Theoretical Population Biology 151 (2023) 44–63

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

# Waiting times in a branching process model of colorectal cancer initiation

Ruibo Zhang, Obinna A. Ukogu, Ivana Bozic *

*Department of Applied Mathematics, University of Washington, United States of America*

A R T I C L E I N F O

*Article history:*
Received 13 September 2022
Available online 24 April 2023

*Keywords:*
Branching process
Waiting times
Cancer

A B S T R A C T

We study a multi-stage model for the development of colorectal cancer from initially healthy tissue. The model incorporates a complex sequence of driver gene alterations, some of which result in immediate growth advantage, while others have initially neutral effects. We derive analytic estimates for the sizes of premalignant subpopulations, and use these results to compute the waiting times to premalignant and malignant genotypes. This work contributes to the quantitative understanding of colorectal tumor evolution and the lifetime risk of colorectal cancer.



## 1. Introduction

Cancer is the result of somatic evolution during which cells accumulate driver mutations required for malignant transformation (Vogelstein and Kinzler, 2004; Jones et al., 2008). While some leukemias and pediatric cancers may be initiated with a single driver mutation, initiation of solid cancers typically requires multiple driver mutations (Vogelstein et al., 2013). Colorectal cancer (CRC) is one of the most common cancers in the United States (Siegel et al., 2021); it has been shown that mutations in three driver genes are sufficient to initiate the development of CRC (Tomasetti et al., 2015; Fearon, 2011; Paterson et al., 2020). Typically, this involves inactivation of two tumor suppressor genes and activation of an oncogene. As inactivation of a tumor suppressor gene (TSG) requires inactivation of both alleles, and activation of an oncogene only requires a mutation in one allele of the gene (Sherr, 2004), this leads to a total of five genetic alterations required for CRC initiation.

Multi-type branching processes, with types corresponding to genotypes or cell states (Antal and Krapivsky, 2011; Durrett and Moseley, 2010), have emerged as a viable model for studying cancer evolution. Multiple aspects of cancer evolution have been modeled, including initiation (Paterson et al., 2020; Meza et al., 2008), progression (Durrett and Moseley, 2010; Bozic et al., 2010; Reiter et al., 2013; Bozic et al., 2019), metastasis (Foo and Leder, 2013; Avanzini and Antal, 2019; Danesh et al., 2012), and resistance to therapy (Komarova and Wodarz, 2005; Komarova, 2006; Bozic et al., 2013; Bozic and Nowak, 2014; Nicholson and Antal, 2019). Evolutionary dynamics are closely related to the relative fitness advantages conferred by individual mutations. Durrett and Moseley (2010) analyzed a scenario in which the clonal growth rate strictly increases after each mutation, and computed distributions for clonal sizes and waiting times. Nicholson and Antal (2019) studied a general framework wherein wild-type individuals have the largest fitness (growth rate), which could be applied to cases involving drug resistance. Random fitness advantages have also been investigated (Durrett et al., 2010; Foo et al., 2014).

Recent work (Paterson et al., 2020) studied a branching process model for the initiation of colorectal cancer that involves the three most commonly mutated driver genes in colorectal cancer: tumor suppressors *APC* and *TP53* and oncogene *KRAS*. The study found that, in the majority of cases, the driver mutations accrue in a specific order, with inactivation of *APC* followed by activation of *KRAS* and inactivation of the *TP53* gene. Following Paterson et al. (2020), we study the mutational pathway to colorectal cancer in which the genetic alteration order is given by *APC*, *KRAS* and *TP53*. We model the dynamics using a multi-type branching process that starts from $N$ wild-type crypts, small tubular assemblies of cells that line the intestinal epithelium (Vermeulen et al., 2013b; Barker et al., 2009). As the process evolves, individual crypts stochastically obtain driver mutations, with mutation rates determined by the genotype of the crypt and the driver gene in question. All mutants are initially derived from a large population of non-dividing wild-type crypts through genetic alterations which may have neutral or advantageous effects on the growth rates of the resulting subpopulations.

In this work, we precisely estimate the time it takes for each altered genotype to occur, and compare the analytic results for the waiting time distributions to exact computer simulations of the process. In addition to studying the case where subsequent types are not strictly increasing in fitness, and presenting multiple approximations that can be useful for the study of waiting

* Corresponding author.
*E-mail address:* ibozic@uw.edu (I. Bozic).

https://doi.org/10.1016/j.tpb.2023.04.001
0040-5809/

**Table 1**
Pathway to CRC initiation.

| Step | Rate | Biological process |
|---|---|---|
| $N_0 \to N_1$ | $u_1$ | Inactivation of 1st copy of *APC* |
| $N_1 \to N_2$ | $u_2$ | Inactivation of 2nd copy of *APC* |
| $N_2 \to N_3$ | $u_3$ | Activation of *KRAS* |
| $N_3 \to N_4$ | $u_4$ | Inactivation of 1st copy of *TP53* |
| $N_4 \to N_5$ | $u_5$ | Inactivation of 2nd copy of *TP53* |

**Table 2**
Parameter values for the model of CRC initiation.

| (a) Non-decreasing crypt growth rates | | | | | |
|---|---|---|---|---|---|
| Crypts | $N_0$ | $N_1$ | $N_2$ | $N_3$ | $N_4$ |
| Birth rate | $\lambda_0 = 0$ | $\lambda_1 = 0$ | $\lambda_2 > 0$ | $\lambda_3 > \lambda_2$ | $\lambda_4 = \lambda_3$ |
| (b) Biologically reasonable range of parameter values | | | | | |
| Number of wild-type crypts | $N$ | $10^7 - 10^8$ | | | |
| Birth rates | $\lambda_i$ | $\lambda_2 = 0.2/y, \lambda_3 = 0.27/y$ | | | |
| Transition rates | $u_i$ | $10^{-7} - 10^{-4}/y$ | | | |

times, our work also differs from previous works in another significant way. In particular, related previous works (Durrett and Moseley, 2010; Cheek and Antal, 2018) consider the case where the model is initiated by an advantageous population that grows exponentially. In contrast, our model's initial populations do not grow in size, as they correspond to healthy tissue that has not yet collected a functional driver alteration that would lead to uncontrolled growth. We also derive exact expressions for the limiting random variables that are used in calculating waiting time distributions, and provide insight into the accuracy of approximating the size of a premalignant population with its corresponding long-time limit.

The approach presented here can be extended to other multi-type branching process models in which the growth rates of subsequent types are non-decreasing. For colorectal cancer, one can use a similar approach to compute the waiting time distributions for other mutational pathways. Our results are also applicable to other multi-hit models of carcinogenesis, as many cancer types are thought to be initiated through a multi-step process that involves inactivation of tumor suppressor genes and activation of oncogenes.

## 2. Model and parameters

Let $N_i(t)$ be the stochastic process that counts the population of type-$i$ crypts at time $t$. The process is started at time 0 with all crypts being type-0, which corresponds to healthy crypts with no driver gene mutations. A type-$i$ crypt can transform into a type-$(i+1)$ crypt by obtaining a driver alteration, which occurs at rate $u_i$. For simplicity, we consider a specific mutational pathway on the way to colorectal cancer (see Table 1) reported in recent work (Paterson et al., 2020). Type-5 crypts represent the final, malignant state.

We note that our model does not account for genetic heterogeneity within individual crypts. This simplifying assumption is a reasonable approximation, as new mutations are either lost or fixated in the crypt, resulting in crypt stabilization (Campbell et al., 1996). Crypt stabilization times have been reported to be one year or less in the colon (Campbell et al., 1996; Vermeulen et al., 2013a).

We assume that independently from mutations, type-$i$ crypts follow a pure birth process with rate $\lambda_i$. When $\lambda_i > 0$, this corresponds to crypt division (fission). The division rate of a crypt is determined by its genotype. For wild-type crypts, crypt fissions are very rare (Nicholson et al., 2018), so we set their division rates to zero. It has been shown that inactivation of *APC* and/or activation of *KRAS* provides a fitness advantage to mutated crypts, leading to clonal expansion though increased crypt division rates (Lamlum et al., 2000; Snippert et al., 2014). In contrast, under normal conditions *TP53* inactivation alone does not provide a fitness advantage (Vermeulen et al., 2013b). It was recently reported that, in addition to crypt fission, crypt fusion also occurs in human colonic crypts (Baker et al., 2019). In wild-type (healthy) tissue, crypt fission and fusion are in balance, with both being very rare (Nicholson et al., 2018; Baker et al., 2019). In mutated crypts, the rate of crypt fission increases, while the rate of crypt fusion remains very small (Olpe et al., 2021), allowing us to neglect crypt fusion. These findings are reflected in the choice of growth parameters in our model (see Table 2).

This branching process model is summarized as

$$N_0(t) \xrightarrow{u_1} N_1(t) \xrightarrow{u_2} \overset{\circlearrowleft \text{divide at rate } \lambda_2}{N_2(t)} \xrightarrow{u_3} \overset{\circlearrowleft \lambda_3}{N_3(t)} \xrightarrow{u_4} \overset{\circlearrowleft \lambda_4 = \lambda_3}{N_4(t)} \xrightarrow{u_5} N_5(t).$$

Let $Z^{(i)}$ represent a single type-$i$ crypt. All type-$i$ crypts independently follow the transition scheme:

$$Z^{(i)} \to \begin{cases} Z^{(i)}Z^{(i)}, & \text{birth rate } \lambda_i \\ Z^{(i+1)}, & \text{mutation rate } u_{i+1}. \end{cases}$$

The system initially consists of $N$ wild-type (type-0) crypts, and we seek to estimate the waiting times for the first type-$i$ crypt which is defined by

$$\tau_i = \inf\{t \geq 0 | N_i(t) > 0\}.$$

To verify our analytic results, we developed Monte Carlo simulations of a multi-type branching process model based on the Gillespie algorithm (Gillespie, 1977). Parameter values for our model come from Paterson et al. (2020), and their typical ranges are listed in Table 2.

## 3. Population dynamics and waiting time for type-i crypts

In this section we analyze the growth dynamics of the premalignant subpopulations on a specific path to CRC initiation, and use these results to derive expressions for the waiting times of premalignant types and as well as the waiting time for the final, malignant, type. We compare results obtained from exact computer simulations of the multi-type branching process with our analytic results. For all figures in this section, parameter values are given in Table 3, and follow estimates from Paterson et al. (2020).

### 3.1. Type-0 and waiting time to type-1

The process of somatic evolution that can lead to colorectal cancer is started with a population of $N$ initially healthy (type-0) crypts. These crypts are wild-type for all three driver genes of interest. Healthy human crypts rarely divide (Nicholson et al., 2018), hence we set the division rate of the type-0 crypts to zero ($\lambda_0 = 0$). Type-0 crypts can inactivate one copy of the *APC* gene and become type-1 crypts, which occurs with rate $u_1$.

In other words, the number of type-0 crypts, $N_0(t)$, follows a pure death process with death rate $u_1$ and initial condition $N_0(0) = N$. The expectation and variance for the number of healthy crypts in the process are $\mathbb{E}[N_0(t)] = Ne^{-u_1 t} \approx N$ and $\mathrm{Var}[N_0(t)] = N(e^{-u_1 t} - e^{-2u_1 t}) \approx Nu_1 t$, where the approximations are made in the $u_1 t \ll 1$ limit. Due to the small variance, we approximate the population of the healthy crypts by its expectation

$$N_0(t) \approx \mathbb{E}[N_0(t)] \approx N.$$

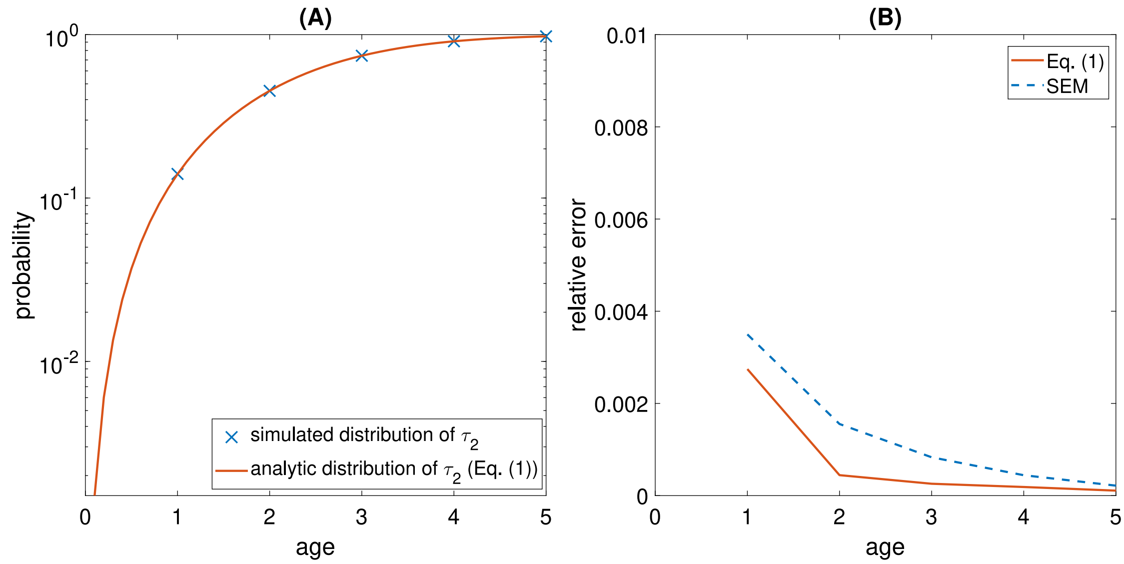


**Fig. 1.** (A) Comparison of the analytic cumulative distribution function of $\tau_2$, the waiting time to the first type-2 crypt (Eq. (1)), and the simulated distribution of $\tau_2$ across $5 \times 10^5$ runs. (B) Dashed line shows the standard error of the mean obtained from simulations. Solid line is the relative error of the analytic result. The relative error at time $t$ is defined by $|\mathbb{P}_s(\tau_2 \le t) - \mathbb{P}_a(\tau_2 \le t)|/|\mathbb{P}_s(\tau_2 \le t)|$, where $\mathbb{P}_s$ is obtained from exact computer simulations of the process, and $\mathbb{P}_a$ is the approximation in Eq. (1).

**Table 3**
Estimates of parameter values for colorectal cancer initiation from Paterson et al. (2020).

| Crypt $N_i$ | $N_0$ | $N_1$ | $N_2$ | $N_3$ | $N_4$ |
|---|---|---|---|---|---|
| Initial population (crypts) | $N = 1 \times 10^8$ | 0 | 0 | 0 | 0 |
| Birth rate $\lambda_i$ (per year) | 0 | 0 | 0.2 | 0.27 | 0.27 |
| Transition rate $u_{i+1}$ (per year) | $2.86 \times 10^{-4}$ | $1.06 \times 10^{-5}$ | $9.00 \times 10^{-7}$ | $1.36 \times 10^{-4}$ | $4.56 \times 10^{-7}$ |

It follows that the waiting time for the first type-1 crypt, $\tau_1 \sim \text{Exponential}(u_1 N)$.

### 3.2. Type-1 and waiting time to type-2

Type-1 crypts have a single copy of the *APC* gene inactivated. This genetic alteration does not immediately lead to increase in crypt division rate (Lamlum et al., 2000), so the division rate of type-1 crypts $\lambda_1 = 0$. A type-1 crypt can incur inactivation of the second allele of the *APC* gene and become a type-2 crypt, which occurs at rate $u_2$. Initially, there are no type-1 crypts, i.e., $N_1(0) = 0$.

Assuming that the loss of type-1 crypts to transition to type-2 is negligible (since $u_2$ is very small), we can approximate the number of type-1 crypts by $N_1(t) \approx N - N_0(t)$. In the small $u_1 t$ limit, we can obtain the expectation and variance of type-1 crypts as $\mathbb{E}[N_1(t)] \approx \text{Var}[N_1(t)] \approx N u_1 t$. For typical parameter values,

$$\frac{\sqrt{\text{Var}[N_1(t)]}}{\mathbb{E}[N_1(t)]} \approx \frac{1}{\sqrt{u_1 N t}} \ll 1,$$

so we can approximate $N_1(t)$ by a deterministic function

$$N_1(t) \approx u_1 N t.$$

Thus the waiting time distribution of type-2 crypts can be obtained as

$$P(\tau_2 \le t) = 1 - \mathbb{E}\left[\exp\left(-u_2 \int_0^t N_1(s) ds\right)\right]$$
$$\approx 1 - \exp\left(-\frac{1}{2} u_1 u_2 N t^2\right). \quad (1)$$

We compare the last expression with the probability distribution of waiting time to type-2 crypts obtained from exact computer simulations of the process in Fig. 1. Our results predict that the first crypt that has both copies of the *APC* gene inactivated will appear within the first five years of life.

### 3.3. Type-2 and waiting time to type-3

Type-2 crypts have both copies of the *APC* gene inactivated. The *APC* inactivation provides a fitness advantage to type-2 crypts (Lamlum et al., 2000), leading to an increased division rate $\lambda_2 > 0$. At time $t = 0$, there are no type-2 crypts, i.e $N_2(0) = 0$. We begin with the expectation of $N_2(t)$. The expected value of $N_i(t)$ can be computed recursively:

$$\mathbb{E}[N_i(t)] = u_i \int_0^t \mathbb{E}[N_{i-1}(s)] e^{\lambda_i (t-s)} \, ds. \quad (2)$$

This follows from the martingale result in Lemma 4.1 (see also equation (18) in Durrett and Moseley (2010)). Using the recursion, we compute

$$\mathbb{E}[N_2(t)] \approx \frac{N u_1 u_2 (e^{\lambda_2 t} - \lambda_2 t - 1)}{\lambda_2^2}.$$

The approximation is obtained by noting that $u_1 \ll \lambda_2 \sim 10^{-1}$ and $u_1 t \ll 1$.

The following large-time asymptotic limit exists for $N_2(t)$.

**Theorem 3.1.** *$e^{-\lambda_2 t} N_2(t) \to W_2$ a.s. and in $L^1$ with*

$$\mathbb{E}[W_2] = \frac{N u_1 u_2}{\lambda_2(\lambda_2 + u_1)} \approx \frac{N u_1 u_2}{\lambda_2^2}.$$

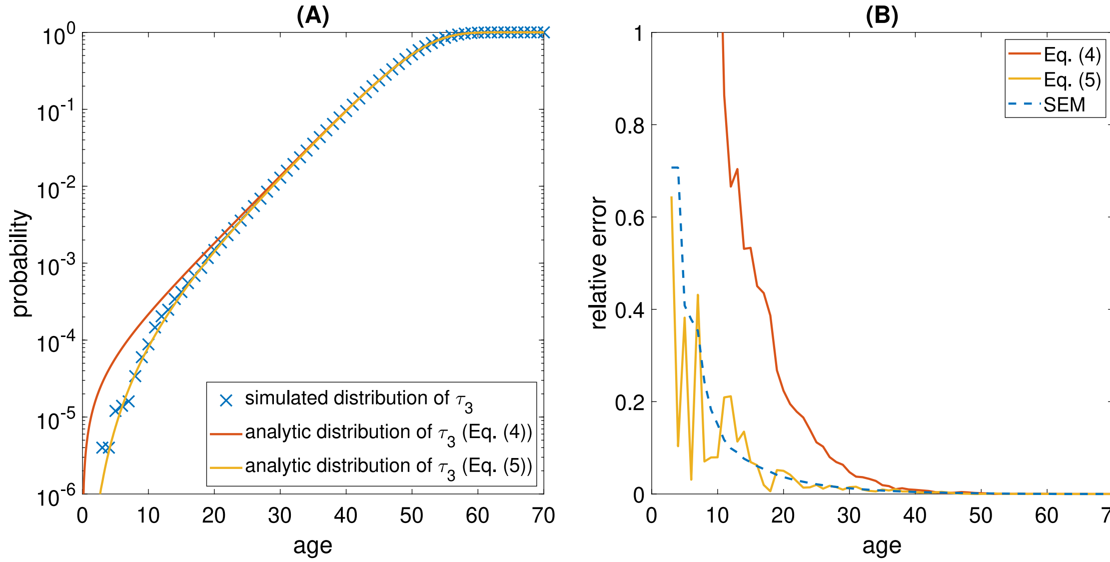


**Fig. 2.** (A) Comparison of the analytic cumulative distribution functions of $\tau_3$, the waiting time to the first type-3 crypt ((4) and (5)), and the distribution of $\tau_3$ across $5 \times 10^5$ simulation runs. In (4), $f_2(t)$ is set to be the exponential function $f_{20}(t) = e^{\lambda_2 t}$, while in (5), $f_2(t) := f_{21}(t) = \mathbb{E}[N_2(t)]/\mathbb{E}[W_2]$. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results. The relative error at time $t$ is defined by $|\mathbb{P}_s(\tau_3 \le t) - \mathbb{P}_a(\tau_3 \le t)|/|\mathbb{P}_s(\tau_3 \le t)|$, where $\mathbb{P}_s$ is obtained from exact computer simulations of the process, and $\mathbb{P}_a$ represents the approximation in Eq. (4) or (5).

