \to \frac{1-\beta}{2\pi\sigma^2}
\end{aligned}$$

as $L \to \infty$. To complete the calculation, (3.5) implies

$$\begin{aligned}
&\frac{1}{\log L}\int_0^{u_L(1+|x|^2)}\exp\left(\frac{-\lambda t}{L^2\log L}\right)\rho_t^L(x,0)\,dt \\
&\qquad \le \frac{1}{\log L}\int_0^{u_L(1+|x|^2)}\frac{C}{(1+|x|)^2}\,dt \qquad (3.10)\\
&\qquad \le \frac{C u_L}{\log L} \to 0
\end{aligned}$$

as $L \to \infty$. Combining (3.7)–(3.10) gives (3.6). □

4. Proof of Theorem 4. Our first step is to show that the random variables $t_0/2NL^2$, under P_x for $x \in \Lambda(L)$, are uniformly integrable. To do this, we first observe that

$$\frac{NL^2}{(L^2 \log L)/\nu} = \frac{N\nu}{\log L} \to \infty \qquad \text{as } L \to \infty. \tag{4.1}$$

On account of this, Theorem 1 implies that, for any $\varepsilon > 0$, $P_\pi(T_0 > \varepsilon NL^2) \to 0$ as $L \to \infty$. Considering the location of $X_{[\varepsilon NL^2]}$, we may write

$$P_x(T_0 > 2\varepsilon NL^2) \le \sum_{y \in \Lambda(L)} P_x\big(X_{[\varepsilon NL^2]} = y\big) P_y(T_0 > \varepsilon NL^2).$$

By (3.3) and (4.1), $P_x(X_{[\varepsilon NL^2]} = y) = (1 + o(1))/L^2$ uniformly in y and ν as $L \to \infty$. Thus,

$$P_x(T_0 > 2\varepsilon NL^2) \le \sum_{y \in \Lambda(L)} \frac{1 + o(1)}{L^2} P_y(T_0 > \varepsilon NL^2) = \big(1 + o(1)\big) P_\pi(T_0 > \varepsilon NL^2)$$

as $L \to \infty$. That is, we have established that, for any $\varepsilon > 0$,

$$\lim_{L \to \infty} \sup_{x \in \Lambda(L)} P_x(T_0 > \varepsilon NL^2) = 0. \tag{4.2}$$

We may therefore choose L_0 such that, for all $L \ge L_0$ and $x \in \Lambda(L)$,

$$P_x(T_0 > NL^2) < 1/12.$$

Since t_0 occurs after T_0, and $P_0(t_0 > 3NL^2) \le E_0 t_0/3NL^2 = 2/3$ by Theorem 3, we have

$$P_x(t_0 > 4NL^2) \le P_x(T_0 > NL^2) + P_0(t_0 > 3NL^2) < 3/4$$

for all $L \ge L_0$ and $x \in \Lambda(L)$. Iterating this inequality, we obtain

$$\sup_{x \in \Lambda(L)} P_x(t_0 > 4nNL^2) < (3/4)^n, \qquad n \ge 1, \tag{4.3}$$

which implies uniform integrability.

Our next step is to argue that

$$\lim_{L \to \infty} P_0(t_0 > 2NL^2 t) = \exp(-t). \tag{4.4}$$

Tightness of the laws of $t_0/2NL^2$ follows from the uniform integrability just established (or, more simply, from Theorem 3). We will show that every subsequential limit is exponential with mean 1, proving convergence of the whole sequence. Let F denote the distribution function of the limit law along some subsequence L_k, which, for notational simplicity, we will write as L, and let s, t be continuity points of F. We will show now that

$$1 - F(s + t) = \big(1 - F(s)\big)\big(1 - F(t)\big). \tag{4.5}$$

By (4.1), we may choose a sequence $\varepsilon_L > 0$ such that

$$\varepsilon_L \to 0 \quad \text{and} \quad \varepsilon_L \frac{N\nu}{\log L} \to \infty \tag{4.6}$$

as $L \to \infty$. We can estimate $P_0(t_0 > 2NL^2(s+t))$ by arguing that, with high probability, the event $t_0 > 2NL^2(s+t)$ occurs in the following way. First, t_0 does not occur before time $2NL^2(s+\varepsilon_L)$, at which time the random walk is uniformly distributed over $\Lambda(L)$. Next, the time T_0 occurs at some time $k \in [2NL^2(s+\varepsilon_L), 2NL^2(s+2\varepsilon_L)]$. Starting at 0 at this time k, t_0 does not occur for another $2NL^2t - k$ units of time. Here are the details.

By (4.1), (4.6) and (3.3), $P_x(X_j = y) = (1+o(1))/L^2$ as $L \to \infty$, uniformly in $j \ge [2NL^2 s]$ and $x, y \in \Lambda(L)$. Therefore, with $I = [[2NL^2 s], 2NL^2(s+\varepsilon_L)]$,

$$P_0(t_0 \in I) \le \frac{1}{2N} \sum_{j \in I} P_0(X_j = 0) \le \frac{(1+o(1))(2NL^2\varepsilon_L + 1)}{NL^2} \to 0 \tag{4.7}$$

as $L \to \infty$. By Theorem 1, we also have

$$P_\pi(T_0 < 2NL^2\varepsilon_L) \to 1. \tag{4.8}$$

By (4.7), (4.8) and the Markov property,

$$\begin{aligned} &P_0\bigl(t_0 > 2NL^2(s+t)\bigr) \\ &\quad = o(1) + \sum_{x,y \in \Lambda(L)} \sum_{k=0}^{[2NL^2\varepsilon_L]} P_0\bigl(t_0 > [2NL^2 s], X_{[2NL^2 s]} = x\bigr) P_x(X_{[2NL^2\varepsilon_L]} = y) \\ &\qquad \times P_y(T_0 = k) P_0(t_0 > u_L - k), \end{aligned}$$

where $u_L = 2NL^2(s+t) - [2NL^2 s] - [2NL^2\varepsilon_L]$. Replacing $P_x(X_{[2NL^2\varepsilon_L]} = y)$ with $(1+o(1))/L^2$ and summing on x and y, we obtain

$$\begin{aligned} &P_0\bigl(t_0 > 2NL^2(s+t)\bigr) \\ &\quad = o(1) + P_0(t_0 > 2NL^2 s) \sum_{k=0}^{[2NL^2\varepsilon_L]} P_\pi(T_0 = k) P_0(t_0 > u_L - k). \end{aligned}$$

Since $(u_L - k)/2NL^2 \to t$ as $L \to \infty$, uniformly for $k \in [0, 2NL^2\varepsilon_L]$, it follows that

$$P_0\bigl(t_0 > 2NL^2(s+t)\bigr) = o(1) + P_0(t_0 > 2NL^2 s) P_0(t_0 > 2NL^2 t)$$

as $L \to \infty$, which proves (4.5).

