Electronic Journal of Probability

Electron. J. Probab. **22** (2017), no. 38, 1–54.
ISSN: `1083-6489` DOI: `10.1214/17-EJP58`

Rigorous results for a population model with selection II: genealogy of the population*

Jason Schweinsberg†

Abstract

We consider a model of a population of fixed size N undergoing selection. Each individual acquires beneficial mutations at rate μ_N, and each beneficial mutation increases the individual's fitness by s_N. Each individual dies at rate one, and when a death occurs, an individual is chosen with probability proportional to the individual's fitness to give birth. Under certain conditions on the parameters μ_N and s_N, we show that the genealogy of the population can be described by the Bolthausen-Sznitman coalescent. This result confirms predictions of Desai, Walczak, and Fisher (2013), and Neher and Hallatschek (2013).

Keywords: population model; selection; Bolthausen-Sznitman coalescent.
AMS MSC 2010: Primary 60J27, Secondary 60J75; 60J80; 92D15; 92D25.
Submitted to EJP on January 28, 2017, final version accepted on April 18, 2017.
Supersedes `arXiv:1507.00394`.

1 Introduction

In population genetics, one is often interested in understanding the genealogical structure of a population. That is, we take a sample of individuals from a population at some time and trace their ancestral lines backwards in time. As we trace the ancestral lines backwards in time, the lineages will merge until eventually all sampled individuals are traced back to one common ancestor. For many standard population models, including the classical Moran model [20], the genealogy of the population is best described by a process known as Kingman's coalescent, which was introduced in [17]. Kingman's coalescent is the coalescent process in which only two lineages ever merge at one time and each pair of lineages merges at rate one.

For populations undergoing selection, Kingman's coalescent does not always provide an adequate description of the genealogy of the population. If one individual acquires a beneficial mutation which then spreads rapidly to a large fraction of the population, many ancestral lines could merge nearly at once because they all get traced back to

---

*Supported in part by NSF Grant DMS-1206195

†University of California, San Diego, E-mail: `jschwein@math.ucsd.edu`

the individual that acquired the beneficial mutation. As a result, the genealogy of the population is best described by a coalescent process that permits more than two lineages to merge at one time. Such processes, known as coalescents with multiple mergers or Λ-coalescents, were introduced by Pitman [23] and Sagitov [27] and have been studied extensively in the probability literature in recent years. For previous work in which coalescents with multiple mergers were used to describe the genealogy of populations undergoing selection, see [4, 8, 9, 12, 16, 22].

In this paper, we will consider the following population model. The population has fixed size N. Each individual independently acquires mutations at times of a Poisson process with rate μ_N. All mutations are assumed to be beneficial, and the fitness of each individual depends on how many mutations the individual has acquired, relative to the mean of the population. More precisely, let $X_j(t)$ be the number of individuals with j mutations at time t, which we call type j individuals, and let

$$M(t) := \frac{1}{N} \sum_{j=0}^{\infty} j X_j(t)$$

be the average number of mutations carried by the individuals in the population at time t. Then the fitness of an individual with j mutations at time t is defined to be

$$\max \big\{ 0, 1 + s_N(j - M(t)) \big\}.$$

Note that the parameter s_N measures the selective advantage that an individual gets from each mutation. As in the Moran model, each individual independently lives for an exponentially distributed time with mean one. When an individual dies, it gets replaced by a new individual whose parent is chosen at random from the population. The probability that a particular individual is chosen as the parent is proportional to that individual's fitness, and the new individual inherits all of its parent's mutations.

This model was studied in great detail using nonrigorous methods by Desai and Fisher [11], who obtained results concerning the rate of adaptation, meaning the rate at which the mean fitness $M(t)$ grows as a function of time, as well as the distribution of the fitnesses of individuals in the population at a given time. See also [10, 25, 29] for related results, and see [29] for a good summary of the literature on this model and closely related models. The genealogy of the population in this model has been studied only within the past few years. Desai, Walczak, and Fisher [12] argued that the genealogy of the population can be described by a process called the Bolthausen-Sznitman coalescent, which we will define precisely in section 2. Neher and Hallatschek [22] arrived at the same conclusion for a slightly different model.

This model was also studied in detail in [28], which contains rigorous proofs of the results of Desai and Fisher [11] concerning the rate of adaptation and the distribution of fitnesses of individuals in the population. In the present paper, which is a sequel to [28], we build on the techniques developed in [28] to provide a mathematically rigorous description of the genealogy of the population. We confirm nonrigorous predictions of Desai, Walczak, and Fisher [12] and show that the genealogy of the population is given by the Bolthausen-Sznitman coalescent, under suitable conditions on the parameters s_N and μ_N.

The rest of this paper is organized as follows. In section 2, we state precisely our assumptions and the main result of the paper, which is Theorem 2.1 below. In section 3, we give a heuristic argument that explains the ideas behind why Theorem 2.1 is true, and we make some connections with other results in the literature. In section 4, we summarize the results from [28] that will be needed in the present paper. The remaining sections are devoted to proving Theorem 2.1.

2 Assumptions and main result

We first define the following two quantities, which were also used in [28] and which are important for scaling the process correctly:

$$k_N := \frac{\log N}{\log(s_N/\mu_N)}, \qquad a_N := \frac{\log(s_N/\mu_N)}{s_N}. \tag{2.1}$$

As we will see below, k_N is the natural scale for the number of mutations because the difference in the number of mutations carried by the fittest individual in the population and an individual of average fitness is typically within a constant multiple of k_N. Also, we will see that a_N is the natural time scale on which to study the process because the time required to trace two randomly chosen individuals back to a common ancestor is typically within a constant multiple of a_N.

We will need the following assumptions on the parameters s_N and μ_N, which are identical to the three assumptions that appeared in [28]:

A1: We have $\lim_{N\to\infty} \frac{k_N}{\log(1/s_N)} = \infty$.

A2: We have $\lim_{N\to\infty} \frac{k_N \log k_N}{\log(s_N/\mu_N)} = 0$.

A3: We have $\lim_{N\to\infty} s_N k_N = 0$.

Dividing A3 by A1, we get

$$\lim_{N\to\infty} s_N = 0. \tag{2.2}$$

Therefore, assumptions A1 and A2 imply that

$$\lim_{N\to\infty} k_N = \lim_{N\to\infty} a_N = \infty. \tag{2.3}$$

Also, as noted in [28], these assumptions imply that for all $a > 0$, we have

$$\lim_{N\to\infty} \frac{\mu_N}{s_N^a} = \lim_{N\to\infty} \frac{1}{\mu_N N^a} = 0, \tag{2.4}$$

which means the mutation rate μ_N tends to zero faster than any power of s_N but more slowly than any power of $1/N$. The three assumptions will be satisfied, for example, if for all N, we have $\mu_N = e^{-(\log N)^b}$ with $1/2 < b < 1$ and $e^{-(\log N)^a} \leq s_N \leq (\log N)^{-1/2}$ with $0 < a < 1 - b$. This example was previously given in [28].

Assumption A1 implies that the difference between the number of mutations carried by the fittest individual and the number carried by an individual of average fitness tends to infinity as $N \to \infty$. Because each additional mutation adds s_N to the fitness of an individual, assumption A3 implies that the difference in fitness between these two individuals tends to zero as $N \to \infty$. As can be seen from Proposition 3.1 of [28], assumption A2 ensures that mutations do not happen too fast for the analysis in this paper and [28] to be valid. Understanding how the population evolves under faster mutation rates is an important question for future work.

Although the parameters μ_N and s_N depend on N, we will drop the subscripts and write μ and s throughout the rest of the paper to lighten notation.

Before stating the main result, we need to define the Bolthausen-Sznitman coalescent, which was introduced in [6]. The Bolthausen-Sznitman coalescent is a continuous-time Markov chain $(\Pi(t), t \geq 0)$ taking its values in the set of partitions of $\{1, \dots, n\}$. It is defined by the property that $\Pi(0) = \{\{1\}, \dots, \{n\}\}$ is the partition of $1, \dots, n$ into singletons, and then whenever the partition has b blocks, each possible transition that involves merging k of the blocks into one, where $2 \leq k \leq b$, happens at rate

$$\lambda_{b,k} = \int_0^1 y^{k-2}(1-y)^{b-k} \, dy, \tag{2.5}$$

and these are the only possible transitions. A more detailed construction of the Bolthausen-Sznitman coalescent will be given shortly in section 3.1.

Theorem 2.1. *Assume A1-A3 hold. Fix positive real numbers t_0 and T such that $t_0 > 0$ and $T > t_0 + 2$. Fix a positive integer n, and sample n individuals at random from the population at time $a_N T$. For $0 \leq u \leq t_0 + 1$, let $\Pi_N(u)$ be the partition of $\{1, \dots, n\}$ such that i and j are in the same block of the partition if and only if the ith and jth sampled individuals have the same ancestor in the population at time $a_N(T-u)$. Then*

$$\lim_{N \to \infty} P(\Pi_N(1) = \{\{1\}, \dots, \{n\}\}) = 1. \tag{2.6}$$

Also, the finite-dimensional distributions of $(\Pi_N(1+u), 0 \leq u \leq t_0)$ converge as $N \to \infty$ to the finite-dimensional distributions of the Bolthausen-Sznitman coalescent.

Note that Theorem 2.1 stipulates that with probability tending to one as $N \to \infty$, the sampled individuals at time $a_N T$ will all be descended from different ancestors at time $a_N(T-1)$. However, as the ancestral lines are traced back further, the merging of these ancestral lines obeys the law of the Bolthausen-Sznitman coalescent. This result also appears in [12], where it was obtained by nonrigorous methods.

Theorem 2.1 establishes only convergence of the finite-dimensional distributions. We do not expect convergence to hold with respect to the usual Skorohod J_1 topology on path space because for finite N, we have many pairwise mergers of ancestral lines happening at approximately the same time, rather than the multiple mergers that appear in the limit.

Our proof of Theorem 2.1 will also yield Proposition 2.2 below concerning the number of mutations carried by the ancestors of the sampled individuals. Before stating this proposition, we introduce some notation. Let $q : [0, \infty) \to [0, \infty)$ be the unique bounded function such that

$$q(t) = \begin{cases} e^t & \text{if } 0 \leq t < 1 \\ \int_{t-1}^t q(u) \, du & \text{if } t \geq 1. \end{cases} \tag{2.7}$$

The existence and uniqueness of q is part of Theorem 1.1 of [28]. Also define the function $m : [0, \infty) \to [0, \infty)$ by

$$m^*(t) := 1 + \int_0^t q(u) \, du.$$

Theorem 1.2 of [28] states that if

$$M^*(t) := \max\{j : X_j(t) > 0\}$$

denotes the number of mutations carried by the fittest individual in the population at time t and S is a compact subset of $(0, \infty)$, then

$$\sup_{t \in S} \left| \frac{M^*(a_N t)}{k_N} - m^*(t) \right| \to_p 0, \tag{2.8}$$

and if S is a compact subset of $(1, \infty)$, we also have

$$\sup_{t \in S} \left| \frac{M(a_N t)}{k_N} - m^*(t-1) \right| \to_p 0, \tag{2.9}$$

where $\to_p$ denotes convergence in probability as $N \to \infty$.

Proposition 2.2. *Fix t_0, T, and n as in Theorem 2.1, and sample n individuals at random from the population at time $a_N T$. For $i \in \{1, \dots, n\}$ and $t \in [0, a_N T]$, let $U_i(t)$ be the* *number of mutations carried by the individual at time t that is the ancestor of the individual labelled i at time $a_N T$. Then*

$$\max_{1 \leq i \leq n} \sup_{T-(t_0+1) \leq u \leq T} \left| \frac{U_i(a_N u)}{k_N} - m^*((T-1) \wedge u) \right| \to_p 0.$$

Proposition 2.2 implies that with high probability, we have $U_i(a_N T) \approx k_N m^*(T-1)$, which from (2.9) is approximately the average number of mutations of the individuals at time $a_N T$. As we follow the ancestral lines backwards in time, the number of mutations does not change until we get back to approximately time $a_N(T-1)$. Then for $u \leq T-1$, the ancestor of the sampled individual at time $a_N u$ is close to being the fittest individual in the population. Note that this picture was already implicit in the work of Desai, Walczak, and Fisher [12]. The proof of Proposition 2.2 is given at the end of section 6.2.

3 Heuristics and background

3.1 The Bolthausen-Sznitman coalescent

Recall that the Bolthausen-Sznitman coalescent is the coalescent process whose transition rates are given by (2.5). Pitman [23] showed how to construct the Bolthausen-Sznitman coalescent from a Poisson process. We give a variation of this construction here. Consider a Poisson process on $[0,\infty) \times (0,1] \times [0,1]^n$ with intensity

$$dt \times y^{-2}\, dy \times dz_1 \times \cdots \times dz_n.$$

Let $\Pi(0) = \{\{1\}, \dots, \{n\}\}$ be the partition of $1, \dots, n$ into singletons. If $(t, y, z_1, \dots, z_n)$ is a point of the Poisson process, and if the blocks of the partition $\Pi(t-)$, ranked in order by their smallest elements, are $B_1, \dots, B_b$, then $\Pi(t)$ is the partition obtained from $\Pi(t-)$ by merging together all of the blocks B_i for which $z_i \leq y$.

Informally, this means that if (t, y) are the first two coordinates of a point of the Poisson process, then at time t we have a so-called y-merger, in which each block independently participates in the merger with probability y. If $\Pi(t-)$ has b blocks, then for $2 \leq k \leq b$, the probability that a particular set of k blocks merges into one is $y^k(1-y)^{b-k}$, which allows us to recover the formula (2.5) for the transition rates.

To see that the construction above is well-defined, note that a point $(t, y, z_1, \dots, z_n)$ of the Poisson process can only produce a merger at time t if at least two of $z_1, \dots, z_n$ are less than or equal to y. The rate at which such points appear is bounded above by

$$\int_0^1 y^{-2} \cdot \binom{n}{2} y^2 \, dy < \infty.$$

Therefore, only finitely many such points will appear in any bounded time interval, and the construction above can be carried out by considering these points in order by their time coordinate.

We now give a heuristic argument to explain when the Bolthausen-Sznitman coalescent should be expected to describe the genealogy of a population. Note that if a population has size S and then a new large family of size Sx suddenly appears, then the fraction of the population belonging to the large family will be $x/(1+x)$. Consequently, if we are tracing ancestral lines backwards in time, approximately a fraction $x/(1+x)$ of the lineages will coalesce around the time that this family appears. That is, we will have a y-merger with $y = x/(1+x)$. For the Bolthausen-Sznitman coalescent, we can see from the Poisson process construction above that y-mergers with $y \geq x/(1+x)$ occur at rate

$$\int_{x/(1+x)}^1 y^{-2}\, dy = x^{-1}. \tag{3.1}$$

Therefore, the Bolthausen-Sznitman coalescent will describe the genealogy of a population when families of size Sx or larger appear at a rate proportional to x^{-1}.

3.2 A heuristic argument for Theorem 2.1

In this subsection, we give a short approximate calculation to suggest why Theorem 2.1 should be true. For $j \in \mathbb{N}$, let

$$\tau_j := \inf\left\{t : X_{j-1}(t) \geq \frac{s}{\mu}\right\} \tag{3.2}$$

be the first time that there are at least s/μ individuals in the population with $j-1$ mutations. It was shown in [28] that typically no individual acquires a jth mutation until after time τ_j. We write for now $q_j = j - M(\tau_j)$, so that $q_j - 1$ is the difference between $j-1$ and the mean number of mutations carried by the individuals in the population at time τ_j. Typically q_j will be of the same order of magnitude as k_N, which means that when N is large, typically q_j will be large and sq_j will be small.

As argued in [11, 28], shortly after time τ_j, the number of type $j-1$ individuals in the population is growing approximately exponentially at the rate $s(q_j - 1)$, which means that when t is slightly larger than τ_j, we have

$$X_{j-1}(t) \approx \frac{s}{\mu} e^{s(q_j-1)(t-\tau_j)}. \tag{3.3}$$

Because each type $j-1$ individual independently acquires mutations at rate μ, at time u we have type j individuals appearing due to a mutation at rate $\mu X_{j-1}(u)$. If such a mutation happens at time u, then because type j individuals have a selective advantage of approximately sq_j over the rest of the population, the expected number of descendants of this mutation alive at time t is approximately $e^{sq_j(t-u)}$. Therefore, using (3.3),

$$X_j(t) \approx \int_{\tau_j}^{t} \mu \cdot \frac{s}{\mu} e^{s(q_j-1)(u-\tau_j)} \cdot e^{sq_j(t-u)}\, du = s e^{sq_j(t-\tau_j)} \int_{\tau_j}^{t} e^{-s(u-\tau_j)}\, du \approx e^{sq_j(t-\tau_j)}. \tag{3.4}$$

Usually, the type j individuals will belong to many small families. That is, many type $j-1$ individuals will acquire mutations, each of which will become the ancestor of only a small fraction of the type j population. In that case, the approximation in (3.4) will be valid. However, occasionally there can be an unusually early mutation, when a type $j-1$ individual acquires a jth mutation much sooner than expected. When this occurs, the descendants of the new type j individual can eventually constitute a significant fraction of the type j individuals in the population. These unusually large families can lead to multiple mergers of ancestral lines, as many lineages get traced back to the individual that got the early mutation.