*The Laplace transform of $W_2$ is given by*

$$\mathcal{L}_{W_2}(\theta) = \left[\frac{1}{u_1 - u_2}\left(u_1\, {}_2F_1\left(1, \frac{u_2}{\lambda_2}; 1 + \frac{u_2}{\lambda_2}; -\theta\right) - u_2\, {}_2F_1\left(1, \frac{u_1}{\lambda_2}; 1 + \frac{u_1}{\lambda_2}; -\theta\right)\right)\right]^N, \tag{3}$$

*where ${}_2F_1(a, b; c, z)$ is the Gauss hypergeometric function (DLMF, 2022, 15.2.1).*

**Corollary 3.2.** *$\frac{N_2(t)}{f_2(t)} \to W_2$ a.s. and in $L^1$ for all*

$$f_2(t) \in F_2 := \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) = 1, f_2(t) \ge 0\},$$

*where $C(\mathbb{R})$ is the space of continuous functions on $\mathbb{R}$.*

The goal of allowing $f_2(t)$ to be potentially different from $e^{\lambda_2 t}$ in Theorem 3.1 is that a suitably chosen $f_2(t)$ can lead to increased accuracy when computing the waiting time to the next type.

Recall that a type-2 crypt can activate the *KRAS* oncogene with rate $u_3$, becoming a type-3 crypt. We can compute the distribution for the waiting time to the first type-3 crypt, $\tau_3$, using

$$P(\tau_3 \le t) \approx 1 - \mathcal{L}_{W_2}\left(u_3 \int_0^t f_2(s)\mathrm{d}s\right), \quad f_2 \in F_2.$$

To compute the waiting time to the first type-3 crypt, we will consider two candidate functions for $f_2$:

$$f_{20}(t) := e^{\lambda_2 t}, \; f_{21}(t) := \frac{\mathbb{E}[N_2(t)]}{\mathbb{E}[W_2]} = e^{\lambda_2 t} - \lambda_2 t - 1.$$

The two candidate functions $f_{20}$ and $f_{21}$ correspond to the following approximate distributions of waiting time to type-3:

$$p_{30}(t) := 1 - \mathcal{L}_{W_2}\left(u_3 \int_0^t f_{20}(s)\mathrm{d}s\right) = 1 - \mathcal{L}_{W_2}\left(u_3 \frac{e^{\lambda_2 t} - 1}{\lambda_2}\right), \tag{4}$$

and

$$p_{31}(t) := 1 - \mathcal{L}_{W_2}\left(u_3 \int_0^t f_{21}(s)\mathrm{d}s\right) = 1 - \mathcal{L}_{W_2}\left(u_3\left(\frac{e^{\lambda_2 t} - 1}{\lambda_2} - \frac{\lambda_2}{2}t^2 - t\right)\right). \tag{5}$$

Here we use $p_{ij}$ to represent a specific approximation, distinguishing it from the exact waiting time distribution. We note that, by design, the first moment of $f_{21}(t)W_2$ is identical to that of $N_2(t)$.

Both (4) and (5) agree with the simulated distributions for $t > 40$ (Fig. 2). However, in the intermediate regime where $t$ is small, one can observe that $p_{31}$ is more accurate than $p_{30}$.

At the end of this section, we present an approximation of the random variable $W_2$ which is denoted by $V_2$. We present a detailed description of $V_2$ including its construction, properties and show that it is in excellent agreement with $W_2$ in Section 6. Compared with $\mathcal{L}_{W_2}(\theta)$, the Laplace transform of $V_2$ is simpler in its form and easier to obtain. Here we present $\mathcal{L}_{V_2}(\theta)$ so that the reader can compare the formula with $\mathcal{L}_{W_2}(\theta)$ (3).

$$\mathcal{L}_{V_2}(\theta) = \mathbb{E}\left[e^{-\theta V_2}\right] = \exp\left(\frac{Nu_1u_2\mathrm{PolyLog}(2, -\theta)}{\lambda_2^2}\right) \tag{6}$$

In the above expression, $\mathrm{PolyLog}(n, z), n \in \mathbb{Z}, n \ge 2, z \in \mathbb{C}$ represents the polylogarithm function defined by the series $\mathrm{PolyLog}(n, z) = \sum_{k=1}^{\infty} \frac{z^k}{k^n}$ when $|z| \le 1$ and the analytic continuation of the series when $|z| > 1$ (DLMF, 2022, 25.12.10).

### 3.4. Type-3 and waiting time to type-4

Type-3 crypts are produced by type-2 crypts through activation of the *KRAS* oncogene, which increases the division rate of mutated crypts (Snippert et al., 2014). Thus the division rate has a positive increment, i.e. $\lambda_3 > \lambda_2$. The initial population is $N_3(0) = 0$. From Eq. (2), the expected value of $N_3(t)$ is given by

$$\mathbb{E}[N_3(t)] \approx \frac{Nu_1u_2u_3\left(\lambda_2^2\left(e^{\lambda_3 t} - \lambda_3 t - 1\right) - \lambda_3^2\left(e^{\lambda_2 t} - \lambda_2 t - 1\right)\right)}{\lambda_2^2\left(\lambda_3 - \lambda_2\right)\lambda_3^2},$$

where the approximation is made by observing $\lambda_i + u_1 \approx \lambda_i$ and $1 - \exp(-u_1 t) \approx u_1 t$.

**Theorem 3.3.** *$e^{-\lambda_3 t}N_3(t) \to W_3$ a.s. and in $L^1$ with*

$$\mathbb{E}[W_3] = \frac{Nu_1u_2u_3}{(\lambda_3 - \lambda_2)\lambda_3(\lambda_3 + u_1)} \approx \frac{Nu_1u_2u_3}{(\lambda_3 - \lambda_2)\lambda_3^2}.$$

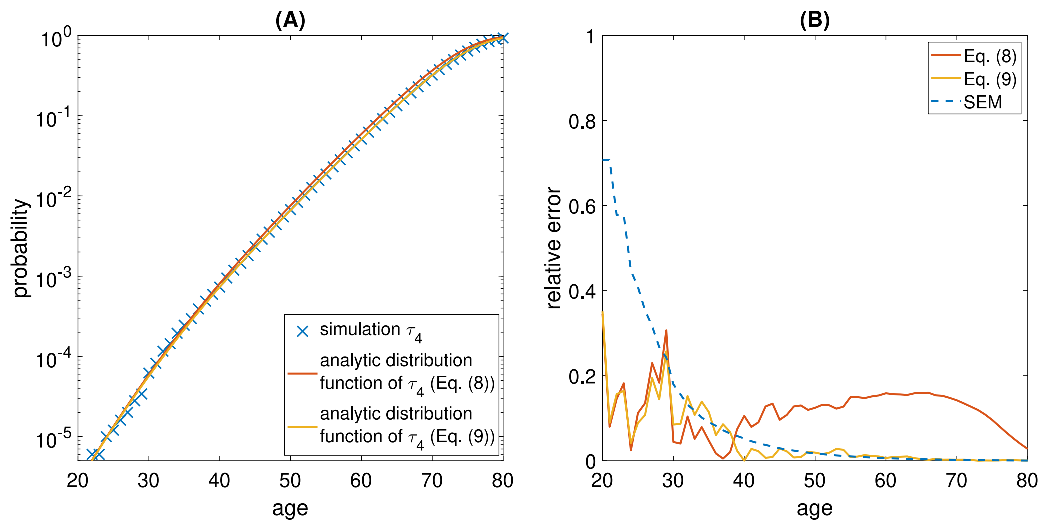


**Fig. 3.** (A) Comparison of analytic cumulative distribution functions of $\tau_4$, the waiting time to the first type-4 crypt (Eqs. (8) and (9)), and the distribution of $\tau_4$ across $5 \times 10^5$ simulation runs. Eq. (8) is derived using the Laplace transform $V_3$. Eq. (9) is derived via skipping $V_3$ and using the Laplace transform of $V_2$. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results. The relative error at time $t$ is defined by $|\mathbb{P}_s(\tau_4 \le t) - \mathbb{P}_a(\tau_4 \le t)|/|\mathbb{P}_s(\tau_4 \le t)|$, where $\mathbb{P}_s$ is obtained from exact computer simulations of the process, and $\mathbb{P}_a$ represents the approximation in Eq. (8) or (9).

*The Laplace transform of $W_3$ is given by*

$$\mathcal{L}_{W_3}(\theta) = \left( \int_0^\infty \mathcal{L}_U(\theta e^{-\lambda_3 x}) \frac{u_1 u_2}{u_2 - u_1}(e^{-u_1 x} - e^{-u_2 x})\mathrm{d}x \right)^N. \tag{7}$$

*Here $U$ is the limiting random variable of the type-3 population $e^{-\lambda_3 t} M_3(t) \to U$ in a two-type process $(M_2(t), M_3(t))$ started with a single type-2 crypt at time 0.*

We derive the Laplace transform of the limiting random variable $U$ in Appendix C.

**Corollary 3.4.** *$\frac{N_3(t)}{f_3(t)} \to W_3$ a.s. and in $L^1$ for all*

$$f_3(t) \in F_3 := \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_3 t} f_3(t) = 1, f_3(t) \ge 0\}.$$

We use the above result to derive the waiting time of type-4 crypts. Recall that type-4 crypts are produced by type-3 crypts with a small rate $u_4$. We consider the following approximation for the waiting time:

$$P(\tau_4 \le t) \approx 1 - \mathcal{L}_{W_3}\left( u_4 \int_0^t f_3(s)\mathrm{d}s \right), \quad f_3 \in F_3.$$

Note that evaluating the Laplace transform of $W_3$ (7) requires numerical integration of a relatively complicated function. Therefore we find $V_3$, which is easier to manipulate, as the approximation of $W_3$. We present the construction of $V_3$ in Section 6. The Laplace transform of $V_3$ is given by (B.6). We compare $L_{W_3}(\theta)$ and $L_{V_3}(\theta)$ in Section 6. As the two Laplace transforms are in excellent agreement, we employ $\mathcal{L}_{V_3}(\theta)$ to compute $P(\tau_4 \le t)$.

Using $f_{i1} = \frac{\mathbb{E}[N_i(t)]}{\mathbb{E}[W_i]}$ typically leads to more accurate computation of waiting times compared to the more simple $f_{i0} = \exp(\lambda_i t)$. We illustrate this phenomenon in Fig. 5. For that reason, we will use $f_2 = f_{21}$ to compute $\mathcal{L}_{V_3}(\theta)$ (since $\mathcal{L}_{V_3}(\theta)$ depends on the choice of $f_2$. See Lemma 6.4). In particular, we have

$$f_{31}(t) = \frac{\mathbb{E}[N_3(t)]}{\mathbb{E}[W_3]} = e^{\lambda_3 t} - \lambda_3 t - 1 - \frac{\lambda_3^2}{\lambda_2^2}(e^{\lambda_2 t} - \lambda_2 t - 1).$$

This leads to the following approximation for the cumulative distribution of the waiting time to type-4 crypts:

$$p_{41}(t) := 1 - \mathcal{L}_{V_3}\left( u_4 \int_0^t f_{31}(s)\mathrm{d}s \right). \tag{8}$$

We note that $p_{41}$ has a closed form expression, which is shown in Eq. (B.3). The comparison of analytic and simulation results for waiting time to type-4 are shown in Fig. 3. The relative error of the approximation (8) is on the order of 20%, showing a tendency to decrease even lower for $t > 70$ years.

We also find that increased accuracy in computing the waiting time of type-$(i+1)$ crypts, in particular at early times, can be achieved by skipping the long-time limit of the entire type-$i$ subpopulation and instead using a long-time limit of individual type-$i$ lineages. Note that a type-$i$ lineage is the type-$i$ offspring of a single type-$i$ crypt that has been mutated from a type-$(i-1)$ crypt. Mathematically, a type-$i$ lineage is the number of type-$i$ crypts in a system initiated by a single type-$i$ crypt. The skipping process is described in more detail in Section 4.3. In the case of waiting time to type-4, we can use this methodology to "skip" the long time limit of type-3 crypts, leading to the following expression for the cumulative distribution for the waiting time of type-4 crypts

$$p_{41}^{s3}(t) := 1 - \mathcal{L}_{V_2}\left( u_4 \int_0^t f_{21}(s)(1 - p_0^{3\to 4}(s,t))\mathrm{d}s \right). \tag{9}$$

Here $p_0^{3\to 4}(s,t)$ is the probability that no type-4 crypt is produced by time $t$ by a lineage started with a single a type-3 crypt at time $s$. We note that the closed form version of $p_{41}^{s3}$ is presented in Eq. (B.11). We observe that skipping $V_3$ improves the accuracy of the results at intermediate times (Fig. 3).

### *3.5. Type-4 and waiting time to type-5*

Type-4 crypts have fully inactivated *APC*, activated *KRAS*, and a single inactivated copy of *TP53*. Compared to type-3 crypts, each of the type-4 crypts has one inactivated copy of a tumor suppressor *TP53*. This mutation does not lead to an increment of the crypt division rate. Thus the division rate of type-4 crypts $\lambda_4 = \lambda_3$. The initial condition for type-4 crypts is $N_4(0) = 0$. The fact that the division rates of type-4 and type-3 crypts are the same leaves us unable to confirm the existence of a large-time limiting random variable for the population of type-4 crypts.

Instead, for $\tau_5$, the waiting time to the first type-5 crypt, we consider an alternative approach: we compute the distribution using the large-time limit of type-3 crypts, $\mathcal{L}_{V_3}(\theta)$, and $p_0^{4\to 5}(s,t)$ (effectively skipping the large-time limit of type-4 crypts). Here

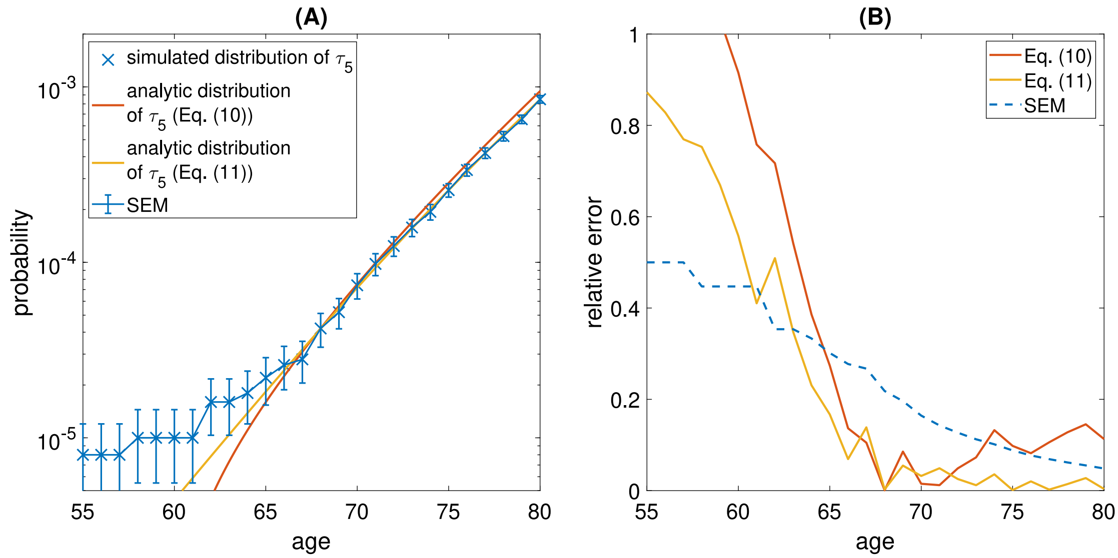


**Fig. 4.** (A) Comparison of the analytic cumulative distribution functions of $\tau_5$, the waiting time to the first type-5 crypt (Eqs. (10) and (11)), and the distribution of $\tau_5$ across $5 \times 10^5$ simulation runs. Eq. (10) is derived via skipping type-4 and using the Laplace transform of $V_3$. Eq. (11) is derived via using the Laplace transform of $V_2$ and skipping the Laplace transforms of type-4 and type-3. The error bars represent the standard error of the mean of the simulation. (B) Dashed line shows the standard error of the mean of the simulation. Solid lines are the relative errors of the analytic results. The relative error at time $t$ is defined by $|\mathbb{P}_s(\tau_5 \leq t) - \mathbb{P}_a(\tau_5 \leq t)|/|\mathbb{P}_s(\tau_5 \leq t)|$, where $\mathbb{P}_s$ is obtained from exact computer simulations of the process, and $\mathbb{P}_a$ represents the approximation in Eq. (10) or (11).

$p_0^{4\to 5}(s,t)$ is the probability that no type-5 crypt is produced by time $t$ by a lineage started with a single type-4 crypt at time $s$. $p_0^{4\to 5}(s,t)$ can be computed using Lemma 4.4 as

$$p_0^{4\to 5}(s,t) = \frac{1}{1 + u_5 \frac{\exp(\lambda_3(t-s))-1}{\lambda_3}}.$$

The corresponding approximation of the distribution of $\tau_5$ resulting from this approach is

$$p_{51}^{s4}(t) := 1 - \mathcal{L}_{V_3}\left(u_4 \int_0^t f_{31}(s)(1 - p_0^{4\to 5}(s,t))ds\right). \tag{10}$$

The above expression has an explicit form given by Eq. (B.12). Comparison of formula (10) with the waiting time for the first type-5 crypt obtained from exact computer simulations (Fig. 4) shows good agreement at intermediate times, but increasing deviation (approaching 0.2 relative error) at $t > 75$. We find that increased accuracy is achieved by skipping large-time limits of both type-4 and type-3 crypts, and computing the distribution using $\mathcal{L}_{V_2}(\theta)$ and $p_0^{3\to 5}(s,t)$. $p_0^{4\to 5}(s,t)$, the probability that no type-5 crypt is produced by time $t$ by a lineage started with a single type-3 crypt at time $s$, is given by (using Lemma 4.4)

$$p_0^{3\to 5}(s,t) = \left(1 + \frac{u_4 u_5\left(\lambda_3 - u_5 + e^{\lambda_3(t-s)}\left(u_5 - \lambda_3 + \lambda_3 \log\left(\frac{\lambda_3 e^{\lambda_3(t-s)}}{\lambda_3 + u_5(e^{\lambda_3(t-s)}-1)}\right)\right)\right)}{(\lambda_3 - u_5)^2 \lambda_3}\right)^{-1}.$$

The corresponding approximation of the distribution of $\tau_5$ resulting from the latter approaches is

$$p_{51}^{s34}(t) := 1 - \mathcal{L}_{V_2}\left(u_3 \int_0^t f_{21}(s)(1 - p_0^{3\to 5}(s,t))ds\right). \tag{11}$$

The result $p_{51}^{s34}$ (explicitly given by Eq. (B.14)), to the best of our knowledge, is not an elementary function or a standard special function. In Fig. 4, we observe that $p_{51}^{s34}$ achieves higher accuracy compared to $p_{51}^{s4}$, especially at later times (above age 70). In other words, compared with the result incorporating the long-time limit of type-4 crypts, skipping this stage gives more accurate results. The intuition behind this is that approximating each subclone by its large time limit is more accurate than approximating the total population by its overall large time limit.

## 4. Multi-type branching process results

In this section we establish a martingale convergence lemma to get possible large time limiting random variables. Next, we state the results needed for approximating the waiting time distribution of type-$i$ using the large time limit of type-$(i-1)$. We generalize these results by employing an approximation that allows us to derive type-$i$ results directly from type-$j$ results and skip intermediate limiting behaviors between $i$ and $j$.