To see that (4.5) implies that F is exponential, we can argue as follows. Since F is monotone, it has only countably many discontinuity points, and there exists a $\theta > 0$ such that all points of the form m/θ^n, for positive integers m, n, are continuity points of F. Since (4.5) must hold for all s, t of this form, using monotonicity again we conclude that there exists a $\lambda \geq 0$ such that $F(t) = 1 - e^{-\lambda t}$ for all $t > 0$. Now the uniform integrability estimate (4.3) implies that $\lambda = 1$, and hence we have established (4.4).

To complete the proof of Theorem 4, we note that, for $x \neq 0$ and $0 < \varepsilon < t$,

$$P_x(t_0 > 2NL^2 t) \leq P_x(T_0 > \varepsilon NL^2 t) + P_0\bigl(t_0 > 2NL^2(t - \varepsilon)\bigr).$$

By (4.2) and (4.4), this implies

$$\lim_{L\to\infty} \sup_{x\in\Lambda(L)} P_x(t_0 > 2NL^2 t) \leq \lim_{L\to\infty} P_0(t_0 > 2NL^2 t) = \exp(-t).$$

On the other hand, $P_x(t_0 > 2NL^2 t) \geq P_0(t_0 > 2NL^2 t)$, so the proof of Theorem 4 is complete. □

5. Proof of Theorem 5.

We begin with some preliminary results for a random walk on $\mathbf{Z}^2$. Let $\bar{X}_n = W_n^1 - W_n^2$ be the difference between the two random walk positions in $\mathbf{Z}^2$ and let $\bar{Y}_s = \bar{X}_{[s/\nu]}$.

LEMMA 5.0. *As $t \to \infty$, uniformly for $\nu \in (0, 1]$,*

$$\int_0^t P_0(\bar{Y}_s = 0)\, ds \sim \frac{\log t}{2\pi\sigma^2}.$$

PROOF. The rescaled random walks have second moments bounded away from 0 and third absolute moments bounded. With this in hand, one can work through one's favorite proof of the local central limit theorem [see, e.g., Durrett (1995), pages 132–134] and get explicit error estimates that only depend on the moment bounds. Alternatively, using the stronger local central limit theorem given in the Appendix, one can use the argument for part (c) of Lemma 3.1 to show that, as $s \to \infty$, uniformly in $\nu \in (0, 1]$,

$$2\pi(2\sigma^2)s P_0(\bar{Y}_s = 0) \to 1,$$

and then integrate this result. □

Now let $\bar{T}_0 = \inf\{n \geq 1 : \bar{X}_n = 0\}$, and define

$$\gamma = P_0(\bar{X}_1 \neq 0) = 2\nu(1-\nu) + \nu^2\Biggl(1 - \sum_{x\in\mathbf{Z}^2} q(x)^2\Biggr) \tag{5.1}$$

[recall $q(0) = 0$].

LEMMA 5.1. *As $t \to \infty$, uniformly for $\nu \in (0, 1]$,*

$$P_0\biggl(\bar{T}_0 > \frac{t}{\nu}\Big|\bar{X}_1 \neq 0\biggr) \sim \frac{2\pi\sigma^2}{\log t} \cdot \frac{\nu}{\gamma}. \tag{5.2}$$

REMARK. We note that $\gamma \sim 2\nu$ as $\nu \to 0$, so ν/γ is bounded and bounded away from 0 for all $\nu \in (0, 1]$.

PROOF OF LEMMA 5.1. Breaking things down according to the last visit to 0 before time $[t/\nu]$,

$$1 = \sum_{r=1}^{[t/\nu]} P_0(\bar{X}_{r-1} = 0)\gamma P_0\bigl(\bar{T}_0 > [t/\nu] - r \mid \bar{X}_1 \neq 0\bigr).$$

Dropping the $-r$, which makes the probability smaller, we have

$$\begin{aligned} P_0\biggl(\bar{T}_0 > \biggl[\frac{t}{\nu}\biggr] \bigg| X_1 \neq 0\biggr) &\le \frac{1}{\gamma \sum_{r=1}^{[t/\nu]} P_0(\bar{X}_{r-1} = 0)} \\ &= \frac{\nu}{\gamma \int_0^{[t/\nu]\nu} P_0(\bar{Y}_s = 0)\,ds} \sim \frac{2\pi\sigma^2}{\log t} \cdot \frac{\nu}{\gamma} \end{aligned} \tag{5.3}$$

as $t \to \infty$ by Lemma 5.0. For a bound in the other direction, consider the last visit to 0 before time $[t(1 + \log t)/\nu]$,

$$1 \le \sum_{r=0}^{[t\log t/\nu]} P_0(\bar{X}_r = 0)\gamma P_0\bigl(\bar{T}_0 > [t/\nu] \mid \bar{X}_1 \neq 0\bigr) + \sum_{r=[t\log t/\nu]}^{[t(1+\log t)/\nu]} P_0(\bar{X}_r = 0)\gamma.$$

Computing as before, we obtain

$$\begin{aligned} P_0\biggl(\bar{T}_0 > \biggl[\frac{t}{\nu}\biggr] \bigg| \bar{X}_1 \neq 0\biggr) &\ge \frac{\nu/\gamma - \int_{[t\log t]}^{[t(1+\log t)]} P_0(\bar{Y}_s = 0)\,ds}{\int_0^{[t\log t]} P_0(\bar{Y}_s = 0)\,ds} \\ &\sim \frac{2\pi\sigma^2\nu/\gamma - \log(1 + 1/\log t)}{\log(t\log t)} \sim \frac{2\pi\sigma^2}{\log t} \cdot \frac{\nu}{\gamma} \end{aligned} \tag{5.4}$$

as $t \to \infty$. Together, (5.3) and (5.4) imply (5.2). □

We will make use of the following abbreviations:

$$\begin{aligned} u_1 &= L^2/(\log L)^2, \\ u_2 &= L^2/\sqrt{\log L}, \\ u_3 &= L^2\sqrt{\log L}. \end{aligned}$$

LEMMA 5.2. *There is a constant C so that, for all $L \ge 3$,*

$$P_0\biggl(\frac{u_1}{\nu} < \bar{T}_0 \le \frac{u_3}{\nu} \bigg| \bar{X}_1 \neq 0\biggr) \le \frac{C \log\log L}{(\log L)^2}. \tag{5.5}$$

PROOF. Let $I(s,t)=\int_s^t P_0(\bar{Y}_r=0)\,dr$. By (5.3) and (5.4), the probability of interest is asymptotically at most

$$\begin{aligned}&\frac{\nu/\gamma}{I(0,u_1)}-\frac{\nu/\gamma-I(u_3\log u_3,u_3(1+\log u_3))}{I(0,u_3\log u_3)}\\&\qquad=\frac{(\nu/\gamma)I(u_1,u_3\log u_3)+I(0,u_1)I(u_3\log u_3,u_3(1+\log u_3))}{I(0,u_1)I(0,u_3\log u_3)}.\end{aligned}$$