To estimate the probability that this happens, we approximate $q_j - 1$ by q_j in (3.3) to see that at time u, mutations from type $j-1$ to type j are occurring at rate approximately $se^{sq_j(u-\tau_j)}$. If such a mutation does occur, then the number of descendants of this mutation behaves like a supercritical branching process with deaths at rate 1 and births at rate $1 + sq_j$. Such a branching process survives with probability approximately sq_j and, conditional on survival, the size of the population after it has evolved for time $t-u$ is approximately

$$\frac{W}{sq_j} e^{sq_j(t-u)},$$

where W has an exponential distribution with mean one. In particular, a successful mutation that occurs at time

$$u = \tau_j + \frac{1}{sq_j} \log\left(\frac{1}{sq_j}\right) + v$$

has approximately

$$We^{-sq_j v}e^{sq_j(t-\tau_j)}$$

descendants in the population at time t. Write $S = e^{sq_j(t-\tau_j)}$, which from (3.4) is approximately the number of type j individuals at time t that do not come from unusually early mutations. By integrating over the possible times when the mutation could occur, we see that the probability that there will be a mutation that is the ancestor of at least Sx type j individuals at time t is approximately

$$\int_{-\infty}^{\infty} se^{sq_j[\log(1/sq_j)/sq_j+v]} \cdot sq_j \cdot P(We^{-sq_j v} > x)\, dv = \int_{-\infty}^{\infty} se^{sq_j v}e^{-xe^{sq_j v}}\, dv = \frac{1}{q_j x}. \quad (3.5)$$

Inside the integral on the left-hand side, the first factor is the total rate of type j mutations, with the number of type $j-1$ individuals being approximated as in (3.3) with q_j in place of $q_j - 1$. The second factor sq_j is the probability that the mutation survives, and the third factor is the probability that there are at least Sx descendants of this mutation at time t. Note that the factor of x^{-1} on the right-hand side of (3.5) matches the right-hand side of (3.1).

Consider now what happens when we sample n individuals from the population at time $a_N T$ and trace their ancestral lines backwards in time. As noted in [28], one type will dominate the population at a typical time, so with high probability, the sampled individuals will all have the same type, which we will call type ℓ. With high probability, the sampled individuals will be descended from distinct type ℓ ancestors at time $\tau_{\ell+1}$. Because we will see that the time between when type ℓ individuals originate and when they become the dominant type in the population is approximately a_N, this means the ancestral lines will most likely not merge when they are traced back from time $a_N T$ to time $a_N(T-1)$, which leads to the result (2.6).

As we trace the lineages further back, with high probability they get traced back to type $\ell - 1$ ancestors at time τ_ℓ, then to type $\ell - 2$ ancestors at time $\tau_{\ell-1}$, and so on. At each stage of this process, there is a small probability that a group of ancestral lines will merge together because they get traced back to an individual that acquired an unusually early mutation. Because of the agreement between (3.1) and (3.5), these mergers follow the same dynamics, in the limit as $N \to \infty$, as the Bolthausen-Sznitman coalescent.

The explanation given here for the appearance of the Bolthausen-Sznitman coalescent is similar to that given by Desai, Walczak, and Fisher [12] and by Neher and Hallatschek [22], though these authors did not work directly from the Poisson process construction of the Bolthausen-Sznitman coalescent.

3.3 Comparison with branching Brownian motion

Theorem 2.1 resembles the main result of [4], which confirmed nonrigorous predictions of Brunet, Derrida, Mueller, and Munier [8, 9] by showing that the Bolthausen-Sznitman coalescent describes the genealogy in a different population model involving selection. In [4], the population was modeled by branching Brownian motion with absorption, in which initially there are particles in $(0, \infty)$, each particle independently moves according to Brownian motion with drift $-\nu_N$, each particle divides into two at rate one, and particles are killed upon reaching the origin. The particles represent individuals, the position of a particle corresponds to the fitness of the individual, branching events represent births, and killing at the origin models the deaths of individuals whose fitness is too low. It was shown in [4] that if the initial configuration of particles and the drift parameter ν_N are chosen so that the number of particles stays comparable to N, then the genealogy of this population is given by the Bolthausen-Sznitman coalescent.

An important difference between the model in [4] and the model studied in the present paper is that for branching Brownian motion with absorption, all individuals have the same birth rate, while individuals with low fitness are killed, which is known as viability selection. In the model considered in the present paper, all individuals have the same death rate, while individuals with higher fitness are more likely to give birth, which is known as fertility selection. In part because of this difference, the two population models behave quite differently in many respects. For example, for branching Brownian motion with absorption, the speed of evolution is measured by the drift ν_N required to maintain a stable population size, which tends to the limiting value $\sqrt{2}$ at the rate of $(\log N)^{-2}$ as $N \to \infty$. This kind of behavior was first observed by Brunet and Derrida [7] and was verified rigorously for other probabilistic models in [3, 19, 21]. However, as shown in [11, 28], the population model studied in the present paper does not have this behavior. Also, for branching Brownian motion with absorption, once the process has evolved for a sufficient time, most particles will be close to the left boundary, as noted in [4, 5]. This is again quite different from the results for the model studied in this paper, where the distribution of fitnesses of individuals in the population has Gaussian-like tails; see, for example, [2, 11, 26, 28, 29]. Finally, for branching Brownian motion with absorption, if two particles are sampled at some time, then the time that one has to go back to find a common ancestor of these two particles is comparable to $(\log N)^3$, as compared with the time scaling by a_N in Theorem 2.1. Yet, in spite of these differences, the Bolthausen-Sznitman coalescent describes the genealogy in both models.

Of course, one could also formulate a discrete population model with viability selection, by adding mutations to the Moran model and making individuals with low fitness more likely to die. One might expect such a population model to behave similarly to the branching Brownian motion model studied in [4, 8, 9]. Likewise, one could formulate a branching Brownian motion model with fertility selection in which particles far from the origin are more likely to branch. Indeed, this is essentially the model studied non-rigorously by Neher and Hallatschek [22], who conclude that the Bolthausen-Sznitman coalescent describes the genealogy of the particles. The model studied by Neher and Hallatschek is thus quite similar to the model studied here, but corresponds to a scenario in which mutations occur more rapidly than what is permitted by assumptions A1-A3. Consequently, we expect that the Bolthausen-Sznitman coalescent may still describe the genealogy of the population even for these faster mutation rates, though much of the heuristic discussion in Section 3.2 would no longer apply.

3.4 Connection with multitype branching processes

Consider a two-type Yule process in which type 1 individuals give birth to type 1 individuals at rate λ and to type 2 individuals at rate μ, and type 2 individuals give birth to type 2 individuals at rate $\lambda + s$. If we say that type 2 individuals belong to the same family when they are descended from the same mutation, then the sizes of type 2 families at some large time t can be approximated by the points of a Poisson process on $(0, \infty)$ with intensity $Cx^{-1-\alpha}$, where C is a constant and $\alpha = \lambda/(\lambda + s)$; see Theorem 3 of [15] and the following corollary. This implies that the total number of type 2 individuals has approximately a stable law of index α. Also, the distribution of the family sizes, normalized to sum to one, is approximately the Poisson-Dirichlet distribution with parameters $(\alpha, 0)$, which was introduced in [24]. Consequently, stable laws appeared in the work of Durrett and Moseley [15] and Durrett, Foo, Leder, Mayberry, and Michor [13], who studied a multitype branching process model for tumor progression, and the Poisson-Dirichlet distribution appeared in the work of Leviyang [18], who studied the coalescence of HIV lineages in a similar model.

If $(\Pi(t), t \geq 0)$ is the Bolthausen-Sznitman coalescent taking its values in the set of partitions of $\{1, \dots, n\}$, then the distribution of the block sizes of $\Pi(t)$, normalized to sum to one, converges as $n \to \infty$ to the Poisson-Dirichlet distribution with parameters

$(e^{-t}, 0)$, as shown in [23]. In particular, the normalized distribution of the type 2 family sizes in the branching process model discussed above is approximately the distribution of $\Pi(\log(1 + s/\lambda))$. This result suggests that in models in which repeated beneficial mutations can occur, but the selective advantage s of a single mutation tends to zero, the Bolthausen-Sznitman coalescent could possibly describe the genealogy of the population, under suitable conditions on the parameters. Theorem 2.1 above is a rigorous formulation of this statement. Indeed, the work [13, 15, 18], which appeared before the work of Desai, Walczak, and Fisher [12] and Neher and Hallatschek [22], served as the original motivation for the present paper.

3.5 Structure of the proof

We summarize here the organization of the proof of Theorem 2.1. In section 4, we review the results from [28] that are needed in the present paper. In section 5, we focus on what happens as the ancestral lines are traced back from time $a_N T$ to time $a_N(T-1)$. First, we show in Lemma 5.1 that with high probability, the individuals sampled from the population at time $a_N T$ will all have one of a small number of types and will have a fitness close to the mean fitness of the population. Then in Lemma 5.2, we show that with high probability, if a sampled individual has ℓ mutations at time $a_N T$, then its ancestor at time $\tau_{\ell+1} \approx a_N(T-1)$ also has ℓ mutations, and its fitness will be close to that of the fittest individual in the population. In Lemma 5.4, we show that it is unlikely that two lineages will coalesce as they are traced back from time $a_N T$ to time $a_N(T-1)$.

In section 6, we trace the lineages further back in time. The main result in this section is Lemma 6.4, in which we show that if $a_N(T-(t_0+1)) \leq \tau_j \leq a_N(T-1)$, then with high probability, all of the ancestors of the sampled individuals at time τ_j have type $j-1$. That is, if the sampled individual has type ℓ, then its ancestor at time τ_ℓ has type $\ell - 1$, its ancestor at time $\tau_{\ell-1}$ has type $\tau_{\ell-2}$, and so on. Because it is also shown in Lemma 6.1 that the random times τ_j can be well approximated by deterministic times, we are able to use Lemma 6.4 to prove Proposition 2 at the end of section 6.2. Then in section 6.3, we show that the only coalescence events that are likely to happen, when lineages are traced back from time τ_j to time τ_{j-1}, are those in which two or more lineages are traced back to an individual that acquired a jth mutation unusually early.

In section 7, we work to understand the coalescence events that occur when an individual acquires a jth mutation unusually early. In section 7.2, we use supercritical branching processes to bound from above and below the number of descendants of the individual that gets the unusually early jth mutation. The probability that some descendants of this individual survive is bounded in Lemma 7.5, and control on the distribution of the number of such surviving individuals is provided by Lemmas 7.6 and 7.7. Note the correspondence between the result of Lemma 7.6 and the right-hand side of (3.5). In Lemma 7.8, we convert these results into a result about the fraction of individuals at time τ_{j+1} that are descended from the individual that received its jth mutation unusually early. Finally, in section 8, we use a point process to couple the genealogy of the sampled individuals with the Bolthausen-Sznitman coalescent, completing the proof of Theorem 2.1.

4 Review of results from [28]

The population model considered in this paper was also studied extensively in [28], and in the present paper, we will make heavy use of some of the results and techniques developed in [28]. In this section, we will state the results from [28] that we will need.

4.1 Evolution of type j individuals

We first present some results summarizing how the type j individuals evolve. Let $\varepsilon > 0$, $\delta > 0$, and $T > 1$. Recall the definition of k_N from (2.1), and let

$$k^* := \left\lceil k_N + \frac{2k_N \log k_N}{\log(s/\mu)} - 1 \right\rceil.$$

Note from assumption A2 that $(2k_N \log k_N)/\log(s/\mu) \to 0$ as $N \to \infty$. As discussed in [28], for $j \leq k^*$, individuals of type j appear in the population very quickly. To understand the evolution of the type j individuals for $j \geq k^* + 1$, define

$$b := \log\left(\frac{24000\,T}{\delta^2 \varepsilon}\right). \tag{4.1}$$

Also, define τ_j as in (3.2), and then set

$$q_j^* := \begin{cases} j - k_N & \text{if } a_N - 2a_N/k_N \leq \tau_j \leq a_N + 2a_N/k_N \\ j - M(\tau_j) & \text{otherwise} \end{cases}$$

and

$$q_j := \max\{1, q_j^*\}. \tag{4.2}$$

Next, let

$$\xi_j := \max\left\{\tau_j,\ \tau_j + \frac{1}{sq_j}\log\left(\frac{1}{sq_j}\right) + \frac{b}{sq_j}\right\}, \tag{4.3}$$

as in [28]. Every type j individual at time t has an ancestor that acquired a jth mutation before time t. If this jth mutation occurred at or before time ξ_j, we call the individual an early type j individual. When an individual gets its jth mutation, we call this a type j mutation, and we call such a mutation an early type j mutation if it occurs at or before time ξ_j. Let $X_{j,1}(t)$ denote the number of early type j individuals at time t, and let $X_{j,2}(t)$ denote the number of other type j individuals at time t, which means

$$X_j(t) = X_{j,1}(t) + X_{j,2}(t).$$

For $t \geq 0$, let

$$G_j(t) := s(j - M(t)) - \mu,$$

which represents the growth rate of the type j individuals in the population at time t. For $j \geq k^* + 1$, let

$$\gamma_j := \tau_j + a_N. \tag{4.4}$$

As noted in Proposition 3.5 of [28], this is approximately the time when type j individuals become the dominant type in the population. Also, let

$$\tilde{\tau}_j := \tau_j + \frac{a_N}{4Tk_N}.$$

Note that this time is called τ_j^* in [28]. Assumptions A1 and A2 imply that $\xi_j \leq \tilde{\tau}_j$ for sufficiently large N. See Figure 1 below.

Proposition 4.1 collects several results related to how the type j individuals evolve. The first four parts of the proposition are identical to Proposition 3.3 of [28], except for the last statement of part 1, which comes instead from Lemma 9.18 of [28]. The first two parts of the proposition describe how the type j individuals emerge before time τ_{j+1}. Part 3 describes the evolution of the type j individuals after time τ_{j+1} but before the type j individuals start to get close to extinction. Part 4 bounds the extinction time for the type j individuals, as well as the size of the type j population as it nears

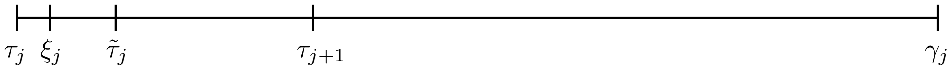

Figure 1: The times τ_j, ξ_j, $\tilde{\tau}_j$, τ_{j+1}, γ_j. Type j mutations occurring before time ξ_j are "early". There are $\lceil s/\mu \rceil$ type j individuals by time τ_{j+1}.

extinction. Part 5 of the proposition, which is Remark 7.8 in [28], demonstrates that nearly all individuals in the population have type j between times γ_j and γ_{j+1}. Finally, part 6, which is a combination of parts 1 and 3 of Proposition 3.6 in [28], bounds the difference between τ_j and τ_{j+1}.

Proposition 4.1. *There exist positive constants C_1 and C_2, depending on δ, ε, and T, such that if N is sufficiently large, then the following statements all hold with probability at least $1-\varepsilon$:*

1. *For all $j \geq k^*+1$ and all $t \in [\tilde{\tau}_j, \tau_{j+1}] \cap [0, a_N T]$, we have*

$$X_{j,1}(t) \leq C_1 \exp\left(\int_{\tau_j}^t G_j(v)\, dv \right). \tag{4.5}$$

Also, $X_{j,1}(t) \leq s/2\mu$ for all $t \leq \tilde{\tau}_j \wedge a_N T$, and no early type j individual acquires a type $j+1$ mutation until after time $\tau_{j+1} \wedge a_N T$. Moreover, no individual that gets a jth mutation at or before time τ_j has a descendant alive in the population at time $\tilde{\tau}_j$.

2. *For all $j \geq k^*+1$ and all $t \in [\tilde{\tau}_j, \tau_{j+1}] \cap [0, a_N T]$, we have*

$$(1-4\delta) \exp\left(\int_{\tau_j}^t G_j(v)\, dv \right) \leq X_{j,2}(t) \leq (1+4\delta) \exp\left(\int_{\tau_j}^t G_j(v)\, dv \right). \tag{4.6}$$

Moreover, the upper bound holds for all $t \in [\xi_j, \tau_{j+1}] \cap [0, a_N T]$.

3. *Let $K = \lfloor k_N/4 \rfloor$. For all $j \geq k^*+1$ and all $t \in [\tau_{j+1}, \gamma_{j+K}] \cap [0, a_N T]$, we have*

$$\frac{(1-\delta)s}{\mu} \exp\left(\int_{\tau_{j+1}}^t G_j(v)\, dv \right) \leq X_j(t) \leq \frac{(1+\delta)s}{\mu} \exp\left(\int_{\tau_{j+1}}^t G_j(v)\, dv \right). \tag{4.7}$$

4. *For all $j \geq k^*+1$, we have*

$$X_j(t) \leq \frac{k_N^2 s}{\mu} \exp\left(\int_{\tau_{j+1}}^t G_j(v)\, dv \right) \tag{4.8}$$

for all $t \in [\gamma_{j+K}, a_N T]$. Also, for all $j \geq k^*+1$ such that $\gamma_{j+\lceil 17k_N \rceil} < a_N T$, we have $X_j(t) = 0$ for all $t \geq \gamma_{j+\lceil 17k_N \rceil}$.

