

## TRAVELING WAVES OF SELECTIVE SWEEPS

BY RICK DURRETT\* AND JOHN MAYBERRY†

*Cornell University*

The goal of cancer genome sequencing projects is to determine the genetic alterations that cause common cancers. Many malignancies arise during the clonal expansion of a benign tumor which motivates the study of recurrent selective sweeps in an exponentially growing population. To better understand this process, Beerenwinkel et al. (2007) consider a Wright-Fisher model in which cells from an exponentially growing population accumulate advantageous mutations. Simulations show a traveling wave in which the time of the first  $k$ -fold mutant,  $\tau_k$ , is approximately linear in  $k$  and heuristics are used to obtain formulas for  $E\tau_k$ . Here, we consider the analogous problem for the Moran model and prove that as the mutation rate  $\mu \rightarrow 0$ ,  $\tau_k \sim c_k \log(1/\mu)$ , where the  $c_k$  can be computed explicitly.

**1. Introduction.** Recent studies have sought to identify the mutations that give rise to common cancers by sequencing protein-coding genes in common tumor types including: breast and colon cancer ([23],[26]), pancreatic cancer ([14]), and glioblastoma ([19], [24]). The last study is part of a one hundred million dollar pilot project of the NIH, which could lead to a 1.5 billion dollar effort. These studies have rediscovered genes known to play a role in cancer (e.g., APC, KRAS and TP53 in colon cancer), but they have also found that tumors contain a large number of mutations. Analysis of 13,023 genes in 11 breast and 11 colorectal cancers in Sjoblom et al. [23] revealed that individual tumors accumulate an average of  $\approx 90$  mutated genes but only a subset of these contribute to the development of cancer.

Follow up work in Wood et al. [26] studied 18,191 distinct genes in the same 22 samples. Any gene that was mutated in a tumor but not normal tissue was analyzed in 24 additional tumors, and selected genes were further analyzed in 96 colorectal cancers. Statistical analysis suggested that most of the  $\approx 80$  mutations in an individual tumor were harmless and that  $< 15$  were likely to be responsible for driving the initiation, progression, or maintenance of the tumor. These two types of mutations are commonly referred to as “drivers” and “passengers”. The latter provide no selective advantage to the growing cancer mass, but are retained by chance during repeated rounds of cell division and clonal expansion (exponential growth).

The results of [23] and [26] contrast with the long held belief that most cancers are the end result of a handful of mutations. Armitage and Doll [1] did log-log plots of cancer mortality versus age and found slopes of 5.18 and 4.97 for colon cancer in men and women. From this they predicted that the occurrence of colon cancer was the result of a six stage process. In

---

\*Partially supported by NSF grant DMS 0704996 from the probability program.

†Partially supported by NSF RTG grant DMS 0739164.

AMS 2000 subject classifications: Primary 60J85, 92D25; Secondary 92C50

Keywords and phrases: cancer models, waiting times, Moran model, selective sweep, branching process

essence their argument is that the density function of the sum of six exponentials with rates  $\mu_i$  is

$$\approx \mu_1 \cdots \mu_6 t^5 / 5! \quad \text{for small } t.$$

This result yields the density of the well-known gamma distribution when all the  $\mu_i$  are equal, but only readers with well developed skills in calculus (or complex variables) will succeed in deriving this result for unequal  $\mu_i$  on their own.

[1] and the subsequent work of Knudson [15], who used statistics to argue that retinoblastoma was the end result of two mutations, sparked a large amount of work, see Knudson [16] and the books by Wodarz and Komarova [25] and Frank [11] for surveys. From this large body of work on multi-stage carcinogenesis, we will only cite two sources. Luebeck and Moolgavakar [17] used multistage models to fit the age-specific incidence of colorectal cancers in the SEER registry, which covers 10% of the U.S. population to conclude that a four stage model gave the best fit. Calabrese et al. [5] used data for 1,022 colorectal cancers to argue that “sporadic” cancers developed after six mutations, but in the subgroup of individuals with strong familial predispositions, only five mutations were required.

There is good reason to doubt some of the conclusions of [23] and [26]. First of all, the statistical methods of [23] have been criticized (see letters on pages 762–763 in the February 9, 2007 issue of *Science*). Furthermore, in [26], a follow-up study on 40 of the 119 highest scoring genes, chosen because they were in pathways of biological interest, showed that 15 of the 40 genes (37.5%) were not mutated in any of the 96 tumors, casting doubt on the claimed 10% false discovery rate. However, the more recent studies ([14], [19], [24]) using well-known and trusted statistical methods have found similar patterns: an average of 63 mutations in pancreatic cancers and 47 in glioblastoma.

To better understand this process by which an exponentially growing cell mass accumulates driver and passenger mutations, and in particular to understand the data in [23], Beerenwinkel et al. [3] considered a Wright-Fisher model with selection and mutation in an exponentially growing population. They assumed that there were 100 potential driver genes and asked for the waiting time until one cell has accumulated  $k$  mutations. Simulations, see their Figure 3, showed that a traveling wave developed in which the time until the first  $k$ -fold mutant was approximately linear in  $k$  and they used heuristic arguments to obtain quantitative predictions for the first time that a cell with  $k$  mutations appears.

Here we will consider this problem for the analogous Moran model, and prove asymptotic results as the mutation rate  $\mu \rightarrow 0$  for the behavior of  $X_k^\mu(t)$  = the number of cells with  $k$  mutations at time  $t$ . A cell with  $k$  mutations will be referred to as a type  $k$  individual. Our main result is Theorem 2 which allows for an exponentially growing population  $N^\mu(t)$  of individuals. The process of fixation of advantageous mutations in a population of constant size has been the subject of much theoretical work (see e.g., Chapter 6 of Durrett [7]), so it is natural to ask how the behavior changes in an exponentially growing population. A second difference from the standard theory of the fixation of a single mutation is that we consider a situation in which new mutations arise before older ones have gone to fixation, a process often referred to as “stochastic tunneling”. The resulting “Hill-Robertson” interference (see e.g., Section 8.3 in [7]) can be analyzed here because only the newest mutation is stochastic while the older mutations behave deterministically. This idea was used by Rouzine et al. in [21], (and later developed in more detail in [4], [20]) as a heuristic, but here it leads to rigorous results.

The rest of the paper is organized as follows. In Section 1.1, we begin with a fixed population size of  $N = \mu^{-\alpha}$  individuals and state Theorem 1 which says that when time is scaled by  $L = \log(1/\mu)$ , the log sizes of  $X_k$ , divided by  $L$ , converge to a limit that is deterministic and piecewise linear and hence, the time the first type  $k$  individual appears is  $O(\log(1/\mu))$ . Since we have assumed the population size is  $\mu^{-\alpha}$ , this time scale agrees with results in Yu et al. [27], [28] which show that the rate of adaptation (defined as the change in the mean fitness of the population) for a related fixed population size Moran model is  $O(\log N)$  and simulations in Desai and Fisher [6] which suggest that the speed of adaptation depends logarithmically on both the mutation rate and the population size. Sections 1.2-1.4 contain examples elucidating the nature of the limit in Theorem 1 and illustrating the traveling wave like behavior of the limit. In Section 1.5, we return to the growing population scenario and state our main result, Theorem 2, which generalizes Theorem 1. Section 2 contains statements of the main tools used to prove Theorem 2 and Sections 3-4 contain the technical details.

1.1. *Fixed Population Size: Main Result.* We begin by considering our Moran model in a fixed population of  $N$  individuals and return to our analysis of the exponentially growing population in Section 1.5. We assume that

- (i) Initially, all individuals are of type 0.
- (ii) Type  $k$  individuals mutate to individuals of type  $k + 1$  at rate  $\mu$ .
- (iii) All individuals die at rate 1 and upon death, are replaced by an individual of type  $k$  with probability

$$\frac{(1 + \gamma)^k X_k^\mu(t)}{W^\mu(t)}$$

where  $(1 + \gamma)^k$  is the relative fitness of type  $k$  individuals compared to type 0, and

$$W^\mu(t) = \sum_{k=0}^{\infty} (1 + \gamma)^k X_k^\mu(t)$$

is the “total fitness” of the population. We assume throughout that  $\gamma > 0$  is fixed (i.e. mutations are advantageous). Approximations of the time the first type  $k$  individual appears have been carried out for the neutral case ( $\gamma = 0$ ) in Iwasa et al. [13], Haeno et al. [12], Durrett et al. [10], and Schweinsberg [22] (and applied to regulatory sequence evolution in Durrett and Schmidt [9]). The case  $\gamma < 0$  is of interest in studying Muller’s Ratchet (Muller [18]), but since deleterious mutation behave much differently from advantageous mutations, we will not consider this case here.

We will suppose throughout that  $N \gg 1/\mu$ , i.e.,  $N\mu \rightarrow \infty$ . If  $N\mu \rightarrow 0$  then the 1’s arise and go almost to fixation before the first mutation to a 2 occurs, so the times between fixations are independent exponentials. We will not here consider the borderline scenario. Let  $T_0^\mu = 0$  and for  $k \geq 1$  define

$$T_k^\mu = \inf\{t \geq 0 : X_k^\mu(t) \geq 1\}$$

$$\tau_k^\mu = T_k^\mu - T_{k-1}^\mu.$$

$T_k^\mu$  is the time of the first appearance of a type  $k$  individual. In order to study the hitting times  $T_k^\mu$  we will prove a limit theorem for the sizes of the  $X_k^\mu(t)$  on a log scale. Let  $\log^+ x = \max\{\log x, 0\}$ ,  $L = \log(1/\mu)$ , and define

$$\gamma_j = (1 + \gamma)^j - 1$$

for all  $j \in \mathbb{Z}$ .

**THEOREM 1.** *Suppose that  $X_0^\mu(0) = N$  and  $N = \mu^{-\alpha}$  for some  $\alpha > 1$ . Then as  $\mu \rightarrow 0$*

$$Y_j^\mu(t) \equiv \frac{1}{L} \log^+(X_j(Lt/\gamma)) \rightarrow y_j(t) \quad \text{in probability}$$

*uniformly on compact subsets of  $(0, \infty)$  and there exist (explicitly calculable) constants  $\beta_k = \beta_k(\gamma, \alpha)$  so that*

$$\frac{\tau_k}{L/\gamma} \rightarrow \beta_k \quad \text{in probability}$$

*as  $\mu \rightarrow 0$  for all  $k \geq 1$ . The limit  $y_j(t)$  is deterministic and piecewise linear and will be described by (a) and (b) below.*

(a) *Initial Behavior.*  $y_j(0) = (\alpha - j)^+$ . The convergence only occurs on  $(0, \infty)$  because we have  $Y_j^\mu(0) = 0$  for all  $j \geq 1$  by assumption, so a discontinuity is created at time 0.

(b) *Inductive step.* Suppose that we have computed the  $y_j(t)$  at times  $t \leq s_n$ , where  $s_0 = 0$ . Let

$$m = m_n = \max\{j : y_j(s_n) = \alpha\}$$

Since the  $X_j^\mu(Ls_n/\gamma)$  sum to  $N = \mu^{-\alpha}$  there is at least one such value. Suppose that

(i) there exists  $k = k_n \geq 0$  such that  $y_j(s_n) = 0$  for  $j > k$ ,  $y_j(s_n) > 0$  for  $m < j \leq k$ ,

and (ii)  $y_{j+1}(s_n) \geq y_j(s_n) - 1$  for  $0 \leq j \leq k$ , so that in particular,  $y_k(s_n) \leq 1$ .

Let  $k^* = k$  if  $y_k(s_n) < 1$ ,  $k^* = k + 1$  if  $y_k(s_n) = 1$ , and let

$$\delta_{n,j} = \begin{cases} (\alpha - y_j(s_n))\gamma/\gamma_{j-m} & m < j < k^* \\ (1 - y_{k^*}(s_n))\gamma/\gamma_{k^*-m} & j = k^* \end{cases}$$

Then if  $\Delta_n = \min\{\delta_{n,j} : m < j \leq k^*\}$  we have for  $t \leq \Delta_n$

$$y_j(s_n + t) = \begin{cases} (y_j(s_n) + t\gamma_{j-m}/\gamma)^+ & j \leq k^* \\ 0 & j > k^* \end{cases}.$$

and we define  $s_{n+1} = s_n + \Delta_n$ .

Our description of the limiting dynamical system can be understood as follows. If type  $m$  is the most fit of the dominant types in the population at time  $s_n$ , then the  $y_j(s_n + t)$ ,  $m \leq j \leq k^*$  grow linearly with slope  $\gamma_{j-m}/\gamma \geq 0$  while the  $y_j(s_n + t)$ ,  $j < m$  decrease

linearly with slope  $\gamma_{j-m}/\gamma < 0$  until they hit zero. These rates remain valid until either  $y_j$  reaches level  $\alpha$  for some  $m < j < k^*$  and there is a change in the most fit dominant type or  $y_{k^*}$  reaches level 1 and individuals of type  $k^* + 1$  are born. These two events correspond to  $\Delta_n = \delta_{n,j}$  and  $\Delta_n = \delta_{n,k^*}$ , respectively. The condition  $y_{j+1}(s_n) \geq y_j(s_n) - 1$  guarantees that after birth, the growth of type  $j$  individuals is driven by selection and not by mutations from type  $j - 1$  individuals. If this condition failed, we would encounter a discontinuity in the limiting dynamics like the one at time 0.

(a) and (b) together describe the limiting dynamical system for all time since by part (a), the assumptions of (b) hold at time  $s_0 = 0$  and it is easy to see from the form of  $y_j(t)$  that if the assumptions hold at time  $s_n$  for  $n \geq 0$ , then they also hold at time  $s_{n+1} = s_n + \Delta_n$ . We have re-scaled time by  $\gamma^{-1}$  since in most cases of interest  $\gamma$  is small, e.g.,  $\gamma < 0.01$  and when  $\gamma$  is small, we have  $\gamma_j/\gamma \approx j$  so that the limit process described above is almost independent of  $\gamma$ .

Note that the form of the limit implies that the birth of type  $k+1$ 's occurs when  $y_k(t) = 1$ , i.e., the  $k$ 's have approximate size  $1/\mu$ , and hence if we define  $t_j$ ,  $j \leq k$ , as the first times satisfying  $y_j(t_{j+1}) \geq 1$ , then  $\beta_k$ ,  $k \geq 0$ , can be calculated via the relationship

$$\sum_{j=1}^{k+1} \beta_j = t_{k+1}.$$

Theorem 1 is very general but not very transparent, so our next task is to give some examples in which more explicit expressions for the  $\beta_k$  are available. Figure 1 shows examples in the first three ‘‘regimes’’ of behavior that we consider. In the  $j$ th regime type  $k + j$  arises (but not type  $k + j + 1$ ) before type  $k$  ‘‘fixates’’, i.e. is of order  $N = \mu^{-\alpha}$ . These regimes closely correspond to the different scenarios considered in Brunet et al. [4] in which the ‘‘stochastic edge’’, i.e. the class of the most fit mutant, is always assumed to be  $q$  fitness classes ahead of the population bulk.  $q$  is referred to as the ‘‘lead’’. In the notation of Theorem 1, the lead is always  $k_n^*$  on the interval  $[s_n, s_{n+1}]$  and in regime  $j$ , the lead is always  $j$ . In all three regimes, we see the traveling wave like behavior observed in the simulations of Beerenwinkel et al. [3] (see also Rouzine et al. [21]), but the wave speed is only constant in the first regime. We now look at this first regime in more detail.

**1.2. Results for regime 1.** Let  $r_2 = 1 + \gamma/\gamma_2$ . The first regime occurs for  $1 < \alpha < r_2$ . If  $\gamma$  is small,  $\gamma/\gamma_2 \approx 1/2$  and the condition is roughly  $\alpha \in (1, 3/2)$ . If  $\gamma > 0$ , then  $\gamma/\gamma_2 = 1/(2 + \gamma) < 1/2$  so  $\alpha < 2$  throughout regime 1.

time	time increment	type 0	type 1	type 2	type 3
0+		$\alpha$	$\alpha - 1$		
$s_1$	$\Delta_0 = 2 - \alpha$	$\alpha$	1	0	
$s_2$	$\Delta_1 = \alpha - 1$	$\alpha$	$\alpha$	$\gamma_2(\alpha - 1)/\gamma$	
$s_3$	$\Delta_2 = 1 - \Delta_1\gamma_2/\gamma$		$\alpha$	1	0
$s_4$	$\Delta_3 = \alpha - 1$		$\alpha$	$\alpha$	$\gamma_2(\alpha - 1)/\gamma$

TABLE 1

*Sizes in Regime 1. Times are given in units of  $L/\gamma$ , entries are the size given as a power of  $1/\mu$ , and 0 indicates when the first of the type is born. The first row comes from (a), the next four from applications of (b).*

Table 1 summarizes the situation. To explain the entries, we note that applying part (a) of the limit description implies that  $y_1(0) = \alpha - 1$  and then part (b) implies that

$$y_1(s) = (\alpha - 1) + s$$

for  $s \leq \Delta_0 = 2 - \alpha$ . Since we have assumed that  $\alpha < r_2$ , we have

$$\Delta_1 = \delta_{1,1} \wedge \delta_{1,2} = (\alpha - 1) \wedge \frac{\gamma_2}{\gamma} = \alpha - 1$$

and applying part (b) tells us that we have  $y_2(s_1 + t) = t\gamma_2/\gamma$  for all  $t \leq \Delta_1$ . Another application of (b) then yields  $\Delta_2 = \delta_{2,2} = 1 - y_2(s_1)$  which gives the additional amount of time needed for  $y_2$  to hit 1. Since the relative sizes of 1's, 2's, and 3's at time  $s_3$  are the same as the relative sizes of 0's, 1's, and 2's at time  $s_1$ , we obtain the following result giving the limiting coefficients of  $\tau_k^\mu$ .

**COROLLARY 1.** *Suppose that  $N = \mu^{-\alpha}$  for some  $\alpha \in (1, r_2)$ . Then as  $\mu \rightarrow 0$*

$$\frac{\tau_1^\mu}{L/\gamma} \rightarrow (2 - \alpha) \quad \text{and for all } k \geq 2 \quad \frac{\tau_k^\mu}{L/\gamma} \rightarrow \beta \quad \text{in probability}$$

$$\text{where } \beta \equiv \Delta_1 + \Delta_2 = (\alpha - 1) + 1 - (\alpha - 1)\frac{\gamma_2}{\gamma} = (2 + \gamma) - (1 + \gamma)\alpha.$$

Figure 1 illustrates the limiting dynamical system in the case when  $\gamma = 0.01$  and  $\alpha = 1.3$ . We can see that in regime 1, the system is characterized by a “traveling wave of selective sweeps” in type space, i.e., the growth and decay of types  $k \geq 2$  occur translated in time by a fixed amount. In Figure 2, we show the distributions of types at the times when type 5, 9, 13, and 17 individuals are born (labeled as  $t_5, t_9, t_{13}$ , and  $t_{17}$ ). As we move from time  $t_k$  to  $t_{k+4}$ , the distribution is shifted by a constant amount.