From Lemma 5.0, it follows that, as $L\to\infty$,

$$\begin{aligned}&(2\pi\sigma^2)^2 I(0,u_1)I(0,u_3\log u_3)\sim(2\log L)^2,\\&\qquad 2\pi\sigma^2 I(u_1,u_3\log t_2)\sim\log u_3+\log\log u_3-\log u_1\sim 4.5\log\log L,\\&2\pi\sigma^2 I\bigl(u_3\log u_3,u_3(1+\log u_3)\bigr)\sim\log\biggl(1+\frac{1}{\log u_3}\biggr)\sim\frac{1}{2\log L}.\end{aligned}$$

Combining these asymptotics gives (5.5). □

Let $R_0=0$ and, for $k\ge 1$, define

$$\begin{aligned}Q_k&=\min\{s>R_{k-1}:\bar{X}_s\ne 0\},\\R_k&=\min\{s>Q_k:\bar{X}_s=0\},\end{aligned}$$

and also $K=\min\{k\ge 1:R_k-Q_k>L^2/\nu\}$. Then K is geometric, with success probability $P_0(\bar{T}_0-1>L^2/\nu\mid\bar{X}_1\ne 0)$. Note that, by Lemma 5.1, there is a constant C such that $EK\le C\log L$. The random variables Q_k-R_{k-1} are iid geometric random variables with success probability γ, independent of K. Now define $\mathcal{O}_K$ to be the number of visits to 0 up to time Q_K and $\mathcal{O}_L$ to be the number of visits to 0 by time u_3/ν,

$$\begin{aligned}\mathcal{O}_L&=\sum_{k\le u_3/\nu}1\{\bar{X}_k=0\},\\\mathcal{O}_K&=\sum_{k=1}^{K}(Q_k-R_{k-1}).\end{aligned}$$

We are interested primarily in $\mathcal{O}_L$, but $\mathcal{O}_K$ is easier to analyze, because it is geometric with success probability $p_L=\gamma P_0(\bar{T}_0-1>L^2/\nu\mid\bar{X}_1\ne 0)$. The next result shows that $\mathcal{O}_L=\mathcal{O}_K$ with high probability.

LEMMA 5.3. *If* $L\to\infty$, *then* (a) $P_0(\bar{X}_k=0$ *for some* $k\in[2u_2/\nu,u_3/\nu])\to 0$ *and* (b) $P_0(\mathcal{O}_K\ne\mathcal{O}_L)\to 0$.

PROOF. We claim that

$$P_0\biggl(Q_K<\frac{2u_2}{\nu},R_K>\frac{u_3}{\nu}\biggr)\to 1\qquad\text{as } L\to\infty.$$

Both (a) and (b) follow from this claim and the observation that $\bar{X}_k \neq 0$ for $k \in [Q_K, R_K)$. We start with an upper bound for R_K. Since $R_K - Q_K > L^2/\nu$, we have, using Lemma 5.2,

$$\begin{aligned}
P_0\biggl(R_K \le \frac{u_3}{\nu}\biggr) &\le \sum_{k=1}^{\infty} P_0\biggl(R_j - Q_j \le \frac{L^2}{\nu} \text{ for } j \le k-1, \frac{L^2}{\nu} < R_k - Q_k \le \frac{u_3}{\nu}\biggr)\\
&\le P_0\biggl(\frac{L^2}{\nu} < R_K - Q_K \le \frac{u_3}{\nu}\biggr)\\
&\le \sum_{k=1}^{\infty} P_0(K > k-1) P_0\biggl(\frac{L^2}{\nu} < \bar{T}_0 - 1 \le \frac{u_3}{\nu} \Big| \bar{X}_1 \neq 0\biggr)\\
&\le \frac{C \log\log L}{(\log L)^2} EK \to 0
\end{aligned}$$

as $L \to \infty$, since $EK \le C \log L$.

To bound Q_K, we first observe that

$$Q_K = \sum_{k=1}^{K} (Q_k - R_{k-1}) + \sum_{k=1}^{K-1} (R_k - Q_k). \tag{5.6}$$

For the first sum on the right-hand side, we have

$$E \sum_{k=1}^{K} (Q_k - R_{k-1}) = \frac{1 + EK}{\gamma} \le \frac{C \log L}{\nu},$$

so by Markov's inequality,

$$P_0\biggl(\sum_{k=1}^{K} (Q_k - R_{k-1}) \ge \frac{u_2}{\nu}\biggr) \le \frac{C(\log L)^{3/2}}{L^2}.$$

For the second sum, we note first that Markov's inequality and the estimate $EK \le C \log L$ imply $P(K \ge (\log L)^{3/2}) \le C/\sqrt{\log L}$. Next, since $R_k - Q_k \le L^2/\nu$ for $k < K$,

$$\begin{aligned}
&P_0\biggl(R_k - Q_k > \frac{u_1}{\nu} \text{ for some } k < K \text{ and } K < (\log L)^{3/2}\biggr)\\
&\qquad \le (\log L)^{3/2} P_0\biggl(\frac{u_1}{\nu} < \bar{T}_0 - 1 < \frac{L^2}{\nu} \Big| \bar{X}_1 \neq 0\biggr)\\
&\qquad \le \frac{C \log\log L}{\sqrt{\log L}}
\end{aligned}$$

by Lemma 5.2. When $K < (\log L)^{3/2}$ and $R_k - Q_k < u_1/\nu$ for $k < K$,

$$\sum_{k=1}^{K-1} (R_k - Q_k) \le (\log L)^{3/2} \biggl(\frac{u_1}{\nu}\biggr) = \frac{u_2}{\nu}.$$

Using the estimates in (5.6), we obtain

$$P\left(Q_K \geq \frac{2u_2}{\nu}\right) \to 0$$

as $L \to \infty$, and we are done. $\square$

We are now ready to prove the key lemma.