5. *For all $j \geq k^*+1$, we have*

$$\frac{1}{N} \sum_{i=j+1}^{\infty} X_i(t) \leq C_2 e^{-s(\gamma_{j+1}-t)} + \frac{s}{N\mu}$$

for all $t \in [(4/s)\log k_N, \gamma_{j+1}] \cap [0, a_N T]$ and

$$\frac{1}{N} \sum_{i=0}^{j-1} X_i(t) \leq C_2 e^{-s(t-\gamma_j)}$$

for all $t \in [\gamma_j, \gamma_{j+K}] \cap [0, a_N T]$.

6. *We have* $\tau_{k^*+1} \leq 2a_N/k_N$*. Also, for all* $j \geq k^*+1$ *such that either* $\tau_j + 2a_N/k_N \leq a_N T$ *or* $\tau_{j+1} \leq a_N T$*, we have*
$$\frac{a_N}{3k_N} \leq \tau_{j+1} - \tau_j \leq \frac{2a_N}{k_N}. \tag{4.9}$$
More precisely,
$$\int_{\tau_j/a_N}^{\tau_{j+1}/a_N} q(t)\, dt \leq \frac{1+2\delta}{k_N}$$
and
$$\int_{\tau_j/a_N}^{\tau_{j+1}/a_N} \big(q(t) + \mathbb{1}_{\{t \in [1, \gamma_{k^*+1}/a_N)\}}\big)\, dt \geq \frac{1-2\delta}{k_N}.$$

Remark 4.2. *Let*
$$J := 3k_N T + k^* + 1. \tag{4.10}$$
As noted in Remark 3.7 of [28], when (4.9) holds, we have
$$\tau_J > \tau_J - \tau_{k^*+1} \geq \frac{a_N}{3k_N}(J - (k^*+1)) \wedge a_N T = a_N T,$$
and furthermore when the statement of part 1 of Proposition 4.1 also holds, no individual of type $J+1$ *or higher can appear until after time* $a_N T$*.*

Remark 4.3. *For* $t \geq 0$*, recall that* $M^*(t) = \max\{j : X_j(t) > 0\}$ *is the number of mutations carried by the fittest individual at time* t*. As noted in Lemma 4.2 of [28], when* $j \geq k^*+1$ *and the conclusion of Proposition 4.1 holds, we have* $j - 1 \leq M^*(t) \leq j+1$ *for all* $t \in [\tau_j, \tau_{j+1})$*.*

The next proposition contains some bounds related to the quantities $G_j(t)$ and q_j that are important for the analysis that follows. The first three parts of the proposition come from Lemma 9.8 of [28]. Note that the times ρ_j which appear in Lemma 9.8 of [28] are not needed below because (3.31) of [28] implies that $\rho_j > a_N T$ with high probability. The fourth part is part of Lemma 4.5 of [28], and the fifth comes from Lemmas 9.25 and 9.26 in [28].

Proposition 4.4. *There is a positive constant* C_3*, depending on* ε*,* δ*, and* T*, such that if* N *is sufficiently large, then the following statements all hold for all* j *such that* $k^*+1 \leq j \leq J$ *with probability at least* $1-\varepsilon$*:*

1. *If* $\tau_j > a_N + 2a_N/k_N$ *and* $t \in [\tau_j, \tau_{j+1} \wedge a_N T]$*, then* $s(q_j - C_3) \leq G_j(t) \leq s(q_j + C_3)$*.*
2. *If* $t \in [\tau_j, \tau_{j+1} \wedge a_N T]$*, then* $(1-2\delta)sk_N \leq G_j(t) \leq G_j(t) + \mu \leq (e+2\delta)sk_N$*.*
3. *If* $\tau_j \leq a_N T$*, then* $(1-2\delta)k_N \leq q_j \leq (e+2\delta)k_N$*.*
4. *If* $\tau_{j+1} \leq a_N T$*, then* $\exp\big(\int_{\tau_j}^{\tau_{j+1}} G_j(v)\, dv\big) \leq 2s/\mu$*.*
5. *If* $j \geq k^*+1+K$*, then*
$$e^{-\int_{\tau_{j+1}}^{u} G_j(v)\, dv} \leq \begin{cases} e^{-sk_N(u-\tau_{j+1})/5} & \text{if } u \in [\tau_{j+1}, \gamma_{j-K}] \cap [0, a_N T] \\ (s/\mu)^{-k_N/241} & \text{if } u \in [\gamma_{j-K}, \gamma_{j+K}] \cap [0, a_N T]. \end{cases}$$

Let Λ be the event that the six statements in Proposition 4.1 and the five statements of Proposition 4.4 all hold. Note that the event Λ depends ε, δ, T, and N. Then Propositions 4.1 and 4.4 imply that
$$P(\Lambda) > 1 - 2\varepsilon \tag{4.11}$$
if N is sufficiently large. We now define a random time ζ, which we interpret as being the first time that one of the statements of Proposition 4.1 or Proposition 4.4 fails to

hold. Write $\mathbf{X}(t) = (X_0(t), X_1(t), \dots)$, and let $(\mathcal{F}_t, t \geq 0)$ denote the natural filtration of the population process $(\mathbf{X}(t), t \geq 0)$. Then define

$$\zeta := \inf\{t : P(\Lambda | \mathcal{F}_t) = 0\}.$$

Since Propositions 4.1 and 4.4 only describe the behavior of the process up to time $a_N T$, the event Λ is equivalent to the event $\{\zeta > a_N T\}$, which in turn is equivalent to the event $\{\zeta = \infty\}$. Note that the definition given here for ζ is not quite the same as the definition in [28] because in [28] some additional properties were listed that are not relevant for the present work, and some of the properties listed above were derived from others. Nevertheless, the idea is the same in both papers. Namely, if $t < \zeta$, then all of the properties specified in Propositions 4.1 and 4.4 hold through time t.

4.2 Selective advantage of the fittest individuals

The result below, which is Theorem 1.1 of [28], gives an asymptotic result for the difference in fitness between the fittest individual in the population and an individual of average fitness.

Proposition 4.5. *For $t \geq 0$, let*

$$Q(t) := M^*(t) - M(t). \tag{4.12}$$

Assume A1-A3 hold. If S is a compact subset of $(0,1) \cup (1, \infty)$, then

$$\sup_{t \in S} \left| \frac{Q(a_N t)}{k_N} - q(t) \right| \to_p 0, \tag{4.13}$$

where q is the function defined in (2.7) and $\to_p$ denotes convergence in probability as $N \to \infty$.

The next proposition collects some properties of the function q. All of these results are part of Lemma 4.1 of [28] except for (4.15), which follows from (4.14) and the definition of q.

Proposition 4.6. *The function q defined in (2.7) is continuous on $[0,1) \cup (1, \infty)$, and*

$$\lim_{t \to \infty} q(t) = 2.$$

Also,

$$1 \leq q(t) \leq e \qquad \text{for all } t \geq 0 \tag{4.14}$$

and if $t < u$ with $1 \notin (t, u]$, then

$$|q(u) - q(t)| \leq e(u - t). \tag{4.15}$$

4.3 A useful martingale

Here we review the construction of a martingale that was central to the analysis in [28] and will be important again in the present paper. As in [28], let $F_j(t)$ be the fitness of a type j individual at time t, which is $\max\{0, 1 + s(j - M(t))\}$, divided by the sum of the fitnesses of all individuals in the population at time t, which is N if every individual's fitness is strictly positive. Remark 4.2 and assumption A3 imply that if N is sufficiently large, then every individual's fitness is strictly positive at time t for all $t < \zeta$, in which case

$$F_j(t) = \frac{1 + s(j - M(t))}{N}. \tag{4.16}$$

To define birth and death rates, we follow closely the discussion in [28] and observe that there are three ways that the number of type j individuals could change at time t:

1. Each type $j-1$ individual acquires a jth mutation at rate μ. Therefore, at time t, the rate at which a type j individual appears due to a mutation is $\mu X_{j-1}(t-)$, where we adopt the convention that $X_{-1}(t) = 0$ for all $t \geq 0$ so that our formulas are valid when $j = 0$.
2. The number of type j individuals could increase by one at time t due to a birth. This happens if one of the $N - X_j(t-)$ other individuals dies at time t, which happens at rate $N - X_j(t-)$ because each individual dies at rate one, and if the new individual born has type j, which happens with probability $X_j(t-)F_j(t-)$. Therefore, we define the birth rate
$$B_j(t) := (N - X_j(t))F_j(t). \tag{4.17}$$
3. The number of type j individuals could decrease at time t due to a mutation or death. The rate at which one of the type j individuals becomes type $j+1$ due to a mutation is $\mu X_j(t-)$. Death events that reduce the number of type j individuals happen at rate $X_j(t-)(1 - X_j(t-)F_j(t-))$ because there are $X_j(t-)$ type j individuals each dying at rate one, and when a death occurs, the probability that the new individual born does not have type j is $1 - X_j(t-)F_j(t-)$. Therefore, we define the death rate
$$D_j(t) := 1 + \mu - X_j(t)F_j(t). \tag{4.18}$$

For all $t \geq 0$ and $j \in \mathbb{Z}^+$, let
$$G_j^*(t) := B_j(t) - D_j(t).$$

One can easily check that whenever (4.16) holds, we have $G_j^*(t) = G_j(t)$. Also, as shown in equation (6.5) of [28], whenever (4.16) holds and $j \leq J$, we can see, using assumption A3, that for sufficiently large N,
$$B_j(t) + D_j(t) = \frac{(N - 2X_j(t))(1 + s(j - M(t)))}{N} + 1 + \mu \leq 2 + sJ + \mu \leq 3. \tag{4.19}$$

The result below is Proposition 5.1 of [28]. The martingale defined in this proposition is similar to the one obtained in section 4 of [14].

Proposition 4.7. *For all $t \geq 0$ and $j \in \mathbb{Z}^+$, let*
$$Z_j(t) := e^{-\int_0^t G_j^*(v)\,dv} X_j(t) - \int_0^t \mu X_{j-1}(u) e^{-\int_0^u G_j^*(v)\,dv}\,du - X_j(0). \tag{4.20}$$

Then $(Z_j(t), t \geq 0)$ is a mean zero martingale with
$$\mathrm{Var}(Z_j(t)) = E\left[\int_0^t e^{-2\int_0^u G_j^*(v)\,dv}(\mu X_{j-1}(u) + B_j(u)X_j(u) + D_j(u)X_j(u))\,du\right].$$

We will sometimes need to apply the result of Proposition 4.7 to only a subset of the type j individuals in the population. If κ and γ are stopping times with respect to $(\mathcal{F}_t, t \geq 0)$ such that $0 \leq \kappa \leq \gamma$, then for $t \geq 0$ and $j \in \mathbb{Z}^+$, let $X_j^{\kappa,\gamma}(t)$ be the number of type j individuals in the population at time t that are descended from individuals that acquired a jth mutation during the time interval $(\kappa, \gamma]$. Let $B_j^{\kappa,\gamma}(t)$ and $D_j^{\kappa,\gamma}(t)$ denote the expressions on the right-hand sides of (4.17) and (4.18) with $X_j^{\kappa,\gamma}(t)$ in place of $X_j(t)$. The result below is Corollary 5.4 of [28].

Corollary 4.8. *Let κ and γ be stopping times with $\kappa \leq \gamma$. For $t \geq \kappa$, let*
$$Z_j^{\kappa,\gamma}(t) := e^{-\int_\kappa^t G_j^*(v)\,dv} X_j^{\kappa,\gamma}(t) - \int_\kappa^{t\wedge\gamma} \mu X_{j-1}(u) e^{-\int_\kappa^u G_j^*(v)\,dv}\,du.$$

Then $(Z_j^{\kappa,\gamma}(\kappa+t), t \geq 0)$ *is a mean zero martingale and*

$$\mathrm{Var}(Z_j^{\kappa,\gamma}(\kappa+t)|\mathcal{F}_\kappa) = E\bigg[\int_\kappa^{\kappa+t} e^{-2\int_\kappa^u G_j^*(v)\,dv}(\mu X_{j-1}(u)\mathbb{1}_{u\in(\kappa,\gamma]} \\ + B_j^{\kappa,\gamma}(u)X_j^{\kappa,\gamma}(u) + D_j^{\kappa,\gamma}(u)X_j^{\kappa,\gamma}(u))\,du\bigg|\mathcal{F}_\kappa\bigg].$$

Furthermore, if τ *is a stopping time with* $\kappa \leq \tau$*, then* $(Z_j^{\kappa,\gamma}((\kappa+t)\wedge\tau), t \geq 0)$ *is a mean zero martingale, and* $\mathrm{Var}(Z_j^{\kappa,\gamma}((\kappa+t)\wedge\tau)|\mathcal{F}_\kappa)$ *is obtained by replacing* $\kappa+t$ *with* $(\kappa+t)\wedge\tau$ *in the integral above.*

Finally, suppose κ is a stopping time with respect to $(\mathcal{F}_t, t \geq 0)$ and S is a set of type j individuals alive at time κ. Then for $t \geq \kappa$, let $X_j^S(t)$ be the number of type j individuals in the population at time t that are descended from one of the individuals in the set S, and let $B_j^S(t)$ and $D_j^S(t)$ the expressions on the right-hand sides of (4.17) and (4.18) with $X_j^S(t)$ in place of $X_j(t)$. Then, the same reasoning used to establish Proposition 4.7 and Corollary 4.8 yields the following corollary.

Corollary 4.9. *Let* κ *be a stopping time, and let* S *be a set of type* j *individuals in the population at time* κ*. For* $t \geq \kappa$*, let*

$$Z_j^S(t) := e^{-\int_\kappa^t G_j^*(v)\,dv}X_j^S(t) - X_j^S(\kappa).$$

Then $(Z_j^S(\kappa+t), t \geq 0)$ *is a mean zero martingale and*

$$\mathrm{Var}(Z_j^S(\kappa+t)|\mathcal{F}_\kappa) = E\bigg[\int_\kappa^{\kappa+t} e^{-2\int_\kappa^u G_j^*(v)\,dv}(B_j^S(u)X_j^S(u) + D_j^S(u)X_j^S(u))\,du\bigg|\mathcal{F}_\kappa\bigg].$$

Furthermore, if τ *is a stopping time with* $\kappa \leq \tau$*, then* $(Z_j^S((\kappa+t)\wedge\tau), t \geq 0)$ *is a mean zero martingale, and* $\mathrm{Var}(Z_j^S((\kappa+t)\wedge\tau)|\mathcal{F}_\kappa)$ *is obtained by replacing* $\kappa+t$ *with* $(\kappa+t)\wedge\tau$ *in the integral above.*

Remark 4.10. *By the Strong Markov Property of the population process* $(\mathbf{X}(t), t \geq 0)$*, the results of Corollaries 4.8 and 4.9 hold even when the type* j *is random, as long as* j *is* $\mathcal{F}_\kappa$*-measurable.*

5 Tracing the ancestral lines back to time $a_N(T-1)$

The rest of the paper is devoted to the proof of Theorem 2.1. Throughout the proof, we will fix $\varepsilon > 0$, $\delta > 0$, $t_0 > 0$, and $T > t_0 + 2$. We will also assume that $\varepsilon < 1$ and

$$\delta < \max\left\{\frac{1}{100}, \frac{T-(t_0+2)}{40T}, \frac{1}{19T}, \varepsilon^3\right\}. \tag{5.1}$$

The event Λ is defined as in section 4 for these choices of ε, δ, and T, and for the constants C_1, C_2, and C_3 from Propositions 4.1 and 4.4.

We sample n individuals at random from the population at time $a_N T$ and randomly label these individuals with the integers $1, \dots, n$. We then trace the ancestral lines of these individuals back to time $a_N(T-(t_0+1))$. Recall that if $0 \leq u \leq t_0 - 1$, then $\Pi_N(u)$ is the partition of $\{1, \dots, n\}$ such that i and j are in the same block of $\Pi_N(u)$ if and only if the individuals in the sample labelled i and j have the same ancestor at time $a_N(T-u)$.

For $1 \leq i \leq n$ and $0 \leq t \leq a_N T$, recall that $U_i(t)$ is the number of mutations carried by the individual at time t that is the ancestor of the individual labelled i at time $a_N T$. For $1 \leq i \leq n$ and $1 \leq j \leq U_i(a_N T)$, let

$$V_{i,j} := \inf\{t : U_i(t) = j\} \tag{5.2}$$

be the time when the jth mutation appears on the ith lineage. For $h, i \in \{1, \dots, n\}$, let

$$T_{h,i} := \sup\{t : \text{ the } h\text{th and } i\text{th sampled individuals have the same ancestor at time } t\} \tag{5.3}$$

denote the coalescence time of h and i.

Throughout the rest of the paper, we use C to denote a positive constant that does not depend on δ, ε, or T but whose value may change from line to line. Recall that the numbered constants C_1, C_2, and C_3 do depend on δ, ε, and T.

5.1 The types of the individuals sampled at time $a_N T$

Part 5 of Proposition 4.1 implies that, between times γ_j and γ_{j+1}, the fraction of individuals in the population having type j is very close to one, except for times very close to the boundary of this interval. Consequently, when we take a sample from the population at time $a_N T$, typically either all individuals will have the same type, or else all individuals will have one of two types. The result below is a weaker form of this statement.