**1.3. Results for regime 2.** Regime 2 occurs for  $r_2 < \alpha < r_3$  with  $r_3 = r_2 + \gamma/\gamma_3$ . When  $\gamma$  is small  $\gamma/\gamma_3 \approx 1/3$  so this regime is roughly  $\alpha \in (3/2, 11/6)$ . In general,  $r_3 < 11/6$  so we have  $\alpha < 2$  throughout this regime. As in the previous section, it is easiest to explain the conclusions of Theorem 1 with a table, see Table 2.

time	time increment	type 1	type 2	type 3
0+		$\alpha - 1$		
$s_1$	$\Delta_0 = 2 - \alpha$	1	0	
$s_2$	$\Delta_1 = \gamma/\gamma_2$	$r_2$	1	0
$s_3$	$\Delta_2 = \alpha - r_2$	$\alpha$	$1 + \Delta_2\gamma_2/\gamma$	$\Delta_2\gamma_3/\gamma$
$s_4$	$\Delta_3 = \frac{\gamma}{\gamma_2} \left(1 - \Delta_2 \frac{\gamma_3}{\gamma}\right)$		$1 + \Delta_2\gamma_2/\gamma + \Delta_3$	1

TABLE 2  
 $\log_{1/\mu}$  sizes in Regime 2. Time in units of  $L/\gamma$ .

Since  $\alpha < 2$  the first two rows are the same as in regime 1, and we again have

$$\tau_1^\mu \sim (2 - \alpha)L/\gamma.$$

However, we now have  $\alpha > r_2$  so that

$$\Delta_1 = \delta_{1,1} \wedge \delta_{1,2} = (\alpha - 1) \wedge (\gamma/\gamma_2) = \gamma/\gamma_2$$

and hence the 2's reach level  $1/\mu$  before the 1's fixate. This yields

$$\tau_2^\mu \sim \frac{\gamma}{\gamma_2} \cdot \frac{L}{\gamma}.$$

Now  $y_1(s_2) = 1 + \gamma/\gamma_2 = r_2$  so the additional time it takes  $y_1$  to reach level  $\alpha$  is  $\delta_{2,1} = \alpha - r_2$ . Since  $\alpha < r_3$ , we have  $(\alpha - r_2)\gamma_3/\gamma < 1$  and hence  $\delta_{2,1} < \delta_{2,3}$ , i.e. the 1's will fixate before the 3's reach level  $1/\mu$ . To show that the 1's fixate before the 2's and conclude that  $\Delta_2 = \delta_{2,1} = \alpha - r_2$ , we need to show that  $(\alpha - r_2) < (\alpha - 1)\gamma/\gamma_2$  which holds if and only if

$$(1.1) \quad \alpha < \frac{2 + \gamma}{1 + \gamma}.$$

But comparing  $r_3 = 1 + \gamma/\gamma_2 + \gamma/\gamma_3$  with the upper bound in (1.1), we can see that

$$\begin{aligned} 1 + \gamma/\gamma_2 + \gamma/\gamma_3 < \frac{2 + \gamma}{1 + \gamma} &\iff \frac{(3 + 3\gamma + \gamma^2) + (2 + \gamma)}{(2 + \gamma)(3 + 3\gamma + \gamma^2)} < \frac{1}{1 + \gamma} \\ &\iff \frac{5 + 9\gamma + 5\gamma^2 + \gamma^3}{6 + 9\gamma + 5\gamma^2 + \gamma^3} < 1. \end{aligned}$$

The last inequality is always true and therefore (1.1) holds throughout regime 2 and  $\Delta_2 = \alpha - r_2$ , justifying the fourth line in Table 2. Finally, to check that the 2's have not yet fixated when the 3's reach level  $1/\mu$  and prove

$$\Delta_3 = \delta_{3,3} = \frac{\gamma}{\gamma_2} \left( 1 - \Delta_2 \frac{\gamma_3}{\gamma} \right),$$

we note that the size of  $y_2(s_3 + \delta_{3,3})$  is

$$\begin{aligned} 1 + \Delta_2 \gamma_2/\gamma + \delta_{3,3} &= 1 + \frac{\gamma_2}{\gamma}(\alpha - r_2) + \frac{\gamma}{\gamma_2} - \frac{\gamma_3}{\gamma_2}(\alpha - r_2) \\ &= 1 + \gamma/\gamma_2 + \ell(\alpha - r_2) \end{aligned}$$

with

$$\ell \equiv \gamma_2/\gamma - \gamma_3/\gamma_2 = \frac{(2 + \gamma)^2 - (3 + 3\gamma + \gamma^2)}{2 + \gamma} = \frac{1 + \gamma}{2 + \gamma} \in (0, 1)$$

and hence  $y_2(s_3 + \delta_{3,3}) \in (r_2, \alpha)$ . This justifies the final line of Table 2 and we conclude that

$$\frac{\tau_3^\mu}{L/\gamma} \rightarrow \Delta_2 + \Delta_3 = \alpha - r_2 + \frac{\gamma(1 - \gamma_3(\alpha - r_2)/\gamma)}{\gamma_2}.$$

In contrast to regime 1, the relative sizes of types when the 3's reach  $1/\mu$  are not exactly the same as the relative sizes when the 2's reach level  $1/\mu$ . To describe this more complicated situation, suppose that type  $k - 2$  individuals have size  $(1/\mu)^x$  at the time type  $k - 1$  individuals reach level  $1/\mu$ . Then if we assume

(2a) type  $k - 2$  reaches fixation before type  $k - 1$ ,

(2b) type  $k - 2$  reaches fixation before  $k$ 's reach  $1/\mu$ ,

(2c) type  $k$  reaches level  $1/\mu$  before type  $k - 1$  reaches fixation,

we can repeat the arithmetic leading to Table 2 to yield Table 3

time increment	type $k - 3$	type $k - 2$	type $k - 1$	type $k$
	$\alpha$	$x$	1	0
$t_k^1 = \alpha - x$	$\alpha$	$\alpha$	$1 + \gamma_2 t_k^1 / \gamma$	$\gamma_3 t_k^1 / \gamma$
$t_k^2 = \frac{\gamma}{\gamma_2} (1 - \gamma_3 (\alpha - x) / \gamma)$		$\alpha$	$f(x)$	1

TABLE 3  
Iteration in Regime 2

where here  $f(x) = 1 + \gamma_2 t_k^1 / \gamma + t_k^2 = r_2 + \ell(\alpha - x)$  with  $\ell = (1 + \gamma) / (2 + \gamma)$  as before.

Since the density of 2's is  $f(r_2)$  when the 3's have reached size  $1/\mu$ , we see that when type  $k \geq 3$  reaches size  $1/\mu$  the density of type  $k - 1$  is  $f^{k-2}(r_2)$ . This leads to the statement of our next result.

COROLLARY 2. *Suppose  $N = \mu^{-\alpha}$  for some  $\alpha \in (r_2, r_3)$ . Then as  $\mu \rightarrow 0$*

$$\frac{T_1^\mu}{L/\gamma} \rightarrow (2 - \alpha) \quad \text{and for all } k \geq 2 \quad \frac{T_k^\mu}{L/\gamma} \rightarrow \beta_k \quad \text{in probability}$$

where  $\beta_2 = \gamma/\gamma_2$  and if we let  $f^0(x) = x$  then for all  $k \geq 3$ , we have

$$\beta_k = t_k^1 + t_k^2 = (\alpha - f^{k-3}(r_2)) + \frac{1 - (3 + 3\gamma + \gamma^2)(\alpha - f^{k-3}(r_2))}{2 + \gamma}.$$

Furthermore, the coefficients  $\beta_k \rightarrow \beta_\infty$  as  $k \rightarrow \infty$  where

$$\beta_\infty = \alpha - r^* + \frac{1 - (3 + 3\gamma + \gamma^2)(\alpha - r^*)}{2 + \gamma}$$

with  $r^* = \lim_{k \rightarrow \infty} f^k(r_2) = (r_2 + \ell\alpha) / (1 + \ell)$ .

PROOF. We need to show that conditions (2a), (2b), and (2c) above are satisfied for any  $k \geq 0$  and that  $f^k(r_2)$  converges. The latter follows from the fact that  $f$  has slope  $-\ell$ , with  $\ell \in (0, 1)$ , so as  $k \rightarrow \infty$

$$f^k(r_2) \rightarrow r^* = \frac{r_2 + \ell\alpha}{1 + \ell},$$

the unique fixed point of  $f$ . It is easy to see that  $\ell \in (0, 1)$  implies that

$$(1.2) \quad r_2 \leq f^k(r_2) < \alpha$$



for all  $k \geq 0$  and (2c) immediately follows. Since  $\alpha < r_3$ , we have  $\gamma_3(\alpha - r_2)/\gamma < 1$ , which, along with (1.2), tells us that (2b) holds for all  $k \geq 0$  as well. Finally, (2a) is equivalent to

$$\frac{\alpha - 1}{\gamma_2} > \frac{\alpha - f^{k-3}(r_2)}{\gamma}$$

and so (1.2) implies that to prove (2a), we need only show that

$$\frac{\alpha - 1}{\gamma_2} > \frac{\alpha - r_2}{\gamma}.$$

Rearranging terms, we obtain the equivalent condition  $\alpha < (2 + \gamma)/(1 + \gamma)$  which holds by (1.1), completing the proof.  $\square$

Again the behavior of the limits  $y_j(t)$  can be read off from Tables 2 and 3. Formulas are messy but it is easy to compute  $y_j(t)$  for a fixed value of  $\alpha$ . As Figure 1 shows, after a short transient phase, the increments between the appearance of successive types settle down into the steady state behavior guaranteed by Corollary 2. Figure 2 shows the distribution of types at various times throughout the evolution of the system, which agree with simulations given in Figure 1 in the Appendix of Beerenwinkel et al. [3].

1.4. *Regime 3.* Regime 3 occurs for  $\alpha \in (r_3, r_4)$  with  $r_4 = r_3 + \gamma/\gamma_4$ . When  $\gamma$  is small  $\gamma/\gamma_4 \approx 1/4$ , so this regime is roughly  $\alpha \in (11/6, 25/12)$ . If  $\alpha < 2$  then the initial phases are similar to Regime 2, but now type 3 reaches  $1/\mu$  before the 1's fixate.

time	time increment	type 1	type 2	type 3
0+		$\alpha - 1$		
$s_1$	$\Delta_0 = 2 - \alpha$	1	0	
$s_2$	$\Delta_1 = \gamma/\gamma_2$	$r_2$	1	0
$s_3$	$\Delta_2 = \gamma/\gamma_3$	$r_3$	$1 + \gamma_2/\gamma_3$	1

TABLE 4

$\log_{1/\mu}$  sizes in Regime 3. Time in units of  $L/\gamma$ .

Now if we assume that

(3a) type  $k - 3$  reaches fixation before types  $k - 2$  and  $k - 1$ ,

(3b) type  $k - 3$  reaches fixation before type  $k$ 's reach  $1/\mu$ ,

and (3c) type  $k$  reaches level  $1/\mu$  before types  $k - 2$  and  $k - 1$  reaches fixation,

then the recursion in Table 3 becomes a pair of equations (see Table 5). To imitate the proof in regime 2 we would have to show that (3a), (3b), (3c) hold for  $x = r_3$  and  $y = 1 + \gamma_2/\gamma$ , and for all of the iterates  $f^k(x, y)$  where  $f \equiv (f_1, f_2)$ . Figure 3 shows that this is true when  $\alpha = 1.95$  and  $\gamma = 0.01$ , however, verifying this algebraically is difficult because  $f(x, y)$  may fail to satisfy the conditions when  $(x, y)$  does. It is also pointless, since our inductive procedure allows us to easily compute the limit and hence, we abandon this algebraic drudgery and move on to a discussion of the growing population model.

time increment	type $k-4$	type $k-3$	type $k-2$	type $k-1$	type $k$
	$\alpha$	$x$	$y$	$1$	
$t_k^1 = \alpha - x$	$\alpha$	$\alpha$	$y + t_k^1 \gamma_2 / \gamma$	$1 + t_k^1 \gamma_3 / \gamma$	$t_k^1 \gamma_4 / \gamma$
$t_k^2 = \frac{\gamma}{\gamma_3} (1 - t_k^1 \gamma_4 / \gamma)$		$\alpha$	$f_1(x, y)$	$f_2(x, y)$	$1.$

TABLE 5  
Iteration in Regime 3

1.5. *Growing Population.* We now consider a growing population of individuals  $N^\mu(t)$ ,  $t \geq 0$  with a random initial population in  $\mathbb{N} = \{1, 2, \dots\}$  distributed according to some measure  $\nu_0$ . At time 0, all individuals are of type 0 and we suppose that in addition to the previously imposed Moran dynamics, at rate  $\rho N^\mu(t)$ ,  $\rho \geq 0$ , new individuals are added and their type is chosen to be  $k$  with probability

$$\frac{(1 + \gamma)^k X_k^\mu(t)}{W^\mu(t)}.$$

As in the case of fixed population size, we are able to derive a limiting, piecewise linear approximation to

$$Y_k^\mu(t) \equiv (1/L) \log^+ X_k^\mu(Lt/\gamma).$$

To determine the correct growth rates, suppose that there are  $x_j$  individuals of type  $j$  and the population size is  $N$ . Then we have the jump rates

$$\begin{aligned} x_j &\mapsto x_j + 1 & \text{rate:} & \quad [(1 + \rho)N - x_j] \frac{(1 + \gamma)^j x_j}{\sum_{i \geq 0} (1 + \gamma)^i x_i} + \mu x_{j-1} \\ x_j &\mapsto x_j - 1 & \text{rate:} & \quad x_j \frac{\sum_{i \neq j} (1 + \gamma)^i x_i}{\sum_{i \geq 0} (1 + \gamma)^i x_i} - \mu x_j \end{aligned}$$

If mutations can be ignored, then the growth rate of type  $j$ 's is

$$\frac{\sum_{i \geq 0} [(1 + \rho)(1 + \gamma)^j - (1 + \gamma)^i] x_i x_j}{\sum_{i \geq 0} (1 + \gamma)^i x_i} \approx [(1 + \rho)(1 + \gamma)^{j-m} - 1] x_j$$

if  $x_i = o(N)$  for  $i \neq m$ . This yields the expression  $\lambda_{j-m} \equiv (1 + \rho)(1 + \gamma)^{j-m} - 1$  for the limiting growth rate of type  $j$ 's in a population dominated by type  $m$ .

If type  $j$  individuals have size  $(1/\mu)^x$  at time 0, are growing at rate  $\lambda_{k(j)}$  for some  $k(j) \geq 1$ , and the initial total population size is  $(1/\mu)^z$ , then type  $j$ 's will achieve fixation at the approximate time  $t$  satisfying

$$(1/\mu)^x e^{\lambda_{k(j)} t} = (1/\mu)^z e^{\rho t} \quad \text{or} \quad t = \frac{z - x}{\lambda_{k(j)} - \rho} \log(1/\mu)$$

This leads to the following result. Theorem 1 is the special case  $\rho = 0$ .

**THEOREM 2.** *Let  $F^\mu(t) = (1/L) \log N^\mu(tL/\gamma)$  and suppose that  $F^\mu(0) \rightarrow \alpha$  in probability for some  $\alpha > 0$ . Then  $F^\mu(t) \rightarrow \alpha + t\rho/\gamma$  and  $Y_j^\mu(t) \rightarrow y_j(t)$  in probability uniformly on compact subsets of  $[0, \infty)$  and  $(0, \infty)$  respectively and there exist (calculable) constants  $\beta_k = \beta_k(\rho, \gamma, \alpha)$  so that*

$$\frac{\tau_k}{L/\gamma} \rightarrow \beta_k \quad \text{in probability}$$

as  $\mu \rightarrow 0$  for all  $k \geq 1$ . The limits  $y_j(t)$  are deterministic and piecewise linear and described by (a) and (b) below.

(a) *Initial Behavior.*  $y_j(0) = (\alpha - j)^+$ .

(b) *Inductive step.* Suppose that we have computed the  $y_j(t)$  at times  $t \leq s_n$ , where  $s_0 = 0$ . Let

$$\alpha_n = \alpha + \rho s_n \quad m = m_n = \max\{j : y_j(s_n) = \alpha_n\}$$

Since the  $X_j^\mu(Ls_n/\gamma)$  sum to  $N^\mu(Ls_n/\gamma)$  there is at least one such value. Suppose that

(i) there exists  $k = k_n > 0$  such that  $y_j(s_n) = 0$  for  $j > k$ ,  $y_j(s_n) > 0$  for  $m < j \leq k$ ,

and (ii)  $y_{j+1}(s_n) \geq y_j(s_n) - 1$  for  $0 \leq j \leq k$  so that in particular,  $y_k(s_n) \leq 1$ .

Let  $k^* = k$  if  $y_k(s_n) < 1$ ,  $k^* = k + 1$  if  $y_k(s_n) = 1$ , and let

$$\delta_{n,j} = \begin{cases} (\alpha_n - y_j(s_n))\gamma/(\lambda_{j-m} - \rho) & m < j < k^* \\ (1 - y_{k^*}(s_n))\gamma/\lambda_{k^*-m} & j = k^* \end{cases}$$

Then if  $\Delta_n = \min\{\delta_{n,j} : m < j \leq k^*\}$  we have for  $t \leq \Delta_n$

$$y_j(s_n + t) = \begin{cases} (y_j(s_n) + t\lambda_{j-m}/\gamma)^+ & m \leq j \leq k^* \\ 0 & j > k^* \end{cases}.$$

and we define  $s_{n+1} = s_n + \Delta_n$ . Note that at time  $s_{n+1}$ , conditions (i) and (ii) are again satisfied so Theorem 2 can be applied inductively to calculate the limit  $y_j(t)$  for all  $t \geq 0$ .