### 4.1. General results for large time limits

**Lemma 4.1.** *Consider a multi-type branching process* $(N_0(t), N_1(t), \ldots)$ *in which* $N_i(t)$ *is the population of type-$i$ crypts. In this process, a single type-$i$ crypt can divide into two crypts with rate* $\lambda_i \geq 0$ *and mutate into a type-$(i+1)$ crypt with rate* $u_{i+1} > 0$. *Then*

$$M(t) = e^{-\lambda_i t} N_i(t) - \int_0^t u_i N_{i-1}(s) e^{-\lambda_i s} ds$$

*is a martingale. If*

$$I_i = \int_0^\infty u_i N_{i-1}(s) e^{-\lambda_i s} ds$$

*has a finite expectation, then*

$$e^{-\lambda_i t} N_i(t) \xrightarrow{a.s.} W_i, \quad \mathbb{E}|W_i| < \infty,$$

*as* $t \to \infty$. *Additionally, if* $\{e^{-\lambda_i t} N_i(t); t \geq 0\}$ *is uniform integrable, then*

$$e^{-\lambda_i t} N_i(t) \xrightarrow{L^1} W_i.$$

*This implies*

$$\mathbb{E}[e^{-\lambda_i t} N_i(t)] \to \mathbb{E}[W_i] = \mathbb{E}[I_i].$$

If the first condition from the statement of the Lemma holds (i.e. if $I_i$ has finite expectation) Lemma 4.1 provides a method of obtaining the long-term behavior of $N_i$ using the limiting random variable $W_i$. In that case, we have $e^{-\lambda_i t}N_i(t) \to W_i$, and for large time $t$, $e^{\lambda_i t}W_i$ should be a good approximation of the stochastic process $N_i(t)$. The importance of $N_i(t) \approx e^{\lambda_i t}W_i$ is that it separates a stochastic process into a deterministic function $e^{\lambda_i t}$ and a time-independent random variable $W_i$.

If, in addition, the second condition ($\{e^{-\lambda_i t}N_i(t); t \geq 0\}$ is uniform integrable) holds, then the expected value of the limiting random variable $W_i$ is obtainable. In that case, we have $E[e^{-\lambda_i t}N_i(t)] \to E[W_i]$, which makes the large time approximation $N_i(t) \approx e^{\lambda_i t}W_i$ reasonable in terms of the first moment.

**Proof of Lemma 4.1.**

**Proof.** The proof follows that of Theorem 2 in Durrett and Moseley (2010). The only difference is that we want to include the cases when $\lambda_{i-1} = \lambda_i$ or $\lambda_i = 0$. By Lemma 1 in (Durrett and Moseley, 2010),

$$M(t) = e^{-\lambda_i t}N_i(t) - \int_0^t u_i N_{i-1}(s)e^{-\lambda_i s}ds$$

is a martingale. If $I_i$ has a finite expectation, then by the martingale convergence theorem (Theorem 4.2.11 in Durrett (2019)), the submartingale $X(t) = -M(t)$ converges a.s. to some integrable limit $X$ as $t \to \infty$. Since

$$I_i(t) = \int_0^t u_i N_{i-1}(s)e^{-\lambda_i s}ds \xrightarrow{a.s.} I_i,$$

we also have

$$e^{-\lambda_i t}N_i(t) \xrightarrow{a.s.} W_i.$$

The martingale starts at zero (i.e. $M(0) = 0$), which implies

$$\mathbb{E}[e^{-\lambda_i t}N_i(t)] = \mathbb{E}[I_i(t)].$$

Suppose $\{e^{-\lambda_i t}N_i(t); t \geq 0\}$ is uniform integrable, we have (Theorem 4.6.3 in Durrett (2019))

$$e^{-\lambda_i t}N_i(t) \xrightarrow{L^1} W_i,$$

which guarantees

$$\mathbb{E}[I_i(t)] = \mathbb{E}[e^{-\lambda_i t}N_i(t)] \to \mathbb{E}[W_i].$$

Thus, we have

$$\mathbb{E}[I_i(t)] \to \mathbb{E}[I_i],$$

and

$$\mathbb{E}[W_i] = \mathbb{E}[I_i]. \quad \square$$

### *4.2. Estimating waiting times using large time limits*

Let $\tau_i, (1 \leq i \leq n)$ be the waiting time of the first type-$i$ individual in a multi-type branching process. At time $s \geq 0$, the arrival rate of type-$i$ individuals is $u_i N_{i-1}(s)$. Conditional on the trajectory of $N_{i-1}(s)$ for $0 \leq s \leq t$, the probability that $\tau_i$ is greater than $t$ is:

$$P(\tau_i > t \mid N_{i-1}(s), 0 \leq s \leq t) = \exp\left(-u_i \int_0^t N_{i-1}(s)ds\right).$$

The functional form of $N_{i-1}(s)$ is generally unknown and potentially complicated. One way of evaluating this integral is to approximate $N_{i-1}(s)$ by the product of a deterministic time-dependent growth and a time independent random variable. For example, let $N_0(t) = Z_0(t)$, a pure birth two-type branching process that starts with a single individual. It is well-known that $e^{-\lambda_0 t}Z_0(t) \to W_0 \sim \text{Exponential}(1)$ (Durrett and Moseley, 2010). A classical approximation is $N_0(s) \approx e^{\lambda_0 t}W_0$ where the deterministic time-dependent growth is characterized by $e^{\lambda_0 t}$ and time independent random variable is $W_0$. Applying this approximation yields

$$P(\tau_1 > t) \approx \mathbb{E}\left[\exp\left(-u_1 \int_0^t e^{\lambda_0 s}W_0 ds\right)\right] \tag{12}$$

$$= \mathcal{L}_{W_0}\left(u_1 \frac{e^{\lambda_0 t} - 1}{\lambda_0}\right), \tag{13}$$

where $\mathcal{L}_{W_0}(\theta) = \frac{1}{1+\theta}$ is the Laplace transform of $W_0$. Let $f_0(t) = e^{\lambda_0 t}$ be the time deterministic function. Then the above approximation also holds if a sub-exponential term is added to $f_0$. In other words, we have many reasonable options for $f_0$. In later sections, we consider two specific deterministic functions,

$$f_{i0}(t) = e^{\lambda_i t}, \text{ and}$$

$$f_{i1}(t) = \frac{\mathbb{E}[N_i(t)]}{\mathbb{E}[W_i]}.$$

In the example mentioned above, we have

$$f_{01}(t) = \frac{\mathbb{E}[Z_0(t)]}{\mathbb{E}[W_0]} = e^{\lambda_0 t}$$
$$= f_{00}(t).$$

However, in our model when $i > 1$, $f_{i0} \neq f_{i1}$. We observe in Fig. 5 that an approximation with $f_{i1}$ is typically more precise than that with $f_{i0}$.

**Proposition 4.2.** *Let $(N_{i-1}(t), N_i(t))$ be a two-type process such that $N_i(0) = 0$ and $N_{i-1}(t) \geq 0$ is right-continuous with $\mathbb{E}|N_{i-1}(t)| < \infty$. In this process, type-$i$ crypts are being produced at rate $u_i N_{i-1}(t) > 0$. A single type-$i$ crypt can divide into two crypts with rate $\lambda_i \geq 0$. Suppose there exists a continuous function $f_{i-1}(t) \geq 0$ and a random variable $W_{i-1}$ such that as $t \to \infty$, $(f_{i-1}(t))^{-1}N_{i-1}(t) \to W_{i-1}$ almost surely and in $L^1$. Then the waiting time distribution of the first type-$i$ individual can be approximated by*

$$P(\tau_i > t) \approx \mathcal{L}_{W_{i-1}}\left(u_i \int_0^t f_{i-1}(s)ds\right)$$

*where $\mathcal{L}_{W_{i-1}}(\theta)$ is the Laplace transform of random variable $W_{i-1}$.*

**Proof.** The right continuity and the integrability of $N_{i-1}(t)$ allow us to write $P(\tau_i > t) = \mathbb{E}\left[\exp\left(-\int_0^t u_i N_{i-1}(s)ds\right)\right]$. Since $(f_{i-1}(t))^{-1}N_{i-1}(t) \to W_{i-1}$, we employ the approximation $N_{i-1}(t) \approx f_{i-1}(t)W_{i-1}$. Plugging in this approximation gives us

$$P(\tau_i > t) \approx \mathbb{E}\left[\exp\left(-W_{i-1}u_i \int_0^t f_{i-1}(s)ds\right)\right]$$
$$= \mathcal{L}_{W_{i-1}}\left(u_i \int_0^t f_{i-1}(s)ds\right). \quad \square$$

We note that the accuracy of the approximate waiting time distribution in Proposition 4.2 depends on the accuracy of the long-time approximation $N_{i-1}(t) \approx f_{i-1}(t)W_{i-1}$. To provide insight into the accuracy of this approximation, we investigate the difference $N_{i-1}(t) - f_{i-1}(t)W_{i-1}$ in the case when $i - 1 = 2, 3$. In particular, we present a representation of $N_2(t) - e^{\lambda_2 t}W_2$ (Theorem 5.1) in the case of the type-2 population. This representation allows us to estimate $N_2(t) - e^{\lambda_2 t}W_2$ in the long-time regime as $t \to \infty$ (Corollary 5.4) and in the short-time regime as $t \to 0$ (Corollary 5.5). Using these results, in Section 5, we show the reason that the scaling function $f_{21}(t)$ obtained from the ratio $\mathbb{E}[N_2]/\mathbb{E}[W_2]$ leads to greater accuracy compared to the

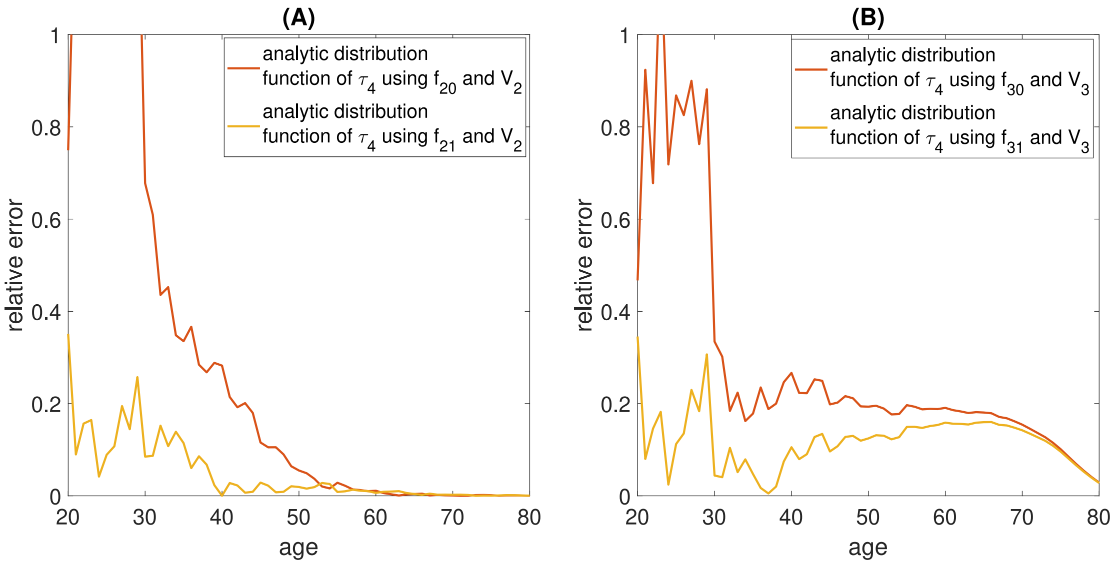


**Fig. 5.** Relative errors of the waiting time distributions of $\tau_4$ obtained using different scaling functions $f_i$. (A) Using $f_{21}$, the distribution function has a higher accuracy compared to the result derived using $f_{20}$. (B) Using $f_{31}$, the distribution function has a higher accuracy compared to the result derived using $f_{30}$.

exponential scaling function $f_{20}(t) = e^{\lambda_2 t}$, when approximating populations at short time. Similarly, we develop the representation theorem for $N_3(t) - e^{\lambda_3 t}W_3$ (Theorem 5.6) and discuss its consequences in Section 5.2.

### 4.3. An inhomogeneous Poisson process approximation

Proposition 4.2 provides an estimate of the arrival time of type-$i$ individuals using the large time limit of the previous type ($W_{i-1}$). However, there are situations when the existence of random variable $W_{i-1}$ cannot be directly inferred from the martingale result. In our model, we are unable to show the existence of $W_4$ due to the fact that $\lambda_3 = \lambda_4$. When dealing with type-4, we found that $\mathbb{E}[N_4(t)] \sim O(te^{\lambda_3 t})$. Thus, one should expect that a non-degenerate limiting distribution may exist when the population is scaled by $t^{-1}e^{-\lambda_3 t}$. However, establishing a square integrable martingale in this case is non-trivial. Therefore, we would like to employ the existing limits of other types to calculate the waiting time to type-5. More generally, we may only have the explicit Laplace transform of the large time limit $W_j$ for some type-$j$ where $j < i$, and we do not have reliable limits for type-($j+1$) through type-($i-1$). To deal with this situation, we introduce a method that uses the large time limit of each independent lineage to "skip" $W_{j+1}$ through $W_{i-1}$.

First, let $Z_{j+1}(t)$ denote a type-($j+1$) lineage started with a single type-($j+1$) individual at time 0. In other words, taking advantage of small mutation rates to neglect outflow, a type-($j+1$) lineage is a simple birth process initiated by a single type-($j+1$) individual that grows at rate $\lambda_{j+1}$. The well-known fact that for each type-($j+1$) lineage

$$e^{-\lambda_{j+1}t}Z_{j+1}(t) \xrightarrow{a.s.} V \sim \text{Exponential}(1) \tag{14}$$

allows us to make an approximation $Z_{j+1}(t) \approx e^{\lambda_{j+1}t}V$. Next, we want to find the likelihood of each type-($j+1$) lineage producing at least a single type-$i$ individual. Suppose we have a type-($j+1$) lineage that was started with a single type-($j+1$) individual at time $s$. We define $p_0^{(j+1)\to i}(s,t)$ to be the probability that no type-$i$ individual is produced by time $t$ by this type-($j+1$) lineage. This implies the following proposition.

**Proposition 4.3.** *Let $(N_j(t), N_{j+1}(t), \ldots, N_i(t))$ be a $(i-j+1)$-type process in which $N_k(0) = 0$ for $j < k \le i$, and $N_j(t) \ge 0$ is right-continuous with $\mathbb{E}|N_j(t)| < \infty$. In this process, type-($j+1$) crypts are produced at rate $u_{j+1}N_j(t) > 0$. A single type-$k$ crypt can divide into two crypts with rate $\lambda_k \ge 0$ and mutate into a type-($i+1$) crypt with rate $u_{i+1} > 0$ for $k$, $j < k \le i$. Suppose that there exists a continuous function $f_j(t) \ge 0$ and a random variable $W_j$ such that as $t \to \infty$, $f_j(t)^{-1}N_j(t) \to W_j$ almost surely and in $L^1$. Then the waiting time distribution of type-$i$ crypts can be approximated by*

$$P(\tau_i > t) \approx \mathcal{L}_{W_j}\left(u_{j+1}\int_0^t f_j(s)\left(1 - p_0^{(j+1)\to i}(s,t)\right)ds\right),$$

*where $\mathcal{L}_{W_j}(\theta)$ is the Laplace transform of $W_j$.*

**Proof.** At time $s$, type-($j+1$) crypts are being produced at rate $u_{j+1}N_j(s) \approx u_{j+1}f_j(s)W_j$. Each type-($j+1$) lineage (present at time $s$) has a probability $1 - p_0^{(j+1)\to i}(s,t)$ to produce at least a single type-$i$ crypt (at time $t$). Therefore, for fixed $t$, we approximate the process of producing type-$i$ individuals as an inhomogeneous Poisson process with rate $u_{j+1}N_j(s)(1-p_0^{(j+1)\to i}(s,t))$ at time $s < t$. The multiplication of rates $u_{j+1}N_j(s)$ and $1-p_0^{(j+1)\to i}(s,t)$ is due to the thinning property of inhomogeneous Poisson processes. Thus we have

$$\begin{aligned} P(\tau_i > t) &\approx \mathbb{E}\left[\exp\left(-W_j u_{j+1}\int_0^t f_j(s)\left(1 - p_0^{(j+1)\to i}(s,t)\right)ds\right)\right] \\ &= \mathcal{L}_{W_j}\left(u_{j+1}\int_0^t f_j(s)\left(1 - p_0^{(j+1)\to i}(s,t)\right)ds\right). \quad \square \end{aligned}$$

Proposition 4.2 is consistent with Proposition 4.3, and one can treat Proposition 4.2 as a special case of Proposition 4.3 when $j = i - 1$. The approximation error in the distribution of $\tau_i$ in Proposition 4.3 comes from the approximation $N_j(t) \approx f_j(t)W_j$, $j < i$. In other words, the waiting time approximation in Proposition 4.3 relies on the long-time limiting random variable $W_j$, "skipping" the long-time limits of subsequent type $j+1, \ldots, i-1$ populations via the nonhomogeneous Poisson approximation. Our numerical results demonstrate that the approximate waiting time distribution function obtained from "skipping" two-steps (i.e. $i - j = 3$) leads to better accuracy compared with "skipping" one-step (i.e. $i-j = 2$). Namely, in Fig. 4 we show two approximations for the distribution function of the waiting time $\tau_5$, obtained from either $W_3$ (one-step skipping approximation), or $W_2$ (two-step skipping approximation). As the two panels in Fig. 4 demonstrate, the two-step skipping approximation has a higher accuracy for most times on the interval $t \in [0, 80]$. Therefore, our results imply that skipping more long-time limits of intermediate populations, i.e. relying on an earlier $W_j$, leads to better accuracy.

To compute $p_0^{j\to i}(s,t), j < i$, we use the iterative relationship between $p_0^{j\to i}(s,t)$ and $p_0^{(j+1)\to i}(s,t)$. This is provided by the following proposition.

**Lemma 4.4.** *Consider a multi-type branching process $(N_0(t), N_1(t), \ldots)$ in which $N_i(t)$ is the population of type-i crypts. In this process, a single type-i crypt can divide into two crypts with rate $\lambda_i \geq 0$ and mutate into a type-$(i+1)$ crypt with rate $u_{i+1} > 0$. Suppose that $\lambda_j > 0$. Then for $i > j$ we have*

$$p_0^{j\to i}(s,t) \approx \int_0^t e^{-v} \exp\left(-v \int_s^t u_{j+1} e^{\lambda_j (r-s)} \times \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr\right) dv$$

*with $p_0^{i\to i}(s,t) = 0$.*

**Proof.** The population at time $r > s$ of a single type-$j$ lineage that appeared at time $s$ can be approximated using its long time limit $Z_j(r) \approx e^{\lambda_j(r-s)}V$, where $V \sim$ Exponential(1). Thus, type-$(j+1)$ individuals are produced from this lineage at rate

$$u_{j+1} Z_j(r) \approx u_{j+1} e^{\lambda_j (r-s)} V.$$

The probability for a new type-$(j+1)$ individual at time $r$ to produce at least one type-$i$ individual by time $t$ is $1 - p_0^{(j+1)\to i}(r,t)$. Thus, conditional on $V$, the expected number of type-$i$ individuals that were produced by a type-$j$ lineage that appeared at time $s$ is

$$\begin{aligned}\Lambda^{j\to i}(s,t,V) &= \int_s^t u_{j+1} Z_j(r) \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr \\ &\approx \int_s^t u_{j+1} e^{\lambda_j (r-s)} V \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr.\end{aligned}$$

Let $X^{j\to i}(s,t)$ be the number of type-$i$ individuals that are produced by a type-$j$ subclone which appeared at time $s$. In the time period $[s,t]$, $X^{j\to i}(\cdot,t)$ follows an inhomogeneous Poisson process with mean $\Lambda(\cdot,t,V)$. Thus the probability that no type-$i$ crypt is produced from this particular type-$(j+1)$ crypt by time $t$ is

$$\begin{aligned}P(X^{j\to i}(s,t) = 0|V) &= \exp(-\Lambda^{j\to i}(s,t,V)) \\ &\approx \exp\left(-V \int_s^t u_{j+1} e^{\lambda_j (r-s)} \times \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr\right).\end{aligned}$$

This implies

$$\begin{aligned}p_0^{j\to i}(s,t) &:= P(X^{j\to i}(s,t) = 0) \\ &= \mathbb{E}\left[P(X^{j\to i}(s,t) = 0|V)\right] \\ &\approx \int_0^t e^{-v} \exp\left(-v \int_s^t u_{j+1} e^{\lambda_j (r-s)} \times \left(1 - p_0^{(j+1)\to i}(r,t)\right) dr\right) dv.\end{aligned}$$

Finally, for $i = j$, notice that the founding individual of a type-$j$ lineage is of type-$j$. This guarantees that at any time $t$ greater than the founding time $s$, the probability of having at least one type-$j$ individual is 1. Thus $p_0^{j\to j}(s,t) = 0$. $\square$

In the case that $i = j+1$, by Lemma 4.4 we have

$$p_0^{i\to(i+1)}(s,t) \approx \frac{1}{1 + u_{i+1} \frac{\exp(\lambda_i(t-s)) - 1}{\lambda_i}}.$$

The right hand side obtained from Lemma 4.4 is in excellent agreement with the exact $p_0^{i\to(i+1)}(s,t)$, the probability of zero type-$(i+1)$ crypts at time $t$ in a two type process initiated by a single type-$i$ crypt at time $s$ (the right hand side above is equal to $p_0^{i\to(i+1)}(s,t)$ in the limit of $u_{i+1} \ll \lambda_i$). In Appendix C.1 we derive $p_0^{2\to 3}(0,t)$, the exact probability that no type-3 crypt is produced by time $t$ by a type-2 lineage started at time 0 (C.6). The exact $p_0^{i\to(i+1)}(s,t)$ can be derived by plugging in $u_3 = u_{i+1}, \lambda_2 = \lambda_i$, and $t = t - s$ into (C.6).