Lemma 5.1. *Let*

$$L := \inf\left\{j : \tau_j \geq a_N(T-1) - \frac{3a_N}{k_N}\right\}. \tag{5.4}$$

Then

$$\lim_{N\to\infty} P\big(\Lambda \cap \big\{U_i(a_N T) \notin \{L, L+1, \dots, L+9\} \text{ for some } i \in \{1, \dots, n\}\big\}\big) = 0.$$

Proof. It follows from equation (4.9) that on the event Λ, we have $\tau_L \leq a_N(T-1) - a_N/k_N$ and $\tau_{L+10} \geq a_N(T-1) + a_N/3k_N$. Therefore, using (4.4), on Λ we have $\gamma_L \leq a_N T - a_N/k_N$ and $\gamma_{L+10} \geq a_N T + a_N/3k_N$. Therefore, by part 5 of Proposition 4.1, on Λ we have

$$\frac{1}{N}\sum_{\ell=L+10}^{\infty} X_\ell(a_N T) \leq C_2 e^{-s(\gamma_{L+10} - a_N T)} + \frac{s}{N\mu} \leq C_2\left(\frac{s}{\mu}\right)^{-1/3k_N} + \frac{s}{N\mu}, \tag{5.5}$$

which tends to zero as $N \to \infty$ because $(1/3k_N)\log(s/\mu) \to \infty$ as $N \to \infty$ by assumption A2, and $s/(N\mu) \to 0$ as $N \to \infty$ by (2.4). Likewise, by part 5 of Proposition 4.1, on Λ we have

$$\frac{1}{N}\sum_{\ell=0}^{L-1} X_\ell(a_N T) \leq C_2 e^{-s(a_N T - \gamma_L)} \leq C_2 e^{-s a_N/k_N} = C_2\left(\frac{s}{\mu}\right)^{-1/k_N}, \tag{5.6}$$

which tends to zero as $N \to \infty$. Because the expressions in (5.5) and (5.6) both tend to zero as $N \to \infty$, we conclude that on Λ, the fraction of individuals in the population at time $a_N T$ having between L and $L+9$ mutations tends to one as $N \to \infty$. Because the n individuals are sampled at random from the population, the result follows. □

5.2 The types of the ancestors at time $a_N(T-1)$

Lemma 5.1 implies that with high probability all individuals sampled at time $a_N T$ will have between L and $L+9$ mutations. The next result, Lemma 5.2 below, shows that for $\ell \in \{L, L+1, \dots, L+9\}$, with high probability the type ℓ individuals in the sample will all be descended from type ℓ individuals at time $\tau_{\ell+1}$.

Lemma 5.2. *We have*

$$\lim_{N\to\infty} P\big(\Lambda \cap \big\{U_i(\tau_{U_i(a_N T)+1}) \neq U_i(a_N T) \text{ for some } i \in \{1, \dots, n\}\big\}\big) = 0.$$

Proof. Choose $\ell \in \{L, L+1, \dots, L+9\}$. Recall from Corollary 4.8 that $X_\ell^{\tau_{\ell+1}, a_N T}(a_N T)$ denotes the number of type ℓ individuals at time $a_N T$ that are descended from an individual that got its ℓth mutation during the time interval $(\tau_{\ell+1}, a_N T]$. Equivalently, this is the number of type ℓ individuals at time $a_N T$ whose ancestor in the population at time $\tau_{\ell+1}$ does not have type ℓ. Because each individual in the population at time $a_N T$ has probability n/N of being in the sample, we therefore have

$$P\big(\Lambda \cap \big\{U_i(\tau_{\ell+1}) \neq U_i(a_N T) = \ell \text{ for some } i \in \{1, \dots, n\}\big\} \big| \mathcal{F}_{a_N T}\big) \leq \frac{n X_\ell^{\tau_{\ell+1}, a_N T}(a_N T) \mathbb{1}_\Lambda}{N}. \tag{5.7}$$

It suffices to show that the expected value of the right-hand side of (5.7) tends to zero as $N \to \infty$.

By Corollary 4.8 and Remark 4.10, on Λ,

$$e^{-\int_{\tau_{\ell+1}}^{a_N T} G_\ell(v)\,dv} X_\ell^{\tau_{\ell+1}, a_N T}(a_N T) = \int_{\tau_{\ell+1}}^{a_N T} \mu X_{\ell-1}(u) e^{-\int_{\tau_{\ell+1}}^{u} G_\ell(v)\,dv}\,du + Z_\ell^{\tau_{\ell+1}, a_N T}(a_N T), \tag{5.8}$$

where $Z_\ell^{\tau_{\ell+1}, a_N T}(\tau_{\ell+1} + t, t \geq 0)$ is a mean zero martingale. Note that (4.9) implies that on Λ, we have $\gamma_{\ell-1+K} > a_N T$ if N is sufficiently large, and therefore from (4.7) and from part 4 of Proposition 4.4, we get for $u \in [\tau_{\ell+1}, a_N T]$,

$$\begin{aligned} \mu X_{\ell-1}(u) e^{-\int_{\tau_{\ell+1}}^{u} G_\ell(v)\,dv} &\leq (1+\delta) s e^{\int_{\tau_\ell}^{u} G_{\ell-1}(v)\,dv} e^{-\int_{\tau_{\ell+1}}^{u} G_\ell(v)\,dv} \\ &= (1+\delta) s e^{\int_{\tau_\ell}^{\tau_{\ell+1}} G_\ell(v)\,dv} e^{-s(u-\tau_\ell)} \\ &\leq \frac{2(1+\delta)s^2}{\mu}\, e^{-s(u-\tau_\ell)}. \end{aligned}$$

It follows that on Λ, if N is sufficiently large,

$$\int_{\tau_{\ell+1}}^{a_N T} \mu X_{\ell-1}(u) e^{-\int_{\tau_{\ell+1}}^{u} G_\ell(v)\,dv}\,du \leq \frac{2(1+\delta)s}{\mu}\, e^{-s(\tau_{\ell+1}-\tau_\ell)}. \tag{5.9}$$

Now on Λ, by (4.9), we have

$$e^{-s(\tau_{\ell+1}-\tau_\ell)} \leq e^{-a_N s/3k_N} = \left(\frac{s}{\mu}\right)^{-1/3k_N}. \tag{5.10}$$

Also, on Λ we have $a_N T \in [\tau_{\ell+1}, \gamma_{\ell+K}]$ if N is sufficiently large and therefore, by (4.7),

$$e^{-\int_{\tau_{\ell+1}}^{a_N T} G_\ell(v)\,dv} X_\ell(a_N T) \geq \frac{(1-\delta)s}{\mu}. \tag{5.11}$$

Combining (5.8), (5.9), (5.10), and (5.11), and using that $X_\ell(a_N T) \leq N$, we get that for sufficiently large N,

$$\begin{aligned} E\left[\frac{X_\ell^{\tau_{\ell+1}, a_N T}(a_N T) \mathbb{1}_\Lambda}{N}\right] &\leq E\left[\frac{e^{-\int_{\tau_{\ell+1}}^{a_N T} G_\ell(v)\,dv} X_\ell^{\tau_{\ell+1}, a_N T}(a_N T) \mathbb{1}_\Lambda}{e^{-\int_{\tau_{\ell+1}}^{a_N T} G_\ell(v)\,dv} X_\ell(a_N T)}\right] \\ &\leq E\left[\frac{2(1+\delta)(s/\mu)^{1-1/3k_N} + Z_\ell^{\tau_{\ell+1}, a_N T}(a_N T)}{(1-\delta)(s/\mu)}\right] \\ &= \frac{2(1+\delta)}{1-\delta} \left(\frac{s}{\mu}\right)^{-1/3k_N}. \end{aligned} \tag{5.12}$$

Because $(1/3k_N)\log(s/\mu) \to \infty$ as $N \to \infty$ by assumption A2, the expression on the right-hand side of (5.12) tends to zero as $N \to \infty$. The lemma follows by taking expectations of both sides in (5.7). ☐

5.3 Coalescence between times $a_N(T-1)$ and a_NT

Our next goal is to show that for $\ell \in \{L, L+1, \dots, L+9\}$, the type ℓ individuals in the sample at time a_NT all come from distinct ancestors at time $\tau_{\ell+1}$ with high probability. That is, the lineages do not coalesce as they are traced back from time a_NT to time $\tau_{\ell+1}$. The precise statement is given in Lemma 5.4 below. Because $\gamma_{\ell+1} - \tau_{\ell+1} = a_N$, this observation is very close to the statement (2.6) that none of the lineages coalesce when they are traced back a_N time units. We first establish the following preliminary lemma, which is more general than what is needed for the proof of Lemma 5.4 but will also be used later to prove Lemma 6.6. Note that the right-hand side of (5.14) below converges to zero as $N \to \infty$ by (2.3) and (2.4).

Lemma 5.3. *Suppose $k^* + 1 + K \leq j \leq J$. Randomly label the type j individuals at time τ_{j+1} by the integers $1, 2, \dots, \lceil s/\mu \rceil$. For $t \geq \tau_{j+1}$, let $X_j^i(t)$ denote the number of type j individuals at time t that are descended from the individual labelled i at time τ_{j+1}. Let $\gamma := \gamma_{j+K} \wedge \zeta \wedge a_NT$, and let*

$$R_{i,j} := \sup_{t \in [\tau_{j+1}, \gamma)} \frac{X_j^i(t)}{X_j(t)}. \tag{5.13}$$

Then

$$E\bigg[\sum_{i=1}^{\lceil s/\mu \rceil} R_{i,j}^2\bigg] \leq \frac{C\mu}{s^2 k_N}. \tag{5.14}$$

Proof. By Corollary 4.9 applied when S consists only of the individual labelled i at time τ_{j+1}, for $i = 1, 2, \dots, \lceil s/\mu \rceil$ and $t \geq \tau_{j+1}$, we have

$$X_j^i(t \wedge \gamma) = e^{\int_{\tau_{j+1}}^{t \wedge \gamma} G_j(v)\, dv}(1 + Z_j^i(t)), \tag{5.15}$$

where $(Z_j^i(\tau_{j+1} + t), t \geq 0)$ is a mean zero martingale. Now suppose $t \in [\tau_{j+1}, \gamma)$. Using (5.15) and (4.7),

$$\left(\frac{X_j^i(t)}{X_j(t)}\right)^2 \leq \frac{\mu^2}{(1-\delta)^2 s^2}(1 + Z_j^i(t))^2.$$

Taking the supremum of both sides over $t \in [\tau_{j+1}, \gamma)$, then taking expectations and using that $(a+b)^2 \leq 2a^2 + 2b^2$, we get

$$E[R_{i,j}^2] \leq \frac{2\mu^2}{(1-\delta)^2 s^2}\bigg(1 + E\bigg[\sup_{t \in [\tau_{j+1}, \gamma)} (Z_j^i(t))^2\bigg]\bigg). \tag{5.16}$$

By the L^2 Maximum Inequality for martingales, Corollary 4.9, and the reasoning used to derive (4.19),

$$E\bigg[\sup_{t \in [\tau_{j+1}, \gamma)} (Z_j^i(t))^2 \bigg| \mathcal{F}_{\tau_{j+1}}\bigg] \leq 4E\bigg[\int_{\tau_{j+1}}^{\gamma} e^{-2\int_{\tau_{j+1}}^{u} G_j(v)\, dv} \cdot 3X_j^i(u)\, du \bigg| \mathcal{F}_{\tau_{j+1}}\bigg].$$

Combining this result with (5.15) gives

$$E\bigg[\sup_{t \in [\tau_{j+1}, \gamma)} (Z_j^i(t))^2 \bigg| \mathcal{F}_{\tau_{j+1}}\bigg] \leq 12E\bigg[\int_{\tau_{j+1}}^{\gamma} e^{-\int_{\tau_{j+1}}^{u} G_j(v)\, dv}(1 + Z_j^i(u))\, du \bigg| \mathcal{F}_{\tau_{j+1}}\bigg]. \tag{5.17}$$

Note that $1 + Z_j^i(u) \geq 0$ for all $u \in [\tau_{j+1}, \gamma)$ by (5.15). Therefore, by part 5 of Proposition

4.4 and the fact that $(Z_j^i(\tau_{j+1}+t), t \geq 0)$ is a mean zero martingale,

$$\begin{aligned}E\bigg[\int_{\tau_{j+1}}^{\gamma_{j-K}\wedge\gamma} e^{-\int_{\tau_{j+1}}^u G_j(v)\,dv}(1+Z_j^i(u))\,du\bigg|\mathcal{F}_{\tau_{j+1}}\bigg]&\\ &\leq E\bigg[\int_{\tau_{j+1}}^{\infty} e^{-sk_N(u-\tau_{j+1})/5}(1+Z_j^i(u))\,du\bigg|\mathcal{F}_{\tau_{j+1}}\bigg]\\ &= \int_{\tau_{j+1}}^{\infty} e^{-sk_N(u-\tau_{j+1})/5}\,du\\ &= \frac{5}{sk_N}. \qquad (5.18)\end{aligned}$$

Also, using part 5 of Proposition 4.4 again and that $\gamma - \gamma_{j-K}\wedge\gamma \leq (2a_N/k_N)(2K) \leq a_N$ for sufficiently large N by (4.9),

$$\begin{aligned}E\bigg[\int_{\gamma_{j-K}\wedge\gamma}^{\gamma} e^{-\int_{\tau_{j+1}}^u G_j(v)\,dv}(1+Z_j^i(u))\,du\bigg|\mathcal{F}_{\tau_{j+1}}\bigg]&\\ &\leq E\bigg[\int_{\gamma_{j-K}\wedge\gamma}^{\gamma} \left(\frac{s}{\mu}\right)^{-k_N/241}(1+Z_j^{(i)}(u))\,du\bigg|\mathcal{F}_{\tau_{j+1}}\bigg]\\ &\leq a_N\left(\frac{s}{\mu}\right)^{-k_N/241}. \qquad (5.19)\end{aligned}$$

Because $sk_N \cdot a_N(s/\mu)^{-k_N/241} \to 0$ as $N \to \infty$, as can easily be seen by taking logarithms, equations (5.17), (5.18), and (5.19) imply that

$$E\bigg[\sup_{t\in[\tau_{j+1},\gamma)}(Z_j^i(t))^2\bigg|\mathcal{F}_{\tau_{j+1}}\bigg] \leq 12\bigg(\frac{5}{sk_N}+a_N\left(\frac{s}{\mu}\right)^{-k_N/241}\bigg) \leq \frac{C}{sk_N}$$

for sufficiently large N. Therefore, using (5.16), we get for sufficiently large N,

$$E\bigg[\sum_{i=1}^{\lceil s/\mu\rceil} R_{i,j}^2\bigg] \leq \left\lceil\frac{s}{\mu}\right\rceil \frac{2\mu^2}{(1-\delta)^2s^2}\cdot\left(1+\frac{C}{sk_N}\right).$$

The result follows because $sk_N \to 0$ as $N \to \infty$ by assumption A3. □

Note that in the statement of Lemma 5.4 below, we consider only the lineages labelled 1 and 2 to simplify notation. This is sufficient because individuals are sampled uniformly at random. To bound the probability that the event in question occurs for some pair of lineages, we may simply multiply the probability that the event occurs for the lineages 1 and 2 by $\binom{n}{2}$.

Lemma 5.4. *We have*

$$\lim_{N\to\infty} P\big(\Lambda \cap \{U_1(a_NT) = U_2(a_NT) = \ell \text{ and } T_{1,2} \geq \tau_{\ell+1} \text{ for some } \ell\}\big) = 0.$$

Proof. We know from Lemma 5.2 that with probability tending to one as $N \to \infty$, on Λ all type ℓ individuals sampled at time a_NT have type ℓ ancestors at time $\tau_{\ell+1}$. Therefore, it suffices to show that

$$\begin{aligned}\lim_{N\to\infty} P\big(\Lambda \cap \{U_1(a_NT) = U_2(a_NT) = U_1(\tau_{\ell+1}) = U_2(\tau_{\ell+1}) = \ell&\\ &\text{and } T_{1,2} \geq \tau_{\ell+1} \text{ for some } \ell\}\big) = 0. \qquad (5.20)\end{aligned}$$

That is, we need to show it is unlikely that the first two individuals in the sample are both type ℓ individuals that are descended from the same type ℓ individual at time $\tau_{\ell+1}$.