An example is given in Figure 4. Since the population size is growing, we progress through the different ‘‘regimes’’ of behavior defined earlier for the fixed population size and the time between successive waves of sweeps decreases. This behavior can also be seen in Figure 3 of Beerenwinkel et al. [3]. Here we are dealing with the small mutation limit so that our waves have sharp peaks.

Motivated by the statistical analysis of cancer data in [23], Beerenwinkel et al. [3] were interested in the time  $T_{20}^\mu$  at which a cell first accumulates 20 mutations. Their choice of the number 20 was inspired by data from [23]. Using heuristics, they obtained the approximation

$$(1.3) \quad T_j^\mu \approx s_j = j \frac{(\log(\gamma/\mu))^2}{\gamma \log(N(0)N(T_{20}^\mu))}$$

for  $j \leq 20$ . Note that the approximation in (1.3) is linear in  $j$  and hence, yields constant estimates for the increments  $\tau_j^\mu = T_j^\mu - T_{j-1}^\mu$  whereas we can see that in the limiting dynamical system, the increments are not constant, but decrease in length as the population size increases. Figure 5 shows a plot of  $j$  vs.  $T_j^\mu$  for the limiting dynamical system and illustrates the non-linearity in  $j$ .

To connect the approximation in (1.3) with reality, we need to assign values to the parameters. Beerenwinkel et al. assume a selective advantage of  $\gamma = 0.01$  and a mutation

rate of  $10^{-7}$ . Since they consider a system with 100 genes this means  $\mu = 10^{-5}$ . They consider an initial benign tumor of mass  $\approx 1$  milligram or  $N(0) = 10^6$  cells that grows to a mass of  $\approx 1$  gram or  $N(T_{20}^\mu) \approx 10^9$  cells. In this case we obtain  $s_j \approx 135j$ . When  $j = 20$  this is 7.4 years, which is consistent with their simulations and clinical observations (see page 2240 of [3]). Their model evolves in discrete time, but the heuristics use only the fact that the drift in the Wright-Fisher Diffusion limit (ignoring mutations) is given by

$$b_j(x) \approx \gamma x_j(j- \langle j \rangle)$$

where  $\langle j \rangle = \sum j x_j$  (see Durrett [7], pg. 253). To get the same drift in continuous time, we need to re-scale time by  $2/N$  as opposed to  $1/N$  and hence we should replace  $\gamma$  by  $\gamma/2$  and  $\mu$  by  $\mu/2$  to obtain the analogous approximations for the Moran Model.

**2. Ideas behind the proof.** We will prove Theorem 2 for generic parameters, i.e., in the inductive step (b) that describes the limit, we suppose that  $\delta_{n,j} \neq \delta_{n,i}$  for  $i \neq j$  so that no two types reach their ‘‘goal’’ at the same time. It is easy to see that all but countably many values of  $\alpha, \gamma, \rho$  are generic, so this should be good enough for applications.

We begin by stating a approximation result for the population size  $N^\mu(t)$  which yields the desired uniform convergence of  $F^\mu(t)$  and also proves useful in other occasions. In what follows,  $C$  will always denote a constant that does not depend on  $\mu$  and whose value may change from line to line.

LEMMA 2.1. *Let  $\zeta, a > 0$ . Then as  $\mu \rightarrow 0$ ,*

$$P \left( \sup_{0 \leq t \leq aL} \left| \frac{N^\mu(t)}{N^\mu(0)e^{\rho t}} - 1 \right| > \zeta \right) \rightarrow 0.$$

PROOF. Let  $N_j(t)$ ,  $0 \leq t \leq aL$  be a family of iid pure birth (Yule) processes in which individuals give birth at rate  $\rho$  and the initial population is  $N_j(0) = 1$ . Then we have

$$N^\mu(t) =^d \sum_{j=1}^{N^\mu(0)} N_j(t).$$

It follows, for example, from [2], page 109, equation (5), that the moments  $m_j^i(t) = E(N_j^i(t))$ ,  $i = 1, 2$  satisfy

$$\begin{aligned} m_j^1(t) &= e^{\rho t} \\ m_j^2(t) &= 2e^{2\rho t}(1 - e^{-\rho t}) \leq Ce^{2\rho t} \end{aligned}$$

and so  $M_j(t) = e^{-\rho t}N_j(t) - 1$ ,  $t \geq 0$ , is a mean zero martingale (Athreya and Ney [2], page 111) with

$$\text{var}(M_j(t)) = \frac{m_j^2(t)}{e^{2\rho t}} - 1 \leq C.$$

Since the  $M_j$  are independent,

$$M(t) = \sum_{j=1}^{N^\mu(0)} M_j(t)$$

(which is itself a mean zero martingale) has

$$E(M_j^2(t)) = \text{var}(M_j(t)) \leq CN^\mu(0)$$

Applying Chebyshev's inequality and the  $L^2$  maximal inequality yields

$$\begin{aligned} P\left(\sup_{0 \leq t \leq aL} \left| \frac{\sum_{j=1}^{N^\mu(0)} N_j(t)}{N^\mu(0)e^{\rho t}} - 1 \right| > \zeta\right) &\leq P\left(\sup_{0 \leq t \leq aL} |M(t)| > \zeta N^\mu(0)\right) \\ &\leq \frac{4}{\zeta^2 N^\mu(0)^2} E(M^2(aL)) \leq \frac{C}{\zeta^2 N^\mu(0)} \rightarrow 0 \quad \text{as } \mu \rightarrow 0. \end{aligned}$$

□

There are four steps to proving the desired convergence of  $Y_j(t)$ ,  $j \geq 0$  in Theorem 2. The first step, taken in Section 3, is to prove a result about the initial behavior of the process.

**PROPOSITION 1.** *Let  $k = \lfloor \alpha \rfloor$  be the largest integer  $\leq \alpha$  and define*

$$\delta_{0,j} = \begin{cases} j\gamma/(\lambda_j - \rho) & j < k \\ (1 - (\alpha - k))\gamma/\lambda_k & j = k \end{cases}$$

*Then for any  $0 < t_1 < t_2 < \Delta_0 \equiv \min\{\delta_{0,j} : j \leq k\}$ ,  $Y_j^\mu(t) \rightarrow y_j(t)$  in probability uniformly on  $[t_1, t_2]$  with*

$$y_j(t) = \begin{cases} (\alpha - j) + t\lambda_j/\gamma & j \leq k \\ 0 & j > k. \end{cases}$$

Proposition 1 yields the correct initial conditions (a). The proof of the inductive step (b) is given in Section 4 and has three main parts that together roughly describe how the limit changes during one iteration of (b), i.e. on the interval  $[s_n, s_{n+1}]$ . Since we wish to apply the results below to  $Y_j^\mu(t)$  at positive times, we consider a version of our Moran model in which we allow for general initial conditions  $X^\mu(0)$  satisfying the following.

**Assumptions.** *As  $\mu \rightarrow 0$ ,  $F^\mu(0) \rightarrow \alpha > 0$  and  $Y_j^\mu(0) \rightarrow y_j^0$  in probability for all  $j \geq 0$ . Furthermore, we suppose that the  $y_j^0$ ,  $j \geq 0$  satisfy the conditions:*

(i) *there is a unique value of  $m$  with  $y_m^0 = \alpha$ ,*

(ii) *there is a  $k > 0$  such that  $y_j^0 = 0$  for all  $j > k$ ,  $y_j^0 > 0$  for  $m < j \leq k$ , and  $y_k^0 < 1$ .*

and (iii)  *$y_{j+1}^0 > y_j^0 - 1$  for  $0 \leq j \leq k$ .*

Define

$$\delta_j^\varepsilon \equiv \begin{cases} (\alpha - y_j^0 - \varepsilon)\gamma/(\lambda_{j-m} - \rho) & m < j < k \\ (1 - y_k^0 - \varepsilon)\gamma/\lambda_{k-m} & j = k \end{cases}$$

and let  $\Delta_\varepsilon \equiv \min\{\delta_j^\varepsilon : m < j \leq k\}$  for  $\varepsilon \geq 0$ . For  $j \geq 0$  and  $t \leq \Delta_0$ , define

$$y_j(t) = \begin{cases} (y_j^0 + t\lambda_{j-m}/\gamma)^+ & j \leq k \\ 0 & j > k \end{cases}.$$

To connect the next three results below back to (b), we will use Proposition 2 to describe the limit on the intervals  $[s_n + \varepsilon, s_{n+1} - \varepsilon']$  for small  $\varepsilon, \varepsilon' > 0$  and use Propositions 3 and 4 to describe the limit on  $[s_{n+1} - \varepsilon', s_{n+1} + \varepsilon]$  depending on which of the following two possible outcomes occurs: (i)  $\Delta_0 = \delta_k^0$  and a new type is born or (ii)  $\Delta_0 = \delta_n^0$  for some  $n \in (m, k)$  and there is a change in the dominant type.

**PROPOSITION 2.** *Let  $\varepsilon > 0$ . Then  $Y_j^\mu(t) \rightarrow y_j(t)$  in probability uniformly on  $[0, \Delta_\varepsilon]$  for all  $j \geq 0$ .*

**PROPOSITION 3. Birth of a new type.** *Suppose that  $\Delta_0 = \delta_k^0$  and for  $t \leq \varepsilon$ , let*

$$y_j(\Delta_0 + t) = \begin{cases} (y_j(\Delta_0) + t\lambda_{j-m}/\gamma)^+ & j \leq k + 1 \\ 0 & j > k + 1. \end{cases}$$

*Then there exists  $\varepsilon_1 = \varepsilon_1(y^0) > 0$  so that for all  $j \neq k + 1$ ,  $Y_j^\mu(t) \rightarrow y_j(t)$  in probability uniformly on  $[\Delta_\varepsilon, \Delta_0 + \varepsilon]$  and*

$$(2.1) \quad P \left( \sup_{\Delta_\varepsilon/2 \leq t \leq \Delta_0 + \varepsilon} Y_{k+1}(t) - (t - \Delta_\varepsilon/2)\lambda_{k+1-m}/\gamma > \varepsilon/2 \right) \rightarrow 0$$

$$P \left( \sup_{\Delta_0 + (\gamma/\lambda_{k-m})\varepsilon/2 \leq t \leq \Delta_0 + \varepsilon} Y_{k+1}(t) - (t - \Delta_0 - (\gamma/\lambda_{k-m})\varepsilon/2)\lambda_{k+1-m}/\gamma < -\varepsilon/2 \right) \rightarrow 0$$

*as  $\mu \rightarrow 0$  provided  $\varepsilon < \varepsilon_1$ .*

**PROPOSITION 4. Change in the dominant species.** *Suppose that  $\Delta_0 = \delta_n^0$  for some  $n \in (m, k)$  and for  $t \leq \varepsilon$ , let*

$$y_j(\Delta_0 + t) = \begin{cases} (y_j(\Delta_0) + t\lambda_{j-n}/\gamma)^+ & j \leq k \\ 0 & j > k \end{cases}.$$

*Then there exists  $\varepsilon_2 = \varepsilon_2(y^0) > 0$  so that  $Y_j^\mu(t) \rightarrow y_j(t)$  in probability uniformly on  $[\Delta_\varepsilon, \Delta_0 + \varepsilon]$  provided  $\varepsilon < \varepsilon_2$ .*

*Proof of Theorem 1 from Propositions 1-4.* Suppose that  $X_0^\mu(0) = N^\mu(0)$ ,  $X_j^\mu(0) = 0$  for all  $j \geq 1$ , and let  $y_j(t)$  denote the dynamical systems described by (a) and (b). Let  $K$  be a compact subset of  $(0, \infty)$ ,  $\zeta > 0$ , and take  $a \in (0, \Delta_0)$ ,  $n(K) \geq 1$  so that  $[a, s_{n(K)}] \supset K$ , where  $s_n$  is as defined in (b). Choose  $\varepsilon > 0$  small enough so that  $\varepsilon < \varepsilon_1(y(s_n)), \varepsilon_2(y(s_n))$  for all  $n \leq n(K)$  where  $\varepsilon_1, \varepsilon_2$  are as in Propositions 3 and 4, respectively. Without loss of

generality, suppose that  $\varepsilon < \zeta/(c+1)$  where  $c = c(\gamma, \rho) > 1$  is defined below. We also set  $s_{n,\varepsilon} = s_n - \varepsilon\gamma/\lambda_{j_n^*-1-m_{n-1}}$  where  $j_n^*$  satisfies  $\Delta_n = \delta_{n,j_n^*}$ .

By Proposition 1, we obtain  $Y_j(t) \rightarrow y_j(t)$  in probability uniformly on  $[a, s_{1,\varepsilon}]$ . Suppose now that we have uniform convergence on  $[a, s_{n,\varepsilon}]$  for some  $n \leq n(K) - 1$ . Then we have two cases to consider. If  $j_n^* = j$  for some  $j \in (m_n, k_n)$ , then applying Proposition 4 up to time  $s_n + \varepsilon$  and then Proposition 2 with  $y_j^0 = y_j(s_n + \varepsilon)$  for all  $j$  up to time  $s_{n+1,\varepsilon}$ , we obtain the result. If  $j_n^* = k_n$ , then Proposition 4 clearly allows us to extend uniform convergence for  $Y_j(t)$ ,  $j \neq k_n + 1$  up to time  $s_n + \varepsilon$ . To do this for  $j = k_n + 1$ , we first apply Proposition 2 to get convergence up to time  $s_{n,\varepsilon/2}$ . Write

$$Y_j(t) - y_j(t) = (Y_j(t) - (t - s_{n,\varepsilon/2})\gamma/\lambda_{k_n+1-m_n}) + ((t - s_{n,\varepsilon/2})\lambda_{k_n+1-m_n}/\gamma - y_j(t)).$$

Recalling that  $y_j(t) = 0$  if  $t \leq s_n$  and  $= (t - s_n)\lambda_{k_n+1-m_n}/\gamma$  if  $s_n \leq t \leq s_n + \varepsilon$ , we can see that

$$(t - s_{n,\varepsilon/2})\lambda_{k_n+1-m_n}/\gamma - y_j(t) \in [0, (\lambda_{k_n+1-m_n}/\lambda_{k_n-m_n})\varepsilon/2] \subset [0, c\varepsilon/2]$$

for all  $s_{n,\varepsilon/2} \leq t \leq s_n + \varepsilon$ , the last inclusion following from the fact that

$$\begin{aligned} \lambda_{k+1}/\lambda_k &= ((1+\rho)(1+\gamma)^{k+1} - 1)/((1+\rho)(1+\gamma)^k - 1) \\ &\leq ((1+\rho)(1+\gamma)^2 - 1)/((1+\rho)(1+\gamma) - 1) \\ &\equiv c \end{aligned}$$

for all  $k \geq 1$ . Since Proposition 4 implies that  $Y_j(t) - (t - s_{n,\varepsilon/2})\gamma/\lambda_{k_n+1-m_n} < \varepsilon/2$  for all  $s_{n,\varepsilon/2} \leq t \leq s_n + \varepsilon$  with high probability and  $c > 1$ , we obtain

$$P\left(\sup_{s_{n,\varepsilon/2} \leq t \leq s_n + \varepsilon} Y_{k+1}^\mu(t) - y_{k+1}(t) > (c+1)\varepsilon/2\right) \rightarrow 0$$

as  $\mu \rightarrow 0$ . To prove the lower bound, we note that  $Y_{k+1}(t) - y_{k+1}(t) \geq 0$  for  $t \leq s_n$ ,  $y_{k+1}(t) \leq c\varepsilon/2$ , for all  $t \leq s_n + (\gamma/\lambda_{k_n-m_n})\varepsilon/2$ , and by a similar argument to the one above using the second equation in (2.1) instead of the first,  $Y_{k+1}(t) - y_{k+1}(t) < -(c+1)\varepsilon/2$  for all  $s_n + (\varepsilon/2)(\gamma/\lambda_{k_n-m_n}) \leq t \leq s_n + \varepsilon$  with high probability. Therefore,

$$P\left(\sup_{s_{n,\varepsilon/2} \leq t \leq s_n + \varepsilon} Y_{k+1}^\mu(t) - y_{k+1}(t) < -(c+1)\varepsilon/2\right) \rightarrow 0$$

Since  $\varepsilon < \zeta/(c+1)$ , we conclude that

$$P\left(\sup_{s_{n,\varepsilon/2} \leq t \leq s_n + \varepsilon} |Y_{k+1}^\mu(t) - y_{k+1}(t)| > \zeta\right) \rightarrow 0$$

as  $\mu \rightarrow 0$  so we have convergence up to time  $s_n + \varepsilon$ . Finally, to complete the proof of the inductive step, apply Proposition 2 with  $y_j^0 = y_j(s_n + \varepsilon)$  to extend the convergence up to time  $s_{n+1,\varepsilon}$ .  $\square$

**3. Initial Behavior.** In this section we prove Proposition 1 about the initial behavior of the limit, but before we can begin, we need to take care of some housekeeping chores. We set  $\mathbb{N}_0 = \{0, 1, \dots\}$  and for  $x = (x_0, x_1, \dots) \in \mathbb{R}^{\mathbb{N}_0}$ , we write  $x^{j,k} = x + e_j - e_k$  where the  $e_j \in \mathbb{R}^{\mathbb{N}_0}$ ,  $j \geq 0$ , are the standard basis vectors. It is useful to note that we can define  $\{(N^\mu(t), X^\mu(t))\}_{t \geq 0}$  as the Markov process with state space

$$\mathcal{S} \equiv \{(N, x) \in \mathbb{N}_0 \times \mathbb{N}_0^{\mathbb{N}_0} : \sum_{j \geq 0} x_j = N\}$$

and initial population  $(N^\mu(0), X^\mu(0)) = (N^\mu(0), (N^\mu(0), 0, 0, \dots))$  with  $N^\mu(0)$  distributed according to  $\nu_0$  in which  $(N, x) \mapsto (N, y)$  at rate  $p_{j,k}(x) + \mu \delta_{j-1,k} x_{j-1}$  if  $y = x^{j,k}$  for some  $j, k \geq 0$ ,  $(N, x) \mapsto (N+1, y)$  at rate  $\rho N(1 + \gamma)^j x_j / w$  if  $y = x + e_j$ , and  $(N, x) \mapsto (M, y)$  at rate 0 otherwise where here,  $\delta_{j,k}$  denotes the Kronecker Delta symbol and

$$p_{j,k}(x) = \frac{(1 + \gamma)^j x_j x_k}{w}, \quad w = \sum_{i \geq 0} (1 + \gamma)^i x_i.$$

We let

$$b_j^0(x) = \rho N(1 + \gamma)^j x_j / w + \sum_{k \neq j} p_{j,k}(x), \quad d_j^0(x) = \sum_{k \neq j} p_{j,k}(x)$$

denote the birth and death rates of type  $j$ 's ignoring mutations, and drop the 0's when the mutation rates are included.  $\mathcal{F}_t = \sigma\{X^\mu(s) : s \leq t\}$  and unless otherwise explicitly stated, when we say a process is a martingale, sub-martingale, etc, it will be with respect to the canonical filtration  $\mathcal{F}_t$ . We will also use the notation

$$\mathcal{S}^N = \{x \in \mathbb{N}_0^{\mathbb{N}_0} : \sum_{j \geq 0} x_j = N\}$$

to denote a particular cross section of our state space  $\mathcal{S}$ .