LEMMA 5.4. *If* $2N\pi\sigma^2\nu/\log L \to \alpha \in [0,\infty)$ *as* $L \to \infty$, *then*

$$\lim_{L\to\infty} P_0\left(t_0 > \frac{u_3}{\nu}\right) = \frac{\alpha}{1+\alpha}. \tag{5.7}$$

PROOF. We consider first the corresponding problem on $\mathbf{Z}^2$, with colonies of size $2N$ located at each point and $\bar{t}_0$ the time that two lineages first coalesce. Since the probability of coalescence when two lines land in the same colony is $1/2N$,

$$P_0\left(\bar{t}_0 > \frac{u_3}{\nu}\right) = E_0\left(1 - \frac{1}{2N}\right)^{\mathcal{O}_L}.$$

Since $P(\mathcal{O}_L \neq \mathcal{O}_K) \to 0$ as $L \to \infty$, it suffices to compute

$$\begin{aligned} E\left(1-\frac{1}{2N}\right)^{\mathcal{O}_K} &= \sum_{k=1}^{\infty} p_L(1-p_L)^{k-1}\left(1-\frac{1}{2N}\right)^{k-1} \\ &= \frac{p_L}{1-(1-p_L)(1-1/2N)} = \frac{2Np_L}{2Np_L+1-p_L}. \end{aligned}$$

By Lemma 5.1, $2Np_L \to \alpha$, so we have proved

$$\lim_{L\to\infty} P_0\left(\bar{t}_0 > \frac{u_3}{\nu}\right) = \frac{\alpha}{\alpha+1}. \tag{5.8}$$

To transfer this result to the torus $\Lambda(L)$, we suppose that our random walks are constructed so that $X_k = \bar{X}_k \bmod L$. In this case $\bar{X}_k = 0$ implies $X_k = 0$. When $\bar{X}_k = 0$ we use a single coin flip to determine if coalescence should occur in the two systems. If we do this, then it follows that $t_0 \leq \bar{t}_0$. In particular, for $\alpha = 0$, (5.8) implies (5.7). For the remainder of the proof, we suppose that $\alpha > 0$. We will argue that $X_k = \bar{X}_k$ for all $k < u_2/\nu$ with high probability and, that X_k does not hit 0 enough times during $[u_2/\nu, u_3/\nu]$ to cause t_0 (or $\bar{t}_0$) to occur during that time interval.

The first step is simple. By the L^2-maximal inequality for martingales and Markov's inequality,

$$P\left(\max_{0\leq k\leq u_2/\nu} |\bar{X}_k| > L/3\right) \leq CE|X_{[u_2/\nu]}|^2/L^2 \leq Cu_2\sigma^2/L^2 \to 0$$

as $L \to \infty$. Consequently, $\bar{X}_k$ stays in $\Lambda(L)$ up to time u_2/ν with high probability, which implies that

$$P_0(t_0 > u_2/\nu) - P_0(\bar{t}_0 > u_2/\nu) \to 0. \tag{5.9}$$

For the second step, we first note that we can bound the probability of coalescence of lineages with the expectation estimate

$$P_0\biggl(t_0 \in \biggl[\frac{u_2}{\nu}, \frac{u_3}{\nu}\biggr]\biggr) \le \frac{1}{2N} \sum_{k=u_2/\nu}^{u_3/\nu} P_0(X_k = 0).$$

By Lemma 3.1, there are constants A and C such that

$$P_0(X_k = 0) \le \begin{cases} C/u_2, & \text{if } k \in [u_2/\nu, AL^2/\nu], \\ 2/L^2, & \text{if } k \ge AL^2/\nu. \end{cases}$$

These estimates imply that

$$\sum_{k=u_2/\nu}^{u_3/\nu} P_0(X_k = 0) \le \frac{C}{u_2}\frac{AL^2}{\nu} + \frac{2}{L^2}\frac{u_3}{\nu} \le \frac{C\sqrt{\log L}}{\nu}$$

for an appropriate constant C. Consequently,

$$P_0\biggl(t_0 \in \biggl[\frac{u_2}{\nu}, \frac{u_3}{\nu}\biggr]\biggr) \le \frac{C\sqrt{\log L}}{2N\nu} \to 0 \tag{5.10}$$

as $L \to \infty$, since, by assumption, $2N\pi\sigma^2\nu/\log L \to \alpha > 0$. Furthermore, $P_0(\bar{X}_k = 0) \le P(X_k = 0)$, so this argument shows that (5.10) holds with $\bar{t}_0$ replacing t_0. By combining this observation with (5.8)–(5.10), we obtain (5.7). $\square$

PROOF OF THEOREM 5. By assumption, $N = N(L)$ and $\nu = \nu(L)$ satisfy

$$\frac{2N\pi\sigma^2\nu}{\log L} \to \alpha \in [0, \infty) \qquad \text{as } L \to \infty. \tag{5.11}$$

We fix $t > 0$ and define

$$\Delta_L = \biggl[\frac{u_3}{\nu}\biggr], \qquad a_L = \frac{L^2 \log L}{2\pi\nu\sigma^2}, \qquad \tau_L = \frac{L^2 \log L}{2\pi\nu\sigma^2} t.$$

Let $\mathcal{E}_1, \mathcal{E}_2, \dots$ be independent exponential mean one random variables and let G be a geometric random variable with parameter $1/(1+\alpha)$, independent of the $\mathcal{E}_i$. We will prove that, under P_π, for $\alpha > 0$,

$$\frac{t_0}{a_L} \Rightarrow \mathcal{E}_1 + \cdots + \mathcal{E}_G \qquad \text{as } L \to \infty. \tag{5.12}$$

Since $\mathcal{E}_1 + \cdots + \mathcal{E}_G$ is exponential with parameter $1/(1+\alpha)$, (5.12) implies (1.6).

Define the sequence of delayed return times S_n by setting $S_1 = T_0$ and

$$S_{n+1} = \inf\{k > S_n + 2\Delta_L : X_k = 0\}, \qquad n \geq 1. \tag{5.13}$$

Intuitively, during each interval $[S_n, S_n + \Delta_L]$, t_0 has probability $\approx 1/(1+\alpha)$ of occurring, and we can ignore the probability of t_0 occurring in $[S_n + \Delta_L, S_n + 2\Delta_L]$. Also, at time $S_n + 2\Delta_L$, the difference between the walks has distribution $\approx \pi$, so

$$\bigl(S_{n+1} - (S_n + 2\Delta_L)\bigr)/a_L \approx \mathcal{E}_n.$$

These facts suggest (5.12) should hold.

To begin to make this rigorous, we note that the time t_0 cannot occur during any of the time intervals $(S_n + 2\Delta_L, S_{n+1})$, so

$$P_\pi(t_0 > \tau_L) = \sum_{n=1}^{\infty} P_\pi\bigl(t_0 > \tau_L, t_0 \in [S_n, S_n + 2\Delta_L]\bigr). \tag{5.14}$$

Using this decomposition and Lemma 5.4, we will prove that, for each $n \geq 1$,

$$\lim_{L\to\infty} P_\pi(t_0 \geq S_n) = \left(\frac{\alpha}{1+\alpha}\right)^{n-1} \tag{5.15}$$

and

$$\lim_{L\to\infty} P_\pi\bigl(t_0 > \tau_L, t_0 \in [S_n, S_n + 2\Delta_L)\bigr) = \frac{\alpha^{n-1}}{(1+\alpha)^n} P(\mathcal{E}_1 + \cdots + \mathcal{E}_n > t). \tag{5.16}$$

For $\alpha > 0$, the results (5.14)–(5.16) imply (5.12). For $\alpha = 0$, they imply that $P_\pi(t_0 > \tau_L) \to P(\mathcal{E}_1 > t) = e^{-t}$, and hence (1.6) holds. So it suffices to prove (5.15) and (5.16).