Randomly label the type ℓ individuals at time $\tau_{\ell+1}$ by the integers $1, 2, \dots, \lceil s/\mu \rceil$. Let $X_\ell^i(t)$ denote the number of type ℓ individuals at time t descended from the ith type ℓ individual in the population at time $\tau_{\ell+1}$. Since each individual at time $a_N T$ is equally likely to be sampled,

$$\begin{aligned} P\big(\Lambda \cap \{U_1(a_N T) = U_2(a_N T) = U_1(\tau_{j+1}) = U_2(\tau_{j+1}) = \ell\} \cap \{T_{1,2} \geq \tau_{\ell+1}\} \big| \mathcal{F}_{a_N T}\big) \\ = \sum_{i=1}^{\lceil s/\mu \rceil} \frac{X_\ell^i(a_N T)(X_\ell^i(a_N T) - 1)\mathbb{1}_\Lambda}{N(N-1)}. \quad (5.21) \end{aligned}$$

By Lemma 5.1, it suffices to consider $\ell \in \{L, L+1, \dots, L+9\}$. Part 6 of Proposition 4.1 implies that on Λ, we have $\tau_{k^*+1+K} \leq 2a_N(K+1)/k_N$, and therefore $L \geq k^* + 1 + K$ for sufficiently large N. Also, in view of (4.9), on Λ we have $\gamma_{L+9+K} \geq a_N T$. Therefore, applying Lemma 5.3, and noting that the probability of a change in the population at exactly time $a_N T$ is zero, for each fixed positive integer ℓ we have

$$E\left[\sum_{i=1}^{\lceil s/\mu \rceil} \frac{X_\ell^i(a_N T)(X_\ell^i(a_N T) - 1)\mathbb{1}_{\{L \leq \ell \leq L+9\} \cap \Lambda}}{N(N-1)}\right] \leq E\left[\sum_{i=1}^{\lceil s/\mu \rceil} R_{i,\ell}^2\right] \leq \frac{C\mu}{s^2 k_N}. \quad (5.22)$$

Taking expectations of both sides of (5.21) and then using (5.22) and the fact that $L + 9 \leq J$ on Λ by Remark 4.2, we get that the probability in (5.20) is bounded above by $CJ\mu/(s^2 k_N)$, which tends to zero as $N \to \infty$ by (2.4). Thus, (5.20) holds, which implies the result of the lemma. ☐

Remark 5.5. *It follows from Lemmas 5.1 and 5.2 that with probability tending to one as $N \to \infty$, we have $U_i(\tau_{L+10}) = U_i(a_N T)$ for all $i \in \{1, \dots, n\}$. Note that $\tau_{L+10} < a_N T$ on Λ for sufficiently large N by (4.9) and (5.4). Because individuals in the population model inherit all of their parents mutations, two lineages can only coalesce if they have the same type. That is, we must have $U_h(T_{h,i}) = U_j(T_{h,i})$ for $h, i \in \{1, \dots, n\}$. It therefore follows from Lemma 5.4 that with probability tending to one as $N \to \infty$, no lineages coalesce as they are traced back from time $a_N T$ to time τ_{L+10}. The fact that the probability of coalescence between times τ_L and τ_{L+10} tends to zero as $N \to \infty$, which would imply (2.6), will be established in Lemma 8.1 below.*

6 Tracing the ancestral lines between times τ_j and τ_{j+1}

Lemmas 5.2 and 5.4 show that the type ℓ individuals in the sample at time $a_N T$ are typically descended from distinct type ℓ ancestors at time $\tau_{\ell+1}$. In this subsection, we consider tracing these ancestral lines back further in time. In particular, we focus on what happens when lineages are traced back from time τ_{j+1} to τ_j. We establish that with high probability, type j individuals at time τ_{j+1} are descended from type $j-1$ individuals at time τ_j, and lineages will only coalesce when many type j lineages in the population are traced back to an individual that acquired its jth mutation before the time ξ_j defined in (4.3).

6.1 Approximating τ_j by the fixed time τ_j^*

We define here some fixed times τ_j^* that approximate the random times τ_j. Let $\tau_{k^*+1}^* = 0$. For integers $j \geq k^* + 1$, let

$$\tau_{j+1}^* := \tau_j^* + \frac{a_N}{k_N q(\tau_j^*/a_N)}, \quad (6.1)$$

where q is the function defined in (2.7). Because $1 \leq q(u) \leq e$ for all $u \geq 0$ by Proposition 4.6, we have

$$\frac{a_N}{e k_N} \leq \tau_{j+1}^* - \tau_j^* \leq \frac{a_N}{k_N}. \quad (6.2)$$

For $u \in (0,T]$, let $j^*(u) := \max\{j : \tau_j^* \leq a_N u\}$ and $j'(u) := \max\{j : \tau_j \leq a_N u\}$. The lemma below shows that τ_j^* is a good approximation to τ_j.

Lemma 6.1. *Fix* $u \in (0,T]$. *On the event* $\{\zeta > a_N u\}$, *we have*

$$|j^*(u) - j'(u)| \leq 9\delta T k_N. \tag{6.3}$$

Likewise, let $j \in \{k^*+1, \dots, J\}$. *On the event* $\{\tau_j < \zeta \wedge a_N T\}$, *we have*

$$|\tau_j^* - \tau_j| \leq 10\delta a_N T. \tag{6.4}$$

Proof. Suppose $j \in \{k^*+1, \dots, J\}$ and $\tau_{j+1} < \zeta \wedge a_N T$. By part 6 of Proposition 4.1,

$$\frac{1-2\delta}{k_N} - \int_{\tau_j/a_N}^{\tau_{j+1}/a_N} \mathbb{1}_{\{u \in [1, \gamma_{k^*+1}/a_N)\}}\, du \leq \int_{\tau_j/a_N}^{\tau_{j+1}/a_N} q(u)\, du \leq \frac{1+2\delta}{k_N}. \tag{6.5}$$

Therefore, if $u \in (0,T]$ and $\zeta > a_N u$, then, using that $\tau_{k^*+1}/a_N \leq 2/k_N$ and $u - \tau_{j'(u)}/a_N \leq 2/k_N$ by part 6 of Proposition 4.1 and that $q(v) \leq e$ for all $v \in [0,u]$ by Proposition 4.6, we have

$$\begin{aligned}\int_0^u q(v)\, dv &\leq \int_0^{\tau_{k^*+1}/a_N} q(u)\, du + \frac{(1+2\delta)(j'(u) - (k^*+1))}{k_N} + \int_{\tau_{j'(u)}/a_N}^u q(v)\, dv \\ &\leq \frac{(1+2\delta)(j'(u) - (k^*+1))}{k_N} + \frac{4e}{k_N}.\end{aligned} \tag{6.6}$$

Likewise, using that $\gamma_{k^*+1}/a_N - 1 \leq 2/k_N$ by part 6 of Proposition 4.1, the lower bound in (6.5) implies that

$$\int_0^u q(v)\, dv \geq \frac{(1-2\delta)(j'(u) - (k^*+1))}{k_N} - \frac{2}{k_N}. \tag{6.7}$$

By definition,

$$\int_{\tau_j^*/a_N}^{\tau_{j+1}^*/a_N} q\left(\frac{\tau_j^*}{a_N}\right) du = \frac{1}{k_N}. \tag{6.8}$$

By (4.15) and (6.2), if $u \in [\tau_j^*/a_N, \tau_{j+1}^*/a_N)$ and $\tau_{j+1}^*/a_N < \zeta$, then

$$\left| q(u) - q\left(\frac{\tau_j^*}{a_N}\right)\right| \leq \frac{e(\tau_{j+1}^* - \tau_j^*)}{a_N} \leq \frac{e}{k_N}$$

unless $1 \in (\tau_j^*/a_N, u]$. Combining this observation with (6.8) and (6.2), we get

$$\left(\frac{1}{k_N} - \frac{e}{k_N^2}\right)\mathbb{1}_{\{1 \notin (\tau_j^*/a_N, \tau_{j+1}^*/a_N]\}} \leq \int_{\tau_j^*/a_N}^{\tau_{j+1}^*/a_N} q(u)\, du \leq \frac{1}{k_N} + \frac{e}{k_N^2}. \tag{6.9}$$

Now (6.2) and (6.9) imply that

$$\begin{aligned}\int_0^u q(v)\, dv &= \int_{\tau_{k^*+1}/a_N}^{\tau_{j^*(u)}/a_N} q(v)\, dv + \int_{\tau_{j^*(u)}/a_N}^u q(v)\, dv \\ &\leq \frac{(1+e/k_N)(j^*(u) - (k^*+1))}{k_N} + \frac{e}{k_N}\end{aligned} \tag{6.10}$$

and

$$\int_0^u q(v)\, dv \geq \frac{(1-e/k_N)(j^*(u) - (k^*+1) - 1)}{k_N}. \tag{6.11}$$

Combining (6.7) and (6.10) gives, for sufficiently large N,

$$\begin{aligned} j'(u) - (k^* + 1) &\leq \frac{(1 + e/k_N)(j^*(u) - (k^* + 1))}{1 - 2\delta} + \frac{e + 2}{1 - 2\delta} \\ &\leq (1 + 3\delta)(j^*(u) - (k^* + 1)) + 5. \end{aligned}$$

Rearranging this expression, and using that $j^*(u) - (k^* + 1) \leq (a_N u)(e k_N / a_N) \leq e T k_N$ by (6.2), we get for sufficiently large N,

$$j'(u) - j^*(u) \leq 3\delta(j^*(u) - (k^* + 1)) + 5 \leq 9\delta T k_N. \tag{6.12}$$

Likewise, combining (6.6) and (6.11), we get for sufficiently large N,

$$\begin{aligned} j'(u) - (k^* + 1) &\geq \frac{(1 - e/k_N)(j^*(u) - (k^* + 1) - 1)}{1 + 2\delta} - \frac{4e}{1 + 2\delta} \\ &\geq (1 - 3\delta)(j^*(u) - (k^* + 1)) - (4e + 1). \end{aligned}$$

Rearranging, and again using that $j^*(u) - (k^* + 1) \leq e T k_N$, we get

$$j'(u) - j^*(u) \geq -3\delta(j^*(u) - (k^* + 1)) - (4e + 1) \geq -9\delta T k_N. \tag{6.13}$$

The result (6.3) follows from (6.12) and (6.13).

Finally, to prove (6.4), note that on the event $\{\tau_j < \zeta \wedge a_N T\}$, we have $j^*(\tau_j^*/a_N) = j$ and $j'(\tau_j/a_N) = j$. Therefore, using (6.3), we have

$$|j^*(\tau_j/a_N) - j^*(\tau_j^*/a_N)| = |j^*(\tau_j/a_N) - j'(\tau_j/a_N)| \leq 9\delta T k_N.$$

Since $|j^*(\tau_j/a_N) - j^*(\tau_j^*/a_N)|$ is the number of points τ_i^* that land between τ_j and τ_j^*, it now follows from (6.2) that for sufficiently large N,

$$|\tau_j - \tau_j^*| \leq (9\delta T k_N + 1) \cdot \frac{a_N}{k_N} \leq 10\delta a_N T,$$

which matches (6.4). □

Define the fixed positive integers

$$j_1 := j^*(T - (t_0 + 1)) - \lfloor 9\delta T k_N \rfloor, \qquad j_2 := j^*(T - 1 + 19/k_N) + \lfloor 9\delta T k_N \rfloor,$$

and let

$$I := \{j \in \mathbb{N} : j_1 \leq j \leq j_2\}. \tag{6.14}$$

The next result shows that, when tracing ancestral lines back from time $a_N(T - 1)$ to time $a_N(T - (t_0 + 1))$, we only need to consider time intervals $[\tau_j, \tau_{j+1}]$ for $j \in I$.

Lemma 6.2. *On the event Λ, for sufficiently large N, we have*

$$a_N(T - (t_0 + 1)) - 10\delta a_N T \leq \tau_{j_1}^* \leq \tau_{j_2}^* \leq a_N(T - 1) + 10\delta a_N T \tag{6.15}$$

and

$$a_N + \frac{2a_N}{k_N} < \tau_{j_1} < a_N(T - (t_0 + 1)). \tag{6.16}$$

Also, $L + 9 \leq j_2 \leq J$ and $\tau_{j_2+1} < a_N T$. Furthermore, the cardinality of I is at most $3Tk_N$.

Proof. Throughout the proof, we will work on the event Λ. Using (6.2), we get that for sufficiently large N,

$$\tau_{j_1}^* \geq a_N(T - (t_0 + 1)) - (9\delta T k_N + 1) \cdot \frac{a_N}{k_N} \geq a_N(T - (t_0 + 1)) - 10\delta a_N T$$

and

$$\tau^*_{j_2} \leq a_N\left(T - 1 + \frac{19}{k_N}\right) + (9\delta T k_N) \cdot \frac{a_N}{k_N} \leq a_N(T-1) + 10\delta a_N T.$$

We have now proved (6.15).

By (6.3), we have $j_1 \leq j'(T - (t_0 + 1))$, and thus $\tau_{j_1} \leq a_N(T - (t_0 + 1))$, which is the upper bound in (6.16). To get the lower bound, note that (6.4) and (6.2) give

$$\tau_{j_1} \geq \tau^*_{j_1} - 10\delta a_N T \geq \tau^*_{j^*(T-(t_0+1))} - (9\delta T k_N) \cdot \frac{a_N}{k_N} - 10\delta a_N T.$$

Since (6.2) implies $\tau^*_{j^*(u)} \geq a_N u - a_N/k_N$ for $u \in (0, T]$, it follows, using (5.1), that for sufficiently large N,

$$\begin{aligned} \tau_{j_1} &\geq a_N(T - (t_0 + 1)) - \frac{a_N}{k_N} - 19\delta a_N T \\ &> a_N + a_N(T - (t_0 + 2) - 20\delta T) \\ &\geq a_N + a_N\left(\frac{T - (t_0 + 2)}{2}\right). \end{aligned} \tag{6.17}$$

The lower bound in (6.16) follows because $\lim_{N\to\infty} k_N = \infty$.

Next, note that by (4.9) and (5.4), we have $\tau_{L+10} \leq a_N(T-1) + 19a_N/k_N$. By (6.3), we have $j_2 \geq j'(T - 1 + 19/k_N)$, which means $\tau_{j_2+1} > a_N(T - 1 + 19/k_N) \geq \tau_{L+10}$ and thus $j_2 \geq L + 9$. Also, by (6.2), the number of times τ^*_i between $a_N(T - 1 + 19/k_N)$ and $a_N T$ is at least $(k_N/a_N)(a_N(1 - 19/k_N)) - 1 = k_N - 20$. Therefore, using (6.3),

$$j'(T) \geq j^*(T) - 9\delta T k_N \geq j^*(T - 1 + 19/k_N) + k_N - 20 - 9\delta T k_N \geq j_2 + k_N - 20 - 18\delta T k_N,$$

which is greater than $j_2 + 1$ for sufficiently large N because $\delta < 1/19T$ by (5.1). It follows that $\tau_{j_2+1} < a_N T$.

Finally, by Remark 4.2, we have $j_2 + 1 \leq J$. Also, we have and $j_1 \geq k^* + 1$ for sufficiently large N by (6.17), so $j_2 - j_1 + 1 \leq 3Tk_N$, which is equivalent to the last statement of the lemma. ☐

6.2 The types of the ancestors at time τ_j

Lemma 6.4 below establishes that with high probability, the type ℓ individuals in the sample get traced back to type $\ell - 1$ individuals at time τ_ℓ, then to type $\ell - 2$ individuals at time $\tau_{\ell-1}$, and so on until we have traced the lineages back to time $a_N(T - (t_0 + 1))$. We begin with the following preliminary result.

Lemma 6.3. *Let $j \in I$. Let K_j be the number of type j individuals in the population at time τ_{j+1} whose ancestor in the population at time τ_j does not have type $j - 1$. Then*

$$E[K_j \mathbb{1}_{\{\tau_{j+1}<\zeta\}}] \leq 5\left(\frac{s}{\mu}\right)^{1-1/3k_N}.$$

Proof. By parts 1 and 6 of Proposition 4.1, on $\{\tau_{j+1} < \zeta\}$, no individual of type j or higher in the population at time τ_j has a descendant alive in the population at time τ_{j+1}. Therefore, K_j is the number of type j individuals at time τ_{j+1} whose ancestor at time τ_j has type less than $j - 1$. Such an individual must be descended from an individual that gets its $(j-1)$st mutation after time τ_j. We will therefore consider the number of type $j - 1$ individuals at times $t \geq \tau_j$ that are descended from individuals that acquired their $(j-1)$st mutation between times τ_j and τ_{j+1}. Following Corollary 4.8, we denote the number of such individuals by $X_{j-1}^{\tau_j,\tau_{j+1}}(t)$. Then, writing $\zeta_j = \zeta \wedge \tau_{j+1}$,

$$\begin{aligned} &e^{-\int_{\tau_j}^{(\tau_j+u)\wedge\zeta_j} G_{j-1}(v)\,dv} X_{j-1}^{\tau_j,\tau_{j+1}}((\tau_j+u)\wedge\zeta_j) \\ &\qquad = \int_{\tau_j}^{(\tau_j+u)\wedge\zeta_j} \mu X_{j-2}(w) e^{-\int_{\tau_j}^{w} G_{j-1}(v)\,dv}\,dw + Z_{j-1}^{\tau_j,\tau_{j+1}}((\tau_j+u)\wedge\zeta_j), \end{aligned} \tag{6.18}$$

where $(Z_j^{\tau_j,\tau_{j+1}}(\tau_j+u), u \geq 0)$ is a mean zero martingale. By (4.7), on the event $\{\zeta_j > \tau_j\}$, we have for $t \geq 0$,

$$\begin{aligned} &\int_{\tau_j}^{(\tau_j+u)\wedge\zeta_j} \mu X_{j-2}(w) e^{-\int_{\tau_j}^{w} G_{j-1}(v)\,dv}\,du \\ &\qquad\qquad \leq \int_{\tau_j}^{(\tau_j+u)\wedge\zeta_j} (1+\delta) s e^{\int_{\tau_{j-1}}^{w} G_{j-2}(v)\,dv} e^{-\int_{\tau_j}^{w} G_{j-1}(v)\,dv}\,dw \\ &\qquad\qquad = (1+\delta) s e^{\int_{\tau_{j-1}}^{\tau_j} G_{j-2}(v)\,dv} \int_{\tau_j}^{(\tau_j+u)\wedge\zeta_j} e^{-s(w-\tau_j)}\,dw \\ &\qquad\qquad \leq (1+\delta) e^{\int_{\tau_{j-1}}^{\tau_j} G_{j-2}(v)\,dv}. \end{aligned} \tag{6.19}$$