For convenience, we will assume for the remainder of this section that  $N^\mu(0) = \mu^{-\alpha}$ . Our first lemma takes care of the limits for  $j \geq k + 1$ . Recall that  $T_{k+1}^\mu = \min\{t : X_{k+1}^\mu(t) > 0\}$ .

LEMMA 3.1. *If  $k = \lfloor \alpha \rfloor$  then  $P(T_{k+1}^\mu < Lt/\gamma) \rightarrow 0$  as  $\mu \rightarrow 0$  for any  $t < \delta_{0,k}$ .*

PROOF. Since type  $j$ 's are born at rate  $b_j(x)$  and die at rate  $d_j(x)$ , we have

$$\frac{d}{dt} EX_j^\mu(t) = E(b_j(X^\mu(t)) - d_j(X^\mu(t)))$$

Using  $\sum X_i^\mu(t) = N^\mu(t)$  and  $(1 + \gamma)^i \geq 1$  for  $i \geq 0$ , we have

$$\begin{aligned} b_j(X^\mu(t)) - d_j(X^\mu(t)) &= \frac{\sum_{i \geq 0} [(1 + \rho)(1 + \gamma)^j - (1 + \gamma)^i] X_i^\mu(t) X_j^\mu(t)}{\sum_{i \geq 0} (1 + \gamma)^i X_i^\mu(t)} \\ &\quad + \mu(X_{j-1}^\mu(t) - X_j^\mu(t)) \\ (3.1) \qquad \qquad \qquad &\leq \lambda_j X_j^\mu(t) + \mu X_{j-1}^\mu(t) \end{aligned}$$

for any  $t \geq 0$ . Thus for  $j \geq 1$ , we obtain

$$\frac{d}{dt} EX_j^\mu(t) \leq \lambda_j EX_j^\mu(t) + \mu EX_{j-1}^\mu(t)$$



so that integrating both sides yields

$$EX_j^\mu(t) \leq \mu \int_0^t EX_{j-1}(s) e^{\lambda_j(t-s)} ds \quad \text{for } j \geq 1.$$

We claim that induction now implies

$$(3.2) \quad EX_j^\mu(t) \leq C_j (1/\mu)^{\alpha-j} e^{\lambda_j t}$$

To prove this, we note that  $EX_0^\mu(t) \leq EN^\mu(t) = (1/\mu)^\alpha e^{\rho t}$  (recall that  $N^\mu(t)$  is just a Yule process), so the result for  $j = 0$  holds with  $C_0 = 1$ . Using the induction hypothesis and integrating

$$\begin{aligned} EX_j^\mu(t) &\leq \mu \int_0^t C_{j-1} (1/\mu)^{\alpha-j+1} e^{\lambda_{j-1}s} e^{\lambda_j(t-s)} ds \\ &\leq C_{j-1} (1/\mu)^{\alpha-j} e^{\lambda_j t} \int_0^t e^{-(\lambda_j - \lambda_{j-1})s} ds \end{aligned}$$

which proves the claim with  $C_j = C_{j-1}/(\lambda_j - \lambda_{j-1})$ .

From (3.2) it follows that

$$\int_0^t EX_j^\mu(s) ds \leq C (1/\mu)^{\alpha-j} e^{\lambda_j t}$$

In particular, taking  $t < \delta_{0,k} = \gamma(1 - (\alpha - k))/\lambda_k$ , we have

$$(3.3) \quad \int_0^{Lt/\gamma} EX_k^\mu(s) ds \leq C (1/\mu)^{1 - (\delta_{0,k} - t)\lambda_k/\gamma}.$$

If we let  $M_k^\mu(t)$  be the number of mutations from  $k$ 's to  $(k+1)$ 's up to time  $Lt/\gamma$ , then

$$EM_k^\mu(t) = \mu \int_0^{Lt/\gamma} EX_k^\mu(s) ds$$

and therefore, Chebyshev's inequality along with (3.3) imply that

$$P(M_k^\mu(t) \geq 1) \leq EM_k^\mu(t) \rightarrow 0$$

as  $\mu \rightarrow 0$ . This completes the proof.  $\square$

To obtain the appropriate limits for  $j \leq k$  and complete the proof of Proposition 1, we will couple  $X_j^\mu(t)$ ,  $j \leq k$  with upper and lower bounding branching processes  $Z_{j,u}^\mu(t)$  and  $Z_{j,\ell}^\mu(t)$  so that  $Z_{j,\ell}^\mu(t) \leq X_j^\mu(t) \leq Z_{j,u}^\mu(t)$  up until some stopping time  $\sigma$  which will be greater than  $Lt/\gamma$  with high probability for any  $t < \Delta_0$  and then show that we have

$$(1/L) \log^+ Z_{j,a}(Lt/\gamma) \rightarrow y_j(t)$$

in probability uniformly on  $[t_1, t_2]$  for any  $0 < t_1 < t_2 < \Delta_0$  (see Lemma 3.4). The coupling is made possible by applying the following result to bound the birth and death rates of type  $j$ 's on the interval  $[0, \Delta_0]$ .

LEMMA 3.2. *Suppose that  $x \in \mathcal{S}^N$  and that there exist  $m, M \in \mathbb{N}_0$ ,  $\eta > 0$  so that (i)  $\sum_{j \neq m} x_j \leq \mu^\eta N$  and (ii)  $x_j = 0$  for all  $j > M$ . Then for all  $j \neq m$ , we have the inequalities*

$$\begin{aligned} \frac{(1 + \rho - \mu^\eta)(1 + \gamma)^{j-m} x_j}{1 + g_\mu} &\leq b_j^0(x) \leq \frac{(1 + \rho)(1 + \gamma)^{j-m} x_j}{1 - h_\mu} \\ \frac{(1 - M\mu^\eta)x_j}{1 + g_\mu} &\leq d_j^0(x) \leq x_j \end{aligned}$$

where  $g_\mu = \gamma_{M-m}(M - m)\mu^\eta$  and  $h_\mu = -\gamma_{-m}m\mu^\eta$

PROOF. From the definition,

$$\begin{aligned} b_j^0(x) &= x_j(1 + \gamma)^j \frac{(1 + \rho)N - x_j}{\sum_i (1 + \gamma)^i x_i} \\ &= x_j(1 + \gamma)^{j-m} \frac{(1 + \rho) - x_j/N}{1 + \sum_i [(1 + \gamma)^{i-m} - 1]x_i/N} \end{aligned}$$

To get the lower bound, drop the terms in the denominator with  $i \leq m$ , which are  $\leq 0$ , and use the fact that  $j \rightarrow \gamma_j$  is increasing. For the upper bound drop the terms with  $i \geq m$ . The death rates are given by

$$d_j^0(x) = x_j \frac{\sum_{i \neq j} (1 + \gamma)^i x_i}{\sum_i (1 + \gamma)^i x_i}$$

so the upper bound is trivial. The lower bound follows in the same way as the lower bound for  $b_j^0(x)$  once we write

$$d_j^0(x) = x_j \frac{x_m + \sum_{i \neq j, m} (1 + \gamma)^{i-m} x_i}{N + \sum_i [(1 + \gamma)^{i-m} - 1]x_i} \geq x_j \frac{N - \sum_{i \neq m} x_i}{N + \sum_i [(1 + \gamma)^{i-m} - 1]x_i}$$

□

We now describe the bounding processes. Let  $0 < t_1 < t_2 < \Delta_0$ ,

$$\eta = \eta(t_2) = \frac{\lambda_1 - \rho}{4\gamma}(\Delta_0 - t_2).$$

The reason for this choice of  $\eta$  is that

$$y_j(t) \leq (\alpha + t\rho/\gamma) - 4\eta$$

for all  $t \leq t_2$ ,  $j \geq 1$ . For our bounding processes, we set  $Z_{0,u}(t) \equiv N^\mu(t)$ ,  $Z_{0,\ell}^\mu(t) \equiv (1 - k\mu^\eta)N^\mu(t)$ , and let  $Z_{j,a}^\mu$ ,  $1 \leq j \leq k$ ,  $a = u, \ell$  be (birth and death) branching processes with rates given in Table 6 taking  $m = 0$ ,  $M = k$ . Note that the birth and death rates are per particle. The extra factor  $\mu$  in the definition of  $d_{j,\ell}^\mu$  takes care of deaths due to mutations. We also set  $\lambda_{j,a}^\mu \equiv b_{j,a}^\mu - d_{j,a}^\mu$  to be the growth rates of  $Z_{j,a}^\mu$ ,  $a = u, \ell$  so that we have  $\lambda_{j,a}^\mu \rightarrow \lambda_j$  as  $\mu \rightarrow 0$  for  $j \geq 1$ ,  $a = u, \ell$ . If we use the convention that  $\lambda_{0,a}^\mu = \rho$  for  $a = u, \ell$ , this also holds for  $j = 0$ .

For the next result, we use the notation  $Z_a^\mu(t) = (Z_{0,a}^\mu(t), Z_{1,a}^\mu(t), \dots, Z_{k,a}^\mu(t), 0, \dots)$ , for  $a = u, \ell$ .

birth rate	$b_{j,u}^\mu \equiv \frac{Z_{j,u}^\mu(s)}{(1+\rho)(1+\gamma)^{j-m}}$	$b_{j,\ell}^\mu \equiv \frac{Z_{j,\ell}^\mu(s)}{(1+\rho-\mu^\eta)(1+\gamma)^{j-m}}$
death rate	$d_{j,u}^\mu \equiv \frac{1-h_\mu^\eta}{1+g_\mu}$	$d_{j,\ell}^\mu \equiv 1 + \mu$
immigration rate	$\mu Z_{j-1,u}^\mu(t)$	$\mu Z_{j-1,\ell}^\mu(t)$

TABLE 6

Rates for the comparison branching processes,  $j \geq 1$ .

LEMMA 3.3. *There exists a coupling of  $X^\mu(t)$  with  $Z_a^\mu(t)$ ,  $a = u, \ell$  so that*

$$Z_{j,\ell}^\mu(t) \leq X_j^\mu(t) \leq Z_{j,u}^\mu(t)$$

for all  $t \leq (L\sigma/\gamma) \wedge T_{k+1}^\mu$ ,  $j \leq k$ , where  $\sigma =$

$$\inf\{t \geq 0 : Y_i^\mu(t) > \alpha + t\rho/\gamma - 2\eta, \text{ for some } i \geq 1 \text{ or } |F^\mu(t) - (\alpha + t\rho/\gamma)| > \eta\}.$$

PROOF. For  $t \leq T_{k+1}^\mu$ , we have  $X_j(t) = 0$  if  $j > k$ . Furthermore, if  $t \leq \sigma L/\gamma$ ,

$$\frac{N^\mu(Lt/\gamma)}{(1/\mu)^{\alpha e^{\rho t}}} \leq (1/\mu)^\eta$$

so that

$$\frac{X_j^\mu(Lt/\gamma)}{N^\mu(Lt/\gamma)} \leq \frac{\mu^2 \eta}{(1/\mu)^\eta} = \mu^\eta$$

for all  $j \geq 1$  and hence we have the bounds on birth and death rates given in Lemma 3.2 with  $m = 0$  and  $M = k$ . The processes can therefore be coupled in an elementary way by matching birth, deaths, and immigrations in the appropriate manner.  $\square$

The task which we will dedicate most of the remainder of the section to proving is the following.

LEMMA 3.4. *Let  $0 < t_1 < t_2 < \delta_0$ . For  $a = u, \ell$  and  $j \leq k$ , we have*

$$(1/L) \log^+ Z_{j,a}^\mu(Lt/\gamma) \rightarrow y_j(t)$$

in probability uniformly on  $[t_1, t_2]$ .

Because  $y(t) \leq (\alpha + \rho t) - 4\eta$  for all  $t \leq t_2$ , Lemma 3.4 implies that

$$P((1/L) \log^+ Z_a^\mu(Lt/\gamma) \leq (\alpha + \rho t) - 2\eta, \forall t \leq t_2, a = u, \ell) \rightarrow 1$$

as  $\mu \rightarrow 0$ . This and Lemma 2.1 imply that  $P(\sigma > t_2) \rightarrow 1$  as  $\mu \rightarrow 0$  and therefore Proposition 1 follows from Lemma 3.4, Lemma 3.3, and Lemma 3.1.

To prove Lemma 3.4, we begin by defining another level of upper and lower bounds  $\hat{Z}_{j,a}^\mu$  in which immigrations occur at deterministic rates. More specifically, for  $a = u, \ell$ , we define  $\hat{Z}_{j,a}^\mu(t)$  as a branching process with the same initial population and birth and death rates as  $Z_{j,a}^\mu(t)$ , but with immigrations at rate  $\mu I_{j,a}^\mu(t)$  where

$$I_{j,u}^\mu(t) \equiv E\hat{Z}_{j-1,u}^\mu(t) + e^{\lambda_{j-1,u}^\mu t} (1/\mu)^{2(\alpha-(j-1))/3}.$$

and

$$I_{j,\ell}^\mu(t) \equiv E\hat{Z}_{j-1,\ell}^\mu(t) - e^{\lambda_{j-1,\ell}^\mu t} (1/\mu)^{2(\alpha-(j-1))/3}.$$

We will use the convention that  $I_{0,a}^\mu(t) \equiv 0$  for all  $t$ . Note that

$$(3.4) \quad E(e^{-\lambda_{j,a}^\mu t} \hat{Z}_{j,a}^\mu(t)) = \mu \int_0^t e^{-\lambda_{j,a}^\mu s} I_{j,a}^\mu(s) ds$$

for all  $j \geq 1$  and  $a = u, \ell$ , an expression which will be used often throughout the remainder of this section.

LEMMA 3.5. *For  $j \geq 0$  and  $a = u, \ell$ ,*

$$M_{j,a}^\mu(t) \equiv e^{-\lambda_{j,a}^\mu t} \hat{Z}_{j,a}^\mu(t) - E(e^{-\lambda_{j,a}^\mu t} \hat{Z}_{j,a}^\mu(t))$$

*is a martingale with respect to the  $\sigma$ -field*

$$\mathcal{G}_{a,t} \equiv \sigma\{\hat{Z}_{i,a}^\mu(s) : 0 \leq i \leq j, s \leq t\}.$$

PROOF. We prove the result for  $a = u$ , the proof for  $a = \ell$  being similar, and drop the subscripts  $u$  on all quantities for the remainder of the proof. It is easy to see that

$$E(\hat{Z}_j^\mu(t+h)|\mathcal{G}_t) = e^{\lambda_j^\mu h} \hat{Z}_j^\mu(t) + E\left(\mu \int_t^{t+h} e^{\lambda_j^\mu(t+h-s)} I_j^\mu(s) ds \middle| \mathcal{G}_t\right)$$

and multiplying by  $e^{-\lambda_j^\mu(t+h)}$  gives

$$E\left(e^{-\lambda_j^\mu(t+h)} \hat{Z}_j^\mu(t+h) - e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t) - \mu \int_t^{t+h} e^{-\lambda_j^\mu s} I_j^\mu(s) ds \middle| \mathcal{G}_t\right) = 0$$

Since (3.4) implies that

$$M_j^\mu(t+h) - M_j^\mu(t) = e^{-\lambda_j^\mu(t+h)} \hat{Z}_j^\mu(t+h) - e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t) - \mu \int_t^{t+h} e^{-\lambda_j^\mu s} I_j^\mu(s) ds$$

for  $j \geq 1$  and the same equality clearly holds for  $j = 0$  as well, the desired result,  $E(M_j^\mu(t+h) - M_j^\mu(t)|\mathcal{G}_t) = 0$ , follows.  $\square$

LEMMA 3.6. *For all  $a = u, \ell$ ,  $T > 0$ , and  $\mu$  sufficiently small, we have*

$$P\left(\sup_{t \leq T} |M_{j,a}^\mu(t)| > (1/\mu)^{2(\alpha-j)/3}\right) \leq C\mu^{(\alpha-j)/3} [1 + \mu^{(\alpha-j+1)/3}]$$

*In particular, for all  $j \leq k$*

$$P(|\hat{Z}_{j,a}^\mu(t) - E\hat{Z}_{j,a}^\mu(t)| > e^{\lambda_{j,a}^\mu t} (1/\mu)^{2(\alpha-j)/3}, \forall t \leq T) \rightarrow 0$$

*as  $\mu \rightarrow 0$*

PROOF. The second part of the result follows directly from the first along with the definition of  $M_{j,a}^\mu(t)$ . To obtain the first part, we suppose for the remainder of the proof that  $u = a$  and drop the subscript  $u$ . The proof for  $a = \ell$  is similar. We will also assume that  $j \geq 1$  and leave the (simpler)  $j = 0$  case to the reader.