## 5. Accuracy of long-time approximations and derivation of $W_i$

In this section, we establish results to measure the distance between the exact process and its long-time approximations employing limiting random variables $W_i$, which we have applied in Section 3. Recall that we have shown that $e^{-\lambda_i} N_i(t) \to W_i$ as $t \to \infty$ for $i = 2, 3$. Thus, in this section we mainly focus on the approximations made for the type-2 and type-3 population. As the waiting time formulas we derived in Section 3 rely on these approximations, the following analysis can provide insight into the discrepancy between the actual waiting time and the waiting times obtained using long-time approximations. Along the way, we also obtain exact expressions for Laplace transforms of $W_2$ and $W_3$. The only assumption (or approximation) we keep here is that, when considering the behavior of the type-2 (type-3) population, we ignore the population loss due to the mutations from type-2 to type-3 (from type-3 to type-4). This assumption is referred to as *"neglecting outflows"*, which is reasonable since mutation rates are much smaller compared to division rates, i.e. $u_3 \ll \lambda_2, u_4 \ll \lambda_3$.

### 5.1. Error of type-2 approximation

To measure the distance between the population $N_2(t)$ and its approximation $e^{\lambda_2 t} W_2$, we rely on the additive property of the branching processes (Athreya and Ney, 2004, p. 201). This property allows us to derive a decomposition of $N_2(t) - e^{\lambda_2 t} W_2$ and calculate the Laplace transform of $W_2$. In addition, we use a generalized central limit theorem to give a long time error estimate of the approximation. We also provide a short time error formula which explains why the scaling function $f_{21}(t) = e^{\lambda_2 t} - \lambda_2 t - 1$ outperforms $f_{20}(t) = e^{\lambda_2 t}$.

Before giving the main results, we briefly recap related results of a one-dimensional process and their consequences (Athreya and Ney, 2004). For a one-dimensional pure birth process $Z(t)$ with birth rate $\lambda > 0$, we have $e^{-\lambda t} Z(t) \xrightarrow{a.s} V \sim$ Exponential(1). Athreya and Ney's representation theorem (Athreya and Ney, 2004, p. 123, Theorem 1) implies that $Z(t) - e^{\lambda t} V$ has the property

$$Z(t) - e^{\lambda t} V = \sum_{j=1}^{Z(t)} \left(1 - V_t^{(j)}\right) \text{ for all } t \geq 0 \text{ a.s.} \tag{15}$$

where $\{V_t^{(j)}; j = 1, 2, 3 \ldots, Z(t,\omega)\}$ are independent identically distributed copies of $V \sim$ Exponential(1) when conditioned on $Z(t)$. The consequence of this result is that one can get a central-limit-type corollary by scaling both sides properly, i.e. (Athreya and Ney, 2004 Chapter III, Section 10, p. 124, Theorem 3.)

$$\frac{Z(t) - e^{\lambda t} V}{\sqrt{Z(t)}} \xrightarrow{d.} \mathcal{N}(0,1)$$

where $\mathcal{N}(0,1)$ is the standard normal distribution. In the following, we will show that similar properties hold for the type-2 population in our process.

Consider the sub-process $(N_0(t), N_1(t), N_2(t))$ with birth rates and mutation rates as described in Section 2. The transition scheme can be summarized by

$$N_0(t) \xrightarrow{u_1} N_1(t) \xrightarrow{u_2} \overset{\circlearrowleft \text{divide at rate } \lambda_2}{N_2(t)} .$$

Recall that in Theorem 3.1 we have established the large time limit for type-2 population, i.e. $e^{-\lambda_2 t} N_2(t) \xrightarrow{a.s.} W_2$. We present a decomposition of $N_2(t) - e^{\lambda_2 t} W_2$ in the following theorem.

**Theorem 5.1.** *There exists a family of random variables* $\{V_t^{(i,j)}, j = 1, 2, \ldots, N_i(t); i = 0, 1, 2\} \cup \{T_{k,t}^{(i,j)}, j = 1, 2, \ldots, N_i(t); i = 0, 1; k = 0, 1\}$ *such that*

(a) *when conditioned on* $N_0(t), N_1(t), N_2(t)$, *the random variables in the family are independent;*
(b) *when conditioned on* $N_0(t), N_1(t), N_2(t)$, $\{V_t^{(i,j)}\}$ *are distributed as* $V \sim \text{Exp}(1)$, *and* $\{T_{k,t}^{(i,j)}\}$ *are distributed as* $T_k \sim \text{Exp}(u_{k+1})$; *and*
(c) *this family satisfies*

$$N_2(t) - e^{\lambda_2 t} W_2 = \sum_{j=1}^{N_2(t)} (1 - V_t^{(2,j)}) - \sum_{j=1}^{N_1(t)} e^{-\lambda_2 T_{1,t}^{(1,j)}} V_t^{(1,j)} - \sum_{j=1}^{N_0(t)} e^{-\lambda_2 \left(T_{0,t}^{(0,j)} + T_{1,t}^{(0,j)}\right)} V_t^{(0,j)}. \tag{16}$$

*where the equality holds for all* $t \geq 0$ *almost surely.*

The proof of this theorem is in Section 7. Roughly speaking, the intuition behind this representation theorem is that the limiting random variable can be obtained by stopping the process at a fixed time $t$ and gathering the "contributions" from existing lineages to the limiting random variable. The $j$th type-$i$ lineage is marked by superscript $(i,j)$. $k$ represents the mutation from type-$k$ to type-$(k+1)$ and $t$ represents the time correlation between the "contribution" from lineages and the total population. The first consequence of Theorem 5.1 is that we can also represent $W_2$ by the independent copies of waiting times $T_0$, $T_1$ and exponential random variable $V$. $T_0$ represents the waiting time for a single type-0 to mutate to type-1. Similarly, $T_1$ represents the waiting time for a single type-1 to mutate to type-2.

**Corollary 5.2** (*Representation of* $W_2$). *There exists a family of independent random variables* $\{T_0^{(j)}, T_1^{(j)}, V^{(j)}, j = 1, 2, \ldots, N\}$ *such that for all j,* $T_0^{(j)} \sim \text{Exp}(u_1)$, $T_1^{(j)} \sim \text{Exp}(u_2)$, $V^{(j)} \sim \text{Exp}(1)$ *and*

$$W_2 = \sum_{j=1}^{N} e^{-\lambda_2 (T_0^{(j)} + T_1^{(j)})} V^{(j)} \quad a.s. \tag{17}$$

**Proof.** Evaluating the result from Theorem 5.1(c) at $t = 0$ gives us the equality (17). □

**Corollary 5.3.** *The Laplace transform of* $W_2$ *is given by*

$$\mathcal{L}_{W_2}(\theta) = \left[ \frac{1}{u_1 - u_2} \left( u_1 \, {}_2F_1 \left( 1, \frac{u_2}{\lambda_2}; 1 + \frac{u_2}{\lambda_2}; -\theta \right) - u_2 \, {}_2F_1 \left( 1, \frac{u_1}{\lambda_2}; 1 + \frac{u_1}{\lambda_2}; -\theta \right) \right) \right]^N . \tag{18}$$

**Proof.** To derive the Laplace transform, we use the representation of $W_2$ to write $W_2 = \sum_{j=1}^{N} X_2^{(j)}$, where $X_2^{(j)} = e^{-\lambda_2 (T_0^{(j)} + T_1^{(j)})} V^{(j)}$. Let $T_0 \sim \text{Exp}(u_1)$, $T_1 \sim \text{Exp}(u_2)$, $V \sim \text{Exp}(1)$. Note that $X_2^{(j)}$ are independent and identically distributed with common distribution $X_2 = e^{-\lambda_2 (T_0 + T_1)} V$. It follows that

$$\mathcal{L}_{W_2}(\theta) = \mathbb{E}\left( e^{-\theta \sum_{j=1}^{N} X_2^{(j)}} \right) = \mathbb{E}\left( \prod_{j=1}^{n} e^{-\theta X_2^{(j)}} \right) = \prod_{j=1}^{n} \mathbb{E}\left( e^{-\theta X_2^{(j)}} \right) = \left( \mathcal{L}_{X_2}(\theta) \right)^N .$$

Hence, our goal becomes calculating $\mathcal{L}_{X_2}(\theta)$. Our first step is to compute the probability density of $T_0 + T_1$. We note that they are independent exponentially distributed random variables with parameters $u_1, u_2$ respectively. Therefore we write

$$p_{T_0+T_1}(x) = \int_0^x u_1 e^{-u_1(x-y)} u_2 e^{-u_2 y} dy = \frac{u_1 u_2}{u_2 - u_1} (e^{-u_1 x} - e^{-u_2 x}).$$

It follows that the joint probability density of $T_0 + T_1$ and $V$ is $p_{T_0+T_1,V}(x,y) = \frac{u_1 u_2}{u_2 - u_1} (e^{-u_1 x} - e^{-u_2 x}) e^{-y}$. Next we compute

$$\begin{aligned} \mathcal{L}_{X_2}(\theta) &= \mathbb{E}(e^{-\theta e^{-\lambda_2 (T_0 + T_1)} V}) \\ &= \int_0^\infty \int_0^\infty e^{-\theta e^{-\lambda_2 x} y} \frac{u_1 u_2}{u_2 - u_1} (e^{-u_1 x} - e^{-u_2 x}) e^{-y} dx\, dy \\ &= \frac{1}{u_1 - u_2} \left( u_1 \, {}_2F_1 \left( 1, \frac{u_2}{\lambda_2}; 1 + \frac{u_2}{\lambda_2}; -\theta \right) - u_2 \, {}_2F_1 \left( 1, \frac{u_1}{\lambda_2}; 1 + \frac{u_1}{\lambda_2}; -\theta \right) \right). \end{aligned}$$ □

The second consequence of Theorem 5.1 is that the long time accuracy of the type-2 approximation $e^{\lambda_2 t} W_2$ follows a central-limit-type theorem.

**Corollary 5.4** (*Type-2 Long Time Error*). *As* $t \to \infty$,

$$\frac{N_2(t) - e^{\lambda_2 t} W_2}{\sqrt{N_2(t)}} \xrightarrow{d.} \mathcal{N}(0, 1), \tag{19}$$

*where* $\mathcal{N}(0,1)$ *is a standard Gaussian random variable with mean* 0 *and variance* 1.

**Proof.** As $t \to \infty$, both type-0 population and type-1 population go extinct because they do not divide and can only mutate to further types. This means that

$$\lim_{t \to \infty} \frac{N_2(t) - e^{\lambda_2 t} W_2}{\sqrt{N_2(t)}} = \lim_{t \to \infty} \frac{1}{\sqrt{N_2(t)}} \sum_{j=1}^{N_2(t)} \left( 1 - V_t^{(2,j)} \right)$$

Due to Anscombe's generalization of the central limit theorem (see related descriptions in Athreya and Ney (2004)) , the weak convergence to a normal distribution holds and its variance is given by the variance of $1 - V_t^{(2,1)}$, which is 1. □

The third consequence of Theorem 5.1 is that we can estimate the short time behavior of the difference $N_2(t) - e^{\lambda_2 t} W_2$ near $t = 0$.

**Corollary 5.5** (*Type-2 Short Time Error*). *In the limit of* $\lambda_2 t \to 0$, *we have*

$$(\lambda_2 t + 1 + o(t)) W_2 = \sum_{j=1}^{N_1(t)} e^{-\lambda_2 T_{1,t}^{(1,j)}} V_t^{(1,j)} + \sum_{j=1}^{N_0(t)} e^{-\lambda_2 \left(T_{0,t}^{(0,j)} + T_{1,t}^{(0,j)}\right)} V_t^{(0,j)}. \tag{20}$$

**Proof.** In Theorem 5.1, the right hand side of equality (16) consists of three sums. As $\lambda_2 t \to 0$, the process approaches its initial condition. Before the first type-2 is produced, we can neglect the term $\sum_{j=1}^{N_2(t)}(1-V_t^{(2,j)})$. Then it follows that

$$e^{\lambda_2 t}W_2 = \sum_{j=1}^{N_1(t)} e^{-\lambda_2 T_{1,t}^{(1,j)}} V_t^{(1,j)} + \sum_{j=1}^{N_0(t)} e^{-\lambda_2\left(T_{0,t}^{(0,j)}+T_{1,t}^{(0,j)}\right)} V_t^{(0,j)}$$

(as $\lambda_2 t \to 0$).

Writing out the Taylor expansion of the left hand side gives us the desired equality. □

Plugging (20) into the right hand side of equality (16) gives us the following short time error estimations of the two scaling functions. In the sense that $\lambda_2 t \to 0$ (not necessarily before the production of the first type-2 crypt), we have

$$N_2(t) - f_{20}(t)W_2 \sim \sum_{j=1}^{N_2(t)} \left(1 - V_t^{(2,j)}\right) - (\lambda_2 t + 1)W_2, \tag{21}$$

$$N_2(t) - f_{21}(t)W_2 \sim \sum_{j=1}^{N_2(t)} \left(1 - V_t^{(2,j)}\right). \tag{22}$$

This indicates that, if we use the exponential as the scaling function, the difference (21) has a nonzero expected value. On the other hand, if we add the lower order terms, the difference (22) has a zero expected value and the variance is also lower. This provides insight into the observation that the scaling function $f_{21}(t) = e^{\lambda_2 t} - \lambda_2 - 1$ makes a better approximation than the scaling function $f_{20}(t) = e^{\lambda_2 t}$ at short times. When solely using the exponential scaling function, the contribution made from type-0 and type-1 lineages to the limiting random variable is ignored, especially in the short-time regime when type-3 lineages are not dominating the whole population.

### 5.2. Error of type-3 approximations

In this section, we measure the difference $N_3(t) - e^{\lambda_3 t}W_3$, i.e the distance between type-3 population and its large time approximation. The methodology we use is same as what we have used to measure the error of type-2 approximations. We recall that the transition scheme from type-0 to type-3 is :

$$N_0(t) \xrightarrow{u_1} N_1(t) \xrightarrow{u_2} \underset{N_2(t)}{\circlearrowleft\text{divide at rate } \lambda_2} \xrightarrow{u_3} \underset{N_3(t)}{\circlearrowleft\text{divide at rate } \lambda_3},$$

where we have used the "neglecting outflows" assumption for type-3. In Theorem 3.3 we have shown that $e^{-\lambda_3 t}N_3(t) \xrightarrow{a.s.} W_3$.

Before presenting a decomposition of $N_3(t) - e^{\lambda_3 t}W_3$, we will first state basic properties of a two-type birth process, which will be needed for the analysis of the behavior of the type-3 population in our model. We consider a two-type supercritical birth process with transition scheme

$$\underset{M_2(t)}{\circlearrowleft\text{divide at rate } \lambda_2} \xrightarrow{u_3} \underset{M_3(t)}{\circlearrowleft\text{divide at rate } \lambda_3}.$$

It is well known that there exists a limiting random variable $U$ such that $e^{-\lambda_3 t}M_3(t) \to U$ almost surely (Durrett and Moseley, 2010). Adapting the results from Antal and Krapivsky (2011), we obtain the Laplace transform of $U$, denoted by $\mathcal{L}_U(\theta)$ (see Appendix C.). An explicit formula of $\mathcal{L}_U(\theta)$ is presented in Eq. (C.8).

We present a decomposition of $N_3(t) - e^{\lambda_3 t}$ in the following theorem:

**Theorem 5.6.** *There exists a family of random variables* $\{V_t^{(j)}, j = 1, 2, \ldots, N_3(t)\} \cup \{U_t^{(i,j)}, j = 1, 2, \ldots, N_i(t); i = 0, 1, 2\} \cup \{T_{k,t}^{(i,j)}, j = 1, 2, \ldots, N_i(t); i = 0, 1; k = 0, 1\}$ *such that*

(a) *when conditioned on* $N_0(t), N_1(t), N_2(t), N_3(t)$, *the random variables in the family are independent;*
(b) *when conditioned on* $N_0(t), N_1(t), N_2(t), N_3(t)$, $\{V_t^j\}$ *are distributed as* $V \sim \mathrm{Exp}(1)$, $\{U_t^{(i,j)}\}$ *are distributed as* $U$ *with known Laplace transform* (C.8), *and* $\{T_{k,t}^{(i,j)}\}$ *are distributed as* $T_k \sim \mathrm{Exp}(u_{k+1})$; *and*
(c) *this family satisfies*

$$N_3(t) - e^{\lambda_3 t}W_3 = \sum_{j=1}^{N_3(t)}(1 - V_t^{(j)}) - \sum_{j=1}^{N_2(t)} U_t^{(2,j)} - \sum_{j=1}^{N_1(t)} e^{-\lambda_3 T_{1,t}^{(1,j)}} U_t^{(1,j)} - \sum_{j=1}^{N_0(t)} e^{-\lambda_3\left(T_{0,t}^{(0,j)}+T_{1,t}^{(0,j)}\right)} U_t^{(0,j)}, \tag{23}$$

*where the equality holds for all* $t \geq 0$ *almost surely.*

Decomposition (23) provides insight into the accuracy of approximation $N_3(t) \approx e^{\lambda_3 t}W_3$. The right hand side of (23) can be demarcated in the following way:

$$\underbrace{\sum_{j=1}^{N_3(t)}(1 - V_t^{(j)})}_{(A)} - \underbrace{\sum_{j=1}^{N_2(t)} U_t^{(2,j)}}_{(B)} \underbrace{- \sum_{j=1}^{N_1(t)} e^{-\lambda_3 T_{1,t}^{(1,j)}} U_t^{(1,j)} - \sum_{j=1}^{N_0(t)} e^{-\lambda_3\left(T_{0,t}^{(0,j)}+T_{1,t}^{(0,j)}\right)} U_t^{(0,j)}}_{(C)}.$$

In the above expression, only $(A)$ has a zero mean value. $(B)$ and $C$ have non-positive first moments at any time $t > 0$. The scaling function $f_{31}$ is designed to cancel the first moment of $(B)$ and $(C)$. We have some knowledge for the behaviors of $(A), (B)$ and $(C)$ in the long-time limit. $N_0(t)$ and $N_1(t)$ are expected to go extinct for sufficient large $t$. Therefore, the term $(C)$ is expected to vanish. Due to Anscombe's generalized central limit theorem, $(A)$ and $(B)$, when scaled by $\sqrt{N_3(t)}$ and $\sqrt{N_2(t)}$ respectively, behave like Gaussian distributions as $t \to \infty$. Since $\lambda_3 > \lambda_2$, the second moment of $(A) + (B)$ is dominated by the type-3 population and the first moment is dominated by the type-2 population.

The following corollary gives us a representation of $W_3$.