By (4.9) and part 4 of Proposition 4.4, on $\{\zeta_j > \tau_j\}$ we have

$$\begin{aligned} e^{\int_{\tau_{j-1}}^{\tau_j} G_{j-2}(v)\,dv} &= e^{-s(\tau_j-\tau_{j-1})} e^{\int_{\tau_{j-1}}^{\tau_j} G_{j-1}(v)\,dv} \\ &\leq e^{-s(a_N/3k_N)} \left(\frac{2s}{\mu}\right) \\ &\leq 2\left(\frac{s}{\mu}\right)^{1-1/3k_N}. \end{aligned} \tag{6.20}$$

Taking conditional expectations on both sides of (6.18) and then using (6.19) and (6.20) gives

$$E\left[e^{-\int_{\tau_j}^{(\tau_j+u)\wedge\zeta_j} G_{j-1}(v)\,dv} X_{j-1}^{\tau_j,\tau_{j+1}}((\tau_j+u)\wedge\zeta)\middle|\mathcal{F}_{\tau_j}\right] \leq 2(1+\delta)\left(\frac{s}{\mu}\right)^{1-1/3k_N}. \tag{6.21}$$

For $u \geq \tau_j$, let $X_j^*(u)$ denote the the number of type j individuals in the population at time u that got their $(j-1)$st mutation after time τ_j. Note that $K_j = X_j^*(\tau_{j+1})$ on $\{\tau_{j+1} < \zeta\}$. By the reasoning that leads to Corollary 4.8, we get

$$e^{-\int_{\tau_j}^{u\wedge\zeta_j} G_j(v)\,dv} X_j^*(u\wedge\zeta_j) = \int_{\tau_j}^{u\wedge\zeta_j} \mu X_{j-1}^{\tau_j,\tau_{j+1}}(w) e^{-\int_{\tau_j}^{w} G_j(v)\,dv}\,dw + Z_j^*(u\wedge\zeta_j), \tag{6.22}$$

where $(Z_j^*(\tau_{j+1}+u), u \geq 0)$ is a mean zero martingale. By part 4 of Proposition 4.4, on $\{\tau_{j+1} < \zeta\}$, we have

$$e^{\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv} \leq \frac{2s}{\mu}.$$

Therefore, using that the expression in (6.22) is nonnegative,

$$X_j^*(\tau_{j+1})\mathbb{1}_{\{\tau_{j+1}<\zeta\}} \leq \frac{2s}{\mu}\left(\int_{\tau_j}^{\infty} \mu X_{j-1}^{\tau_j,\tau_{j+1}}(w) e^{-\int_{\tau_j}^{w} G_{j-1}(v)\,dv} e^{-s(w-\tau_j)} \mathbb{1}_{\{w\leq\zeta_j\}}\,dw + Z_j^*(\zeta_j)\right).$$

Taking conditional expectations of both sides and using Fubini's Theorem and (6.21),

$$\begin{aligned} E[X_j^*(\tau_{j+1})\mathbb{1}_{\{\tau_{j+1}<\zeta\}}|\mathcal{F}_{\tau_j}] &\leq \frac{2s}{\mu}\left(\int_{\tau_j}^{\infty} 2(1+\delta)\mu\left(\frac{s}{\mu}\right)^{1-1/3k_N} e^{-s(w-\tau_j)}\,dw\right) \\ &= 4(1+\delta)\left(\frac{s}{\mu}\right)^{1-1/3k_N}. \end{aligned}$$

Taking expectations of both sides gives the result of the lemma. □

Lemma 6.4. *Recall the definition of I from (6.14). We have*

$$\lim_{N\to\infty} P\big(\Lambda \cap \big\{U_i(\tau_j) \neq j-1 \text{ for some } i \in \{1,\dots,n\} \text{ and } j \in I \text{ with } j \leq U_i(a_N T)+1\big\}\big) = 0.$$

Proof. Fix $i \in \{1,\dots,n\}$. Suppose Λ occurs and $U_i(\tau_j) \neq j-1$ for some $j \in I$ with $j \leq U_i(a_N T)+1$. Then either $U_i(\tau_{U_i(a_N T)+1}) \neq U_i(a_N T)$, an event whose probability tends to zero as $N \to \infty$ by Lemma 5.2, or else there is an integer $j \in I$ with $j \leq U_i(a_N T)$ such that $U_i(\tau_j) \neq j-1$ and $U_i(\tau_{j+1}) = j$. Therefore, to prove the lemma, it suffices to show that

$$\lim_{N\to\infty} \sum_{j\in I} P\big(\Lambda \cap \{U_i(\tau_j) \neq j-1\} \cap \{U_i(\tau_{j+1}) = j\}\big) = 0. \tag{6.23}$$

Fix $j \in I$. Recall from Lemma 6.3 that K_j is the number of type j individuals in the population at time τ_{j+1} whose ancestor in the population at time τ_j does not have type $j-1$. Note that the probability, conditional on $\mathcal{F}_{\tau_{j+1}}$, that a randomly chosen type j individual at time τ_{j+1} is not descended from a type $j-1$ individual at time τ_j is $K_j/\lceil s/\mu \rceil$. Also, conditional on $\mathcal{F}_{\tau_{j+1}}$, the $\lceil s/\mu \rceil$ type j individuals at time τ_{j+1} are equally likely to be the ancestor of the ith individual in the sample taken at time $a_N T$. Therefore, since K_j is $\mathcal{F}_{\tau_{j+1}}$-measurable, on the event $\tau_{j+1} < \zeta$ we have

$$P\big(\{U_i(\tau_j) \neq j-1\} \cap \{U_i(\tau_{j+1}) = j\} | \mathcal{F}_{\tau_{j+1}}\big) = P(U_i(\tau_{j+1}) = j | \mathcal{F}_{\tau_{j+1}}) \cdot \frac{K_j}{\lceil s/\mu \rceil} \leq \frac{\mu K_j}{s}.$$

Therefore, multiplying both sides by $\mathbb{1}_{\{\tau_{j+1}<\zeta\}}$, taking expectations, and using Lemma 6.3, we get

$$P\big(\{U_i(\tau_j) \neq j-1\} \cap \{U_i(\tau_{j+1}) = j\} \cap \{\tau_{j+1} < \zeta\}\big) \leq \frac{\mu}{s} E[K_j \mathbb{1}_{\{\tau_{j+1}<\zeta\}}] \leq 5\left(\frac{s}{\mu}\right)^{-1/3k_N}.$$

Since the cardinality of I is at most $3Tk_N$ by Lemma 6.2, it follows that the sum of the probabilities on the left-hand side of (6.23) is at most

$$15Tk_N\left(\frac{s}{\mu}\right)^{-1/3k_N}.$$

To check that this expression goes to zero as $N \to \infty$, we consider the logarithm. Note that $\log(k_N(s/\mu)^{-1/3k_N}) = \log k_N - (1/3k_N)\log(s/\mu)$, which tends to $-\infty$ as $N \to \infty$ by assumption A2. In view of the discussion before equation (6.23), the result of the lemma follows. □

We have now developed enough tools to prove Proposition 2.2.

Proof of Proposition 2.2. It suffices to prove the result for $n = 1$. Define L as in (5.4). As noted in the first sentence of the proof of Lemma 5.1, we have $\tau_L \leq a_N(T-1) \leq \tau_{L+10}$ on the event Λ. Therefore, using Remark 4.3, we have $L - 1 \leq M^*(a_N(T-1)) \leq L+11$ on Λ. It then follows from Lemma 5.1 that

$$\lim_{N\to\infty} P\big(\Lambda \cap \{|U_1(a_N T) - M^*(a_N(T-1))| > 11\}\big) = 0.$$

Combining this observation with Lemma 5.2, we get

$$\lim_{N\to\infty} P\big(\Lambda \cap \{|U_1(t) - M^*(a_N(T-1))| > 11 \text{ for some } t \in [\tau_{U_1(a_N T)+1}, a_N T]\}\big) = 0. \tag{6.24}$$

We next apply Lemma 6.4. Note that if $U_1(\tau_j) = j-1$ for all $j \in I$ with $j \leq U_1(a_N T)+1$, then because the function $t \mapsto U_1(t)$ is increasing, we have $U_1(t) \in \{j-1, j\}$ for all

$t \in [\tau_j, \tau_{j+1}]$ and all $j \in I$ with $j \le U_1(a_N T)$. On Λ, in view of Lemma 6.2 and Remark 4.3, this statement implies that $|U_1(t) - M^*(t)| \le 2$ for all $t \in [a_N(T-(t+1)), \tau_{U_1(a_N T)+1}]$. Therefore, Lemma 6.4 yields

$$\lim_{N\to\infty} P\big(\Lambda \cap \{|U_1(t) - M^*(t)| > 2 \text{ for some } t \in [a_N(T-(t_0+1)), \tau_{U_1(a_N T)+1}]\}\big) = 0. \quad (6.25)$$

On Λ, we have $\tau_L \le a_N(T-1) \le \tau_{L+10}$ and therefore $|M^*(t) - M^*(a_N(T-1))| \le 12$ when $\tau_L \le t \le \tau_{L+10}$. Combining this observation with (6.24), (6.25), and Lemma 5.1 yields

$$\lim_{N\to\infty} P\big(\Lambda \cap \{|U_1(t) - M^*(t \wedge a_N(T-1))| > 14 \text{ for some } t \in [a_N(T-(t_0+1)), a_N T]\}\big) = 0.$$

Thus, in view of (4.11) and the fact that $\lim_{N\to\infty} k_N = \infty$, we have

$$\sup_{T-(t+1)\le u\le T} \left| \frac{U_1(a_N u)}{k_N} - \frac{M^*(a_N(u \wedge (T-1)))}{k_N} \right| \to_p 0. \quad (6.26)$$

The result follows from (6.26) and (2.8). ☐

6.3 Coalescence between times τ_j and τ_{j+1}

We next consider the merging of ancestral lines between times τ_j and τ_{j+1}. It will suffice to consider the lineages labelled 1 and 2. In view of Lemma 6.4, we may also assume these lineages have type j at time τ_{j+1} and type $j-1$ at time τ_j, which will occur with high probability. Recall the definitions of $V_{i,j}$ and $T_{h,i}$ from (5.2) and (5.3). Also, let $V_j = \min\{V_{1,j}, V_{2,j}\}$ and $V_j^* = \max\{V_{1,j}, V_{2,j}\}$. Because only lineages of the same type can coalesce, there are only three ways that these lineages could coalesce between times τ_j and τ_{j+1} (see Figure 2):

1. Two lineages at time τ_{j+1} could be traced back to one individual that acquires its jth mutation between times ξ_j and τ_{j+1}. That is, $\xi_j < V_{1,j} = V_{2,j} < T_{1,2} \le \tau_{j+1}$.
2. Two lineages at time τ_{j+1} could be traced back to one individual that acquires its jth mutation before time ξ_j. That is, $\tau_j < V_{1,j} = V_{2,j} < \xi_j$ and $V_{1,j} = V_{2,j} < T_{1,2} \le \tau_{j+1}$.
3. Two lineages at time τ_{j+1} could be descended from different type j mutations between times τ_j and τ_{j+1}, but then the two type $j-1$ lineages could coalesce before time τ_j. That is, $\tau_j \le T_{1,2} < V_j < V_j^* < \tau_{j+1}$.

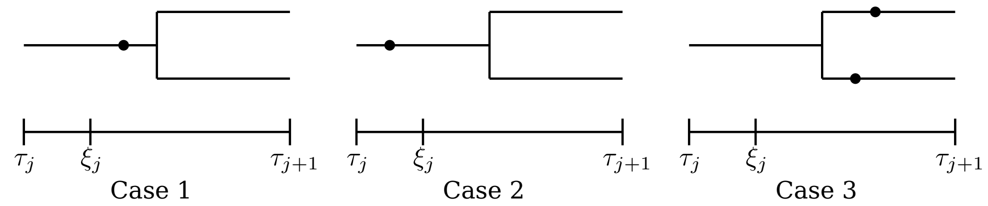

Figure 2: The three ways lineages can coalesce between times τ_j and τ_{j+1}. Dots represent mutations from type $j-1$ to type j.

We will now show that only coalescence events of the second type need to be considered. Lemma 6.5 rules out Case 1 above, and Lemma 6.6 rules out Case 3.

Lemma 6.5. *Define the event*

$$\begin{aligned} A_j' = {} & \{U_1(\tau_{j+1}) = U_2(\tau_{j+1}) = j\} \\ & \cap \{U_1(\tau_j) = U_2(\tau_j) = j-1\} \\ & \cap \{\xi_j < V_{1,j} = V_{2,j} < T_{1,2} < \tau_{j+1}\}. \end{aligned}$$

For sufficiently large N, we have

$$P\biggl(\Lambda \cap \bigcup_{j \in I} A'_j\biggr) \leq CTe^{-b}.$$

Proof. Fix $j \in I$. Let H_j be the number of type j mutations between times ξ_j and τ_{j+1}. Let $0 < \kappa_1 < \kappa_2 < \cdots < \kappa_{H_j}$ denote the times at which these mutations occur. Recall that $X_{2,j}(t)$ is the number of type j individuals at time t that are not early. Let $X_{j,2,i}(u)$ be the number of type j individuals at time u descended from the individual that acquires its jth mutation at time κ_i. This means that

$$X_{j,2}(\tau_{j+1}) = \sum_{i=1}^{H_j} X_{j,2,i}(\tau_{j+1}).$$

Conditional on $X_{j,2,1}(\tau_{j+1}), \dots, X_{j,2,H_j}(\tau_{j+1})$, the probability that two randomly chosen individuals at time τ_{j+1} are descended from the same individual that gets its jth mutation between times ξ_j and τ_{j+1} is

$$\frac{1}{\lceil s/\mu \rceil (\lceil s/\mu \rceil - 1)} \sum_{i=1}^{H_j} X_{j,2,i}(\tau_{j+1})(X_{j,2,i}(\tau_{j+1}) - 1) \leq \frac{\mu^2}{s^2} \sum_{i=1}^{H_j} X_{j,2,i}(\tau_{j+1})^2.$$

Since, conditional on $\mathcal{F}_{\tau_{j+1}}$, each of the $\lceil s/\mu \rceil$ type j individuals at time τ_{j+1} is equally likely to be the ancestor of an individual in our sample at time $a_N T$, it follows that on $\tau_{j+1} < \zeta$,

$$P(A'_j | \mathcal{F}_{\tau_{j+1}}) \leq \frac{\mu^2}{s^2} \sum_{i=1}^{H_j} X_{j,2,i}(\tau_{j+1})^2. \tag{6.27}$$

Therefore, multiplying both sides by $\mathbb{1}_{\{\tau_{j+1} < \zeta\}}$ and taking expectations,

$$P(A'_j \cap \{\tau_{j+1} < \zeta\}) \leq \frac{\mu^2}{s^2} E\biggl[\biggl(\sum_{i=1}^{H_j} X_{j,2,i}(\tau_{j+1})^2\biggr) \mathbb{1}_{\{\tau_{j+1} < \zeta\}}\biggr]. \tag{6.28}$$

We now bound the expectation on the right-hand side of (6.28). Write $\zeta_j := \zeta \wedge \tau_{j+1}$. By Corollary 4.9 applied with κ_i playing the role of κ and the single type j individual that acquires its jth mutation at time κ_i playing the role of S, we get

$$X_{j,2,i}(u \wedge \zeta_j) = e^{\int_{\kappa_i}^{u \wedge \zeta_j} G_j(v)\, dv}(1 + Z_{i,j}(u)), \tag{6.29}$$

where $(Z_{i,j}(\kappa_i + u), u \geq 0)$ is a mean zero martingale. Therefore, using part 4 of Proposition 4.4, we get that on $\{\kappa_i < \tau_{j+1} < \zeta\}$,

$$\begin{aligned} X_{j,2,i}(\tau_{j+1}) &= e^{\int_{\tau_j}^{\tau_{j+1}} G_j(v)\, dv} e^{-\int_{\tau_j}^{\kappa_i} G_j(v)\, dv}(1 + Z_{i,j}(\tau_{j+1})) \\ &\leq \frac{2s}{\mu} e^{-\int_{\tau_j}^{\kappa_i} G_j(v)\, dv}(1 + Z_{i,j}(\tau_{j+1})). \end{aligned} \tag{6.30}$$

Corollary 4.9 combined with (4.19) and (6.29) gives that on $\{\kappa_i < \tau_{j+1}\}$,

$$\begin{aligned} \mathrm{Var}(Z_{i,j}(\tau_{j+1}) | \mathcal{F}_{\kappa_i}) &\leq 3E\biggl[\int_{\kappa_i}^{\tau_{j+1} \wedge \zeta_j} e^{-2\int_{\kappa_i}^{u} G_j(v)\, dv} X_{j,2,i}(u)\, du \biggm| \mathcal{F}_{\kappa_i}\biggr] \\ &= 3E\biggl[\int_{\kappa_i}^{\tau_{j+1} \wedge \zeta_j} e^{-\int_{\kappa_i}^{u} G_j(v)\, dv}(1 + Z_{i,j}(u))\, du \biggm| \mathcal{F}_{\kappa_i}\biggr]. \end{aligned}$$