We begin with a preliminary estimate, which shows that t_0 will not occur during any of the time intervals $[S_n + \Delta_L, S_n + 2\Delta_L]$ with significant probability. By the strong Markov property, an obvious inequality and (3.3), as $L \to \infty$,

$$\begin{aligned} P_\pi\bigl(t_0 \in [S_n + \Delta_L, S_n + 2\Delta_L)\bigr) &\leq P_0\bigl(t_0 \in [\Delta_L, 2\Delta_L]\bigr) \\ &\leq \frac{1}{2N} \sum_{k\in[\Delta_L, 2\Delta_L]} P(X_k = 0) \leq \frac{1+o(1)}{2NL^2}(\Delta_L + 1). \end{aligned}$$

By (5.11), the last expression above tends to 0 as $L \to \infty$. That is,

$$\lim_{L\to\infty} P_\pi\bigl(t_0 \in [S_n + \Delta_L, S_n + 2\Delta_L]\bigr) = 0, \qquad n \geq 1. \tag{5.17}$$

We now prove (5.15). By the strong Markov property and (5.17),

$$\begin{aligned} P_\pi(t_0 \geq S_n) &= o(1) + P_\pi\bigl(t_0 \notin [S_k, S_k + \Delta_L], 1 \leq k < n\bigr) \\ &= o(1) + P_\pi\bigl(t_0 \notin [S_k, S_k + \Delta_L], 1 \leq k < n-1\bigr)\frac{2N-1}{2N} P_0(t_0 \geq \Delta_L). \end{aligned}$$

Applying Lemma 5.4 gives

$$P_\pi(t_0 \ge S_n) = o(1) + \frac{\alpha}{1+\alpha}\big(P_0(t_0 \ge S_{n-1})\big)$$

as $L \to \infty$. Iteration of this argument establishes (5.15).

Now we prepare for the proof of (5.16). The estimate (5.17) implies that we may replace $P_\pi(t_0 > \tau_L, t_0 \in [S_n, S_n + 2\Delta_L])$ in the decomposition (5.14) with $P_\pi(t_0 > \tau_L, t_0 \in [S_n, S_n + \Delta_L])$. We would like to replace the latter quantity with $P_\pi(S_n > \tau_L, t_0 \in [S_n, S_n + \Delta_L])$. To see that this is possible, we first observe that

$$\begin{aligned} &P_\pi\big(t_0 > \tau_L, t_0 \in [S_n, S_n + \Delta_L]\big) \\ &\qquad = P_\pi\big(S_n > \tau_L, t_0 \in [S_n, S_n + \Delta_L]\big) + P_\pi(S_n \le \tau_L < t_0 \le S_n + \Delta_L). \end{aligned} \tag{5.18}$$

We claim that

$$\lim_{L\to\infty} P_\pi\big(\tau_L \in [S_n, S_n + \Delta_L]\big) = 0, \tag{5.19}$$

which certainly implies that the second term on the right-hand side of (5.18) tends to 0. The $n = 1$ case of (5.19) is straightforward. Since $\Delta_L = o(\tau_L)$ as $L \to \infty$,

$$P_\pi\big(\tau_L \in [S_1, S_1 + \Delta_L]\big) = P_\pi\big(T_0 \in [\tau_L - \Delta_L, \tau_L]\big) \to e^{-t} - e^{-t} = 0$$

by Theorem 1. For $n > 1$, by decomposing according to the time S_{n-1} and positions $X_{S_{n-1}+\Delta_L}$ and $X_{S_{n-1}+2\Delta_L}$ and using the Markov property, we have

$$\begin{aligned} P_\pi\big(\tau_L \in [S_n, S_n + \Delta_L]\big) = \sum_j \sum_{x,y\in\Lambda(L)} &P_\pi(S_{n-1} = j, X_{j+\Delta_L} = x) P_x(X_{\Delta_L} = y) \\ &\times P_y\big(\tau_L - (j + 2\Delta_L) \in [T_0, T_0 + \Delta_L]\big). \end{aligned}$$

Using (3.3) to replace $P_x(X_{\Delta_L} = y)$ with $(1 + o(1))/L^2$, summing on x and y, we obtain

$$\begin{aligned} &P_\pi\big(\tau_L \in [S_n, S_n + \Delta_L]\big) \\ &\qquad = o(1) + \sum_j P_\pi(S_{n-1} = j) P_\pi\big(\tau_L - (j + 2\Delta_L) \in [T_0, T_0 + \Delta_L]\big) \end{aligned} \tag{5.20}$$

as $L \to \infty$. By Theorem 1, for fixed j, since $\Delta_L = o(\tau_L)$ as $L \to \infty$,

$$\lim_{L\to\infty} P_\pi\big(\tau_L - (j + 2\Delta_L) \in [T_0, T_0 + \Delta_L]\big) = 0.$$

Substitution into (5.20) shows that (5.19) holds.

In view of (5.14), (5.17), (5.18) and (5.19), we now have

$$P_\pi(t_0 > \tau_L) = o(1) + \sum_{n=1}^{\infty} P_\pi\big(S_n > \tau_L, t_0 \in [S_n, S_n + \Delta_L]\big). \tag{5.21}$$

To add up the $o(1)$'s, we note that (5.3) implies that there is a $\delta > 0$ independent of L so that $P_\pi(t_0 \geq S_n) \leq (1-\delta)^{n-1}$ for all n, and we use the dominated convergence theorem. By the strong Markov property and Lemma 5.4,

$$\begin{aligned}P_\pi\bigl(S_n > \tau_L, t_0 \in [S_n, S_n + \Delta_L]\bigr) &= P_\pi(\tau_L < S_n \leq t_0) P_0(t_0 < \Delta_L)\\ &= o(1) + P_\pi(\tau_L \leq S_n \leq t_0)\frac{1}{1+\alpha}.\end{aligned}$$

Inserting this into (5.21) gives

$$P_\pi(t_0 > \tau_L) = o(1) + \frac{1}{1+\alpha}\sum_{n=1}^{\infty} P_\pi(\tau_L < S_n \leq t_0). \tag{5.22}$$

Our final task is to show that

$$\begin{aligned}&\lim_{L\to\infty} P_\pi(\tau_L < S_n \leq t_0)\\ &\qquad = \left(\frac{\alpha}{1+\alpha}\right)^{n-1} P(\mathcal{E}_1 + \mathcal{E}_2 + \cdots + \mathcal{E}_n > t), \qquad n \geq 1,\end{aligned}\tag{5.23}$$

because (5.16) follows from (5.15), (5.22) and (5.23).