We proceed by calculating the variance of  $e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t)$  and then using the  $L^2$  maximum inequality to bound the second moment of  $M_j^\mu(t)$  uniformly on  $[0, T]$ . To begin, we claim that provided we choose  $\mu$  small enough so that  $\lambda_i^\mu > \lambda_{i-1}^\mu$  for all  $1 \leq i \leq j$ , we have

$$(3.5) \quad g(t)(1/\mu)^{(\alpha-j)} \leq E(e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t)) \leq C[(1/\mu)^{(\alpha-j)} + \mu^{1/3}(1/\mu)^{2(\alpha-j)/3}]$$

where  $g(t)$  is continuous on  $[0, \infty)$  and positive on  $(0, \infty)$ . To see this, we note that  $E\hat{Z}_0(t) = EN^\mu(t) = (1/\mu)^\alpha e^{\rho t}$  so the result clearly holds for  $j = 1$  by (3.4) and the general case follows by induction on  $j$ . Now,  $\frac{d}{dt}E(e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t))^2$  is

$$(3.6) \quad \begin{aligned} &= -2\lambda_j^\mu E(e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t))^2 + e^{-2\lambda_j^\mu t} E[b_j^\mu \hat{Z}_j^\mu(t)(2\hat{Z}_j^\mu(t) + 1)] \\ &\quad - d_j^\mu e^{-2\lambda_j^\mu t} E[\hat{Z}_j^\mu(t)(2\hat{Z}_j^\mu(t) - 1)] + \mu I_j^\mu(t) e^{-2\lambda_j^\mu t} E[2\hat{Z}_j^\mu(t) + 1] \\ &= (b_j^\mu + d_j^\mu) e^{-2\lambda_j^\mu t} E\hat{Z}_j^\mu(t) + \mu I_j^\mu(t) e^{-2\lambda_j^\mu t} + 2\mu I_j^\mu(t) e^{-2\lambda_j^\mu t} E\hat{Z}_j^\mu(t). \end{aligned}$$

(3.4) implies that

$$\begin{aligned} &\int_0^t 2\mu I_j^\mu(s) e^{-2\lambda_j^\mu s} E\hat{Z}_j^\mu(s) ds \\ &= 2 \int_0^t \mu I_j^\mu(s) e^{-\lambda_j^\mu s} \int_0^s \mu I_j^\mu(r) e^{-\lambda_j^\mu r} dr ds = [E(e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t))]^2 \end{aligned}$$

so that integrating both sides of (3.6) and applying (3.5) yields

$$(3.7) \quad \begin{aligned} \text{var}(e^{-\lambda_j^\mu t} \hat{Z}_j^\mu(t)) &\leq (b_j^\mu + d_j^\mu) \int_0^t e^{-2\lambda_j^\mu s} E\hat{Z}_j^\mu(s) ds + \int_0^t \mu I_j^\mu(s) e^{-2\lambda_j^\mu s} ds \\ &\leq C[(1/\mu)^{(\alpha-j)} + \mu^{1/3}(1/\mu)^{2(\alpha-j)/3}]. \end{aligned}$$

By Lemma 3.5  $M_j^\mu$  is a martingale with respect to  $\mathcal{G}_t$  and so the  $L^2$  maximum inequality implies that

$$E\left(\sup_{t \leq T} (M_j^\mu(t))^2\right) \leq 4E(M_j^\mu(T))^2 = 4\text{var}(e^{-\lambda_j^\mu T} \hat{Z}_j^\mu(T))$$

the second equality following from the definition of  $M_j^\mu$ . Applying Chebyshev's Inequality and (3.7) then yields

$$P\left(\sup_{t \leq t_2} |M_j^\mu(t)| > (1/\mu)^{2(\alpha-j)/3}\right) \leq C\mu^{(\alpha-j)/3} [1 + \mu^{(\alpha-j+1)/3}]$$

completing the proof.  $\square$

COROLLARY 3. For  $a = u, \ell$ , there exists a coupling of the process  $Z_a^\mu$  with  $\hat{Z}_a^\mu$  so that

$$\hat{Z}_{j,\ell}^\mu(t) \leq Z_{j,\ell}^\mu(t) \leq Z_{j,u}^\mu(t) \leq \hat{Z}_{j,u}^\mu(t)$$

for all  $t \leq \hat{\sigma}$  where

$$\hat{\sigma} \equiv \inf\{t \geq 0 : \hat{Z}_{j-1,u}^\mu(t) > I_{j,u}^\mu(t) \text{ or } \hat{Z}_{j-1,\ell}^\mu(t) < I_{j,\ell}^\mu(t) \text{ for some } j \geq 1\}.$$

Furthermore,  $P(\hat{\sigma} \leq t_2) \rightarrow 0$  as  $\mu \rightarrow 0$ .

PROOF. Arguing inductively, we can see that the immigration rates for type  $j$ 's in  $\hat{Z}_\ell^\mu, Z_\ell^\mu, Z_u^\mu$  and  $\hat{Z}_u^\mu$ , respectively, satisfy

$$\mu I_{j,\ell}^\mu(t) \leq \mu Z_{j-1,\ell}^\mu(t) \leq \mu Z_{j-1,u}^\mu(t) \leq \mu I_{j,u}^\mu(t)$$

for  $t \leq \hat{\sigma}$ . Therefore, we define a coupling for the two processes by coupling births, deaths, and immigrations. The fact that  $P(\hat{\sigma} \leq t_2) \rightarrow 0$  follows from Lemma 3.6.  $\square$

Define  $\hat{Y}_{j,a}^\mu(t) \equiv (1/L) \log^+ \hat{Z}_{j,a}^\mu(Lt/\gamma)$ . Lemma 3.4 follows from Corollary 3 along with our next result.

LEMMA 3.7. Let  $j \leq k$ ,  $a = u, \ell$ . Then  $\hat{Y}_{j,a}^\mu(t) \rightarrow y_j(t)$  in probability uniformly on  $[t_1, t_2]$ .

PROOF. Again, we only prove the result for  $u = a$  and drop the  $u$  subscript. Let  $j \leq k$  and write  $\hat{Y}_j^\mu(t) - y_j(t)$

$$(3.8) \quad \begin{aligned} &= \left( (1/L) \log^+ [e^{-\lambda_j^\mu Lt/\gamma} \hat{Z}_j^\mu(Lt/\gamma)] - (1/L) \log^+ E[e^{-\lambda_j^\mu Lt/\gamma} \hat{Z}_j^\mu(Lt/\gamma)] \right) \\ &+ \left( (1/L) \log^+ E[e^{-\lambda_j^\mu Lt/\gamma} \hat{Z}_j^\mu(Lt/\gamma)] - (\alpha - j) \right) + (\lambda_j^\mu - \lambda_j)t/\gamma \end{aligned}$$

By Lemma 3.6,

$$P\left(|\hat{Z}_{j,u}^\mu(t) - E\hat{Z}_{j,u}^\mu(t)| \leq e^{\lambda_j^\mu t} (1/\mu)^{2(\alpha-j)/3}, \forall t \leq t_2\right) \rightarrow 1$$

and on set where

$$|\hat{Z}_{j,u}^\mu(t) - E\hat{Z}_{j,u}^\mu(t)| \leq e^{\lambda_j^\mu s} (1/\mu)^{2(\alpha-j)/3}, \quad \forall t \leq t_2$$

we have

$$\begin{aligned} &(1/L) \log^+ [e^{-\lambda_j^\mu sL/\gamma} \hat{Z}_j^\mu(Lt/\gamma)] - (1/L) \log^+ E[e^{-\lambda_j^\mu Lt/\gamma} \hat{Z}_j^\mu(Lt/\gamma)] \\ &= \frac{1}{L} \log \left( 1 + \frac{\hat{Z}_j^\mu(Lt/\gamma) - E\hat{Z}_j^\mu(Lt/\gamma)}{E\hat{Z}_j^\mu(Lt/\gamma)} \right) \\ &\leq \frac{C}{L} \frac{|\hat{Z}_j^\mu(Lt/\gamma) - E\hat{Z}_j^\mu(Lt/\gamma)|}{E\hat{Z}_j^\mu(Lt/\gamma)} \leq \frac{C}{L} \frac{(1/\mu)^{2(\alpha-j)/3}}{(1/\mu)^{\alpha-j}} \rightarrow 0 \end{aligned}$$

uniformly on  $[t_1, t_2]$  as  $\mu \rightarrow 0$ , the last inequality following from (3.5) and the fact that  $g(t)$  is bounded away from 0 on  $[t_1, t_2]$ . Therefore, the absolute value of the first term on the right of (3.8) goes to zero uniformly on  $[t_1, t_2]$ . It is clear from (3.5) that the second term goes to 0 as well and since  $\lambda_j^\mu \rightarrow \lambda_j$  as  $\mu \rightarrow 0$ , the result follows.  $\square$

**4. Inductive step.** In this section, we prove Propositions 2-4. We begin with the proof of Proposition 2. The reader should refer to the statement of that result for notation used throughout this section.

4.1. *Interior convergence.* Let  $\varepsilon > 0$ , set  $a_j(t) \equiv \alpha + t\rho/\gamma$  for  $j \neq k$ ,  $a_k(t) \equiv 1$  and choose  $\eta = \eta(\varepsilon) > 0$  so that (i)  $y_j(t) < a_j(t) - 2\eta$ ,  $\forall t \leq \Delta_\varepsilon$ ,  $j \neq m$  and (ii)  $y_{j-1}(t) - y_j(t) < 1 - 2\eta$  for all  $t \leq \Delta_\varepsilon$ ,  $j \geq 0$ . Given  $\zeta > 0$ , we define the stopping times

$$\begin{aligned} \sigma_0(j) &\equiv \gamma T_j^\mu / L \\ \sigma_1(j) &\equiv \inf\{t \geq 0 : Y_j^\mu(t) \geq a_j(t) - \eta\} \\ \sigma_1 &\equiv \inf_{j \neq m} \sigma_1(j) \\ \sigma_1' &\equiv \inf_{j < m} \sigma_1(j) \\ \sigma_2(j) &\equiv \inf\{t \geq 0 : Y_{i-1}^\mu(t) - Y_i^\mu(t) \geq 1 - \eta, \text{ for some } 1 \leq i \leq j\} \\ \sigma_3(j) &\equiv \inf\{t \geq 0 : Y_j^\mu(t) \leq \zeta\}. \end{aligned}$$

For the remainder of this section, set  $\sigma_0 = \sigma_0(k+1)$ . We shall prove convergence of  $Y_j^\mu(t)$  up to time  $\sigma(j) \equiv \sigma_0 \wedge \sigma_1 \wedge \sigma_2(j) \wedge \sigma_3(j)$ . For types  $j \leq k$ , this will essentially amount to controlling the infinitesimal variance of  $Y_j^\mu$  (Lemma 4.1) and then showing that the infinitesimal mean converges to the appropriate limit (Lemma 4.2) while for types  $j > k$ , we will simply show that they are unlikely to be born before time  $L\Delta_\varepsilon/\gamma$ , i.e.,  $\sigma_0 > \Delta_\varepsilon$  with high probability (Lemma 4.3). We then complete the proof of Proposition 2 by using the structure of the limit  $y_j(t)$  to extend convergence up to time  $\Delta_\varepsilon$  as required. If  $y_j(t)$  is bounded away from 0, this is easy since our choice of  $\eta$  implies that  $\sigma_1, \sigma_2(j)$  are unlikely to occur before time  $\Delta_\varepsilon$  and if  $y_j(t)$  is not bounded away from 0 (which can only happen if  $j < m$ ), we will essentially show that  $Y_j^\mu(t)$  is a supermartingale to conclude that once it drops below a certain level, it will never climb up again.

The first step is to calculate infinitesimal means and variances. Writing  $y_j = (1/L) \log(x_j)$ ,  $y = (y_0, y_1, \dots)$ , and  $N = \sum e^{Ly_i}$ , noting the time re-scaling, and using the fact that the change in  $y_j$  when  $x_j$  jumps to  $x_j \pm 1$  is  $(1/L) \log(1 \pm x_j^{-1})$ , we can write the infinitesimal mean of  $Y_j^\mu(t)$  as  $B_j(y) = B_{j,r}(y) + B_{j,m}(y)$  where

$$\begin{aligned} B_{j,r}(y) &= \gamma^{-1} \frac{[(1+\rho)N - e^{Ly_j}](1+\gamma)^j e^{Ly_j}}{\sum_{i \geq 0} (1+\gamma)^i e^{Ly_i}} \log^+(1 + e^{-Ly_j}) \\ &\quad + \gamma^{-1} \frac{\sum_{i \neq j} (1+\gamma)^i e^{Ly_i}}{\sum_{i \geq 0} (1+\gamma)^i e^{Ly_i}} e^{Ly_j} \log^+(1 - e^{-Ly_j}) \\ B_{j,\mu} &= \mu \gamma^{-1} e^{Ly_{j-1}} \log^+(1 + e^{-Ly_j}) + \mu \gamma^{-1} e^{Ly_j} \log^+(1 - e^{-Ly_j}) \end{aligned}$$

In words,  $B_{j,r}(y)$  is the rate of change due to death and subsequent replacement, while

$B_{j,\mu}(y)$  is the rate of change due to mutations. Similarly, the infinitesimal variance is

$$\begin{aligned} A_j(y) = & (1/L) \left[ \gamma^{-1} \frac{[(1+\rho)N - e^{Ly_j}](1+\gamma)^j e^{Ly_j}}{\sum_{i \geq 0} (1+\gamma)^i e^{Ly_i}} (\log^+(1 + e^{-Ly_j}))^2 \right. \\ & + \gamma^{-1} \frac{\sum_{i \neq j} (1+\gamma)^i e^{Ly_i}}{\sum_{i \geq 0} (1+\gamma)^i e^{Ly_i}} e^{Ly_j} (\log^+(1 - e^{-Ly_j}))^2 \\ & \left. + \mu \gamma^{-1} e^{Ly_{j-i}} (\log^+(1 + e^{-Ly_j}))^2 + \mu \gamma^{-1} e^{Ly_j} (\log^+(1 - e^{-Ly_j}))^2 \right] \end{aligned}$$

Introducing  $f_1(x) \equiv x \log^+(1 + x^{-1})$ ,  $f_2(x) \equiv x \log^+(1 - x^{-1})$ ,

$$\begin{aligned} g_{j,1}(y) & \equiv \frac{(1+\gamma)^j [(1+\rho)N - e^{Ly_j}]}{\gamma \sum_{i \geq 0} (1+\gamma)^i e^{Ly_i}} \\ \text{and } g_{j,2}(y) & \equiv \frac{1 \sum_{i \neq j} (1+\gamma)^i e^{Ly_i}}{\gamma \sum_{i \geq 0} (1+\gamma)^i e^{Ly_i}}, \end{aligned}$$

we can write

$$\begin{aligned} A_j(y) = & (1/L) \left[ e^{-Ly_j} g_{j,1}(y) f_1^2(e^{Ly_j}) + e^{-Ly_j} g_{j,2}(y) f_2^2(e^{Ly_j}) \right. \\ & \left. + \mu \gamma^{-1} f_1^2(e^{Ly_j}) e^{Ly_{j-i-2Ly_j}} + \mu \gamma^{-1} f_2^2(e^{Ly_j}) e^{-Ly_j} \right]. \end{aligned}$$

Since  $g_{j,1}(x) \leq (1+\rho)(1+\gamma)^j/\gamma$ ,  $g_{j,2}(x) \leq 1/\gamma$ , and  $f_2(x) \leq f_1(x) \leq 1$  for all  $x \in [0, \infty)$ , we obtain the bound

$$(4.1) \quad A_j(y) \leq (C/L)((1+\mu)e^{-Ly_j} + \mu e^{Ly_{j-i-2Ly_j}}).$$

Define

$$M_j(t) = Y_j^\mu(t) - Y_j^\mu(0) - \int_0^t B_j(Y^\mu(s)) ds.$$

LEMMA 4.1. *For any  $\xi > 0$ ,*

$$P \left( \sup_{t \leq \sigma_2(j)} |M_j(t)| > \xi \right) \rightarrow 0$$

PROOF. Since  $M_j^2(t) - \int_0^t A_j(Y^\mu(s)) ds$  is a martingale, the  $L^2$  maximal inequality and (4.1) yield

$$E \left( \sup_{t \leq \sigma_2(j)} M_j^2(t) \right) \leq (C/L) E \left( \int_0^{\sigma_2(j)} (1+\mu) e^{-LY_j^\mu(t)} + \mu e^{LY_{j-i}^\mu(t) - 2LY_j^\mu(t)} dt \right).$$

It follows from Chebyshev's inequality that

$$\begin{aligned} (4.2) \quad P \left( \sup_{t \leq \sigma_2(j)} |M_j(t)| > \xi \right) & \leq \frac{C(1+\mu)}{L\xi^2} E \left( \int_0^{\sigma_2(j)} e^{-LY_j^\mu(t)} dt \right) \\ & \quad + \frac{C\mu}{L\xi^2} E \left( \int_0^{\sigma_2(j)} e^{LY_{j-i}^\mu(t) - 2LY_j^\mu(t)} dt \right). \end{aligned}$$



Since  $Y_{j-1}^\mu(t) - Y_j^\mu(t) < 1 - \eta$  for  $t \leq \sigma_2$ , we have

$$e^{-LY_j^\mu(t)} \leq 1 \quad \text{and} \quad \mu e^{LY_{j-i}^\mu(t) - 2LY_j^\mu(t)} \leq C\mu^\eta$$

and therefore, both terms on the right-hand side of (4.2)  $\rightarrow 0$  as  $\mu \rightarrow 0$ .  $\square$

Our next step is to show that the infinitesimal means converge to the appropriate limit. In the proof, we will derive a series of facts that will be useful several times in what follows and tease out the exact requirements for convergence of the various components of the drift.

LEMMA 4.2. *If  $j \neq m$  and  $\xi > 0$  then as  $\mu \rightarrow 0$ ,*

$$P \left( \sup_{t \leq \sigma(j)} \left| \int_0^t B_j(Y_j^\mu(s)) ds - \lambda_{j-m} t / \gamma \right| > \xi \right) \rightarrow 0.$$

PROOF. Using the definition of  $f_i, g_{j,i}, i = 1, 2$ , we write

$$B_{j,r}(y) = f_1(e^{Ly_j})g_{j,1}(y) + f_2(e^{Ly_j})g_{j,2}(y)$$

and

$$B_{j,\mu}(y) = \mu\gamma^{-1}[f_1(e^{Ly_j})e^{L(y_{j-1}-y_j)} + f_2(e^{Ly_j})].$$

We will complete the proof by proving the following four facts.