Because $k^* + 1 \le j \le J$ by Lemma 6.2, it follows from part 2 of Proposition 4.4 that for sufficiently large N, if $v \in [\tau_j, \tau_{j+1}]$ and $v < \zeta$, then $G_j(v) \ge (1-2\delta)sk_N$. Therefore, on $\{\kappa_i < \tau_{j+1}\}$,

$$\begin{aligned} \mathrm{Var}(Z_{i,j}(\tau_{j+1})|\mathcal{F}_{\kappa_i}) &\le 3E\bigg[\int_{\kappa_i}^{\infty} e^{-sk_N(1-2\delta)(u-\kappa_i)}(1+Z_{i,j}(u))\,du\bigg|\mathcal{F}_{\kappa_i}\bigg] \\ &= \frac{3}{sk_N(1-2\delta)}. \end{aligned} \tag{6.31}$$

From (6.30) and (6.31), we get that on $\{\kappa_i < \tau_{j+1}\}$,

$$E\big[X_{j,2,i}^2(\tau_{j+1})\mathbb{1}_{\{\tau_{j+1}<\zeta\}}\big|\mathcal{F}_{\kappa_i}\big] \le \frac{4s^2}{\mu^2}e^{-2\int_{\tau_j}^{\kappa_i} G_j(v)\,dv}\bigg(1 + \frac{3}{sk_N(1-2\delta)}\bigg).$$

By assumption A3, the second term inside the parentheses dominates when N is large. Also, by part 1 of Proposition 4.4, we have $s(q_j - C_3) \le G_j(v) \le s(q_j + C_3)$ if $\tau_j \le v < \zeta_j$. Therefore, for sufficiently large N, on $\{\kappa_i < \tau_{j+1}\}$,

$$E\big[X_{j,2,i}^2(\tau_{j+1})\mathbb{1}_{\{\tau_{j+1}<\zeta\}}\big|\mathcal{F}_{\kappa_i}\big] \le \frac{Cs}{\mu^2 k_N}e^{-2s(q_j-C_3)(\kappa_i-\tau_j)}. \tag{6.32}$$

Next, we condition on $\mathcal{F}_{\tau_j}$. Using (4.7) followed by part 1 of Proposition 4.4, the rate at which type j mutations are appearing at time u, provided that $\tau_j \le u < \zeta_j$, is

$$\mu X_{j-1}(u) \le (1+\delta)se^{\int_{\tau_j}^{u} G_{j-1}(v)\,dv} \le (1+\delta)se^{s(q_j+C_3-1)(u-\tau_j)}. \tag{6.33}$$

Therefore, using (6.32), (6.33), and (4.9), we get that on $\{\tau_j < \zeta\}$,

$$\begin{aligned} &E\bigg[\bigg(\sum_{j=1}^{H_j} X_{j,2,i}(\tau_{j+1})^2\bigg)\mathbb{1}_{\{\tau_{j+1}<\zeta\}}\bigg|\mathcal{F}_{\tau_j}\bigg] \\ &\qquad \le \int_{\xi_j}^{(\tau_j+2a_N/k_N)\wedge\zeta} (1+\delta)se^{s(q_j+C_3-1)(u-\tau_j)} \cdot \frac{Cs}{\mu^2 k_N}e^{-2s(q_j-C_3)(u-\tau_j)}\,du \\ &\qquad \le \frac{Cs^2}{\mu^2 k_N}\int_{\xi_j}^{(\tau_j+2a_N/k_N)\wedge\zeta} e^{-s(q_j-3C_3+1)(u-\tau_j)}\,du \\ &\qquad \le \frac{Cs}{\mu^2 k_N(q_j-3C_3+1)} \cdot e^{-s(q_j-3C_3+1)(\xi_j-\tau_j)}. \end{aligned} \tag{6.34}$$

Note that $e^{-sq_j(\xi_j-\tau_j)} = sq_je^{-b}$ for sufficiently large N by (4.3). Also, $q_j \ge (1-2\delta)k_N$ on $\{\tau_j < \zeta\}$ for sufficiently large N by part 3 of Proposition 4.4, so

$$s(\xi_j - \tau_j) = \frac{1}{q_j}\bigg(\log\bigg(\frac{1}{sq_j}\bigg) + b\bigg) \le \frac{1}{(1-2\delta)k_N}\bigg(\log\bigg(\frac{1}{s}\bigg) + b\bigg) \to 0 \tag{6.35}$$

as $N \to \infty$ by assumption A1. Therefore, $e^{(3C_3-1)s(\xi_j-\tau_j)} \to 1$ as $N \to \infty$. Combining these observations with (6.34), and using that $\tau_{j+1} < \zeta$ implies $\tau_j < \zeta$ in view of (4.9), we get that for sufficiently large N,

$$E\bigg[\bigg(\sum_{j=1}^{H_j} X_{j,2,i}(\tau_{j+1})^2\bigg)\mathbb{1}_{\{\tau_{j+1}<\zeta\}}\bigg|\mathcal{F}_{\tau_j}\bigg] \le \frac{Cs^2e^{-b}}{\mu^2 k_N}. \tag{6.36}$$

Finally, we can take expectations of both sides in (6.36) and combine the result with (6.28) and the fact that the cardinality of I is at most $3k_NT$ by Lemma 6.2 to obtain the result of the lemma. ☐

Lemma 6.6. *Recall that* $V_j = \min\{V_{1,j}, V_{2,j}\}$. *Define the event*

$$A_j^* := \{U_1(\tau_j) = U_2(\tau_j) = j-1\} \cap \{\tau_j \leq T_{1,2} < V_j < \tau_{j+1}\}.$$

Then

$$\lim_{N\to\infty} P\bigg(\Lambda \cap \bigcup_{j\in I} A_j^*\bigg) = 0.$$

Proof. Fix $j \in I$. Randomly label the type $j-1$ individuals at time τ_j by $1, 2, \dots, \lceil s/\mu \rceil$. For $t \geq \tau_j$, let $X_{j-1}^i(t)$ denote the number of type $j-1$ individuals at time t descended from the type $j-1$ individual labelled i at time τ_j.

Let $\mathcal{C}_j$ be the σ-field generated by the the event $\{V_j < \zeta\}$ and the random variables $V_{1,j}, V_{2,j}, X_{j-1}^1(V_j-), \dots, X_{j-1}^{\lceil s/\mu\rceil}(V_j-)$. The only way that A_j^* can occur is if the first two lineages get traced back to distinct type $j-1$ ancestors at time V_j- and then merge between times τ_j and V_j-. Conditional on $\mathcal{C}_j$, we know that one of the type $j-1$ individuals at time V_j- will get a jth mutation at time V_j, but all of the type $j-1$ individuals at time V_j- are equally likely to be ancestors of individuals in our sample at time $a_N T$. Therefore, using the notation from (5.13),

$$P(A_j^* \mathbb{1}_{\{V_j<\zeta\}} | \mathcal{C}_j) \leq \bigg(\sum_{i=1}^{\lceil s/\mu\rceil} \frac{X_{j-1}^i(V_j-)(X_{j-1}^i(V_j-)-1)}{X_{j-1}(V_j-)(X_{j-1}(V_j-)-1)} \bigg) \mathbb{1}_{\{V_j<\zeta\}} \leq \sum_{i=1}^{\lceil s/\mu\rceil} R_{i,j-1}^2.$$

It follows from part 6 of Proposition 4.1 that $\tau_{k^*+K+1} < (K+1)(2a_N/k_N) \leq a_N$ for sufficiently large N, and therefore by (6.16), we have $j_1 - 1 \geq k^* + 1 + K$. Thus, summing over $j \in I$, taking expectations of both sides, and applying Lemma 5.3 and Lemma 6.2, we get

$$\sum_{j\in I} P(A_j^* \cap \{V_j < \zeta\}) \leq \sum_{j\in I} \frac{C\mu}{s^2 k_N} \leq \frac{C\mu T}{s^2}. \tag{6.37}$$

The right-hand side of (6.37) tends to zero as $N \to \infty$ by (2.4). The result of the lemma follows because if $\Lambda \cap A_j^*$ occurs for some $j \in I$, then $V_j < \tau_{j+1} < \zeta$ by Lemma 6.2. ☐

7 Coupling with a branching process between times τ_j and τ_{j+1}

Recall from Lemmas 6.5 and 6.6 and the discussion before Lemma 6.5 that we have shown that all possible coalescence events have low probability, except for the possibility that type j lineages at time τ_{j+1} could be descended from the same type j mutation between times τ_j and ξ_j. In this section, we study these early type j mutations in depth. The strategy here will be to couple the descendants of these mutations with a supercritical branching process.

7.1 Review of results on continuous-time branching processes

Consider a continuous-time birth and death process $(Z(t), t \geq 0)$ in which each individual independently dies at rate $\nu > 0$ and gives birth to a new individual at rate $\lambda > \nu$. Assume $Z(0) = 1$. Using results in [1], one can show that

$$P(Z(t) > 0) = \frac{\lambda - \nu}{\lambda - \nu e^{-(\lambda-\nu)t}}, \tag{7.1}$$

which is also stated as part of Lemma 9.16 of [28]. Let r denote the probability that the population goes extinct by time t. By letting $t \to \infty$ in (7.1), we get

$$1 - r = \frac{\lambda - \nu}{\lambda}. \tag{7.2}$$

Let $W(t) = e^{-(\lambda-\nu)t} Z(t)$. It is well-known (see, for example, section 7 of Chapter III in [1]) that $(W(t), t \geq 0)$ is a martingale, and there is a random variable W such that

$$\lim_{t \to \infty} W(t) = W \quad \text{a.s.}, \tag{7.3}$$

where W is zero on the event that the branching process goes extinct and is almost surely strictly positive on the event that the branching process survives forever. In this instance, it is also known that the conditional distribution of W given that the branching process survives forever is the exponential distribution with rate parameter $1 - r$, so that if $x \geq 0$, then

$$P(W > x) = (1 - r) e^{-(1-r)x}. \tag{7.4}$$

This can be derived from results in [1] and is also worked out, for example, in [15]. Recall that if S has an exponential distribution with parameter λ, then $E[S] = 1/\lambda$ and $E[S^2] = 2/\lambda^2$. Because $P(W > 0) = 1 - r$, it follows that $E[W] = 1$ and $\text{Var}(W) \leq E[W^2] = 2/(1-r)$. We will need the following result concerning the rate of convergence of $W(t)$ to W.

Lemma 7.1. *For all $\eta > 0$ and $t > 0$, we have*

$$P(|W(t) - W| > \eta) \leq \frac{2e^{-(\lambda-\nu)t}}{\eta^2 (1-r)}.$$

Proof. Conditional on $Z(t)$, we can consider separately the descendants of the $Z(t)$ individuals at time t to see that

$$W = e^{-(\lambda-\nu)t} \sum_{i=1}^{Z(t)} W_i,$$

where the random variables $W_1, \dots, W_{Z(t)}$ are independent and have the same distribution as W (see section 10 in Chapter III of [1]). It follows that

$$E[W|Z(t)] = e^{-(\lambda-\nu)t} Z(t) E[W] = W(t)$$

and

$$\text{Var}(W|Z(t)) = e^{-2(\lambda-\nu)t} Z(t) \text{Var}(W) \leq \frac{2e^{-(\lambda-\nu)t} W(t)}{1-r}.$$

Therefore, by Chebyshev's Inequality,

$$P(|W - W(t)| > \eta | Z(t)) \leq \frac{\text{Var}(W|Z(t))}{\eta^2} \leq \frac{2e^{-(\lambda-\nu)t} W(t)}{\eta^2 (1-r)}. \tag{7.5}$$

Because $E[W(t)] = 1$, the result follows by taking expectations of both sides in (7.5). □

7.2 A branching process coupling between times τ_j and τ_{j+1}

We will assume now that $j \in I$, where I was defined in (6.14). By Lemma 6.2, this ensures that $\tau_{j+1} < a_N T$ on Λ. Recalling Corollary 4.8, we will let

$$X'_j(t) := X_j^{\tau_j, \xi_j}(t)$$

denote the number of type j individuals at time t that are descended from individuals that acquired a jth mutation during the time interval $(\tau_j, \xi_j]$. Note that $X'_j(t) = X_{j,1}(t)$, as long as there are no type j mutations before time τ_j. We say there is a pure birth event at time t if $X'_j(t) = X'_j(t-) + 1$ and a pure death event at time t if $X'_j(t) = X'_j(t-) - 1$. We say there is a birth and death event at time t if one of the $X'_j(t-)$ individuals at time

$t-$ gives birth and another dies, so that $X_j'(t) = X_j'(t-)$. Let $B_j'(t)$ and $D_j'(t)$ denote the expressions in (4.17) and (4.18) respectively with $X_j'(t)$ in place of $X_j(t)$. Recall from the discussion surrounding (4.16), (4.17), and (4.18) that if $t \in [\tau_j, \tau_{j+1} \wedge \zeta)$, then the rate at which a particular type j individual gives birth as part of a pure birth event is

$$B_j'(t) = \left(1 - \frac{X_j'(t)}{N}\right)(1 + s(j - M(t))), \tag{7.6}$$

while the rate at which a particular type j individual is involved in a pure death event is

$$D_j'(t) = 1 + \mu - \frac{X_j'(t)}{N}(1 + s(j - M(t))). \tag{7.7}$$

Also, the rate at which a particular type j individual gives birth as part of a birth and death event and the rate at which a particular type j individual dies as part of a birth and death event are both equal to

$$O_j(t) = \frac{X_j'(t)}{N}(1 + s(j - M(t))). \tag{7.8}$$

We write $B_j^*(t) := B_j'(t) + O_j(t) = 1 + s(j - M(t))$ and $D_j^*(t) := D_j'(t) + O_j(t) = 1 + \mu$ for the total birth and death rates respectively. The following lemma gives upper and lower bounds on these birth and death rates. The lemma also gives a bound on the rate of type j mutations, which will correspond to immigration in our branching process.

Lemma 7.2. *There is a positive constant C_4 such that for sufficiently large N, if $X_j'(t) \leq s/2\mu$ and $t \in [\tau_j, \tau_{j+1} \wedge \zeta)$, then the following hold:*

$$1 - s \leq D_j'(t) \leq D_j^*(t) = 1 + \mu, \tag{7.9}$$

$$1 + sq_j - C_4 s \leq B_j'(t) \leq B_j^*(t) \leq 1 + sq_j + C_4 s, \tag{7.10}$$

$$(1 - \delta)se^{s(q_j - C_4)(t - \tau_j)} \leq \mu X_{j-1}(t) \leq (1 + \delta)se^{s(q_j + C_4)(t - \tau_j)}. \tag{7.11}$$

Proof. Suppose $X_j'(t) \leq s/2\mu$ and $t \in [\tau_j, \tau_{j+1} \wedge \zeta)$. By (2.4), assumption A3, and the fact that $j \leq J$ by Lemma 6.2, for sufficiently large N we have

$$O_j(t) \leq \left(\frac{s}{2\mu N}\right)(1 + s(j - M(t))) \leq \left(\frac{s}{2\mu N}\right)(1 + sJ) \leq s. \tag{7.12}$$

The result (7.9) follows immediately from equations (7.7), (7.8), and (7.12).

To bound the birth rate, note that since $G_j(t) = s(j - M(t)) - \mu$, we have

$$1 + \mu + G_j(t) - O_j(t) = B_j'(t) \leq B_j^*(t) = 1 + \mu + G_j(t).$$

Since $s(q_j - C_3) \leq G_j(t) \leq s(q_j + C_3)$ for sufficiently large N by (6.16) and part 1 of Proposition 4.4, the inequality (7.10) now follows from (7.12) and (2.4).

Finally, if $t \in [\tau_j, \tau_{j+1} \wedge \zeta)$, then since $G_{j-1}(t) = G_j(t) - s$, part 1 of Proposition 4.4 gives $s(q_j - C_3 - 1) \leq G_{j-1}(t) \leq s(q_j + C_3 - 1)$. Now (7.11) follows from this observation and (4.7). ☐

We will use the bounds in Lemma 7.2 to obtain a coupling in which $(X_j'(t), t \geq \tau_j)$ is bounded between two branching processes with immigration. More specifically, we will construct processes $(X_j^+(t), t \geq 0)$ and $(X_j^-(t), t \geq 0)$ such that

$$X_j^-(t) \leq X_j'(t + \tau_j) \leq X_j^+(t) \tag{7.13}$$

for $t < \kappa_j$, where

$$\kappa_j := \inf\left\{u : X_j^+(u) \geq \frac{s}{2\mu}\right\} \wedge \big((\tau_{j+1} \wedge \zeta) - \tau_j\big). \tag{7.14}$$

The processes $(X_j^+(t), t \geq 0)$ and $(X_j^-(t), t \geq 0)$ evolve according to the following rules. First, $X_j^+(t)$ is the size at time t of a population for which, at time t:

- New immigrants appear at rate $\phi_j^+(t) = (1+\delta)se^{s(q_j+C_4)t}\mathbb{1}_{\{t \leq \xi_j - \tau_j\}}$.
- Each individual gives birth to a new individual at rate $\lambda_j^+ = 1 + s(q_j + C_4)$.
- Each individual dies at rate $\nu_j^+ = 1 - s$.