The $n = 1$ case of (5.23) is an immediate consequence of Theorem 1, since

$$P_\pi(\tau_L < S_1 \leq t_0) = P_\pi(T_0 > \tau_L) \to P(\mathcal{E}_1 > t) \qquad \text{as } L \to \infty.$$

For $n > 1$, we use a decomposition similar to the one used in the proof of (5.19). As $L \to \infty$, by Lemma 3.1,

$$\begin{aligned}P_\pi(\tau_L < S_n \leq t_0) = o(1) + \sum_j \sum_{x,y\in\Lambda(L)} &P_\pi(S_{n-1} = j, t_0 > j) P_0(t_0 > \Delta_L, X_{\Delta_L} = x)\\ &\times P_x(X_{\Delta_L} = y) P_y\bigl(T_0 > \tau_L - (j + 2\Delta_L)\bigr).\end{aligned}$$

Replacing $P_x(X_{\Delta_L} = y)$ with $(1 + o(1))/L^2$ and summing over x and y, the right-hand side above becomes

$$o(1) + \sum_j P_\pi(S_{n-1} = j, t_0 > j) P_0(t_0 > \Delta_L) P_\pi\bigl(T_0 > \tau_L - (j + 2\Delta_L)\bigr).$$

Consequently, by Lemma 5.4,

$$P_\pi(\tau_L < S_n \leq t_0) = o(1) + P_\pi(S_{n-1} + T_0^1 > \tau_L - 2\Delta_L, t_0 > S_{n-1})\frac{\alpha}{1+\alpha}$$

as $L \to \infty$, where T_0^1 is independent of the walk X_j and has the same law as T_0 under P_π. Iterating this argument, we see that, as $L \to \infty$,

$$P_\pi(\tau_L < S_n \leq t_0) = o(1) + P_\pi(T_0^n + \cdots + T_0^1 > \tau_L - 2n\Delta_L)\left(\frac{\alpha}{1+\alpha}\right)^{n-1},$$

where the T_0^i are independent with the same law as T_0 under P_π. By Theorem 1, since $\Delta_L = o(\tau_L)$,

$$\lim_{L\to\infty} P_\pi(T_0^n + \cdots + T_0^1 > \tau_L - 2n\Delta_L) = P(\mathcal{E}_1 + \cdots + \mathcal{E}_n > t), \qquad n \ge 1,$$

and the proof of (5.23) is finished.

To prove (1.7) now, we let $\Gamma_L = \varepsilon_L L^2 \log L/\nu$. By Theorem 2, if $\varepsilon_L \to 0$ slowly enough, $P_x(T_0 > \Gamma_L) \to \beta$. Lemma 5.4 implies that $P_0(t_0 > \Delta_L) \to \alpha/(1+\alpha)$. When $T_0 > \Gamma_L$, or $T_0 \le \Gamma_L$ and $t_0 > T_0 + \Delta_L$, an event of probability $\approx \beta + (1-\beta)\alpha/(1+\alpha)$, X_n is approximately uniformly distributed at time $\Gamma_L + 2\Delta_L = o(L^2 \log L/\nu)$. These observations having been made, the rest of the proof is straightforward using the techniques above. Details are left to the reader. ☐

APPENDIX

PROOF OF LEMMA 3.1. We first prove all four results for the case $\nu = 1$, that is, for $p = q$. As the reader will see, the extension to $\nu \in (0,1)$ can be obtained by using the observation that ρ_u^L corresponds to a binomial$(2[u/\nu], \nu)$ number of steps according to q. We will sometimes write $q(x,y)$ for $q(y-x)$. Let q_n be the nth iterate of q and let q_n^L be the corresponding kernel on the torus,

$$q_n^L(x,y) = \sum_{z\in\mathbf{Z}^2} q_n(x, y+Lz), \qquad x, y \in \Lambda(L). \tag{A.1}$$

Note that the symmetry of q implies $\rho_t^L(x,y) = q_{2[t]}^L(x,y)$ for $\nu = 1$, and by the right-hand side we need only be concerned with q_n^L when n is even.

For (a), we begin by observing that

$$q_{m+n}^L(x,0) = \sum_{y\in\Lambda(L)} q_m^L(x,y) q_n^L(y,0) \le \sup_{y\in\Lambda(L)} q_n^L(y,0). \tag{A.2}$$

Next, let $\hat{q}_n^L(k) = \sum_{x\in\Lambda(L)} e^{ik\cdot x} q_n^L(0,x)$ be the Fourier transform of q_n^L. The inversion formula tells us that

$$q_n^L(0,x) = \frac{1}{L^2} \sum_{k\in 2\pi\Lambda(L)/L} e^{-ik\cdot x} \hat{q}_n^L(k).$$

The symmetry of q implies that $\hat{q}_n^L(k)$ is real. When n is even the Fourier coefficients are positive, so the above implies that, when n is even,

$$q_n^L(0,x) \le q_n^L(0,0) \qquad \text{for all } x \in \Lambda(L). \tag{A.3}$$

To estimate q_n^L, we will apply a local central limit theorem for q_n from Bhattacharya and Rao (1976), in the form given in (2.10)–(2.11) of Cox (1989):

$$q_n(0,x) = \phi_n(x) + \psi_n(x). \tag{A.4}$$

Here $\phi_n(x) = (2\pi\sigma^2 n)^{-1}\exp(-|x|^2/2\sigma^2 n)$, and

$$\psi_n(x) = \phi_n(x)\sum_{r=1}^{2} n^{-r/2} B_r(x/\sqrt{n}) + e(x,n),$$

where each $B_r(x)$ is a polynomial (depending on q) of degree at most r, and

$$n \sum_{x\in\mathbf{Z}^2} |e(x,n)| \to 0 \qquad \text{as } n \to \infty.$$

By straightforward calculation, one can check that, as $L \to \infty$:

(i) $\varepsilon_L L^2 \phi_{[\varepsilon_L L^2]}(0) \to 1/2\pi\sigma^2$,
(ii) $\varepsilon_L L^2 \sum_{z\in\mathbf{Z}^2\setminus\{0\}} \phi_{[\varepsilon_L L^2]}(Lz) \to 0$,
(iii) $\varepsilon_L L^2 \sum_{z\in\mathbf{Z}^2} \psi_{[\varepsilon_L L^2]}(Lz) \to 0$

(recall $\varepsilon_L \to 0$). See pages 1342–1343 of Cox (1989) for similar computations. Consequently, (A.1) and (A.4) imply

$$\varepsilon_L L^2 q^L_{[\varepsilon_L L^2]}(0,0) \to \frac{1}{2\pi\sigma^2} \qquad \text{as } L \to \infty. \tag{A.5}$$

Combining (A.2), (A.3) and (A.5), we obtain, for some constant C,

$$\sup_{n\ge\varepsilon_L L^2, x\in\Lambda(L)} \varepsilon_L L^2 q^L_n(x,0) \le C, \tag{A.6}$$

from which the $\nu = 1$ case of (3.2) easily follows.