- (I) For any  $\zeta > 0$ ,  $f_1(e^{LY_j^\mu(t)}) \rightarrow 1$  and  $f_2(e^{LY_j^\mu(t)}) \rightarrow -1$  in probability uniformly on  $[0, \sigma_3(j, \zeta)]$ .
- (II)  $g_{j,2}(Y_j^\mu(t)) \rightarrow 1/\gamma$  in probability uniformly on  $[0, \sigma_1(j)]$
- (III) For any  $\zeta > 0$ ,

$$P \left( \sup_{t \leq \sigma'_1} g_{j,1}(Y_j^\mu(t)) > (1 + \rho)(1 + \gamma)^{j-m} / \gamma + \zeta \right) \rightarrow 0$$

and furthermore,  $g_{j,1}(Y_j^\mu(t)) \rightarrow (1 + \rho)(1 + \gamma)^{j-m} / \gamma$  in probability uniformly on  $[0, \sigma_0 \wedge \sigma_1]$

- (IV)  $B_{j,\mu}(Y^\mu(t)) \rightarrow 0$  in probability uniformly on  $[0, \sigma_2(j)]$

(I) follows immediately since  $f_1(x) \rightarrow 1, f_2(x) \rightarrow -1$  as  $x \rightarrow \infty$  and  $Y_j^\mu(t) \geq \zeta$  on  $[0, \sigma_3(j, \zeta)]$ . To prove (II), write

$$g_{j,2}(y) = \frac{1}{\gamma} \left( 1 - \frac{(1 + \gamma)^j e^{Ly_j}}{\sum_{i \geq 0} (1 + \gamma)^i e^{Ly_i}} \right)$$

and note that if  $\sum e^{Ly_i} = N$  and  $y_i \leq a$ , then

$$0 \leq \frac{(1 + \gamma)^j e^{Ly_j}}{\sum_{i \geq 0} (1 + \gamma)^i e^{Ly_i}} \leq (1 + \gamma)^j e^{Ly_j} / N \leq C(1/\mu)^a / N.$$

Now Lemma 2.1 and the assumption that  $F^\mu(0) \rightarrow \alpha$  imply that  $N^\mu(Lt/\gamma) \geq (1/\mu)^{\alpha+t\rho/\gamma-\eta/2}$  for all  $t \leq \Delta_0$  with high probability so that since  $\sum e^{LY_i^\mu(t)} = N^\mu(t)$  and  $Y_j^\mu(t) \leq \alpha+t\rho/\gamma-\eta$

if  $t \leq \sigma_1(j)$ , (II) follows. For (III), we note that if  $\sum e^{Ly_i} = N$ , then using the definition of  $\gamma_j = (1 + \gamma)^j - 1$ , we have

$$g_{j,1}(y) = \frac{(1 + \gamma)^{j-m}}{\gamma} \left( \frac{1 + \rho - e^{Ly_j}/N}{1 + \sum_{i \neq m} \gamma_{i-m} e^{Ly_i}/N} \right)$$

The first part of (III) then follows from the fact that

$$\sum_{i \neq m} \gamma_{i-m} e^{Ly_i}/N \geq \sum_{i < m} \gamma_{i-m} e^{Ly_i}/N \geq \gamma_{-m} (1/\mu)^a / N$$

if  $y_i \leq a$  for all  $i < m$  while the second part follows from the fact that we also have

$$\left( \frac{1 + \rho - e^{Ly_j}/N}{1 + \sum_{i \neq m} \gamma_{i-m} e^{Ly_i}/N} \right) \geq \frac{1 + \rho - (1/\mu)^a / N}{1 + (k-m)\gamma_{k-m}(1/\mu)^a / N}$$

if  $y_j \leq a$  for all  $j \leq k$  and  $y_j = 0$  for all  $j > k$ . Finally, to prove (IV), we use the bound

$$B_{j,\mu}(y) \leq C\mu[f_1(e^{Ly_j})e^{L(y_{j-1}-y_j)} + f_2(e^{Ly_j})]$$

so that since  $f_2(x), f_1(x) \leq 1$  for all  $x \geq 0$ , the result follows from the fact that

$$\mu e^{L(y_{j-1}-y_j)} \leq \mu^\eta$$

if  $y_{j-1} - y_j < 1 - \eta$ . □

The final ingredient we need to prove Proposition 2 is a result to tell us that type  $k+1$ 's are unlikely to be born before the  $k$ 's reach level 1.

LEMMA 4.3.  $P(\sigma_0(k+1) \leq \sigma_1(k)) = P(T_{k+1}^\mu \leq L\sigma_1(k)/\gamma) \rightarrow 0$  as  $\mu \rightarrow 0$ .

PROOF. Let  $M_k^\mu(t)$  denote the number of mutations from  $k$ 's to  $(k+1)$ 's by time  $t$ . Since these mutations occur at rate  $\mu X_k^\mu(t)$  and  $Y_k^\mu(t) \leq 1 - \eta$  implies that  $X_k^\mu(Lt/\gamma) \leq (1/\mu)^{1-\eta}$ , we have

$$EM_k^\mu(L(t \wedge \sigma_1(k))/\gamma) \leq tL\mu^\eta/\gamma \rightarrow 0$$

as  $\mu \rightarrow 0$  and therefore, Chebyshev's Inequality implies that

$$P(M_k^\mu(L(t \wedge \sigma_1(k))/\gamma) \geq 1) \leq EM_k^\mu(L(t \wedge \sigma_1(k))/\gamma) \rightarrow 0$$

as  $\mu \rightarrow 0$ . □

*Proof of Proposition 2.* Suppose first that  $j \leq k$ . Lemmas 4.1 and 4.2 and the assumption that  $Y_j^\mu(0) \rightarrow y_j^0$  in probability imply that  $Y_j(t) \rightarrow y_j(t)$  in probability uniformly on  $[0, \sigma(j)]$  for any  $\zeta > 0$ . We will show that the convergence is uniform on  $[0, \Delta_\varepsilon \wedge \sigma_2(j)]$ . Proposition 2 follows since  $y_{j-1}(t) - y_j(t) < 1 - 2\eta$  for all  $t \leq \Delta_\varepsilon, j \geq 0$ .

Suppose first that  $j \geq m$ . Then there exists  $\zeta > 0$  so that  $y_j(t) \geq \zeta$  for all  $t \leq \Delta_\varepsilon$  and since  $y_i(t) \leq a_i(t) - 2\eta$  for all  $t \leq \Delta_\varepsilon$ ,  $i \neq m$  by our choice of  $\eta$ , we obtain uniform convergence on  $[0, \Delta_\varepsilon \wedge \sigma_2(j)]$ . The same argument applies if  $j < m$  and  $y_j(t)$  is bounded away from 0 on  $[0, \Delta_\varepsilon]$ . If  $j < m$  and  $y_j(t)$  is not bounded away from 0 on  $[0, \Delta_\varepsilon]$ , let  $\zeta, \xi > 0$  be small and choose a time  $t_0 < \Delta_\varepsilon$  so that  $y_j(t_0) = \zeta\xi/4$ . Let  $\sigma' = \sigma_0 \wedge \sigma_1 \wedge \sigma_2(j)$ . Then  $y_j(s) \leq \zeta\xi/4$  for all  $t_0 \leq s \leq \Delta_\varepsilon$  and  $P\left(\sup_{t \leq \Delta_\varepsilon \wedge \sigma'} |Y_j(t) - y_j(t)| > \zeta\right)$

$$\begin{aligned}
&\leq P\left(\sup_{t \leq t_0 \wedge \sigma'} |Y_j(t) - y_j(t)| > \zeta\right) + P\left(\sup_{t_0 \leq t \leq \Delta_\varepsilon} Y_j(t \wedge \sigma') > \zeta(1 + \xi/4)\right) \\
&\leq P\left(\sup_{t \leq t_0 \wedge \sigma'} |Y_j(t) - y_j(t)| > \zeta\right) + P(Y_j(t_0 \wedge \sigma') > \zeta\xi/2) \\
(4.3) \quad &+ P\left(\sup_{t_0 \leq t \leq \Delta_\varepsilon} Y_j^\mu(t \wedge \sigma') > \zeta(1 + \xi/4) \mid Y_j^\mu(t_0 \wedge \sigma') \leq \zeta\xi/2\right)
\end{aligned}$$

Our previous argument can then be used to show that  $Y_j^\mu(t) \rightarrow y_j(t)$  uniformly on  $[0, t_0 \wedge \sigma_2]$  and hence the first and second terms on the right hand side of (4.3) are each  $< \xi/4$  for all  $\mu$  sufficiently small. To control the third term, we note that (II)-(IV) from the proof of Lemma 4.2 along with the bounds  $f_1(x) \leq 1$ ,  $f_2(x) \leq -1$  for all  $x \geq 0$  imply that if  $\mu$  is sufficiently small, then  $B_j(Y^\mu(t)) \leq 0$  for all  $t \leq \sigma'$  with high probability so that  $Y^\mu(t \wedge \sigma')$  is a supermartingale. Therefore,

$$P\left(\max_{t_0 \leq t \leq \Delta_0} Y_j^\mu(t \wedge \sigma') > \zeta(1 + \xi/4) \mid Y_j^\mu(t_0 \wedge \sigma') \leq \zeta\xi/2\right) \leq \frac{\xi/2}{1 + \xi/4} \leq \xi/2.$$

Since  $\zeta, \xi$  were arbitrary, this proves that  $Y_j(t) \rightarrow y_j(t)$  in probability uniformly on  $[0, \Delta_\varepsilon \wedge \sigma']$ . But since  $y_j(t) \leq \alpha - 2\eta$  for all  $t \leq \Delta_\varepsilon$  and  $P(\sigma_0 \leq \sigma_1) \rightarrow 0$  by Lemma 4.3, we can extend this convergence to  $[0, \Delta_\varepsilon \wedge \sigma_2(j)]$ , completing the proof.

The result for  $j > k$  follows from Lemma 4.3. □

**4.2. Birth of a New Type.** In this section, we prove Proposition 3. Note that  $y_k(\Delta_\varepsilon) = 1 - \varepsilon$  for small  $\varepsilon$  since  $\Delta_0 = \delta_k^0$  and choose  $\bar{\varepsilon} = \bar{\varepsilon}(y^0)$ , small enough so that the limiting dynamical system satisfies  $y_j(\Delta_0 + t) < \alpha + t\rho/\gamma - 2\eta$ ,  $j \neq m$  and  $y_{j-1}(t) - y_j(t) < 1 - 2\eta$  for all  $j \geq 0$ ,  $t \leq \bar{\varepsilon}$ , and  $\eta$  sufficiently small. Since the result for  $j \neq k$  follows from the arguments used to prove Proposition 2, we only need to prove (2.1). To explain these inequalities, we note that our limiting process has  $y_{k+1}(\Delta_0 + t) = \lambda_{k+1-m}t/\gamma$  and  $y_{k+1}(\Delta_0 - t) = 0$  for small  $t$ . But when  $t$  is small, the number of type  $k+1$  individuals is small and deterministic approximations are not valid. The best we can do is say that  $Y_{k+1}^\mu(t)$  cannot get too far above the line with slope  $\lambda_{k+1-m}/\gamma$  that starts just before time  $\Delta_0$  (the first inequality in (2.1)) or too far below the line with slope  $\lambda_{k+1-m}/\gamma$  that starts just after time  $\Delta_0$  (the second inequality).

We begin by defining branching processes  $Z_{k+1,a}^\mu(t)$ ,  $a = u, \ell$  with initial populations  $Z_{k+1,a}^\mu(0) = 0$  and per particle birth and death rates given by Table 6, but with immigrations at rate  $e^{\lambda_{k-m}t}$ . The methods used in the proof of the next result closely parallel the methods used to prove Lemmas 3.5- 3.7 in Section 3.

LEMMA 4.4. *Let  $0 < t_1 < t_2$ . Then*

$$Y_{k+1,a}^\mu(t) \equiv (1/L) \log^+ Z_{k+1,a}(Lt/\gamma) \rightarrow t\lambda_{k+1-m}/\gamma$$

*in probability uniformly on  $[t_1, t_2]$  as  $\mu \rightarrow 0$  for  $a = u, \ell$ .*

PROOF. We prove the result for  $a = u$  and drop the subscripts  $u$  on all quantities. For ease of notation, we will also write  $Z(t) = Z_{k+1}^\mu(t)$  but leave the  $\mu$  superscript on  $\lambda_{k+1}^\mu$  to distinguish it from  $\lambda_{k+1} = (1 + \rho)(1 + \gamma)^k - 1$ . Notice that

$$\lambda_{k+1}^\mu = \frac{(1 + \rho)(1 + \gamma)^{j-m}}{1 + \gamma_{-m} m \mu^\eta} - \frac{1 - (k + 1)\mu^\eta}{1 + \gamma_{k+1-m}(k + 1 - m)\mu^\eta} \rightarrow \lambda_{k+1-m}$$

as  $\mu \rightarrow 0$ .

Define  $M(t) = e^{-\lambda_{k+1}^\mu t} Z(t) - E(e^{-\lambda_{k+1}^\mu t} Z(t))$ . Then the same argument in the proof of Lemma 3.5 implies that  $M(t)$  is a martingale (with respect to the  $\sigma$ -algebra generated by  $Z(s), s \leq t$ ). Furthermore, we have

$$(4.4) \quad E(e^{-\lambda_{k+1}^\mu t} Z(t)) = (1 - e^{-(\lambda_{k+1}^\mu - \lambda_{k-m})t}) / (\lambda_{k+1}^\mu - \lambda_{k-m})$$

and a similar argument to the one used to prove (3.7) in Section 1 implies that

$$\text{var}(e^{-\lambda_{k+1}^\mu t} Z(t)) \leq C.$$

From the  $L_2$  maximum inequality and Chebyshev's inequality we can conclude that

$$P\left(\sup_{0 \leq s \leq t_2} M(s) > L^{1/2}\right) \rightarrow 0$$

as  $\mu \rightarrow 0$ . This yields a result analogous to Lemma 3.6 in Section 3. The conclusion of Lemma 4.4 then follows using the same argument in the proof of Lemma 3.7 since  $\lambda_{k+1}^\mu \rightarrow \lambda_{k+1-m} > \lambda_{k-m}$  as  $\mu \rightarrow 0$  and so (4.4) implies that there exists  $c_1, c_2 > 0$  so that

$$c_1 \leq \sup_{t \leq s} E(e^{-\lambda_{k+1}^\mu s} Z(s)) \leq c_2.$$

for all  $t > 0$  if  $\mu$  sufficiently small. □

*Proof of Proposition 3.* Suppose that  $\varepsilon < \varepsilon_1 = \bar{\varepsilon} \wedge 1/(2(\lambda_{k+1-m}/\lambda_{k-m} + \lambda_{k+1-m}/\gamma))$ . Let  $\eta > 0$  and define,

$$\begin{aligned} A_1 &= \{|Y_k^\mu(t) - y_k(t)| \leq \varepsilon/4, Y_j^\mu(t) < \alpha + \rho t - \eta, \forall j \leq k, j \neq m, t \leq \Delta_0 + \varepsilon\} \\ A_2 &= \{T_{k+1}^\mu \geq \Delta_{3\varepsilon/4} L/\gamma\} \\ A_3 &= \{T_{k+2}^\mu \geq (\Delta_0 + \varepsilon)L/\gamma, Y_{k+1}^\mu(t) < 1 - \eta, \forall t \leq \Delta_0 + \varepsilon\} \\ A &= A_1 \cap A_2 \cap A_3. \end{aligned}$$

Note that  $P(A_1^c) \rightarrow 0$  by Proposition 2 and Proposition 3 applied to  $j \neq k$  while  $P(A_2^c) \rightarrow 0$  by Lemma 4.3. Recalling that

$$y_k(\Delta_{3\varepsilon/4} + t) = 1 - 3\varepsilon/4 + t\lambda_{k-m}/\gamma,$$

we have

$$X_k^\mu(L(\Delta_{3\varepsilon/4} + t)/\gamma)e^{-\lambda_{k-m}Lt/\gamma} \leq 1/\mu$$

for all  $t \leq (\Delta_0 + \varepsilon) - \Delta_{3\varepsilon/4}$  on  $A$ . Therefore, using the bounds on the birth and death rates given in Lemma 3.2, we can couple  $X_{k+1}^\mu(L(\Delta_{3\varepsilon/4} + t)/\gamma)$  with  $Z_{k+1,u}^\mu(Lt/\gamma)$ ,  $a = u$  in a similar manner to Lemma 3.3 so that on  $A$ , we have

$$X_{k+1}^\mu(L(\Delta_{3\varepsilon/4} + t)/\gamma) \leq Z_{k+1,u}^\mu(Lt/\gamma)$$

for all  $t \leq \Delta_0 + \varepsilon - \Delta_{3\varepsilon/4}$ . Lemma 4.4 then yields the first limit in (2.1) provided we are on  $A$ . But then

$$P(\{Y_{k+1}^\mu(t) \leq 1 - \eta, \forall t \leq \Delta_0 + \varepsilon\} \cap A) \rightarrow 1$$

and it follows from this that  $P(A_3^c) \rightarrow 0$  for small  $\eta$  which proves the first limit in (2.1) holds. To prove the second limit, we use the fact that

$$y_k(\Delta_0 + (\varepsilon/4)(\gamma/\lambda_{k-m}) + t) = 1 + \varepsilon/4 + t\lambda_{k-m}/\gamma,$$

to conclude that

$$X_k^\mu(L(\Delta_0 + (\varepsilon/4)(\gamma/\lambda_{k-m}) + t)/\gamma)e^{-\lambda_{k-m}Lt/\gamma} \geq 1/\mu$$

for all  $t \leq \varepsilon(1 - \gamma/(4\lambda_{k-m}))$  on  $A$ . Hence, we can couple  $X_{k+1}^\mu(L(\Delta_0 + (\varepsilon/4)(\gamma/\lambda_{k-m}) + t)/\gamma)$  with  $Z_{k+1,\ell}^\mu(Lt/\gamma)$  so that

$$X_{k+1}^\mu(L(\Delta_0 + (\varepsilon/4)(\gamma/\lambda_{k-m}) + t)/\gamma) \geq Z_{k+1,\ell}^\mu(Lt/\gamma)$$

for all  $t \leq \varepsilon(1 - \gamma/(4\lambda_{k-m}))$  on  $A$  and the second part of (2.1) again follows from Lemma 4.4.  $\square$

**4.3. Change in Dominant Type.** In this section, we prove Proposition 4. We begin with some notation. Let

$$\hat{y}_j(t) \equiv \begin{cases} (y_j^0 + \lambda_{j-m}t/\gamma)^+ & \text{if } j \leq k \\ 0 & \text{if } j > k \end{cases}.$$

Note that  $y_j(t) \leq \hat{y}_j(t)$  for all  $t$  with equality if  $t \leq \Delta_0$ . As in Section 4.1, let  $a_j(t) = \alpha + \rho t$  if  $j \neq k$ ,  $a_k = 1$  and choose  $\bar{\varepsilon} = \bar{\varepsilon}(y^0)$  and  $\eta > 0$  so that (i)  $\hat{y}_j(t) < a_j(t) - 2\eta$  for all  $t \leq \Delta_0 + \bar{\varepsilon}$ ,  $j \neq m, n$  (ii)  $y_{j-1}(t) - y_j(t) < 1 - 2\eta$  for all  $t \leq \Delta_0 + \bar{\varepsilon}$ , and (iii)  $y_j(t) \geq \alpha + \rho t - \eta/4$  for all  $\Delta_{\bar{\varepsilon}} \leq t \leq \Delta_0 + \bar{\varepsilon}$ ,  $j = m, n$ . Without loss of generality, suppose that  $\eta < (\alpha + \Delta_{\bar{\varepsilon}}\rho/\gamma)/4$ . Let  $\sigma_i(j)$ ,  $i = 0, 1, 2, 3$  be as in Section 4.1 and set  $\bar{\sigma}(j) = (\min_{j \neq m, n} \sigma_1(j)) \wedge \sigma_2(j) \wedge \sigma_3(j)$ . Our first lemma sets the stage for the battle between types  $m$  and  $n$  by showing that all the other types remain smaller than these two.