**Corollary 5.7** (*Representation of* $W_3$). *There exists a family of independent random variables* $\{T_0^{(j)}, T_1^{(j)}, U^{(j)}, j = 1, 2, \ldots, N\}$ *such that for all j,* $T_0^{(j)} \sim \mathrm{Exp}(u_1)$, $T_1^{(j)} \sim \mathrm{Exp}(u_2)$, $U^{(j)}$ *are independent and identically distributed with Laplace transform* (C.8) *and*

$$W_3 = \sum_{j=1}^{N} e^{-\lambda_3(T_0^{(j)}+T_1^{(j)})}U^{(j)} \quad a.s. \tag{24}$$

**Proof.** Evaluating the equality in (c) of Theorem 5.6 at $t = 0$ gives us Eq. (24). □

**Corollary 5.8.** *The Laplace transform of* $W_3$ *is given by*

$$\mathcal{L}_{W_3}(\theta) = \left(\int_0^\infty \mathcal{L}_U(\theta e^{-\lambda_3 x})\frac{u_1 u_2}{u_2 - u_1}(e^{-u_1 x} - e^{-u_2 x})\mathrm{d}x\right)^N. \tag{25}$$

**Proof.** Let $T_0 := T_0^{(1)} \sim \mathrm{Exp}(u_1)$, $T_1 := T_1^{(1)} \sim \mathrm{Exp}(u_2)$, $U := U^{(1)}$. To compute the Laplace transform, we use the representation theorem and write $W_3 = \sum_{j=1}^N X_3^{(j)}$, where $\{X_3^{(j)}, j = 1, 2, \ldots, N\}$ is a collection of independent and identically distributed random variables with common distribution $X_3 := e^{-\lambda_3(T_0+T_1)}U$. Then it

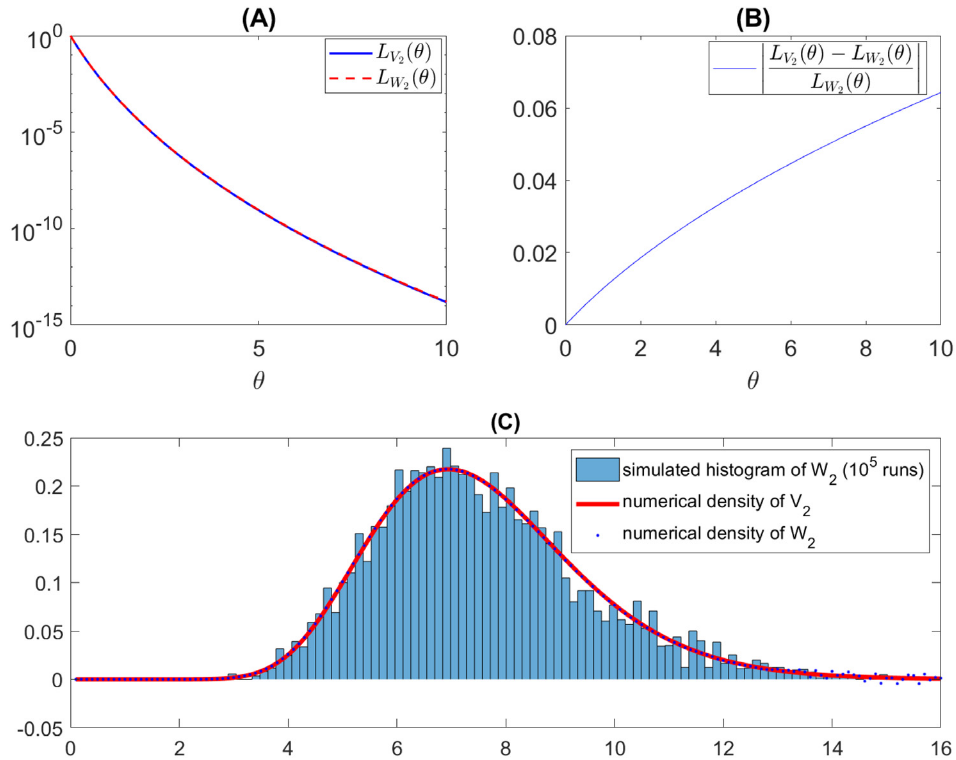


**Fig. 6.** (A) Comparison of $\mathcal{L}_{W_2}(\theta)$ (Eq. (3)) and $\mathcal{L}_{V_2}(\theta)$ (Eq. (6)) on the interval $[0, 10]$. (B) Relative difference of $\mathcal{L}_{W_2}(\theta)$ and $\mathcal{L}_{V_2}(\theta)$ which is defined by $|\mathcal{L}_{W_2}(\theta) - \mathcal{L}_{V_2}(\theta)|/|\mathcal{L}_{W_2}(\theta)|$. (C) Comparison of the numerical probability density functions of $W_2$ and $V_2$ as well as the simulated histogram of $W_2$.

follows that

$$\mathcal{L}_{W_3}(\theta) = \mathbb{E}(e^{-\theta W_3}) = \prod_{j=1}^{N} \mathcal{L}_{X_3^{(j)}}(\theta) = \left(\mathcal{L}_{X_3}(\theta)\right)^N .$$

Next, we evaluate the Laplace transform of $X_3$.

$$\begin{aligned}\mathcal{L}_{X_3}(\theta) &= \mathbb{E}\left(\mathbb{E}(\exp(-\theta U \exp(-\lambda_3(T_0+T_1)))|T_0+T_1)\right)\\ &= \mathbb{E}\left(\mathcal{L}_U(\theta \exp(-\lambda_3(T_0+T_1)))\right)\\ &= \int_0^\infty \mathcal{L}_U(\theta e^{-\lambda_3 x})\frac{u_1 u_2}{u_2-u_1}(e^{-u_1 x}-e^{-u_2 x})\mathrm{d}x\end{aligned}$$

Finally, the Laplace transform of $W_3$ can be written as

$$\mathcal{L}_{W_3}(\theta) = \left(\int_0^\infty \mathcal{L}_U(\theta e^{-\lambda_3 x})\frac{u_1 u_2}{u_2-u_1}(e^{-u_1 x}-e^{-u_2 x})\mathrm{d}x\right)^N . \qquad \square$$

## 6. Approximating the limiting random variables

In this Section, we formally derive random variables $V_i$, which serve as approximations of long-time limiting random variables $W_i$. The usefulness of random variables $V_i$ comes from the fact that their Laplace transforms are easier to obtain and manipulate when evaluating expressions for waiting times. To obtain $V_i$, we use auxiliary processes $N_{i-1}^*$ and $\tilde{N}_i$, which approximate the original processes $N_{i-1}$ and $N_i$. For example, for $i = 2$, we show in Section 3 that $N_1^*(t) = u_1 N t$ is a reasonable approximation for $N_1(t)$. Thus, we construct an auxiliary process $\tilde{N}_2$, in which new type-2 crypts are produced with rate $u_2 N_1^*(t)$. Then, we obtain $V_2$ as the limiting random variable of the process $\tilde{N}_2$, namely we show that $e^{-\lambda_2 t}\tilde{N}_2 \to V_2$ as $t \to \infty$. Similarly, we start with $N_2^* = e^{\lambda_2 t} V_2$ as the process that produces the auxiliary type-3 process, $\tilde{N}_3$, and obtain $V_3$ as the limiting random variable of $\tilde{N}_3$.

We start by considering $\tilde{N}_2(t)$, the population of type-2 crypts produced by $N_1^*(t) := u_1 N t$ with mutation rate $u_2$, and its long time behavior. Recall that type-2 crypts can divide with rate $\lambda_2$.

**Theorem 6.1.** *$e^{-\lambda_2 t}\tilde{N}_2(t) \to V_2$ a.s. and in $L^1$ with*

$$\mathbb{E}[V_2] = \frac{N u_1 u_2}{\lambda_2^2}$$

*and*

$$\mathcal{L}_{V_2}(\theta) = \mathbb{E}\left[e^{-\theta V_2}\right] = \exp\left(\frac{N u_1 u_2 \mathrm{PolyLog}(2, -\theta)}{\lambda_2^2}\right). \tag{26}$$

We recall that $\mathrm{PolyLog}(n, z)$, $n \in \mathbb{Z}$, $n \geq 2$, $z \in \mathbb{C}$ represents the polylogarithm function (DLMF, 2022, 25.12.10).

**Corollary 6.2.** *$\frac{\tilde{N}_2(t)}{f_2(t)} \to V_2$ a.s. and in $L^1$ for all $f_2(t) \in F_2$.*

We recall that $F_2 = \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) = 1, f_2(t) \geq 0\}$.

To measure the distance between $W_2$ and $V_2$, we present $\mathcal{L}_{W_2}(\theta)$, $\mathcal{L}_{V_2}(\theta)$ and their relative difference on $\theta \in [0, 10]$ in panels (A) and (B) of Fig. 6. We choose this domain for $\theta$ since the functions values are negligible when $\theta > 10$ (panel (A) of Fig. 6). Furthermore, we numerically compare the density functions of $W_2$, $V_2$ and the histogram of $e^{-\lambda_2 t}N_2(t)$ at $t = 60$ obtained from exact computer simulations of the type-2 process (panel (C) of Fig. 6). The densities are obtained through a numerical inverse Laplace transform using the Talbot method (Abate and Whitt, 2006), and are in good agreement.

Next, we discuss the construction of $V_3$. Consider a system $(N_2^*(t), \tilde{N}_3(t))$ with $N_2^*(t) := f_2(t)V_2, f_2(t) \in F_2$ and $\tilde{N}_3(t)$ denoting the number of type-3 crypts. In the process, the type-3 crypts are

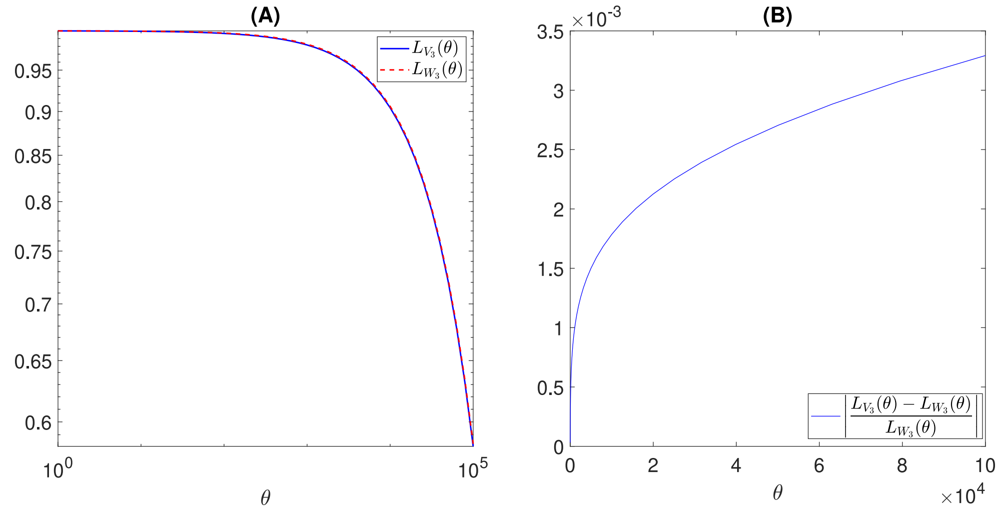


**Fig. 7.** (A) Comparison of $\mathcal{L}_{W_3}(\theta)$ (Eq. (7)) and $\mathcal{L}_{V_2}(\theta)$ (Eq. (B.6)) on the interval $[0, 10^5]$. (B) Relative difference of $\mathcal{L}_{W_3}(\theta)$ and $\mathcal{L}_{V_3}(\theta)$ which is defined by $|\mathcal{L}_{W_3}(\theta) - \mathcal{L}_{V_3}(\theta)|/|\mathcal{L}_{W_3}(\theta)|$.

produced at rate $u_3 N_2^*(t)$ and a single type-3 crypt can divide with rate $\lambda_3$. The following large-time limit holds:

**Theorem 6.3.** *$\frac{\tilde{N}_3(t)}{f_3(t)} \to V_3$ a.s. and in $L^1$ with*

$$\mathbb{E}[V_3] = \frac{N u_1 u_2 u_3}{\lambda_2^2} \int_0^\infty f_2(s) e^{-\lambda_3 s} \mathrm{d}s,$$

*and*

$$\mathcal{L}_{V_3}(\theta) = \mathbb{E}[e^{-\theta V_3}] = \mathcal{L}_{V_2}\left(u_i \int_0^\infty \frac{\theta f_2(s)}{\theta + e^{\lambda_i s}}\, \mathrm{d}s\right),$$

*for all $f_3(t) \in F_3 = \{f \in C(\mathbb{R}) | \lim_{t\to\infty} e^{-\lambda_3 t} f_3(t) = 1, f_3(t) \geq 0\}$.*

As we have mentioned in Section 3, we employ the scaling function $f_{21}(t)$ to compute $\mathcal{L}_{V_3}(\theta)$. In Fig. 7, we numerically compare this Laplace transform with $\mathcal{L}_{W_3}(\theta)$ on $\theta \in [0, 10^5]$. This domain of $\theta$ results from evaluating $\mathbb{P}(\tau_4 \leq t)$, the distribution function of the waiting time to the first type-4 crypt, in the lifespan $t \in [0, 80]$. We recall that when approximating $\mathbb{P}(\tau_4 \leq t)$, we need to evaluate $\mathcal{L}_{V_3}(\theta)$ at $\theta(t) = u_4 \int_0^t f_3(s)ds$. Plugging in the parameters in our model results in $\theta(t) \in [0, 10^5]$ for $t \in [0, 80]$. Both panels (A) and (B) in Fig. 7 show that $\mathcal{L}_{W_3}(\theta)$ and $\mathcal{L}_{V_3}(\theta)$ are in a good agreement on the domain of interest.

The previous discussions imply that there is a recursive relationship involved in the sequence of $(N_{i-1}^*, \tilde{N}_i)$ approximations. After obtaining the limiting random variable $V_i$ in $(N_{i-1}^*, \tilde{N}_i)$, $N_i^*(t) = f_i(t)V_i$ can be employed when moving to the next two-type process $(N_i^*(t), \tilde{N}_{i+1}(t))$. This recursion allows us to present an iterative method for computing the Laplace transforms of random variables $V_i$.

**Lemma 6.4.** *Consider $\tilde{N}_i(t)$, the population of type-i individuals produced by $N_{i-1}^*(t) = f_{i-1}(t)V_{i-1}$ with mutation rate $u_i > 0$. Each type-i individual can divide with rate $\lambda_i > 0$. Suppose $e^{-\lambda_i t}\tilde{N}_i(t) \xrightarrow{a.s.} V_i$ and $f_{i-1}(t)e^{-\lambda_i t}$ is integrable on $[0, \infty)$, then*

$$\mathcal{L}_{V_i}(\theta) := \mathbb{E}\left[e^{-\theta V_i}\right] = \mathcal{L}_{V_{i-1}}\left(u_i \int_0^\infty \frac{\theta f_{i-1}(s)}{\theta + e^{\lambda_i s}}\, \mathrm{d}s\right).$$

**Proof.** We start with a lemma that provides the Laplace transform of $N_i(t)$ conditional on the population of its precursor, $N_{i-1}(t)$.

**Lemma 6.5.** *Let $Z_i(t)$ be the number of type-i individuals in a pure-birth process that starts with $Z_i(0) = 1$ individuals at time $t = 0$. Then*

$$\mathbb{E}\left[e^{-\theta N_i(t)}\middle| N_{i-1}(s), s \leq t\right] = \exp\left(-u_i \int_0^t N_{i-1}(s)(1 - \phi_i(\theta, t-s))\mathrm{d}s\right), \quad (27)$$

*where $\phi_i(\theta, t) := \mathbb{E}\left[e^{-\theta Z_i(t)}\right]$.*

One can prove the above lemma by following the procedure of Lemma 2 in Durrett and Moseley (2010), replacing the start time with $s = 0$.

Now going back to the proof of Lemma 6.4, we consider the star process approximation

$$N_{i-1}^*(t) = f_{i-1}(t)V_{i-1}.$$

Applying Lemma 6.5 the 2-type process $\left(N_{i-1}^*(t), \tilde{N}_i(t)\right)$ gives us

$$\mathbb{E}[e^{-\theta \tilde{N}_i(t)} | V_{i-1}] = \exp\left(-u_i \int_0^t f_{i-1}(s)V_{i-1}(1 - \phi_i(\theta, t-s))\mathrm{d}s\right).$$

Replacing $\theta$ with $\theta e^{-\lambda_i t}$,

$$\mathbb{E}[e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)} | V_{i-1}] = \exp\left(-u_i \int_0^t f_{i-1}(s)V_{i-1}(1 - \phi_i(e^{-\lambda_i t}\theta, t-s))\mathrm{d}s\right), \quad (28)$$

then for each subclone of $N_i$, by Eq. (14), we have

$$Z_i(t-s)e^{-\lambda_i(t-s)} \to \text{Exponential}(1)\ a.s.$$

Thus, it follows that

$$Z_i(t-s)e^{-\lambda_i t} \to \text{Exponential}(e^{\lambda_i s}).$$

Considering the following limit involving terms on the right hand side of (28), we have

$$\begin{aligned}\lim_{t\to\infty} 1 - \phi_i(\theta e^{-\lambda_i t}, t-s) &= 1 - \int_0^\infty e^{-\theta x} e^{\lambda_i s} \exp(-x e^{\lambda_i s})\mathrm{d}x \\ &= \int_0^\infty (1 - e^{-\theta x}) e^{\lambda_i s} \exp(-x e^{\lambda_i s})\mathrm{d}x \\ &= \frac{\theta}{\theta + e^{\lambda_i s}}.\end{aligned}$$

Note that as $t \to \infty$, the left hand side of (28) yields

$$\lim_{t\to\infty} \mathbb{E}[e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)}|V_{i-1}] = \mathbb{E}[\lim_{t\to\infty} e^{-\theta e^{-\lambda_i t}\tilde{N}_i(t)}|V_{i-1}] = \mathbb{E}[e^{-\theta V_i}|V_{i-1}],$$

as switching the limit and the integration is allowed by the dominated convergence theorem. Thus, we can write

$$\mathbb{E}[e^{-\theta V_i}|V_{i-1}] = \exp\left(-u_i V_{i-1}\int_0^\infty \frac{\theta f_{i-1}(s)}{\theta + e^{\lambda_i s}}\mathrm{d}s\right).$$

Taking expectation on both sides gives

$$\mathbb{E}[e^{-\theta V_i}] = \mathcal{L}_{V_{i-1}}\left(u_i\int_0^\infty \frac{\theta f_{i-1}(s)}{\theta + e^{\lambda_i s}}\,\mathrm{d}s\right). \quad \square$$

## 7. Proofs

In this section, we prove Theorems 3.1, 3.3, 5.1, 5.6, 6.1, 6.3 and Corollaries 3.2, 3.4, 6.2.

**Proof of Theorem 3.1.** By Lemma 4.1, consider

$$I_2 = \int_0^\infty u_2 N_1(s)e^{-\lambda_2 s}\mathrm{d}s.$$

Since

$$\begin{aligned}\mathbb{E}[I_2] =& \mathbb{E}\left[\int_0^\infty u_2 e^{-\lambda_2 s}N_1(s)\mathrm{d}s\right]\\ \text{(By Tonelli's theorem)} =& \int_0^\infty u_2\mathbb{E}[N_1(s)]e^{-\lambda_2 s}\mathrm{d}s\\ \le& \int_0^\infty u_2 u_1 N s e^{-\lambda_2 s}\mathrm{d}s\\ =& Nu_1u_2\int_0^\infty s e^{-\lambda_2 s}\mathrm{d}s\\ =& \frac{Nu_1u_2}{\lambda_2^2} < \infty,\end{aligned}$$

there exists $W_2$ s.t.

$$e^{-\lambda_2 t}N_2(t) \to W_2 \ a.s. \quad \text{as } t \to \infty.$$

Next, we show uniform integrability so that $\mathbb{E}[W_2]$ is well-defined. We prove this for $N_1, N_2$ and $N_3$ in Lemma A.1. Since $L^1$ convergence is guaranteed, $\mathbb{E}[W_2] = \mathbb{E}[I_2]$, so that

$$\begin{aligned}\mathbb{E}[W_2] &= \int_0^\infty u_2\mathbb{E}[N_1(s)]e^{-\lambda_2 s}\mathrm{d}s\\ &= \int_0^\infty u_2 N(1-e^{-u_1 s})e^{-\lambda_2 s}\mathrm{d}s\\ &= \frac{Nu_1u_2}{\lambda_2(\lambda_2+u_1)}.\end{aligned}$$

The Laplace transform of $W_3$ is derived in Corollary 5.3.