For (b), we start with the observation that

$$\begin{aligned} |q^L_{m+n}(x,0) - L^{-2}| &\le \sum_{y\in\Lambda(L)} q^L_m(x,y)|q^L_n(y,0) - L^{-2}| \\ &\le \sup_{y\in\Lambda(L)} |q^L_n(y,0) - L^{-2}|. \end{aligned} \tag{A.7}$$

Next, as in pages 1342–1343 of Cox (1989), one can check that, as $L \to \infty$, uniformly for $x \in \Lambda(L)$:

(i) $L^2 \sum_{y\in\mathbf{Z}^2} \phi_{[t_L L^2]}(x + Ly) \to 1$,
(ii) $L^2 \sum_{y\in\mathbf{Z}^2} \psi_{[t_L L^2]}(x + Ly) \to 0$

(recall that $t_L \to \infty$). Therefore, applying (A.1), (A.4), and (A.7), we have

$$\lim_{L\to\infty} \sup_{n\ge t_L L^2} \sup_{x\in\Lambda(L)} L^2|q^L_n(x,0) - L^{-2}| = 0 \tag{A.8}$$

from which the $\nu = 1$ version of (3.3) easily follows.

For (c), one first checks that, as $L \to \infty$, uniformly in $x \in \Lambda(L)$:

(i) $\sup_{u_L(1+|x|^2)\le k\le 3\varepsilon_L L^2} |(2\pi\sigma^2 k)\phi_k(x) - 1| \to 0$,
(ii) $\sup_{u_L(1+|x|^2)\le k\le 3\varepsilon_L L^2} k \sum_{z\in\mathbf{Z}^2\setminus\{0\}} \psi_k(x + zL) \to 0$

(recall $u_L \to \infty$). Combining these estimates with (A.1) and (A.4) gives

$$\sup_{\frac{1}{2}u_L(1+|x|^2)\le k\le 2\varepsilon_L L^2} |(2\pi\sigma^2 k)q_k^L(x,0) - 1| \to 0, \tag{A.9}$$

and the $\nu = 1$ case of (3.4) follows.

The fourth result (d) is closely related to (2.9) in Cox (1989). To prove it, we repeat the calculation on page 1344 of Cox (1989), replacing a_N^2 there by $(1 + |x|)^2$. The second term there does not go to 0 now as $L \to \infty$, but since $\limsup(1 + |x|^2)/L^2 \le 1/4$, that term is bounded. That is enough to prove that, for a finite constant C,

$$\sup_{n\ge 0, x\in\Lambda(L)} (1+x)^2 q_n^L(x,0) \le C, \tag{A.10}$$

from which the $\nu = 1$ case of (3.5) easily follows.

Extension to $\nu \in (0, 1)$. Let B be a binomial random variable with number of trials equal to $2[t/\nu]$ and success probability ν. It is easy to see that

$$\rho_t^L(0,x) = \sum_{k=0}^{2[t/\nu]} P(B = k) q_k^L(0,x). \tag{A.11}$$

To prepare for our proofs, let $Z = \xi_1 + \cdots + \xi_n$, where the ξ_i are iid Bernoulli random variables with success probability p. A simple calculation shows that

$$E(Z - EZ)^4 = \binom{n}{2}\binom{4}{2}\big(E(\xi_1 - p)^2\big)^2 + nE(\xi_1 - p)^4.$$

Since $E(\xi_1 - p)^2 = p(1 - p) \le p$ and $E(\xi_1 - p)^4 = p(1 - p)^4 + (1 - p)p^4 \le 2p$, it follows that, if $np \ge 1$, then

$$E(Z - EZ)^4 \le 5(np)^2 = 5(EZ)^2.$$

For B, setting $n = 2[t/\nu]$ and $p = \nu$ and using the last inequality with $r = 7/8$, we obtain

$$|EB|^{4r} P(|B - EB| \ge |EB|^r) \le E|B - EB|^4 \le 5|EB|^2$$

(when $EB \ge 1$). For $t \ge 3/2$, $2t \ge EB \ge 2t - 2 \ge 1$ and $2t - 2 \ge 2t - (2t)^{7/8}$, so it follows that

$$P\big(|B - 2t| \ge 2(2t)^{7/8}\big) \le \frac{5}{(2t - (2t)^{7/8})^{3/2}}. \tag{A.12}$$

To prove (a), we now take $t = \varepsilon_L L^2$ (recall $\varepsilon_L = 1/\sqrt{\log L}$) and note that, for large L, $2(2t)^{7/8} \le t$ and $2t - 2(2t)^{7/8} \ge \varepsilon_L L^2$. Thus, using (A.11) and (A.12), we have

$$\begin{aligned}\rho_{\varepsilon_L L^2}^L(x,0) &\le P\big(|B - 2t| \ge 2(2t)^{7/8}\big) + \sup_{k\ge\varepsilon_L L^2} q_k^L(x,0)\\ &\le \frac{5}{(\varepsilon_L L^2)^{3/2}} + \sup_{k\ge\varepsilon_L L^2,\, x\in\Lambda(L)} q_k^L(x,0).\end{aligned}$$

In view of (A.6), this shows that, for a finite constant C,

$$\sup_{x\in\Lambda(L)} \varepsilon_L L^2 \rho^L_{\varepsilon_L L^2}(x,0) \le C.$$

This is enough to prove (3.2), since, if $t < u$,

$$\sup_{x\in\Lambda(L)} \rho^L_u(x,0) \le \sup_{x\in\Lambda(L)} \rho^L_t(x,0).$$

To prove (b), we set $t = t_L L^2$ (recall $t_L \to \infty$) and note that, if L is large enough, then $2(2t)^{7/8} \le t$ and (A.12) implies that

$$\begin{aligned}|\rho^L_{t_L L^2}(0,x) - L^{-2}| &\le \sum_{k=0}^{2[t/\nu]} P(B=k)|q^L_k(x,0) - L^{-2}|\\ &\le \frac{5}{(t_L L^2)^{3/2}} + \sup_{k\ge t_L L^2} |q^L_k(0,x) - L^{-2}|.\end{aligned}$$

By (A.8), it follows that

$$\sup_{x\in\Lambda(L)} L^2|\rho_{t_L L^2}(x,0) - L^{-2}| \to 0$$

as $L\to\infty$. This is enough to prove (3.3), since, if $t < u$, $\sup_{x\in\Lambda(L)} |\rho^L_u(x,0) - L^{-2}| \le \sup_{x\in\Lambda(L)} |\rho^L_t(x,0) - L^{-2}|$.