LEMMA 4.5. *If  $j \neq m, n$  and  $\xi > 0$  then*

$$P \left( \sup_{t \leq \bar{\sigma}(j)} (Y_j^\mu(t) - \hat{y}_j(t)) > \xi \right) \rightarrow 0.$$

PROOF. This follows directly from Lemma 4.1 and (I)-(IV) in Lemma 4.2.  $\square$

Note that Lemma 4.5, Lemma 2.1, and our choice of  $\bar{\varepsilon}$  imply that

$$(4.5) \quad \sup_{t \leq \Delta_0 + \bar{\varepsilon}} \frac{X_j(Lt/\gamma)}{N^\mu(t)} \leq \mu^\eta$$

for all  $j \neq n, m$  with high probability. Furthermore, we can control the time that type  $j > k$  individuals begin to be born.

LEMMA 4.6.  *$P(T_{k+1}^\mu \leq (\Delta_0 + \bar{\varepsilon})L/\gamma) \rightarrow 0$  as  $\mu \rightarrow 0$*

PROOF. Once we note that Lemma 4.5 and our choice of  $\bar{\varepsilon}$  imply that  $X_k(tL/\gamma) < (1/\mu)^{1-\eta}$  for all  $t \leq \Delta_0 + \bar{\varepsilon}$  with high probability, the rest of the proof is the same as the proof of Lemma 4.3.  $\square$

Let

$$R_j^\mu(t) \equiv \frac{X_j^\mu(L\Delta_\varepsilon/\gamma + t)}{N^\mu(t)}$$

be the fraction of  $j$ 's in the population at times greater than  $L\Delta_\varepsilon/\gamma$ . Then as a consequence of (4.5) and Lemma 4.3, we have

$$(4.6) \quad 0 \leq 1 - (R_m(Lt/\gamma) + R_n(Lt/\gamma)) \leq (k+1)\mu^\eta$$

for all  $t \leq \Delta_0 + \bar{\varepsilon} - \Delta_\varepsilon$  on a set  $A$  with  $P(A^c) \rightarrow 0$ . Our next result concerns the change of power from  $m$ 's to  $n$ 's. To state the result, let

$$f(r) \equiv r(1-r) \frac{\lambda_{n-m}}{1 + \gamma_{n-m}r}$$

and define  $r_j^\mu(t)$ ,  $j = n, m$  as the (random) solutions to the initial value problem

$$\begin{aligned} \frac{dr_m^\mu}{dt} &= -f(1 - r_m^\mu) \equiv f_m(r_m^\mu) \\ \frac{dr_n^\mu}{dt} &= f(r_n^\mu) \equiv f_n(r_n^\mu) \end{aligned}$$

with initial conditions  $r_j^\mu(0) = R_j^\mu(0)$ ,  $j = n, m$ .

LEMMA 4.7. *There exists an  $\varepsilon_2 = \varepsilon_2(y^0) > 0$  so that for  $j = n, m$ ,*

$$P \left( \sup_{t \leq \Delta_0 - \Delta_\varepsilon + \varepsilon} \left| R_j^\mu(Lt/\gamma) - r_j^\mu(Lt/\gamma) \right| > \mu^{\eta/2} \right) \rightarrow 0$$

as  $\mu \rightarrow 0$  for all  $\varepsilon < \varepsilon_2$ .

PROOF. We will prove the result by calculating infinitesimal means and variances. Without loss of generality, we assume that  $\varepsilon < \bar{\varepsilon}$  so that by (4.5), Lemma 4.6, and (4.6), we have

$$(4.7) \quad \sum_{j \neq n, m} R_j^\mu(t) \leq (k+1)\mu^\eta$$

$$0 \leq 1 - (R_n^\mu(t) + R_m^\mu(t)) \leq (k+1)\mu^\eta$$

for all  $t \leq L(\Delta_0 + \varepsilon - \Delta_\varepsilon)/\gamma$  on a set  $A$  with  $P(A^c) \rightarrow 0$  as  $\mu \rightarrow 0$ . Note also that Lemma 2.1, the fact that  $N^\mu$  is non-decreasing, and our choice of  $\eta < (\alpha + \Delta_\varepsilon \rho/\gamma)/4$  imply that

$$(4.8) \quad N^\mu(t) \geq C(1/\mu)^{\alpha + \Delta_\varepsilon \rho/\gamma - \eta} \geq C(1/\mu)^{3\eta}, \quad \forall t \geq L\Delta_\varepsilon/\gamma$$

on a set  $A$  with  $P(A^c) \rightarrow 0$  as  $\mu \rightarrow 0$ . We will therefore assume that the inequalities in (4.7) and (4.8) hold for the remainder of the proof and write  $O(\mu^\eta)$  for any quantity whose absolute value is bounded above by  $C\mu^\eta$  uniformly for  $t \leq L(\Delta_0 + \varepsilon - \Delta_\varepsilon)/\gamma$  on a set  $A$  with  $P(A^c) \rightarrow 0$  as  $\mu \rightarrow 0$ . It is also convenient to write

$$c_\varepsilon = \Delta_0 + \varepsilon - \Delta_\varepsilon = (1 + \gamma(\lambda_{n-m} - \rho))^{-1}\varepsilon.$$

By looking at the rates for the chain  $(N^\mu(t), X^\mu(t))$ , the fraction  $R_j^\mu(t)$  has the following jump rates corresponding to the events  $x_j/N \mapsto (x_j + 1)/N$ ,  $x_j/N \mapsto (x_j - 1)/N$ ,  $x_j/N \mapsto (x_j + 1)/(N + 1)$ , and  $x_j/N \mapsto x_j/(N + 1)$ , respectively:

$$\begin{aligned} r_j \mapsto r_j + 1/N & \quad \text{rate:} \quad N(1 - r_j) \frac{(1 + \gamma)^j r_j}{w} + \mu N r_{j-1} \\ r_j \mapsto r_j - 1/N & \quad \text{rate:} \quad N r_j \frac{w - (1 + \gamma)^j r_j}{w} + \mu N r_j \\ r_j \mapsto r_j + (1 - r_j)/(N + 1) & \quad \text{rate:} \quad \rho N \frac{(1 + \gamma)^j r_j}{w} \\ r_j \mapsto r_j - r_j/(N + 1) & \quad \text{rate:} \quad \rho N \frac{w - (1 + \gamma)^j r_j}{w} \end{aligned}$$

where  $w \equiv \sum_{i \geq 0} (1 + \gamma)^i r_i$ . From these expressions for the rates, we can see that the infinitesimal mean of  $R_j^\mu$  is given for  $r \in \mathcal{S}^N/N$  by

$$\begin{aligned} B_j(r) &= \frac{(1 + \gamma)^j r_j (1 - r_j)}{w} + \mu r_{j-1} - \left( \frac{r_j (w - (1 + \gamma)^j r_j)}{w} + \mu r_j \right) \\ &+ \frac{\rho N}{N + 1} \frac{(1 + \gamma)^j r_j (1 - r_j)}{w} - \frac{\rho N}{N + 1} \frac{r_j (w - (1 + \gamma)^j r_j)}{w} \\ &= \left( 1 + \frac{\rho N}{N + 1} \right) \left( \frac{r_j ((1 + \gamma)^j - w)}{w} \right) + \mu (r_{j-1} - r_j). \end{aligned}$$

Similarly the infinitesimal variance is given by

$$\begin{aligned}
A_j(r) &= \frac{1}{N} \left( \frac{(1+\gamma)^j r_j (1-r_j)}{w} + \mu r_{j-1} + \frac{r_j (w - (1+\gamma)^j r_j)}{w} + \mu r_j \right) \\
&\quad + \frac{\rho N}{(N+1)^2} \left( \frac{(1+\gamma)^j r_j (1-r_j)^2}{w} + \frac{r_j^2 (w - (1+\gamma)^j r_j)}{w} \right) \\
&= \frac{1}{N} \left( \left( 1 + \frac{\rho N^2}{(N+1)^2} \right) \frac{r_j ((1+\gamma)^j - 2(1+\gamma)^j r_j + w)}{w} \right. \\
&\quad \left. - \frac{\rho r_j (1-r_j) N^2}{(N+1)^2} + \mu (r_{j-1} + r_j) \right)
\end{aligned}$$

where in the second line we have added and subtracted  $\rho r_j N / (N+1)^2$  from the first. (4.7) implies that

$$\begin{aligned}
w(R^\mu(s)) &= (1+\gamma)^m R_m^\mu(s) + (1+\gamma)^n R_n^\mu(s) + O(\mu^\eta) \\
&= (1+\gamma)^m [1 + \gamma_{n-m} r_n^\mu(s)] + O(\mu^\eta)
\end{aligned}$$

for all  $s \leq Lc_\varepsilon/\gamma$  and hence  $B_n(R^\mu(s))$

$$\begin{aligned}
&= \left( 1 + \frac{\rho N^\mu (L\Delta_\varepsilon/\gamma + s)}{N^\mu (L\Delta_\varepsilon/\gamma + s) + 1} \right) R_n^\mu(s) \left( \frac{(1+\gamma)^n - (1+\gamma)^m [1 + \gamma_{n-m} R_n^\mu(s)] + O(\mu^\eta)}{(1+\gamma)^m [1 + \gamma_{n-m} R_n^\mu(s)] + O(\mu^\eta)} \right) \\
&\quad + O(\mu^\eta) \\
&= \left( 1 + \frac{\rho N^\mu (L\Delta_\varepsilon/\gamma + s)}{N^\mu (L\Delta_\varepsilon/\gamma + s) + 1} \right) R_n^\mu(s) \left( \frac{\gamma_{n-m} - \gamma_{n-m} R_n^\mu(s)}{1 + \gamma_{n-m} R_n^\mu(s)} \right) + O(\mu^\eta) \\
&= f_n(R_n^\mu(s)) + O(\mu^\eta)
\end{aligned}$$

for all  $s \leq Lc_\varepsilon/\gamma$ , the last equality following from (4.8) and the definition of  $f_n$ . Similarly, writing

$$\begin{aligned}
w(R^\mu) &= (1+\gamma)^m R_m^\mu + (1+\gamma)^n R_n^\mu + O(\mu^\eta) \\
&= (1+\gamma)^m [1 + \gamma_{n-m} (1 - R_m^\mu)] + O(\mu^\eta)
\end{aligned}$$

we obtain

$$B_m(R^\mu(s)) = f_m(R_m^\mu(s)) + O(\mu^\eta)$$

for all  $s \leq Lc_\varepsilon/\gamma$ .

Now we know that

$$M_j^\mu(s) \equiv R_j^\mu(s) - \int_0^s B_j^\mu(R^\mu(u)) du$$

and

$$(M_j^\mu)^2(s) - \int_0^s A_j^\mu(R^\mu(u)) du$$

are both martingales for all  $i, j$ . By the  $L^2$  maximal inequality, (4.8), and the fact that  $r_j \in [0, 1]$ , we have  $A_j^\mu(R^\mu(s)) \leq C\mu^{3\eta}$  for all  $s \leq L\Delta_\varepsilon/\gamma$  yielding

$$\sup_{s \leq T} E |M_j^\mu|^2(s) \leq CT\mu^{3\eta}$$



for any  $j$  and so Chebyshev's inequality implies that for any  $a > 0$ ,

$$P \left( \sup_{s \leq T} |M_j^\mu(s)| > a \right) \leq \frac{CT\mu^{3\eta}}{a^2}.$$

Therefore as  $\mu \rightarrow 0$

$$P \left( \sup_{s \leq Lc_\varepsilon/\gamma} |M_j^\mu(s)| > \mu^\eta \right) \leq CL\mu^\eta \rightarrow 0.$$

But on the set where  $\sup_{s \leq T} |M_j^\mu(s)| \leq \mu^\eta$ , using the fact that  $|f'(r_j)| \leq \gamma(1 + \gamma)$  for all  $r_j \in [0, 1]$ , we have

$$\begin{aligned} |R_j^\mu(s) - r_j^\mu(s)| &\leq |M_j^\mu(s)| + \int_0^s |B_j^\mu(R^\mu(u)) - f_j^\mu(R_j^\mu(u))| du \\ &\quad + \int_0^s |f_j^\mu(R_j^\mu(u)) - f_j(r_j^\mu(u))| du \\ &\leq \mu^\eta + Ct\mu^\eta + \gamma(1 + \gamma) \int_0^s |R_j^\mu(u) - r_j^\mu(u)| du \end{aligned}$$

for any  $j \geq 0$ ,  $s \leq T$  so that Gronwall's inequality gives

$$\sup_{s \leq T} |R_j^\mu(s) - r_j^\mu(s)| \leq \mu^\eta(1 + CT)e^{\gamma(1+\gamma)T}.$$

Substituting  $T = Lc_\varepsilon/\gamma$  into this expression yields

$$\sup_{s \leq L(\Delta_0 - \Delta_\varepsilon + \varepsilon)/\gamma} |R_j^\mu(s) - r_j^\mu(s)| \leq \mu^{\eta - (1+\gamma)c_\varepsilon}(1 + CL\varepsilon) \leq \mu^{\eta/2}$$

for all  $\mu < \mu_0$  provided  $\varepsilon$  is sufficiently small (how small depends on  $\eta$ , which in turn depends on  $y^0$ ). This completes the proof  $\square$

The next step is to analyze the differential equations for  $j = m, n$  in Lemma 4.7. We will do the analysis for  $j = n$ . For  $j = m$ , apply the below analysis to  $1 - r_m^\mu(s)$ . To begin, write

$$r_n^\mu(t) = \frac{X_n^\mu(L\Delta_\varepsilon/\gamma)}{N^\mu(L\Delta_\varepsilon/\gamma)} \exp \left\{ \int_0^t g_n(r_n^\mu(s)) ds \right\}$$

with

$$g_n(r) \equiv \frac{\lambda_{n-m}(1-r)}{1 + \gamma_{n-m}r}.$$

Note that we have the following set of bounds on the growth rate  $g_n$ :

$$(4.9) \quad \begin{array}{lll} (1 - L^{-2}) \frac{\lambda_{n-m}}{1 + \gamma_{n-m}L^{-2}} & \leq g_n(r_n^\mu) \leq & \lambda_{n-m} \quad \text{when } r_n^\mu < L^{-2} \\ (1 - r_n^\mu) \frac{\lambda_{n-m}}{1 + \gamma_{n-m}} & \leq g_n(r_n^\mu) \leq & (1 - r_n^\mu)\lambda_{n-m} \quad \text{when } L^{-2} \leq r_n^\mu \leq 1 - L^{-2} \\ 0 & \leq g_n(r_n^\mu) \leq & L^{-2}\lambda_{n-m} \quad \text{when } r_n^\mu \geq 1 - L^{-2} \end{array}$$

LEMMA 4.8. *Let  $s_1^\mu = \inf\{s : r_n^\mu(s) \geq L^{-2}\}$  and  $s_2^\mu = \inf\{s : r_n^\mu(s) \geq 1 - L^{-2}\}$  then we have  $s_i^\mu/(L/\gamma) \rightarrow \Delta_0$  for  $i = 1, 2$  and  $(s_2^\mu - s_1^\mu)/L \rightarrow 0$  as  $\mu \rightarrow 0$ .*

PROOF. Let

$$r_\ell^\mu(s) = \frac{X_n^\mu(L\Delta_\varepsilon/\gamma)}{N^\mu(L\Delta_\varepsilon/\gamma)} e^{\lambda_{n-m} c_\mu s} \quad \text{and} \quad r_u^\mu(s) = \frac{X_n^\mu(L\Delta_\varepsilon/\gamma)}{N^\mu(L\Delta_\varepsilon/\gamma)} e^{\lambda_{n-m} s}$$

where  $c_\mu = (1 - L^{-2})(1 + \gamma_{n-m} L^{-2})^{-1}$ . Then it is clear from the first bound in (4.9) that

$$r_\ell^\mu(s) \leq r_n^\mu(s) \leq r_u^\mu(s)$$

for all  $s \leq s_1^\mu$ . Since  $Y_n(\Delta_\varepsilon) \rightarrow y_n(\Delta_\varepsilon)$ ,  $F^\mu(\Delta_\varepsilon) \rightarrow \alpha + \Delta_\varepsilon \rho/\gamma$  by Proposition 2, letting  $s_\ell^\mu$  and  $s_u^\mu$  be the times that  $r_\ell^\mu$  and  $r_u^\mu$  hit  $L^{-2}$ , we have  $s_a^\mu/(L/\gamma) \rightarrow \Delta_0$  as  $\mu \rightarrow 0$  for  $a = \ell, u$  which proves the result for  $i = 1$ . To prove the result for  $i = 2$ , we use the bounds in the second line of (4.9) along with the fact that the logistic  $dx/dt = \beta x(1 - x)$  rises from  $L^{-2}$  to  $1 - L^{-2}$  in time  $(4/\beta) \log L$  to conclude that

$$\frac{s_2^\mu - s_1^\mu}{L} \leq \frac{C \log L}{L} \rightarrow 0$$

as  $\mu \rightarrow 0$  which completes the proof.  $\square$

LEMMA 4.9.

$$(1/L) \log^+[N^\mu(Lt/\gamma)r_j^\mu(Lt/\gamma)] \rightarrow y_j(\Delta_\varepsilon + t)$$

uniformly on  $[0, T]$  for any  $T > 0$ ,  $j = n, m$ .