**Proof of Corollary 3.2.** The proof follows directly from Theorem 3.1. Observe that

$$\forall\ \omega \in \{\omega : \lim_{t\to\infty} e^{-\lambda_2 t}N_2(\omega, t) = W_2(\omega)\},$$

we have

$$\lim_{t\to\infty}\frac{N_2(\omega,t)}{f_2(t)} = \lim_{t\to\infty} e^{-\lambda_2 t}f_2(t)\lim_{t\to\infty}\frac{N_2(\omega,t)}{f_2(t)} = V_2(\omega).$$

Then since $f_2 \in F_2$, for $t > 0$ sufficiently large, we have $f_2(t) > 0$ and $e^{\lambda_2 t}/f_2(t) < M$. Thus, without loss of generality, we can assume that $f_2(t) > 0$ for $t \ge 0$. In this case, we have

$$\sup_t \mathbb{E}\left(\frac{N_2(t)}{f_2(t)}\right)^2 \le \left(\sup_t\left(e^{\lambda_2 t}/f_2(t)\right)\right)^2 \sup_t \mathbb{E}\left[\left(N_2(t)e^{-\lambda_2 t}\right)^2\right].$$

By Lemma A.1 $\sup_t \mathbb{E}\left[\left(N_2(t)e^{-\lambda_2 t}\right)^2\right]$ is bounded. Therefore the expression on the left hand side is bounded. Hence $\{N_2(t)/f_2(t), t \ge 0\}$ is square integrable and the convergence is in $L^1$.

**Proof of Theorem 3.3.** By Lemmas 4.1 and A.1, we need to verify that $I_3$ has finite expectation,

$$\begin{aligned}\mathbb{E}[I_3] &= \int_0^\infty u_3\mathbb{E}[N_2(s)]e^{-\lambda_3 s}\mathrm{d}s\\ &\le \int_0^\infty u_3\frac{u_1u_2N}{\lambda_2^2}(e^{\lambda_2 s} - \lambda_2 s - 1)e^{-\lambda_3 s}\mathrm{d}s\\ &= \frac{Nu_1u_2u_3}{\lambda_2^2}\int_0^\infty (e^{(\lambda_2-\lambda_3)s} - \lambda_2 s e^{-\lambda_3 s} - e^{-\lambda_3 s})\mathrm{d}s\\ &= \frac{Nu_1u_2u_3}{(\lambda_3-\lambda_2)\lambda_3^2} < \infty.\end{aligned}$$

And the expected value of $W_3$ is given by the expected value of $I_3$,

$$\begin{aligned}\mathbb{E}[W_3] &= \mathbb{E}[I_3]\\ &= \int_0^\infty u_3\mathbb{E}[N_2(s)]e^{-\lambda_3 s}\mathrm{d}s\\ &= \int_0^\infty u_3 N u_2\frac{u_1(e^{\lambda_2 s}-1) + \lambda_2(e^{-u_1 s}-1)}{\lambda_2(\lambda_2+u_1)}e^{-\lambda_3 s}\mathrm{d}s\\ &= \frac{Nu_1u_2u_3}{(\lambda_3-\lambda_2)\lambda_3(\lambda_3+u_1)}.\end{aligned}$$

The Laplace transform of $W_3$ is derived in Corollary 5.8.

**Proof of Corollary 3.4.** The outline of this proof is similar to the proof of Corollary 3.2. Almost sure convergence holds since for fixed $\omega$,

$$\lim_{t\to\infty} N_3/f_3(t) = \lim_{t\to\infty} e^{-\lambda_3 t}f_3(t)\lim_{t\to\infty} N_3/f_3(t) = \lim_{t\to\infty} e^{-\lambda_3 t}N_3.$$

Convergence in $L^1$ is a consequence of the square integrability of $\{N_3(t)/f_3(t), t \ge 0\}$.

**Proof of Theorem 5.1.** Before proving the theorem, we first introduce some new notations. Let $N_i^{(a,b,c)}(t)$ be the population of type-$i$ crypts in the multi-type branching process

$$N_0(t) \xrightarrow{u_1} N_1(t) \xrightarrow{u_2} \overset{\circlearrowleft\text{divide at rate } \lambda_2}{N_2(t)},$$

where initially $(N_0(0), N_1(0), N_2(0)) = (a, b, c)$. We observe that the $N_i(t)$ in the model (without the superscript) can be written as $N_i(t) = N_i^{(N,0,0)}(t)$. Then we define the *type*-2 *lineage* which was initiated by a single type-$i$ crypt as:

$$\begin{aligned}Z_{0,2}(t) &:= N_2^{(1,0,0)}(t) \quad t \ge 0,\\ Z_{1,2}(t) &:= N_2^{(0,1,0)}(t) \quad t \ge 0,\\ Z_{2,2}(t) &:= N_2^{(0,0,1)}(t) \quad t \ge 0.\end{aligned}$$

For our convenience, we allow these lineages to be defined on the negative time axis, i.e. $Z_{i,2}(t) = 0,\ \forall t < 0$. Then $\{Z_{2,2}(t), t \ge 0\}$ is a supercritical pure birth process with birth rate $\lambda_2$. Thus it follows that

$$e^{-\lambda_2 t}Z_{2,2}(t) = V \sim \mathrm{Exp}(1).$$

The other type-2 lineages can be represented by the following lemma.

**Lemma 7.1.** *There exist $T_0 \sim \text{Exp}(u_1)$ and $T_1 \sim \text{Exp}(u_2)$ such that*

*(a) $T_0$, $T_1$ and $\{Z_{2,2}(t)\}$ are independent;*
*(b) $Z_{1,2}(t) = Z_{2,2}(t - T_1)$, $Z_{0,2}(t) = Z_{2,2}(t - T_0 - T_1)$; and*
*(c) the following equations hold almost surely*

$$\lim_{t\to\infty} e^{-\lambda_2 t} Z_{1,2}(t) = e^{-\lambda_2 T_1} \lim_{t\to\infty} e^{-\lambda_2(t-T_1)} Z_{2,2}(t - T_1) = e^{-\lambda_2 T_1} V, \tag{29}$$

$$\lim_{t\to\infty} e^{-\lambda_2 t} Z_{0,2}(t) = e^{-\lambda_2(T_0+T_1)} \lim_{t\to\infty} e^{-\lambda_2(t-T_0-T_1)} \times Z_{2,2}(t - T_0 - T_1) = e^{-\lambda_2(T_0+T_1)} V, \tag{30}$$

*where $V \sim \text{Exp}(1)$.*

**Proof.** We use the *minimal process* to construct our multi-type branching process. For a detailed description of the construction of minimal process, please refer to Chapter V, Section 7 of Athreya and Ney (2004). In our model, a type-1 crypt can only mutate into a type-2 crypt. Thus, after the process $Z_{1,2}(t) = N_2^{(0,1,0)}(t)$ incurs its first mutation after an exponentially distributed waiting time $T_0$, it corresponds to $Z_{2,2}(t) = N_2^{(0,0,1)}(t)$. This enables us to write

$Z_{1,2}(t) = Z_{2,2}(t - T_1)$.

Similarly, for $Z_{0,2}(t)$, we need one mutation for this process to become $Z_{1,2}(t)$. Thus it follows that $Z_{0,2}(t) = Z_{2,2}(t - T_0 - T_1)$, where $T_1 \sim \text{Exp}(u_2)$ is the waiting time of the mutation from type-1 to type-2. For (c), note that $\mathbb{P}(T_1 < \infty) = 1, \mathbb{P}(T_1 + T_2 < \infty) = 1$. Thus $t - T_1 \to \infty$ almost surely and $t - T_0 - T_1 \to \infty$ almost surely. Therefore (29) and (30) holds almost surely. □

Now we return to the proof of Theorem 5.1.

**Proof.** Since $e^{-\lambda_2 t} N_2(t) \to W_2$ almost surely, there exists $\hat{A}$ such that $\lim_{t\to\infty} e^{-\lambda_2 t} N_2(t,\omega) = W_2(\omega)$ for all $\omega \in \hat{A}$ and $\mathbb{P}(\hat{A}) = 1$. This limit remains if we shift the time by a finite value, thus

$$\lim_{s\to\infty} e^{-\lambda_2(t+s)} N_2(t+s,\omega) = W_2(\omega) \tag{31}$$

$$\lim_{s\to\infty} e^{-\lambda_2 s} N_2(t+s,\omega) = e^{\lambda_2 t} W_2(\omega), \ \ \forall \omega \in \hat{A}. \tag{32}$$

On the other hand, given the information at time $t$, i.e. $(N_0(t), N_1(t), N_2(t))$, we can represent the population $N_2(t+s)$ by the additive property of a multi-type branching process (Athreya and Ney, 2004),

$$N_2(t+s) = N_2^{(N_0(t),N_1(t),N_2(t))}(s) = \sum_{j=1}^{N_2(t)} Z_{2,2,t}^{(j)}(s) + \sum_{j=1}^{N_1(t)} Z_{1,2,t}^{(j)}(s) + \sum_{j=1}^{N_0(t)} Z_{0,2,t}^{(j)}(s),$$

in which $\{Z_{i,2,t}^{(j)}(s), i = 0,1,2\}$, when conditioned on $N_2(t), N_1(t), N_0(t)$ are independent copies of lineages $Z_{i,2}(s), i = 0, 1, 2$. Then we can use another approach to compute the left hand side of Eq. (32). For fixed $t$, we multiply both sides of the above decomposition by $e^{-\lambda_2 s}$ and take limit as $s \to \infty$.

$$e^{-\lambda_2 s} N_2(t+s) = e^{-\lambda_2 s}\left(\sum_{j=1}^{N_2(t)} Z_{2,2,t}^{(j)}(s) + \sum_{j=1}^{N_1(t)} Z_{1,2,t}^{(j)}(s) + \sum_{j=1}^{N_0(t)} Z_{0,2,t}^{(j)}(s)\right)$$

$$\text{(by (29) and (30))} \xrightarrow{a.s.} \sum_{j=1}^{N_2(t)} V_t^{(2,j)} + \sum_{j=1}^{N_1(t)} e^{-\lambda_2 T_{1,t}^{(1,j)}} V_t^{(1,j)} + \sum_{j=1}^{N_0(t)} e^{-\lambda_2\left(T_{0,t}^{(0,j)} + T_{1,t}^{(0,j)}\right)} V_t^{(0,j)},$$

where $\{V_t^{(i,j)}, i = 0, 1, 2\} \cup \{T_{k,t}^{(i,j)}, i = 0, 1, k = 0, 1\}$, when conditioned on $N_0(t), N_1(t), N_2(t)$ is a family of independent random variables in which $\{V_t^{(i,j)}, i = 0, 1, 2\}$ are distributed as $V \sim \text{Exp}(1)$, $\{T_{0,t}^{(i,j)}, i = 0, 1\}$ are distributed as $T_0 \sim \text{Exp}(u_1)$ and $\{T_{1,t}^{(i,j)}, i = 0, 1\}$ are distributed as $T_1 \sim \text{Exp}(u_2)$. Since the convergence holds almost surely, for each $t$ we can find $A_t$ such that $\mathbb{P}(A_t) = 1$ and

$$\lim_{s\to\infty} e^{-\lambda_2 s} N_2(t+s,\omega) = \sum_{j=1}^{N_2(t,\omega)} V_t^{(2,j)}(\omega) + \sum_{j=1}^{N_1(t,\omega)} e^{-\lambda_2 T_{1,t}^{(1,j)}(\omega)} V_t^{(1,j)}(\omega) + \sum_{j=1}^{N_0(t,\omega)} e^{-\lambda_2\left(T_{0,t}^{(0,j)}(\omega) + T_{1,t}^{(0,j)}(\omega)\right)} V_t^{(0,j)}(\omega)$$

for all $\omega \in A_t$. Now let $\{t_k; k = 1, 2, \dots\}$ be the set of non-negative rationals and define $A_q = \hat{A} \cap \left(\bigcap_{k=1}^{\infty} A_{t_k}\right)$. Then it follows that $\mathbb{P}(A_q) = 1$ and on $A_q$ we have

$$e^{\lambda_2 t} W_2(\omega) = \sum_{j=1}^{N_2(t,\omega)} V_t^{(2,j)}(\omega) + \sum_{j=1}^{N_1(t,\omega)} e^{-\lambda_2 T_{1,t}^{(1,j)}(\omega)} V_t^{(1,j)}(\omega) + \sum_{j=1}^{N_0(t,\omega)} e^{-\lambda_2\left(T_{0,t}^{(0,j)}(\omega) + T_{1,t}^{(0,j)}(\omega)\right)} V_t^{(0,j)}(\omega)$$

for all non-negative rational times. Finally, by the right-continuity of the process, for all $t \geq 0$ the equality holds almost surely. This implies

$$N_2(t) - e^{\lambda_2 t} W_2 = \sum_{j=1}^{N_2(t)} (1 - V_t^{(2,j)}) - \sum_{j=1}^{N_1(t)} e^{-\lambda_2 T_{1,t}^{(1,j)}} V_t^{(1,j)} - \sum_{j=1}^{N_0(t)} e^{-\lambda_2\left(T_{0,t}^{(0,j)} + T_{1,t}^{(0,j)}\right)} V_t^{(0,j)}$$

where the equality holds for all $t \geq 0$ almost surely. □

**Proof of Theorem 5.6.** The outline of proving this theorem is similar to the proof of Theorem 5.1. Let $N_i^{(a,b,c,d)}(t)$ be the population of type-$i$ ($i \leq 3$) crypts in the multi-type branching process

$$N_0(t) \xrightarrow{u_1} N_1(t) \xrightarrow{u_2} \overset{\circlearrowleft \text{divide at rate } \lambda_2}{N_2(t)} \xrightarrow{u_3} \overset{\circlearrowleft \text{divide at rate } \lambda_3}{N_3(t)},$$

where initially $(N_0(0), N_1(0), N_2(0), N_3(0)) = (a, b, c, d)$. By using this notation, $N_i(t)$ can be written as $N_i(t) = N_i^{(N,0,0,0)}(t)$. To describe the type-3 population initiated by a single type-$i$ crypt, we define $Z_{i,3}(t)$ to be the *type*-3 *lineage* started with a type-$i$ crypt. Note that these lineages are allowed to be defined on the negative time axis, i.e. $Z_{i,3}(t) = 0, \ \forall t < 0$. The behavior of each type-3 lineage is clear: $\{Z_{3,3}(t), t \geq 0\}$ is a supercritical pure birth process with birth rate $\lambda_3$; $Z_{2,3}(t)$ is the second type in a two-type process which we discuss in Appendix C; $Z_{1,3}(t)$ and $Z_{0,3}(t)$ can be treated as $Z_{2,3}(t)$ after one jump or two jumps respectively. Thus, it follows that there exist $T_0 \sim \text{Exp}(u_1)$ and $T_1 \sim \text{Exp}(u_2)$ such that

$$e^{-\lambda_3 t} Z_{3,3}(t) = V \sim \text{Exp}(1)$$
$$e^{-\lambda_3 t} Z_{2,3}(t) = U$$

$$e^{-\lambda_3 t} Z_{2,3}(t) = e^{-\lambda_3 T_1} U$$
$$e^{-\lambda_3 t} Z_{1,3}(t) = e^{-\lambda_3 (T_0+T_1)} U.$$

In the above equations, $U$ has an explicit Laplace transform (C.8). Now we go back to the proof for Theorem 5.6.

**Proof.** Since $e^{-\lambda_3 t} N_3(t) \to W_3$ almost surely, $\lim_{s\to\infty} e^{-\lambda_3 s} N_3(t+s) = e^{\lambda_3 t} W_3$ for a fixed $t$. Instead of taking the limit as $t \to \infty$ directly, given $(N_0(t), N_1(t), N_2(t), N_3(t))$, we can represent the population $N_3(t+s)$ by the additive property of a multi-type branching process:

$$\begin{aligned} N_3(t+s) &= N_3^{(N_0(t),N_1(t),N_2(t),N_3(t))}(s) \\ &= \sum_{j=1}^{N_3(t)} Z_{3,3,t}^{(j)}(s) + \sum_{j=1}^{N_2(t)} Z_{2,3,t}^{(j)}(s) \\ &\quad + \sum_{j=1}^{N_1(t)} Z_{1,3,t}^{(j)}(s) + \sum_{j=1}^{N_0(t)} Z_{0,3,t}^{(j)}(s), \end{aligned}$$

in which $\{Z_{i,3,t}^{(j)}(s), i = 0,1,2,3\}$, when conditioned on $N_0(t)$, $N_1(t), N_2(t), N_3(t)$ are independent copies of type-3 lineages $Z_{i,3}(s), i = 0,1,2,3$. Multiplying both sides of the above decomposition by $e^{-\lambda_3 s}$ and taking limit as $s \to \infty$ gives us

$$\begin{aligned} e^{\lambda_3 t} W_3 &\overset{a.s.}{=} \sum_{j=1}^{N_3(t)} V_t^{(j)} + \sum_{j=1}^{N_2(t)} U_t^{(2,j)} + \sum_{j=1}^{N_1(t)} e^{-\lambda_3 T_{1,t}^{(1,j)}} U_t^{(1,j)} \\ &\quad + \sum_{j=1}^{N_0(t)} e^{-\lambda_3\left(T_{0,t}^{(0,j)} + T_{1,t}^{(0,j)}\right)} U_t^{(0,j)}, \end{aligned}$$

where $V_t^j, U_t^{(i,j)}, T_{k,t}^{(i,j)}$, when conditioned on $N_0(t), N_1(t), N_2(t)$ are independent copies with known Laplace transforms. Finally, by the right-continuity of the process, we are able to find a measure-1 set in the sample space such that the equality holds almost surely for all $t \geq 0$. □

**Proof of Theorem 6.1.** By Lemmas 4.1 and A.2, we need to verify that $I_2^*$ has finite expectation. We note Lemma 4.1 still holds true if the initial type $N_0(t)$ is replaced by $N_0^*(t)$, a non-negative right continuous process with $\mathbb{E}|N_0^*(t)| < \infty$. In this case, $I_2^*$ is deterministic and has a finite expected value

$$\mathbb{E}[I_2^*] = I_2^* = \int_0^\infty u_2 u_1 N s e^{-\lambda_2 s} \mathrm{d}s = \frac{N u_1 u_2}{\lambda_2^2} < \infty.$$

By Lemma 4.1, we must have $e^{-\lambda_2 t}\tilde{N}_2(t) \xrightarrow{a.s.} V_2$. Next, we show uniform integrability which guarantees $L^1$ convergence. It is shown in Lemma A.2 that all "tilde" processes in this paper are uniform integrable. This implies $e^{-\lambda_2 t}\tilde{N}_2(t) \xrightarrow{L^1} V_2$ and $\mathbb{E}[V_2] = \mathbb{E}[I_2^*] = \frac{N u_1 u_2}{\lambda_2^2}$. Finally, to compute the Laplace transform of $V_2$, we plug $f_2(t) = e^{\lambda_2 t}$ into the formula in 6.4.

**Proof of Corollary 6.2.** The convergence directly follows Theorem 6.1 by the fact that for fixed

$$\omega \in \{\omega : \lim_{t\to\infty} e^{-\lambda_2 t}\tilde{N}_2(\omega, t) = V_2(\omega)\},$$

$$\lim_{t\to\infty} \frac{\tilde{N}_2(\omega,t)}{f_2(t)} = \lim_{t\to\infty} e^{-\lambda_2 t} f_2(t) \lim_{t\to\infty} \frac{\tilde{N}_2(\omega,t)}{f_2(t)} = V_2(\omega).$$

Similar to the proof of Corollary 3.2, Lemma A.2 and that fact $f_2 \in F_2$ together imply the square integrability of $\{\tilde{N}_2(t)/f_2(t), t \geq 0\}$. Thus the convergence is in $L^1$.

**Proof of Theorem 6.3.** Consider a system with $N_2^*(t) := f_2(t) V_2$ and let $\tilde{N}_3(t)$ denote the number of type-3 crypts in this system. In the beginning we compute

$$\begin{aligned} \mathbb{E}[I_3^*] &= \int_0^\infty u_3 \mathbb{E}[N_2^*(s)] e^{-\lambda_3 s} \mathrm{d}s \\ &= \int_0^\infty u_3 \frac{u_1 u_2 N}{\lambda_2^2} f_2(s) e^{-\lambda_3 s} \mathrm{d}s \\ &= \frac{N u_1 u_2 u_3}{\lambda_2^2} \int_0^\infty f_2(s) e^{-\lambda_3 s} \mathrm{d}s. \end{aligned}$$

By observing that

$$\frac{f_2(s) e^{-\lambda_3 s}}{e^{-(\lambda_3-\lambda_2)s}} = f_2(s) e^{-\lambda_2 s} \to 1 < \infty,$$

we see that the improper integral $\int_0^\infty f_2(s) e^{-\lambda_3 s} \mathrm{d}s$ converges. It follows that $\mathbb{E}[I_3^*] < \infty$. Then by Lemmas 4.1 and A.2, we see that there exists a random variable $V_3$ such that $e^{-\lambda_3 t}\tilde{N}_3(t) \to V_3$ a.s. and in $L^1$.