The proof of (c) is similar to that of (a) and (b), but now requires the full strength of (A.12). Let L be large enough so that, for all $t \in [u_L(1+|x|^2), \varepsilon_L L^2]$, $2(2t)^{7/8} \le t$ (recall $u_L \to \infty$). For t in this interval, let $K(t)$ be the interval $[2t - 2(2t)^{7/8}, 2t + 2(2t)^{7/8}]$ and note that, if $k \in K(t)$, then $k \in [u_L(1+|x|^2), 3\varepsilon_L L^2]$ and $2t/k \to 1$ uniformly as $L\to\infty$. Consequently, using (A.11) and (A.12),

$$\begin{aligned}&|2\pi(2\sigma^2)t\rho^L_t(x,0) - 1|\\ &\quad\le \sum_{k=0}^{2[t/\nu]} P(B=k)|2\pi(2\sigma^2)tq^L_k(x,0) - 1|\\ &\quad\le \bigl(2\pi(2\sigma^2)t + 1\bigr)P\bigl(|B-2t| \ge 2(2t)^{7/8}\bigr) + \sup_{k\in K(t)} |2\pi(2\sigma^2)tq^L_k(x,0) - 1|\\ &\quad\le \frac{2\pi(2\sigma^2)t+1}{t^{3/2}} + \sup_{k\in[u_L(1+|x|^2), 3\varepsilon_L L^2]} \left|\frac{2t}{k}2\pi\sigma^2 k q^L_k(x,0) - 1\right| \to 0\end{aligned}$$

as $L\to\infty$, uniformly for $x\in\Lambda(L)$ and $t\in[u_L(1+|x|)^2, \varepsilon_L L^2]$.

The proof of (d) is the simplest:

$$(1+|x|^2)\rho^L_t(x,0) = \sum_{j=0}^{2[t/\nu]} P(B=j)(1+|x|^2)q^L_j(x,0) \le C,$$

the last inequality by (A.10). $\square$

REFERENCES

BHATTACHARYA, R. N. and RAO, R. R. (1976). *Normal Aproximation and Asymptotic Expansions*. Wiley, New York.

CLIFFORD, P. and SUDBURY, A. (1973). A model of spatial conflict. *Biometrika* **60** 581–588.

COX, J. T. (1989). Coalescing random walks and voter model consensus times on the torus in $\mathbf{Z}^d$. *Ann. Probab.* **17** 1333–1366.

COX, J. T. and GREVEN, A. (1991). On the long term behavior of finite particle systems: a critical dimension example. In *Random Walks, Brownian Motion and Interacting Particle Systems* (R. Durrett and H. Kesten, eds.) 203–213. Birkhäuser, Boston.

COX, J. T. and GRIFFEATH, D. (1986). Diffusive clustering in the two dimensional voter model. *Ann. Probab.* **14** 347–370.

CROW, J. F. and AOKI, K. (1984). Group selection for a polygenic trait: estimating the degree of population subdivision. *Proc. Nat. Acad. Sci. U.S.A.* **81** 6073–6077.

DIACONIS, P. and DURRETT, R. (2001). Chutes and ladders in Markov chains. *J. Theoret. Probab.* **14** 899–926.

DURRETT, R. (1996). *Probability: Theory and Examples*, 2nd ed. Duxbury, North Scituate, MA.

ENDLER, J. A. (1977). *Geographic Variation, Speciation, and Clines.* Princeton Univ. Press.

FLATTO, L., ODLYZKO, A. M. and WALES, D. B. (1985). Random shuttles and group representations. *Ann. Probab.* **13** 154–178.

HEY, J. (1991). A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. *Theoret. Population Biol.* **39** 30–48.

HOLLEY, R. A. and LIGGETT, T. M. (1975). Ergodic theorems for weakly interacting systems and the voter model. *Ann. Probab.* **3** 643–663.

HUDSON, R. R. (1990). Gene genaologies and the coalescent process. In *Oxford Surveys in Evolutionary Biology* (D. Futuyama and J. Antovics, eds.) **7** 1–44.

KIMURA, M. (1953). "Stepping stone" model of population. *Ann. Rep. Nat. Inst. Genet. Japan* **3** 62–63.

KIMURA, M. and MARUYAMA, T. (1971). Patterns of neutral polymorphism in a geographically structured population. *Genet. Res. Cambridge* **18** 125–131.

KIMURA, M. and WEISS, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49** 561–576.

MALÉCOT, G. (1967). Identical loci and relationship. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 317–332. Univ. California Press, Berkeley.

MALÉCOT, G. (1969). *The Mathematics of Heredity.* Freeman, San Francisco.

MALÉCOT, G. (1975). Heterozygosity and relationship in regularly subdivided populations. *Theoret. Population Biol.* **8** 212–241.

MARUYAMA, T. (1970). Effective number of alleles in a subdivided population. *Theoret. Population Biol.* **1** 273–306.

MARUYAMA, T. (1971). Analysis of population structure. II. Two dimensional stepping stone models of finite length and other geographically structured populations. *Ann. Hum. Genet. London* **35** 179–196.

MARUYAMA, T. (1972). The rate of decrease of genetic variability in a two dimensional continuous population of finite size. *Genetics* **70** 639–651.

NAGYLAKI, T. (1974). The decay of genetic variability in geographically structured populations. *Proc. Nat. Acad. Sci. U.S.A.* **71** 2932–2936.

NEI, M. (1975). *Molecular Population Genetics and Evolution.* North-Holland, Amsterdam.

NEI, M. and TAKAHATA, N. (1993). Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.* **37** 240–244.

SAWYER, S. (1976). Results for the stepping stone model in population genetics. *Ann. Probab.* **4** 699–728.

SAWYER, S. (1979) A limit theorem for patch sizes in a selectively neutral migration model. *J. Appl. Probab.* **16** 482–495.

SEIELSTAD, M. T., MINCH, E. and CAVALLI-SFORZA, L. L. (1998). Genetic evidence for a higher female mutation rate in humans. *Nat. Genet.* **20** 278–280.

SLATKIN, M. (1991). Inbreeding coefficients and coalescence times. *Genet. Res. Cambridge* **58** 167–175.

SLATKIN, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47** 264–279.

SLATKIN, M. and BARTON, N. H. (1989). A comparison of three indirect methods of estimating average levels of gene flow. *Evolution* **43** 1349–1368.

STROBECK, C. (1987). Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* **117** 149–153.

TAKAHATA, N. (1983). Gene identity and genetic differentiation of populations in the finite island model. *Genetics* **104** 497–512.

WATSON, G. N. (1966). *A Treatise on the Theory of Bessel Functions.* Cambridge Univ. Press. [This is a paperback version of his classic book (1922).]

WEISS, G. H. and KIMURA, M. (1965). A mathematical analysis of the stepping stone model of genetic correlation. *J. Appl. Probab.* **2** 129–149.

WILKINSON-HERBOTS, H. M. (1998). Genealogy and subpopulation differentiation under various models of population structure. *J. Math. Biol.* **37** 535–585.

WRIGHT, S. (1943). Isolation by distance. *Genetics* **28** 114–156.

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

DEPARTMENT OF MATHEMATICS
SYRACUSE UNIVERSITY
SYRACUSE, NEW YORK 13244
E-MAIL: jtcox@gumby.syr.edu

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
E-MAIL: rtd1@cornell.edu