PROOF. We prove the result for  $j = n$ . Write

$$(1/L) \log^+[N^\mu(Lt/\gamma)r_n^\mu(Lt/\gamma)] - y_n(\Delta_\varepsilon + t) = [Y_n^\mu(\Delta_\varepsilon) - y_n(\Delta_\varepsilon)] \\ + (1/L) \left[ \int_0^{Lt/\gamma} (g_n(r_n^\mu(s)) - \ell_n(s)) ds \right]$$

where  $\ell_n(s) = \lambda_{n-m} \mathbf{1}_{s \leq (\Delta_0 - \Delta_\varepsilon)}$ . The first term in brackets converges to 0 in probability by Proposition 2. To control the second term, split up the integral as

$$\int_0^{tL/\gamma} = \int_0^{s_1^\mu \wedge t} + \int_{s_1^\mu \wedge t}^{s_2^\mu \wedge t} + \int_{s_2^\mu \wedge t}^t.$$

Using the bounds in (4.9) and applying Lemma 4.8, we conclude that each of these integrals is  $o(L)$  which yields the result.  $\square$

*Proof of Proposition 4.* Let  $\varepsilon < \varepsilon_2$  and suppose first that  $j = m, n$ . Writing  $Y_j^\mu(t) - y_j(t)$

$$= (1/L) \log^+ R_j^\mu(Lt/\gamma) - (1/L) \log^+ r_j^\mu(Lt/\gamma) + (1/L) \log^+[N^\mu(Lt/\gamma)r_j^\mu(Lt/\gamma)] - y_n(t),$$

we can see that since

$$r_j^\mu(0) \geq \frac{X_n^\mu(L\Delta_\varepsilon/\gamma)}{N^\mu(L\Delta_\varepsilon/\gamma)} \geq \mu^{\eta/3}$$

for all  $t \geq 0$  with high probability as  $\mu \rightarrow 0$  by our choice of  $\bar{\varepsilon}$ , the result follows from Lemmas 4.7 and 4.9. Suppose now that  $j \neq m, n$ . If  $j > k$ , the result follows from Lemma

4.6 so it remains to prove the result for  $j \leq k$ . In view of Lemma 4.1, it suffices to show that

$$P \left( \sup_{t \leq (\Delta_0 + \varepsilon) \wedge \bar{\sigma}(j)} \left| y_j^0 + \int_0^t B_j(Y_j^\mu(s)) ds - y_j(t) \right| > \xi \right) \rightarrow 0$$

and then follow the argument from the proof of Proposition 2 to yield the result. But now that we have proved Proposition 4 holds for  $j = m, n$ , we can argue as in the proof of Lemma 4.2 to conclude that

$$g_{j,1}(Y^\mu(t)) \rightarrow \begin{cases} \frac{(1+\rho)(1+\gamma)^{j-m}}{\gamma} & 0 \leq t < \Delta_0 \\ \frac{(1+\rho)(1+\gamma)^{j-n}}{\gamma} & \Delta_0 < t \leq (\Delta_0 + \varepsilon) \end{cases}$$

uniformly on compact subsets of  $[0, (\Delta_0 + \varepsilon) \wedge \bar{\sigma}(j)] - \{\Delta_0\}$ . This replaces the second part of (IV) from the proof of Lemma 4.2 and the result follows after using (I)-(III) from the proof of Lemma 4.2. □

## REFERENCES

- [1] Armitage, P., and Doll, R. (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *Brit. J. Cancer.* 8, 1–12
- [2] Athreya, K., and Ney, P.E. (1972) *Branching Processes*. Springer-Verlag, New York
- [3] Beerenwinkel, N., Antal, T., Dingli, D. Traulsen, A., Kinzler, K.W., Velculescu, V., Vogelstein, B., and Nowak, M.A. (2007) Genetic progression and the waiting time to cancer. *PLoS Comp. Bio.* 3, 2239–2246
- [4] Brunet, E., Rouzine, I., and Wilke, C. (2008) The stochastic edge in adaptive evolution. *Genetics*, 179, 603–620.
- [5] Calabrese, P., Mecklin, J-P., Järvinen, H.J., Aaltonen, L.A., Tavaré, S.A., and Shibata, D. (2005) Numbers of mutations to different types of colorectal cancer. *BMC Cancer* 5: paper 126
- [6] Desai, M. and Fisher, D. (2007) Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*, 176:175998.
- [7] Durrett, R. (2008) *Probability Models of DNA Sequence Evolution*. 2nd edition. Springer-Verlag, New York
- [8] Durrett, R. and Moseley, S. (2009) Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor. Pop. Biol.*, to appear.
- [9] Durrett, R. and Schmidt, D. (2008) Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics* 180, 1501–1509
- [10] Durrett, R., Schmidt, D., and Schweinsberg, J. (2009). A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Prob.* 19, 676–718
- [11] Frank, Steven A. (2007) *Dynamics of Cancer: Incidence, Inheritance and Evolution*. Princeton Series in Evolutionary Biology
- [12] Haeno, H., Iwasa, Y., and Michor, F. (2007) The evolution of two mutations during clonal expansion. *Genetics*. 177, 2209–2221
- [13] Iwasa, Y., Nowak, M.A., and Michor, F. (2006) Evolution of resistance during clonal expansion. *Genetics*. 172, 2557–2566
- [14] Jones, S. et al. (2008) Core signalling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, 321, 1801–1806
- [15] Knudson, A.G. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* 68, 820–823
- [16] Knudson, A.G. (2001) Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*. 1, 157–162
- [17] Multistage carcinogenesis and the incidence of colorectal cancer. *Proc. Natl. Acad. Sci. USA* 99, 15095–15100

- [18] Muller, H.J. (1932). Some genetic aspects of sex. *Am Natur* 66, 118-138.
- [19] Parsons, D.W., et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321, 1807–1812
- [20] Rouzine, I.M., Brunet, E., and Wilke, C.O. (2008) The traveling-wave approach to asexual evolution: Muller’s ratchet and speed of adaptation. *Theor. Popul. Biol.* 73, 24–46.
- [21] Rouzine, I.M., Wakeley, J., and Coffin, J.M. (2003) The solitary wave of asexual evolution. *Proc Natl Acad Sci* 100, 587–592
- [22] Schweinsberg, J. (2008) The waiting time for  $m$  mutations. *Electron. J. Probab.* 13, 1442–1478
- [23] Sjoblom, et al. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274
- [24] The Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068
- [25] Wodarz, D., and Komarova, N.L. (2005) *Computational Biology of Cancer*. World-Scientific.
- [26] Wood, L., et al. (2007) The genomic landscape of human breast and colorectal cancers. *Science* 318, 1108–1113
- [27] Yu, F. and Etheridge, A. (2008) Rate of adaptation of large populations. *Evolutionary Biology from Concept to Application*. Springer Berlin Heidelberg, 3–27
- [28] Yu, F., Etheridge, A., and Cuthbertson, C. (2009) Asymptotic Behaviour of the Rate of Adaptation, arXiv:0708.3453v3

RICK DURRETT  
DEPARTMENT OF MATHEMATICS  
CORNELL UNIVERSITY  
310 MALOTT HALL  
ITHACA, NY 14853-4201, U.S.A.  
E-MAIL: rtd1@cornell.edu

JOHN MAYBERRY  
DEPARTMENT OF MATHEMATICS  
CORNELL UNIVERSITY  
310 MALOTT HALL  
ITHACA, NY 14853-4201, U.S.A.  
E-MAIL: jm858@cornell.edu

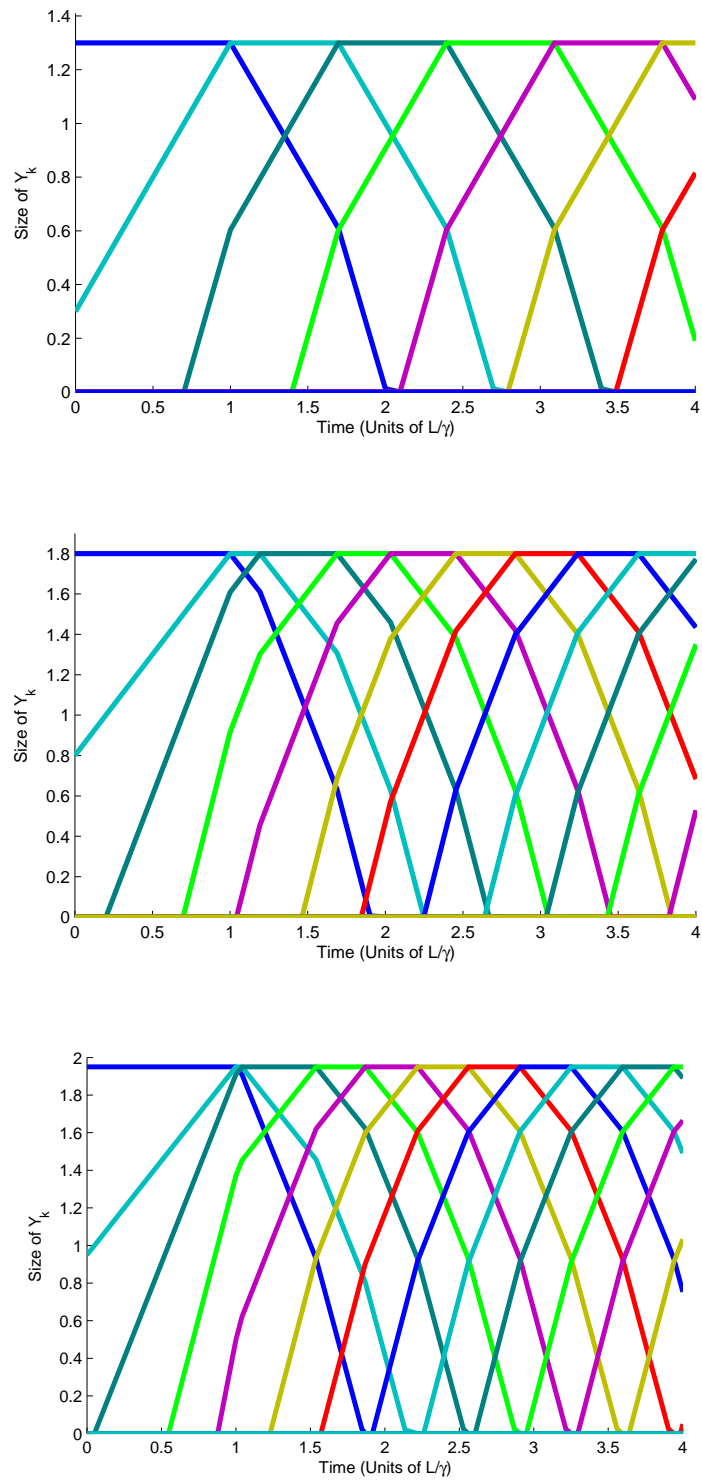


FIG 1. Plot of limiting dynamical system for  $\gamma = 0.01$ ,  $\alpha = 1.3, 1.8, 1.95$  (top to bottom).

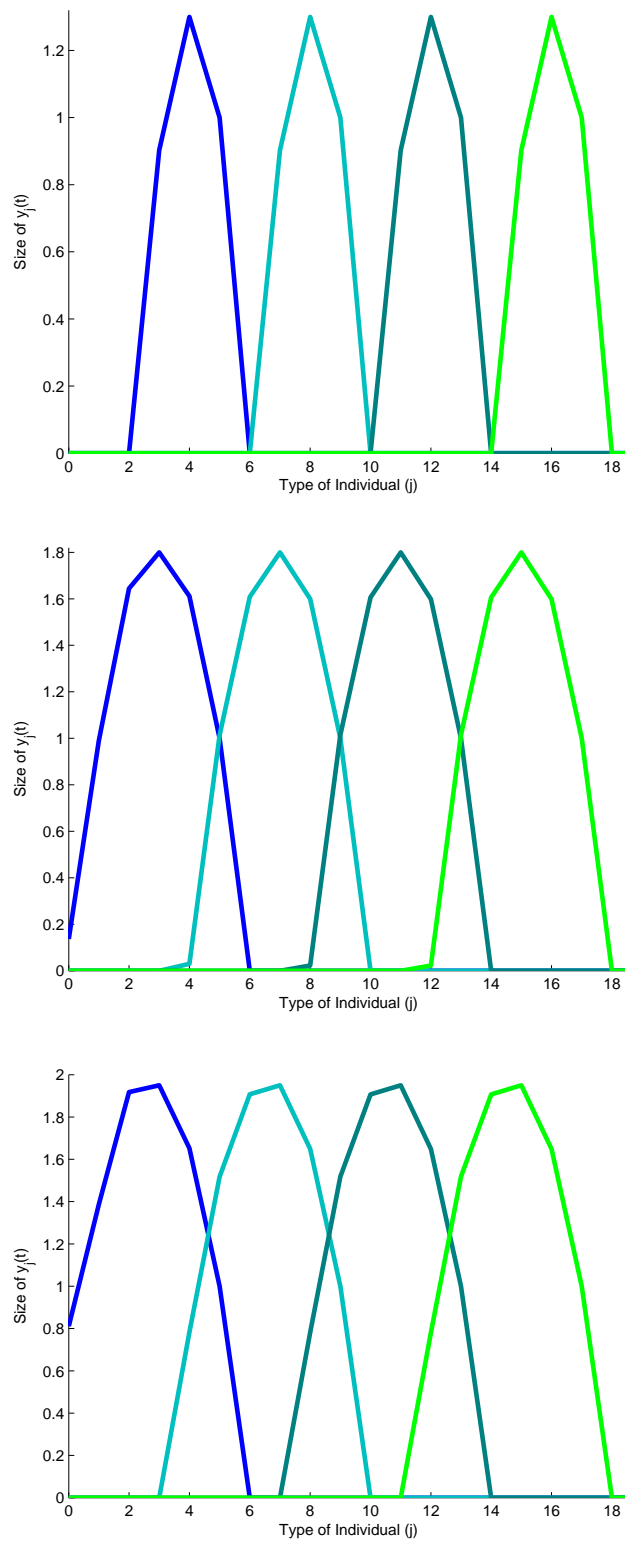


FIG 2. *Distribution of types at the times  $t_5, t_9, t_{13}$  and  $t_{17}$  (from left to right) with parameters  $\gamma = 0.01$  and  $\alpha = 1.3, 1.8, 1.95$  (top to bottom)*

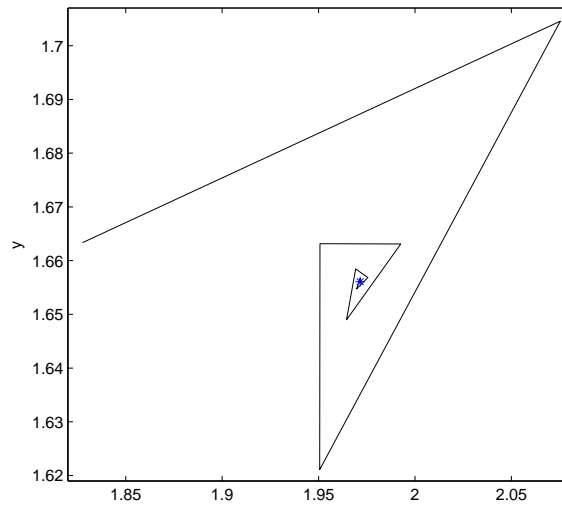


FIG 3. Successive iterates of the 2-D map  $(x, y) \mapsto f(x, y) = (f_1(x, y), f_2(x, y))$  given in Table 5 started with initial conditions  $(x_0, y_0) = (1 + \gamma/\gamma_2 + \gamma/\gamma_3, 1 + \gamma_2/\gamma_3)$  and parameters  $\alpha = 1.95$ ,  $\gamma = .01$ . The star denotes the fixed point of  $f$ .

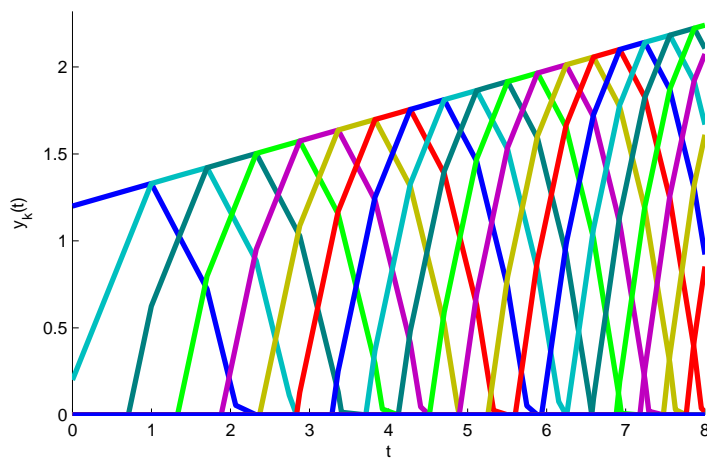


FIG 4. Plot of limiting dynamical system for growing population with  $\gamma = .01$ ,  $\rho = .0013$ ,  $\mu = 10^{-5}$ , and  $N(0) = 10^6$ . Plotted up until just after the first type 20 individual is born. Time is given in units of  $L/\gamma$ .

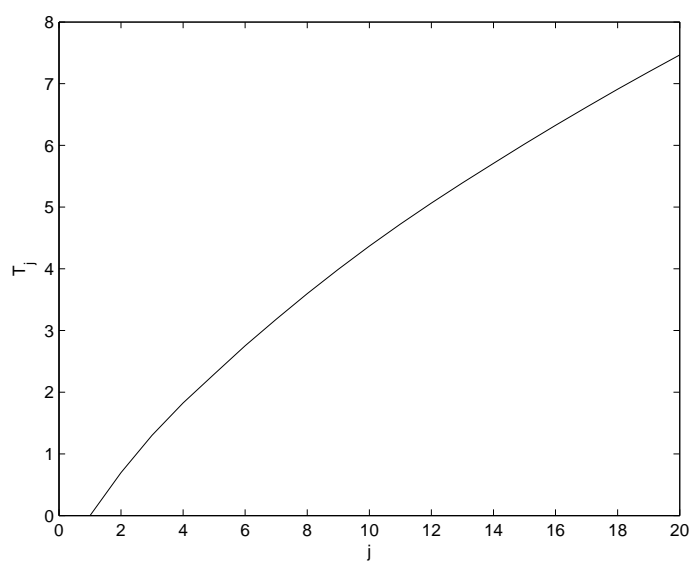


FIG 5. Plot of  $j$  vs.  $T_j^\mu$  for the limiting dynamical system. Same parameter values as Figure 4 and time is again in units of  $L/\gamma$ .