Next, we have

$$\forall\, \omega \in \{\omega : \lim_{t\to\infty} e^{-\lambda_3 t}\tilde{N}_3(\omega, t) = V_3(\omega)\},$$

$$\lim_{t\to\infty} \frac{\tilde{N}_3(\omega,t)}{f_3(t)} = \lim_{t\to\infty} e^{-\lambda_3 t} f_3(t) \lim_{t\to\infty} \frac{\tilde{N}_3(\omega,t)}{f_3(t)} = V_3(\omega).$$

Lemma 6.4 approximates the Laplace transform of $V_3$. Here we note that changing the scaling function $f_2$ does not change the limiting random variable $V_3$. However, the approximation in Lemma 6.4 is made using $f_2$. Hence $f_2$ changes the approximation of the Laplace transform of $V_3$.

## Data availability

No data was used for the research described in the article.

## Acknowledgments

The authors would like to thank Rick Durrett, Chay Paterson and Konstantinos Mamis for helpful comments on the manuscript. This work is supported by the National Science Foundation, United States grant DMS-2045166.

## Appendix A. Auxiliary lemmas

**Lemma A.1.** *For $i \leq 3$, $\sup_t \mathbb{E}\left[(e^{-\lambda_i t} N_i(t))^2\right] < \infty$.*

**Proof.** By Lemma 5 in Durrett and Moseley (2010), we know inductively that if $\sup_t \mathbb{E}[e^{-\lambda_i t} N_i(t)]^2 < \infty$ and $\lambda_i < \lambda_{i+1}$ holds, then $\sup_t (e^{-\lambda_{i+1} t} N_{i+1})^2 < \infty$. In this case, since $0 = \lambda_0 = \lambda_1 < \lambda_2 < \lambda_3$, we only need to show that

$$\sup_t \mathbb{E}(N_0(t))^2 < \infty \text{ and, } \sup_t \mathbb{E}(N_1(t))^2 < \infty.$$

Note that by our transition scheme, one can have

$$N_0(t) + N_1(t) \leq N.$$

Therefore

$$\max\left(\sup_t N_0^2(t), \sup_t N_1^2(t)\right) \leq N^2 < \infty. \quad \square$$

**Lemma A.2.** *For $i \in \{2,3\}$, $\sup_t \mathbb{E}\left[(e^{-\lambda_i t}\tilde{N}_i(t))^2\right] < \infty$.*

**Proof.** For $i=2$, recall that $\tilde{N}_2(t)$ is the second type in a two-type branching process where the first type is $N_1^*(t)=u_1Nt$. And $\tilde{N}_2(t)$ is produced at rate $u_2N_1^*(t)$. By manipulating the master equation of this two-type process, we obtain the following differential equation of $\mathbb{E}\left[\tilde{N}_2(t)^2\right]$,

$$\frac{d\mathbb{E}\left[\tilde{N}_2(t)^2\right]}{dt}=2\lambda_2\mathbb{E}\left[\tilde{N}_2(t)^2\right]+(\lambda_2+2u_2u_1Nt)\mathbb{E}[\tilde{N}_2(t)]+u_2u_1Nt$$

subject to $\mathbb{E}\left[\tilde{N}_2(0)^2\right]=0$. The solution is

$$\mathbb{E}\left[\tilde{N}_2(t)^2\right]=e^{2\lambda_2 t}\int_0^t e^{-2\lambda_2 s}\times\left((\lambda_2+2u_2u_1Ns)\mathbb{E}[\tilde{N}_2(s)]+u_2u_1Ns\right)\mathrm{d}s.$$

Note that

$$\mathbb{E}[\tilde{N}_2(t)]=\frac{Nu_1u_2(e^{\lambda_2 t}-\lambda_2 t-1)}{\lambda_2^2}\le\frac{u_2u_1N}{\lambda_2^2}e^{\lambda_2 t}.$$

Thus,

$$\begin{aligned}\mathbb{E}\left[\tilde{N}_2(t)^2\right]&\le e^{2\lambda_2 t}\int_0^t\left(\frac{2u_2^2u_1^2N^2}{\lambda_2^2}se^{-\lambda_2 s}+\frac{u_1u_2N}{\lambda_2}e^{-\lambda_2 s}+u_2u_1Nse^{-2\lambda_2 s}\right)\mathrm{d}s\\&=e^{2\lambda_2 t}\left(\frac{2u_2^2u_1^2N^2}{\lambda_2^2}\frac{1-e^{-\lambda_2 t}(\lambda_2 t+1)}{\lambda_2^2}+\frac{u_1u_2N}{\lambda_2}\frac{1-e^{-\lambda_2 t}}{\lambda_2}+u_2u_1\frac{1-e^{-2\lambda_2 t}(2\lambda_2 t+1)}{4\lambda_2^2}\right).\end{aligned}$$

It follows that

$$\mathbb{E}\left[(e^{-\lambda_2 t}\tilde{N}_2(t))^2\right]\le\frac{2u_2^2u_1^2N^2}{\lambda_2^4}+\frac{u_1u_2N}{\lambda_2^2}+\frac{u_2u_1}{4\lambda_2^2}.$$

Since the right hand side is not time-dependent, we get $\sup_t\mathbb{E}\left[(e^{-\lambda_2 t}\tilde{N}_2(t))^2\right]<\infty$.

For $i=3$, $\tilde{N}_3(t)$ is produced at rate $u_3N_2^*(t)=u_3e^{\lambda_2 t}V_2$. From the master equation we get the following differential equation:

$$\frac{d\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right]}{dt}=2\lambda_3\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right]+(\lambda_3+2u_3e^{\lambda_2 t}V_2)\mathbb{E}[\tilde{N}_3(t)|V_2]+u_3e^{\lambda_2 t}V_2$$

subject to $\mathbb{E}\left[\tilde{N}_3(0)^2|V_2\right]=0$. The solution is

$$\mathbb{E}\left[\tilde{N}_3(t)^2|V_2\right]=e^{2\lambda_3 t}\int_0^t e^{-2\lambda_3 s}\times\left((\lambda_3+2u_3e^{\lambda_2 s}V_2)\mathbb{E}[\tilde{N}_3(s)|V_2]+u_3e^{\lambda_2 s}V_2\right)\mathrm{d}s.$$

Note that

$$\mathbb{E}[\tilde{N}_3(t)|V_2]=\frac{u_3V_2}{\lambda_3-\lambda_2}(e^{\lambda_3 t}-e^{\lambda_2 t})\le\frac{u_3V_2}{\lambda_3-\lambda_2}e^{\lambda_3 t}.$$

Thus, we have

$$\begin{aligned}E\left[\tilde{N}_3(t)^2|V_2\right]&\le e^{2\lambda_3 t}\int_0^t\left(\frac{2u_3^2V_2^2}{\lambda_3-\lambda_2}e^{-(\lambda_3-\lambda_2)s}+\frac{\lambda_3u_3V_2}{\lambda_3-\lambda_2}e^{-\lambda_3 s}+u_3e^{-(2\lambda_3-\lambda_2)s}V_2\right)\mathrm{d}s\\&\le e^{2\lambda_3 t}\left(\frac{2u_3^2V_2^2}{(\lambda_3-\lambda_2)^2}+\frac{\lambda_3u_3V_2}{\lambda_3(\lambda_3-\lambda_2)}+\frac{u_3V_2}{(2\lambda_3-\lambda_2)}\right).\end{aligned}$$

By Lemma 6.4, one can compute the moments of $V_2$ from its Laplace transform. Then we get

$$\mathbb{E}[V_2]=\frac{Nu_1u_2}{\lambda_2^2},\quad\mathbb{E}[V_2^2]=\frac{Nu_1u_2(2Nu_1u_2+\lambda_2^2)}{\lambda_2^4}.$$

Hence, we conclude that $\sup_t\mathbb{E}\left[(e^{-\lambda_3 t}\tilde{N}_3(t))^2\right]<\infty$. □

## Appendix B. Closed form formulas of analytic distributions

In the main text, we have omitted a few cumbersome formulas to increase readability. Here we present their closed form expressions. We begin with the analytic probability distribution functions of $\tau_4$:

$$p_{40}(t)=1-\mathcal{L}_{V_3}\left(u_4\int_0^t f_{30}(s)ds\right)=1-\mathcal{L}_{V_3}\left(\frac{u_4}{\lambda_3}(e^{\lambda_3 t}-1)\right),\tag{B.1}$$

$$p_{41}(t)=1-\mathcal{L}_{V_3}\left(u_4\int_0^t f_{31}(s)ds\right)\tag{B.2}$$

$$=1-\mathcal{L}_{V_3}\left(u_4\left(\frac{1}{\lambda_3}(e^{\lambda_3 t}-1)-\frac{\lambda_3^2}{2}t^2-t-\frac{\lambda_3^2}{\lambda_2^2}(\frac{1}{\lambda_2}(e^{\lambda_2 t}-1)-\frac{\lambda_2}{2}t^2-t)\right)\right)\tag{B.3}$$

where

$$\mathcal{L}_{V_3}(\theta)=\mathcal{L}_{V_2}\left(u_3\int_0^\infty\frac{f_{21}(s)\theta}{\theta+e^{\lambda_3 s}}\mathrm{d}s\right)\tag{B.4}$$

$$=\exp\left(\frac{Nu_1u_2}{\lambda_2^2}\mathrm{PolyLog}\left(2,-u_3\left(\frac{1}{\lambda_3-\lambda_2}\times\theta\,{}_2F_1(1-\frac{\lambda_2}{\lambda_3},1;2-\frac{\lambda_2}{\lambda_3};-\theta)\right.\right.\right.\tag{B.5}$$

$$\left.\left.\left.-\lambda_2\frac{\mathrm{PolyLog}(2,-\theta)}{\lambda_3^2}-\frac{1}{\lambda_3}\log(1+\theta)\right)\right)\right).\tag{B.6}$$

The two results of skipping type-4 are

$$p_{40}^{s3}(t)=1-\mathcal{L}_{V_2}\left(u_4\int_0^t f_{20}(s)(1-p_0^{3\to4}(s,t))ds\right)\tag{B.7}$$

$$=1-\mathcal{L}_{V_2}\left(\frac{u_4}{\lambda_2(u_4-\lambda_3)}\left(-u_4+\lambda_3\,{}_2F_1(1,\frac{\lambda_2}{\lambda_3},\frac{\lambda_2}{\lambda_3}+1,\frac{u_4-\lambda_3}{u_4}e^{-\lambda_2 t})\right.\right.\tag{B.8}$$

$$\left.\left.+e^{\lambda_2 t}(u_4-\lambda_3\,{}_2F_1(1,\frac{\lambda_2}{\lambda_3},\frac{\lambda_2}{\lambda_3}+1,1-\frac{\lambda_3}{u_4}))\right)\right),\tag{B.9}$$

$$p_{41}^{s3}(t)=1-\mathcal{L}_{V_2}\left(u_4\int_0^t f_{21}(s)(1-p_0^{3\to4}(s,t))ds\right)\tag{B.10}$$

$$=1-\mathcal{L}_{V_2}\left(u_4\left(L_{\lambda_2}^{(4)}(t)-\lambda_2L_l^{(4)}(t)-L_c^{(4)}(t)\right)\right),\tag{B.11}$$

where

$$p_0^{3\to4}(s,t)=\frac{1}{1+u_4\frac{\exp(\lambda_3(t-s))-1}{\lambda_3}},$$

$$\mathcal{L}_{V_2}(\theta)=\exp\left(Nu_1u_2\frac{\mathrm{PolyLog}(2,-\theta)}{\lambda_2^2}\right),$$

$$L_{\lambda_2}^{(4)}(t):=\int_0^t e^{\lambda_2 s}(1-p_0^{3\to4}(s,t))ds=\frac{1}{\lambda_2(u_4-\lambda_3)}\times\left(-u_4+\lambda_3\,{}_2F_1(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;\frac{u_4-\lambda_3}{u_4}e^{-\lambda_3 t})+e^{\lambda_2 t}(u_4-\lambda_3\,{}_2F_1(1,\frac{\lambda_2}{\lambda_3};\frac{\lambda_2}{\lambda_3}+1;1-\frac{\lambda_3}{u_4}))\right),$$

$$L_l^{(4)}(t) := \int_0^t s(1-p_0^{3\to 4}(s,t))ds$$
$$= \frac{t^2}{2}$$
$$+ \frac{-\lambda_3 t\log(\frac{\lambda_3}{u_4}) + \mathrm{PolyLog}(2, \frac{-(\lambda_3-u_4)e^{-\lambda_3 t}}{u_4}) - \mathrm{PolyLog}(2, 1-\frac{\lambda_3}{u_4})}{\lambda_3(\lambda_3-u_4)},$$
$$L_c^{(4)}(t) := \int_0^t (1-p_0^{3\to 4}(s,t))ds$$
$$= \frac{\log(u_4(e^{\lambda_3 t}-1)+\lambda_3) - \log(\lambda_3) - u_4 t}{\lambda_3 - u_4}.$$

For waiting time distributions of the first type-5 crypt, our estimations are $p_{51}^{s4}(t)$ and $p_{51}^{s34}(t)$. $p_{51}^{s4}(t)$ can be expressed explicitly

$$p_{51}^{s4}(t) = 1 - \mathcal{L}_{V_3}\left(u_4\int_0^t f_{31}(s)(1-p_0^{4\to 5}(s,t))ds\right) \tag{B.12}$$
$$= 1 - \mathcal{L}_{V_3}\left(u_4\left(L_{\lambda_3}^{(5)}(t) - CL_{\lambda_2}^{(5)}(t)\right.\right.$$
$$\left.\left. -(\lambda_3 - C\lambda_2)L_l^{(5)}(t) + (C-1)L_c^{(5)}(t)\right)\right), \tag{B.13}$$

where

$$L_{\lambda_3}^{(5)}(t) := \int_0^t e^{\lambda_3 s}(1-p_0^{4\to 5}(s,t))ds$$
$$= \frac{u_5}{\lambda_3(\lambda_3-u_5)^2}\left(\lambda_3 - u_5 + e^{\lambda_3 t}\left(-\lambda_3 + u_5\right.\right.$$
$$\left.\left. +\lambda_3\log(\lambda_3 e^{\lambda_3 t}) - \lambda_3\log(\lambda_3 + u_5(e^{\lambda_3 t}-1))\right)\right),$$
$$L_{\lambda_2}^{(5)}(t) := \int_0^t e^{\lambda_2 s}(1-p_0^{4\to 5}(s,t))ds = \frac{1}{\lambda_2(u_5-\lambda_3)}$$
$$\times\left(-u_5 + \lambda_3\ {}_2F_1(1, \frac{\lambda_2}{\lambda_3}; \frac{\lambda_2}{\lambda_3}+1; \frac{u_5-\lambda_3}{u_5}e^{-\lambda_2 t})\right.$$
$$\left. + e^{\lambda_2 t}(u_5 - \lambda_3\ {}_2F_1(1, \frac{\lambda_2}{\lambda_3}; \frac{\lambda_2}{\lambda_3}+1; 1-\frac{\lambda_3}{u_5}))\right),$$

$$L_l^{(5)}(t) := \int_0^t s(1-p_0^{4\to 5}(s,t))ds$$
$$= \frac{t^2}{2}$$
$$+ \frac{-\lambda_3 t\log(\frac{\lambda_3}{u_4}) + \mathrm{PolyLog}(2, \frac{-(\lambda_3-u_4)e^{-\lambda_3 t}}{u_4}) - \mathrm{PolyLog}(2, 1-\frac{\lambda_3}{u_4})}{\lambda_3(\lambda_3-u_4)},$$
$$L_c^{(5)}(t) := \int_0^t (1-p_0^{4\to 5}(s,t))ds$$
$$= \frac{\log(u_4(e^{\lambda_3 t}-1)+\lambda_3) - \log(\lambda_3) - u_4 t}{\lambda_3 - u_4},$$
$$C := \frac{\lambda_3^2}{\lambda_2^2}.$$

Define

$$I(t) = \int_0^t (e^{\lambda_2 s} - \lambda_2 s - 1)\left(1 - \left(1 + \frac{u_4 u_5(\lambda_3 - u_5 + e^{\lambda_3(t-s)}(u_5 - \lambda_3 + \lambda_3\log(\frac{\lambda_3 e^{\lambda_3(t-s)}}{\lambda_3 + u_5(e^{\lambda_3(t-s)}-1)})))}{(\lambda_3-u_5)^2\lambda_3}\right)^{-1}\right)ds.$$

Then we can have

$$p_{51}^{s34}(t) := 1 - \mathcal{L}_{V_2}(u_3 I(t)). \tag{B.14}$$

Unfortunately, we cannot provide an explicit solution to the integral $I(t)$. Nevertheless, we have computed this value numerically.

## Appendix C. Exact solution of a supercritical two-type pure birth model and its consequences

To measure the distance between $N_3$ and its approximations, it is important to understand the two-type system initiated by a single type-2 crypt. In this section, we consider the branching process generated by the following transition scheme

$$\underset{M_2(t)}{\circlearrowleft \text{divide at rate } \lambda_2} \xrightarrow{u_3} \underset{M_3(t)}{\circlearrowleft \text{divide at rate } \lambda_3}$$

subject to the initial condition $(M_2(0), M_3(0)) = (1, 0)$. To formulate the backward Kolmogorov equations, we define the joint probability generating functions as

$$G_1(x,y,t) := \mathbb{E}(x^{M_2(t)}y^{M_3(t)}|(M_2(0), M_3(0)) = (1,0)),$$
$$G_2(x,y,t) := \mathbb{E}(x^{M_2(t)}y^{M_3(t)}|(M_2(0), M_3(0)) = (0,1)).$$

The corresponding system of ordinary differential equations with respect to variable $t$ is

$$\frac{d}{dt}G_1 = \lambda_2 G_1^2 - (\lambda_2 + u_3)G_1 + u_3 G_2, \qquad G_1(x,y,t=0) = x,$$
$$\frac{d}{dt}G_2 = \lambda_3 G_2^2 - \lambda_3 G_2, \qquad G_2(x,y,t=0) = y.$$

We first rescale the time by $\lambda_2$. Let $t(s) = \frac{s}{\lambda_2}$ and $g_i(x,y,s) := G_i(x,y,t(s))$. Then under the new variables, the equations become

$$\frac{d}{ds}g_1 = g_1^2 - (1+\nu)g_1 + \nu g_2, \qquad g_1(x,y,s=0) = x, \tag{C.1}$$
$$\frac{d}{ds}g_2 = \mu_2 g_2^2 - \mu_2 g_2, \qquad g_2(x,y,s=0) = y, \tag{C.2}$$

where $\nu = \frac{u_3}{\lambda_2}$, $\mu_2 = \frac{\lambda_3}{\lambda_2}$. The solution of the second equation is found to be

$$g_2(x,y,s) = \frac{y}{(1-y)e^{\mu_2 s} + y}.$$

The solution to the first equation has a rather complex form. Similar equations have been previously considered (Kessler and Levine, 2013, 2015). The most general model with death rates is solved by Antal and Krapivsky (AK) (Antal and Krapivsky, 2011). We adapt AK's solution and take into account that the death rates are zero in our model. AK's solution to (C.1) reads

$$g_1(x,y,s) = 1 + \frac{u_3}{\lambda_2} + \frac{\lambda_3}{\lambda_2}\Psi(C_1(x, z_0(y)), z_1(y,s)), \tag{C.3}$$

where

$$\Psi(C,z) := \frac{z^c F_3(z) + C(1-c)F_2(z) + CzF_4(z)}{z^{c-1}F_1(z) + CF_2(z)},$$
$$z_1(y,s) := \left[1 - \frac{1}{1-y}\right]e^{-\frac{\lambda_3}{\lambda_2}s},$$
$$C_1(x,z_0) := z_0^{c-1}\frac{(a - b(x-1))F_1(z_0) + z_0 F_3(z_0)}{(b - b(x-1))F_2(z_0) - z_0 F_4(z_0)},$$
$$z_0(y) := 1 - \frac{1}{1-y},$$
$$F_1(z) := {}_2F_1(a,b;c;z), \quad F_2(z) := {}_2F_1(-a,-b;2-c;z),$$
$$F_3(z) := \frac{ab}{c}{}_2F_1(1+a, 1+b; 1+c; z),$$
$$F_4(z) := \frac{ab}{2-c}{}_2F_1(1-a, 1-b; 3-c; z),$$

with constants

$$a = \frac{u_3}{\lambda_3}, b = \frac{\lambda_2}{\lambda_3}, c = 1 + \frac{\lambda_2 + u_3}{\lambda_3}.$$

We are interested in the probability generating function of the second type $G(y,t) := \mathbb{E}(y^{M_3(t)}|M_2(0) = 1, M_3(0) = 0)$ on the original time scale. We find that

$$G(y,t) = \lim_{x\to 1} g(x,y,s(t)) \\ = 1 + \frac{u_3}{\lambda_2} + \frac{\lambda_3}{\lambda_2}\Psi(C(x,z_0(y)), z_1(y,s(t))), \quad s(t) = \lambda_2 t. \tag{C.4}$$

As $x \to 1$, we observe that

$$C_1(x,z_0) \to C(z_0) := z_0^{c-1}\frac{aF_1(z_0)+z_0F_3(z_0)}{bF_2(z_0)-z_0F_4(z_0)}.$$

Next, the time rescaling only affects $z_1(y,s)$. Hence we define $z(y,t) = z_1(y,s(t)) = \left[1-\frac{1}{1-y}\right]e^{-\lambda_3 t}$. Finally $G(y,t)$ can be written as

$$G(y,t) = 1 + \frac{u_3}{\lambda_2} + \frac{\lambda_3}{\lambda_2}\Psi(C(z_0(y)), z(y,t)). \tag{C.5}$$

*C.1. The probability of having no type-3 crypts at fixed time*

Here we investigate the probability of having no type-3 at time $t$, which we denoted as $p_0(t) := \mathbb{P}(M_3(t)=0)$. By the definition of the generating function $p_0(t) = \lim_{y\to 0} G(0,t)$. We observe that as $y \to 0$, $z_0 \to 0$. First, we compute $\lim_{z_0\to 0} C(z_0)$. We recall that the definition of hypergeometric function gives us that

$${}_2F_1(a,b;c;z) = 1 + \frac{ab}{c}z + \frac{a(a+1)b(b+1)}{c(c+1)2!}z^2 + O(z^3), \quad \text{as } z \to 0.$$

Therefore as $z_0 \to 0$

$$\begin{aligned}
F_1(z_0) &= 1 + \frac{ab}{c}z_0 + \frac{a(a+1)b(b+1)}{c(c+1)2!}z_0^2 + O(z_0^3),\\
F_2(z_0) &= 1 + \frac{ab}{2-c}z_0 + \frac{a(1-a)b(1-b)}{(2-c)(3-c)2!}z_0^2 + O(z_0^3),\\
F_3(z_0) &= \frac{ab}{c} + \frac{a(a+1)b(b+1)}{c(c+1)}z_0 \\
&\quad + \frac{a(a+1)(a+2)b(b+1)(b+2)}{c(c+1)(c+2)2!}z_0^2 + O(z_0^3),\\
F_4(z_0) &= \frac{ab}{2-c} + \frac{a(1-a)b(1-b)}{(2-c)(3-c)}z_0 \\
&\quad + \frac{a(1-a)(2-a)b(1-b)(2-b)}{(2-c)(3-c)(4-c)2!}z_0^2 + O(z_0^3),\\
C(z_0) &= z_0^{c-1}\frac{aF_1(z_0)+z_0F_3(z_0)}{bF_2(z_0)-z_0F_4(z_0)} = \frac{a}{b}z_0^{c-1} + O(z_0^c).
\end{aligned}$$

Note that $z(y,t) = z_0(y)e^{-\lambda_3 t}$. As $z_0 \to 0$, we have

$$\Psi(C(z_0), z_0e^{-\lambda_3 t}) = \frac{\frac{a}{b}(1-c)+O(z_0)}{e^{-(c-1)\lambda_3 t} + \frac{a}{b} + O(z_0)} \\ \to -\frac{u_3(\lambda_2+u_3)}{\lambda_3(\lambda_2 e^{-(\lambda_2+u_3)t}+u_3)}.$$

Thus, we conclude that

$$p_0(t) = 1 + \frac{u_3}{\lambda_2} - \frac{u_3(\lambda_2+u_3)}{\lambda_2(\lambda_2 e^{-(\lambda_2+u_3)t}+u_3)} \approx \frac{1}{1+\frac{u_3}{\lambda_2}\exp(\lambda_2 t)}, \tag{C.6}$$

where the approximation is the result of $u_3 \ll \lambda_2$.

*C.2. The Laplace transform of the limiting scaled type-3 population*

We note that there exists a limiting random variable $U$ such that $e^{-\lambda_3 t}M_3(t) \to U$ almost surely (Durrett and Moseley, 2010). To find the Laplace transform of $U$ we consider the Laplace transform of the scaled population, which is

$$\mathcal{L}_3(\theta,t) := \mathbb{E}\left(e^{-\theta e^{-\lambda_3 t}M_3(t)}\right) = G(e^{-\theta e^{-\lambda_3 t}}, t).$$

Let $y(\theta,t) = e^{-\theta e^{-\lambda_3 t}}$. We have that

$$\mathcal{L}_U(\theta) := \mathbb{E}(e^{-\theta U}) = \lim_{t\to\infty} G(y(\theta,t),t).$$

Here we recall that

$$G(y,t) = 1 + \frac{u_3}{\lambda_2} + \frac{\lambda_3}{\lambda_2}\Psi(C(z_0(y)), z(y,t)). \tag{C.7}$$

We adapt AK's solution (see equation (56) in Antal and Krapivsky (2011)) and take into account that there is no death in our model. This results in the following Laplace transform

$$L_U(\theta) = 1 + \frac{u_3}{\lambda_2} + \frac{\lambda_3}{\lambda_2}\Psi(C^*, -\frac{1}{\theta})$$

where

$$C^* = \lim_{t\to\infty} C(z_0(y(\theta,t))) = (-1)^{a+b}\frac{b}{a}\frac{\Gamma(c)}{\Gamma(2-c)}\left[\frac{\Gamma(-b)}{\Gamma(a)}\right]^2.$$

We note that $\Gamma(z)$ represents the Gamma function. Here the expression is undefined for $\theta = 0$ and also involves evaluating complex numbers in the intermediate steps. These two facts make it hard to compute its value precisely especially when doing related numerical integration, which motivates us to do some transformations and use an alternative expression.

Firstly, we employ the Pfaff transformations (DLMF, 2022, (15.8.1)):

$$\begin{aligned}
{}_2F_1(a,b;c;z) &= (1-z)^{-a}{}_2F_1(a,c-b;c;\frac{z}{z-1}),\\
&= (1-z)^{-b}{}_2F_1(c-a,b;c;\frac{z}{z-1}),\\
&= (1-z)^{c-a-b}{}_2F_1(c-a,c-b;c;z).
\end{aligned}$$

The above equations holds if $\arg(1-z) < \pi$ where $\arg(z)$ is the argument of the complex number $z$. We note that here in the limiting process $-1/\theta$ is negative and real so the criterion is satisfied. We transform all the four hypergeometric functions. The strategy here is to choose the transformation such that the last argument in the hypergeometric function has the form of $\frac{z}{z-1}$ and $\mathrm{Re}(c) > \mathrm{Re}(a+b)$. We find that

$$\begin{aligned}
F_1(z) &= (1-z)^{-a}{}_2F_1(a,1+a;c;\frac{z}{z-1}),\\
F_2(z) &= (1-z)^{b}{}_2F_1(-b,1-b;2-c;\frac{z}{z-1}),\\
F_3(z) &= \frac{ab}{c}(1-z)^{-1-a}{}_2F_1(1+a,1+a;1+c;\frac{z}{z-1}),\\
F_4(z) &= \frac{ab}{2-c}(1-z)^{b-1}{}_2F_1(1-b,1-b;3-c;\frac{z}{z-1}).
\end{aligned}$$

It follows that

$$\begin{aligned}
F_1(-\frac{1}{\theta}) &= \left(\frac{1+\theta}{\theta}\right)^{-a}{}_2F_1\left(a,1+a;c;\frac{1}{1+\theta}\right)\\
&\equiv \left(\frac{1+\theta}{\theta}\right)^{-a}H_1(\theta),\\
F_2(-\frac{1}{\theta}) &= \left(\frac{1+\theta}{\theta}\right)^{b}{}_2F_1\left(-b,1-b;2-c;\frac{1}{1+\theta}\right)\\
&\equiv \left(\frac{1+\theta}{\theta}\right)^{b}H_2(\theta),\\
F_3(-\frac{1}{\theta}) &= \frac{ab}{c}\left(\frac{1+\theta}{\theta}\right)^{-1-a}{}_2F_1\left(1+a,1+a;1+c;\frac{1}{1+\theta}\right)\\
&\equiv \left(\frac{1+\theta}{\theta}\right)^{-1-a}H_3(\theta),
\end{aligned}$$

$$F_4(-\frac{1}{\theta}) = \frac{ab}{2-c}\left(\frac{1+\theta}{\theta}\right)^{b-1} {}_2F_1\left(1-b, 1-b; 3-c; \frac{1}{1+\theta}\right)$$
$$\equiv \left(\frac{1+\theta}{\theta}\right)^{b-1} H_4(\theta).$$

By plugging them into the $\Psi(C, z)$ function, we find

$$\Psi^*(\theta) := \Psi(C^*, -\frac{1}{\theta})$$
$$= \frac{(-1)^c H_3(\theta) - C^*(a+b)(1+\theta)^c H_2(\theta) - C^*(1+\theta)^{a+b} H_4(\theta)}{(-1)^{c-1}(1+\theta)H_1(\theta) + C^*(1+\theta)^c H_2(\theta)}.$$

Secondly, we would like to cancel all the complex parts in the numerator and the denominator. Let

$$d = \frac{b}{a}\frac{\Gamma(c)}{\Gamma(2-c)}\left[\frac{\Gamma(-b)}{\Gamma(a)}\right]^2 \in \mathbb{R}.$$

Then $\Psi^*$ can be rewritten as

$$\Psi^*(\theta) = \frac{-H_3(\theta) - d(a+b)(1+\theta)^c H_2(\theta) - d(1+\theta)^{a+b} H_4(\theta)}{(1+\theta)H_1(\theta) + d(1+\theta)^c H_2(\theta)}.$$

In the end, the Laplace transform of the limiting random variable $U$ reads

$$\mathcal{L}_U(\theta) = 1 + \frac{u_3}{\lambda_2} + \frac{\lambda_3}{\lambda_2}\Psi^*(\theta) \tag{C.8}$$

which is real and well defined at $\theta = 0$.

## References


Abate, J., Whitt, W., 2006. A unified framework for numerically inverting Laplace transforms. INFORMS J. Comput. 18 (4), 408–421. http://dx.doi.org/10.1287/ijoc.1050.0137.

Antal, T., Krapivsky, P., 2011. Exact solution of a two-type branching process: models of tumor progression. J. Stat. Mech. Theory Exp. 2011 (08), P08018. http://dx.doi.org/10.1088/1742-5468/2011/08/p08018.

Athreya, K.B., Ney, P.E., 2004. Branching Processes. Dover Publications, New York.

Avanzini, S., Antal, T., 2019. Cancer recurrence times from a branching process model. PLoS Comput. Biol. 15 (11), e1007423. http://dx.doi.org/10.1371/journal.pcbi.1007423.

Baker, A.-M., Gabbutt, C., Williams, M.J., Cereser, B., Jawad, N., et al., 2019. Crypt fusion as a homeostatic mechanism in the human colon. Gut 68 (11), 1986–1993. http://dx.doi.org/10.1136/gutjnl-2018-317540.

Barker, N., Ridgway, R.A., van Es, J.H., et al., 2009. Crypt stem cells as the cells-of-origin of intestinal cancer. Nature 457 (7229), 608–611. http://dx.doi.org/10.1038/nature07602.

Bozic, I., Antal, T., et al., 2010. Accumulation of driver and passenger mutations during tumor progression. Proc. Natl. Acad. Sci. USA 107 (43), 18545–18550. http://dx.doi.org/10.1073/pnas.1010978107.

Bozic, I., Nowak, M.A., 2014. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. Proc. Natl. Acad. Sci. USA 111 (45), 15964–15968. http://dx.doi.org/10.1073/pnas.1412075111.

Bozic, I., Paterson, C., Waclaw, B., 2019. On measuring selection in cancer from subclonal mutation frequencies. PLoS Comput. Biol. 15 (9), 1–15. http://dx.doi.org/10.1371/journal.pcbi.1007368.

Bozic, I., Reiter, J.G., Allen, B., Antal, T., Chatterjee, K., et al., 2013. Evolutionary dynamics of cancer in response to targeted combination therapy. In: Bergstrom, C.T. (Ed.), eLife 2, e00747. http://dx.doi.org/10.7554/eLife.00747.

Campbell, F., Williams, G.T., Appleton, M.A., Dixon, M.F., Harris, M., Williams, E.D., 1996. Post-irradiation somatic mutation and clonal stabilisation time in the human colon.. Gut 39 (4), 569–573. http://dx.doi.org/10.1136/gut.39.4.569.

Cheek, D., Antal, T., 2018. Mutation frequencies in a birth–death branching process. Ann. Appl. Probab. 28 (6), 3922–3947. http://dx.doi.org/10.1214/18-AAP1413.

Danesh, K., Durrett, R., Havrilesky, L.J., Myers, E., 2012. A branching process model of ovarian cancer. J. Theoret. Biol. 314, 10–15. http://dx.doi.org/10.1016/j.jtbi.2012.08.025.

DLMF, 2022. In: Olver, F.W.J., Olde Daalhuis, A.B., Lozier, D.W., Schneider, B.I., Boisvert, R.F., Clark, C.W., Miller, B.R., Saunders, B.V., Cohl, H.S., McClain, M.A. (Eds.), NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.1.8 of 2022-12-15 URL http://dlmf.nist.gov/.

Durrett, R., 2019. Probability: Theory and Examples, fifth ed. Cambridge University Press, Cambridge; New York.

Durrett, R., Foo, J., Leder, K., Mayberry, J., Michor, F., 2010. Evolutionary dynamics of tumor progression with random fitness values. Theor. Popul. Biol. (ISSN: 0040-5809) 78 (1), 54–66. http://dx.doi.org/10.1016/j.tpb.2010.05.001.

Durrett, R., Moseley, S., 2010. Evolution of resistance and progression to disease during clonal expansion of cancer. Theor. Popul. Biol. 77 (1), 42–48. http://dx.doi.org/10.1016/j.tpb.2009.10.008.

Fearon, E.R., 2011. Molecular genetics of colorectal cancer. Annu. Rev. Pathol. Mech. Dis. 6 (1), 479–507. http://dx.doi.org/10.1146/annurev-pathol-011110-130235.

Foo, J., Leder, K., 2013. Dynamics of cancer recurrence. Ann. Appl. Probab. 23 (4), 1437–1468. http://dx.doi.org/10.1214/12-AAP876.

Foo, J., Leder, K., Zhu, J., 2014. Escape times for branching processes with random mutational fitness effects. Stoch. Process. Appl. 124 (11), 3661–3697. http://dx.doi.org/10.1016/j.spa.2014.06.003.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81 (25), 2340–2361. http://dx.doi.org/10.1021/j100540a008.

Jones, S., dong Chen, W., Parmigiani, G., Diehl, F., Beerenwinkel, N., et al., 2008. Comparative lesion sequencing provides insights into tumor evolution. Proc. Natl. Acad. Sci. USA 105 (11), 4283–4288. http://dx.doi.org/10.1073/pnas.0712345105.

Kessler, D.A., Levine, H., 2013. Large population solution of the stochastic luria–delbrück evolution model. Proc. Natl. Acad. Sci. U.S.A. 110 (29), 11682–11687. http://dx.doi.org/10.1073/pnas.1309667110.

Kessler, D.A., Levine, H., 2015. Scaling solution in the large population limit of the general asymmetric stochastic Luria–Delbrück evolution process. J. Stat. Phys. 158 (4), 783–805. http://dx.doi.org/10.1007/s10955-014-1143-3.

Komarova, N., 2006. Stochastic modeling of drug resistance in cancer. J. Theoret. Biol. 239 (3), 351–366. http://dx.doi.org/10.1016/j.jtbi.2005.08.003.

Komarova, N., Wodarz, D., 2005. Drug resistance in cancer: Principles of emergence and prevention. Proc. Natl. Acad. Sci. USA 102 (27), 9714–9719. http://dx.doi.org/10.1073/pnas.0501870102.

Lamlum, H., Papadopoulou, A., Ilyas, M., Rowan, A., Gillet, C., et al., 2000. APC mutations are sufficient for the growth of early colorectal adenomas. Proc. Natl. Acad. Sci. USA 97 (5), 2225–2228. http://dx.doi.org/10.1073/pnas.040564697.

Meza, R., Jeon, J., Moolgavkar, S.H., Luebeck, E.G., 2008. Age-specific incidence of cancer: Phases, transitions, and biological implications. Proc. Natl. Acad. Sci. USA 105 (42), 16284–16289. http://dx.doi.org/10.1073/pnas.0801151105.

Nicholson, M.D., Antal, T., 2019. Competing evolutionary paths in growing populations with applications to multidrug resistance. PLoS Comput. Biol. 15 (4), e1006866. http://dx.doi.org/10.1371/journal.pcbi.1006866.

Nicholson, A.M., Olpe, C., Hoyle, A., Thorsen, A.-S., Rus, T., et al., 2018. Fixation and spread of somatic mutations in adult human colonic epithelium. Cell Stem Cell 22 (6), 909–918.e8. http://dx.doi.org/10.1016/j.stem.2018.04.020.

Olpe, C., Khamis, D., Chukanova, M., Skoufou-Papoutsaki, N., Kemp, R., et al., 2021. A diffusion-like process accommodates new crypts during clonal expansion in human colonic epithelium. Gastroenterology 161 (2), 548–559.e23. http://dx.doi.org/10.1053/j.gastro.2021.04.035.

Paterson, C., Clevers, H., Bozic, I., 2020. Mathematical model of colorectal cancer initiation. Proc. Natl. Acad. Sci. USA 117 (34), 20681–20688. http://dx.doi.org/10.1073/pnas.2003771117.

Reiter, J.G., Bozic, I., Allen, B., Chatterjee, K., Nowak, M.A., 2013. The effect of one additional driver mutation on tumor progression. Evol. Appl. 6 (1), 34–45. http://dx.doi.org/10.1111/eva.12020.

Sherr, C.J., 2004. Principles of tumor suppression. Cell 116 (2), 235–246. http://dx.doi.org/10.1016/S0092-8674(03)01075-4.

Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A., 2021. Cancer statistics, 2021. CA Cancer J. Clin. 71 (1), 7–33. http://dx.doi.org/10.3322/caac.21654.

Snippert, H.J., Schepers, A.G., van Es, J.H., Simons, B.D., Clevers, H., 2014. Biased competition between Lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion. EMBO Rep. 15 (1), 62–69. http://dx.doi.org/10.1002/embr.201337799.

Tomasetti, C., Marchionni, L., Nowak, M.A., Parmigiani, G., Vogelstein, B., 2015. Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc. Natl. Acad. Sci. USA 112 (1), 118–123. http://dx.doi.org/10.1073/pnas.1421839112.

Vermeulen, L., Morrissey, E., van der Heijden, M., Nicholson, A.M., Sottoriva, A., et al., 2013b. Defining stem cell dynamics in models of intestinal tumor initiation. Science 342 (6161), 995–998. http://dx.doi.org/10.1126/science.1243148.

Vermeulen, L., Morrissey, E., van der Heijden, M., Nicholson, A.M., Sottoriva, A., et al., 2013a. Defining stem cell dynamics in models of intestinal tumor initiation. Science 342 (6161), 995–998. http://dx.doi.org/10.1126/science.1243148.

Vogelstein, B., Kinzler, K.W., 2004. Cancer genes and the pathways they control. Nature Med. 10 (8), 789–799. http://dx.doi.org/10.1038/nm1087.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Kinzler, K.W., 2013. Cancer genome landscapes. Science 339 (6127), 1546–1558. http://dx.doi.org/10.1126/science.1235122.