Likewise, for the process $(X_j^-(t), t \geq 0)$, at time t:

- New immigrants appear at rate $\phi_j^-(t) = (1-\delta)se^{s(q_j-C_4)t}\mathbb{1}_{\{t \leq \xi_j - \tau_j\}}$.
- Each individual gives birth to a new individual at rate $\lambda_j^- = 1 + s(q_j - C_4)$.
- Each individual dies at rate $\nu_j^- = 1 + \mu$.

To establish that a coupling can be achieved so that (7.13) holds, we will give an explicit construction of the processes $(X_j^+(t), t \geq 0)$ and $(X_j^-(t), t \geq 0)$. To do this, we will construct a population in which individuals are colored red, yellow, and blue. We will let $X_j^+(t)$ be the total number of individuals at time t, and we will let $X_j^-(t)$ be the total number of red individuals at time t. For $t < \kappa_j$, the number of individuals at time t that are red or yellow will equal $X_j'(\tau_j + t)$, which we will refer to as the number of individuals in the "original population". We will number the individuals in our population by the order in which they were born.

The construction will require the original population process $(\mathbf{X}(t), t \geq 0)$, as well as additional Poisson processes. For each $i \in \mathbb{N}$, we will have Poisson processes $N_{b,i,j}$ and $N_{d,i,j}$ to help construct births and deaths and an additional Poisson processes $N_{m,j}$ to handle immigration. These will be Poisson processes on $[0,\infty) \times [0,\infty)$ with Lebesgue intensity, which will be independent of one another and of the original population process. We will also need a sequence $(\beta_{\ell,j})_{\ell=1}^\infty$ of independent random variables which are uniformly distributed on $(0,1)$ and are independent of $(\mathbf{X}(t), t \geq 0)$ and the above Poisson processes.

We first construct our population up to time κ_j. Observe, as we go through the construction, that the red population has immigration, birth, and death rates of $\phi_j^-(t)$, λ_j^-, and ν_j^- respectively, the total population has immigration, birth, and death rates of $\phi_j^+(t)$, λ_j^+, and ν_j^+ respectively, and the red and yellow individuals stay in one-to-one correspondence with the original population. This construction is well-defined because Lemma 7.2 ensures that the rates described below are positive and the probabilities indicated below are between zero and one.

- If a type j mutation occurs in the original population at time $\tau_j + t$, then an immigrant appears at time t. This will be the ℓth change in the population for some positive integer ℓ. We color this immigrant red if $\beta_{\ell,j} \leq \phi_j^-(t-)/(\mu X_{j-1}(t-))$, and otherwise we color it yellow. A blue immigrant appears at time t if the Poisson process $N_{m,j}$ has a point (t,x) with $x \leq \phi_j^+(t-) - \mu X_{j-1}(t-)$.
- If the ith individual at time $t-$ is blue, then it gives birth to a blue individual at time t if the Poisson process $N_{b,i,j}$ has a point (t,x) with $x \leq \lambda_j^+$ and dies at time t if there is a point (t,x) in $N_{d,i,j}$ with $x \leq \nu_j^+$.

- Suppose the ith individual at time $t-$ is red. If the corresponding individual in the original population gives birth at time $\tau_j + t$ as part of a pure birth event, then the ith individual gives birth at time t. This will be the ℓth change in the population for some ℓ, and the new individual born will be red if $\beta_{\ell,j} \leq \lambda_j^- / B_j'(t-)$ and otherwise will be yellow. If the corresponding individual in the original population gives birth at time $\tau_j + t$ as part of a birth and death event, then the ith individual gives birth to a yellow individual at time t. The ith individual also gives birth to a blue individual at time t if the Poisson process $N_{b,i,j}$ has a point at (t,x) with $x \leq \lambda_j^+ - B_j^*(t-)$.

  If the corresponding individual in the original population dies at time $\tau_j + t$ as part of a pure death event, then this will lead to the ℓth change in the population for some ℓ, and the ith individual dies at time t if $\beta_{\ell,j} \leq \nu_j^+ / D_j'(t-)$ and otherwise turns blue. If the corresponding individual in the original population dies at time t as part of a birth and death event, then the ith individual turns blue at time t. The ith individual also turns yellow at time t if $N_{d,i,j}$ has a point at (t,x) with $x \leq \nu_j^- - D_j^*(t-)$.
- Suppose the ith individual at time $t-$ is yellow. If the corresponding individual in the original population gives birth at time $\tau_j + t$ as part of either a pure birth or a birth and death event, then a new yellow individual is born at time t. The ith individual also gives birth to a blue individual at time t if the Poisson process $N_{b,i,j}$ has a point at (t,x) with $x \leq \lambda_j^+ - B_j^*(t-)$.

  If the corresponding individual in the original population dies at time t as part of a pure death event, then this will be the ℓth change in the population for some ℓ, and the ith individual dies at time t if $\beta_\ell \leq \nu_j^+ / D_j'(t-)$ and otherwise turns blue. If the corresponding individual dies at time t as part of a birth and death event, then the ith individual turns blue at time t.

At time κ_j, the coupling with the original population is broken, and we make all yellow individuals blue. After time κ_j, the process evolves as follows:

- If $\kappa_j < t \leq \xi_j$, then a red immigrant appears at time t if there is a point (t,x) of $N_{m,j}$ with $x \leq \phi_j^-(t-)$ and a blue immigrant appears at time t if there is a point (t,x) of $N_{m,j}$ with $\phi_j^-(t) < x \leq \phi_j^+(t)$.
- If the ith individual is blue, it gives birth to a blue individual at time t if $N_{b,i,j}$ has a point (t,x) with $x \leq \lambda_j^+$ and dies at time t if there is a point (t,x) in $N_{d,i,j}$ with $x \leq \nu_j^+$.
- Suppose the ith individual is red. Then the ith individual gives birth to a red individual at time t if the Poisson process $N_{b,i,j}$ has a point (t,x) with $x \leq \lambda_j^-$ and to a blue individual at time t if $N_{b,i,j}$ has a point (t,x) with $\lambda_j^- < x \leq \lambda_j^+$. Also, the ith individual dies at time t if the Poisson process $N_{d,i,j}$ has a point (t,x) with $x \leq \nu_j^+$ and turns blue at time t if $N_{d,i,j}$ has a point (t,x) with $\nu_j^+ < x \leq \nu_j^-$.

For $j \in I$, let $\mathcal{H}_j$ be the σ-field generated by $\mathcal{F}_{\tau_j}$ along with the Poisson processes $N_{b,i,h}$, $N_{d,i,h}$, and $N_{m,h}$ and the random variables $\beta_{\ell,h}$ for $h < j$. Because the immigration, birth, and death rates ϕ_j^+, ϕ_j^-, λ_j^+, λ_j^-, ν_j^+, and ν_j^- are all $\mathcal{H}_j$-measurable, conditional on $\mathcal{H}_j$, the processes $(X_j^+(t), t \geq 0)$ and $(X_j^-(t), t \geq 0)$ are continuous-time branching processes with immigration, in which the immigration rate varies with time.

Let

$$\tau_j' := \tau_j + \frac{3}{sq_j} \log\left(\frac{1}{sq_j}\right).$$

Note that $\tau_j < \xi_j < \tau_j'$ for sufficiently large N. In view of (4.9) and part 3 of Proposition 4.4, along with the fact that $\log(s/\mu)/\log(1/sk_N) \to \infty$ as $N \to \infty$ by (2.4), we have

$\tau_j' < \tau_{j+1}$ on $\{\zeta > \tau_j'\}$ if N is sufficiently large. Lemma 7.4 below helps to bound the probability that $\kappa_j < \tau_j' - \tau_j$ and therefore helps to ensure that with high probability, (7.13) holds up to time $\tau_j' - \tau_j$. We will need the following bound on the mean of the branching process.

Lemma 7.3. *For sufficiently large N, on $\{\tau_j < \zeta\}$, we have*

$$E[X_j^+(\tau_j' - \tau_j)|\mathcal{F}_{\tau_j}] \leq \frac{C}{s^3 k_N^4} \log\left(\frac{1}{sk_N}\right).$$

Proof. Standard calculations involving supercritical branching processes give

$$\begin{aligned} E[X_j^+(\tau_j' - \tau_j)|\mathcal{F}_{\tau_j}] &= \int_0^{\tau_j' - \tau_j} \phi_j^+(u) e^{(\lambda_j^+ - \nu_j^+)(\tau_j' - \tau_j - u)}\,du \\ &= (1+\delta)s \int_0^{\xi_j - \tau_j} e^{s(q_j + C_4)u} e^{s(q_j + C_4 + 1)(\tau_j' - \tau_j - u)}\,du \\ &= (1+\delta)s e^{s(q_j + C_4 + 1)(\tau_j' - \tau_j)} \int_0^{\xi_j - \tau_j} e^{-su}\,du. \end{aligned}$$

Now $s(C_4+1)(\tau_j' - \tau_j) \to 0$ as $N \to \infty$ by the reasoning in (6.35), and $e^{sq_j(\tau_j' - \tau_j)} = (sq_j)^{-3}$. Also,

$$\int_0^{\xi_j - \tau_j} e^{-su}\,du = \frac{1 - e^{-s(\xi_j - \tau_j)}}{s} \leq \xi_j - \tau_j = \frac{1}{sq_j} \log\left(\frac{1}{sq_j}\right) + \frac{b}{sq_j}.$$

Since $q_j \geq (1 - 2\delta)k_N$ on $\{\tau_j < \zeta\}$ by part 3 of Proposition 4.4, the result follows. ☐

Lemma 7.4. *We have*

$$\lim_{N \to \infty} P\left(\Lambda \cap \bigcup_{j \in I} \{\kappa_j \leq \tau_j' - \tau_j\}\right) = 0.$$

Proof. In view of Lemma 6.2, for $j \in I$ we have $\tau_j' < \tau_{j+1} < \zeta$ on Λ. Therefore, for $j \in I$, on Λ the only way to have $\kappa_j \leq \tau_j' - \tau_j$ would be to have $X_j^+(t) > s/2\mu$ for some $t \leq \tau_j' - \tau_j$. Because $(X_j^+(t), t \geq 0)$ is a submartingale, it follows from Doob's Maximal Inequality and Lemma 7.3 that

$$\begin{aligned} P(\Lambda \cap \{\kappa_j \leq \tau_j' - \tau_j\}|\mathcal{F}_{\tau_j}) &\leq P\left(\sup_{0 \leq t \leq \tau_j' - \tau_j} X_j^+(t) > \frac{s}{2\mu}\middle|\mathcal{F}_{\tau_j}\right)\mathbb{1}_{\{\zeta > \tau_j\}} \\ &\leq \frac{2\mu}{s} E[X_j^+(\tau_j' - \tau_j)|\mathcal{F}_{\tau_j}]\mathbb{1}_{\{\zeta > \tau_j\}} \\ &\leq \frac{C\mu}{s^4 k_N^4} \log\left(\frac{1}{sk_N}\right). \end{aligned} \tag{7.15}$$

Summing over $j \in I$, and then using (2.4) and the fact that the cardinality of I is at most $3Tk_N$ by Lemma 6.2, we obtain the result. ☐

7.3 The probability that a family survives

Here we use the branching process coupling introduced in the previous subsection to obtain upper and lower bounds on the probability that an individual will acquire a jth mutation before time ξ_j and have descendants surviving a long time into the future. Note that the right-hand sides of (7.16) and (7.17) below converge to zero as $N \to \infty$ on Λ by (2.3) and part 3 of Proposition 4.4.

Lemma 7.5. *Suppose $j \in I$, where j is possibly random, and τ_j is a stopping time. Define $\mathcal{H}_j$ as in subsection 7.2. On the event $\{\tau_j < \zeta\}$, we have for sufficiently large N,*

$$\frac{(1-2\delta)e^b}{q_j} \leq P(X_j^-(\tau_j' - \tau_j) > 0 | \mathcal{H}_j) \leq P(X_j^+(\tau_j' - \tau_j) > 0 | \mathcal{H}_j) \leq \frac{(1+2\delta)e^b}{q_j}. \tag{7.16}$$

Also, letting L_j^- and L_j^+ denote the numbers of immigrants in $(X_j^-(t), t \geq 0)$ and $(X_j^+(t), t \geq 0)$ respectively that have descendants alive at time $\tau_j' - \tau_j$, for sufficiently large N on $\{\tau_j < \zeta\}$ we have

$$P(L_j^+ \geq 2 | \mathcal{H}_j) \leq \frac{2e^{2b}}{q_j^2}. \tag{7.17}$$

Proof. Throughout the proof, we work on the event $\{\tau_j < \zeta\}$. Because $X_j^-(t) \leq X_j^+(t)$ for all $t \geq 0$, the second inequality in (7.16) is obvious. We now prove the third inequality. By (7.1), the probability that an immigrant in the branching process $(X_j^+(t), t \geq 0)$ at time u has descendants that survive until time $\tau_j' - \tau_j$ is

$$\frac{\lambda_j^+ - \nu_j^+}{\lambda_j^+ - \nu_j^+ e^{-(\lambda_j^+ - \nu_j^+)(\tau_j' - \tau_j - u)}}.$$

Now $\lambda_j^+ - \nu_j^+ = s(q_j + C_4 + 1)$. Also, for sufficiently large N,

$$\tau_j' - \xi_j = \frac{2}{sq_j} \log\left(\frac{1}{sq_j}\right) - \frac{b}{sq_j} \geq \frac{3}{2sq_j} \log\left(\frac{1}{sq_j}\right).$$

Therefore, if $u < \xi_j - \tau_j$, then

$$\nu_j^+ e^{-(\lambda_j^+ - \nu_j^+)(\tau_j' - \tau_j - u)} \leq (1-s) e^{-s(q_j + C_4 + 1)(\tau_j' - \xi_j)} \leq e^{-sq_j(\tau_j' - \xi_j)} \leq (sq_j)^{3/2},$$

which, in view of part 3 of Proposition 4.4 and assumption A3, implies that for sufficiently large N,

$$\frac{\lambda_j^+ - \nu_j^+}{\lambda_j^+ - \nu_j^+ e^{-(\lambda_j^+ - \nu_j^+)(\tau_j' - \tau_j - u)}} \leq \frac{s(q_j + C_4 + 1)}{1 + s(q_j + C_4) - (sq_j)^{3/2}} \leq s(q_j + C_4 + 1).$$

Therefore,

$$\begin{aligned} E[L_j^+ | \mathcal{H}_j] &= \int_0^{\xi_j - \tau_j} \phi_j^+(u) \cdot \frac{\lambda_j^+ - \nu_j^+}{\lambda_j^+ - \nu_j^+ e^{-(\lambda_j^+ - \nu_j^+)(\tau_j' - \tau_j - u)}} \, du \\ &\leq \int_0^{\xi_j - \tau_j} (1+\delta) s e^{s(q_j + C_4)u} \cdot s(q_j + C_4 + 1) \, du \\ &= (1+\delta) s^2 (q_j + C_4 + 1) \left(\frac{e^{s(q_j + C_4)(\xi_j - \tau_j)} - 1}{s(q_j + C_4)} \right) \\ &\leq (1+\delta) \left(\frac{q_j + C_4 + 1}{q_j + C_4} \right) s e^{s(q_j + C_4)(\xi_j - \tau_j)}. \end{aligned}$$

Because $e^{sq_j(\xi_j - \tau_j)} = e^b/(sq_j)$ and $C_4 s(\xi_j - \tau_j) \to 0$ as $N \to \infty$ by (6.35), it follows that for sufficiently large N,

$$E[L_j^+ | \mathcal{H}_j] \leq \frac{(1+2\delta)e^b}{q_j}. \tag{7.18}$$

The conditional Markov's Inequality now gives the third inequality in (7.16). Because the conditional distribution of L_j^+ given $\mathcal{H}_j$ is Poisson, we have $P(L_j^+ \geq 2 | \mathcal{H}_j) \leq (E[L_j^+ | \mathcal{H}_j])^2$. Therefore, (7.17) also follows from (7.18).

It remains to prove the first inequality in (7.16). The argument is similar to that for the third inequality, but we will need a lower bound on the expectation. For sufficiently large N,

$$\begin{aligned} E[L_j^-|\mathcal{H}_j] &= \int_0^{\xi_j-\tau_j} \phi_j^-(u)\cdot \frac{\lambda_j^- - \nu_j^-}{\lambda_j^- - \nu_j^- e^{-(\lambda_j^- - \nu_j^-)(\tau_j' - \tau_j - u)}}\,du \\ &\geq \int_0^{\xi_j-\tau_j} \phi_j^-(u)\cdot \frac{\lambda_j^- - \nu_j^-}{\lambda_j^-}\,du \\ &= \int_0^{\xi_j-\tau_j} (1-\delta)se^{s(q_j-C_4)u}\cdot \frac{s(q_j-C_4)-\mu}{1+s(q_j-C_4)}\,du \\ &= \frac{(1-\delta)s^2(q_j-C_4-\mu/s)}{1+s(q_j-C_4)}\left(\frac{e^{s(q_j-C_4)(\xi_j-\tau_j)}-1}{s(q_j-C_4)}\right) \\ &\geq \frac{(1-(3/2)\delta)e^b}{q_j}. \end{aligned} \tag{7.19}$$

Because the conditional distribution of L_j^- given $\mathcal{H}_j$ is Poisson, we have

$$P(X_j^-(\tau_j'-\tau_j)>0|\mathcal{H}_j) = P(L_j^->0|\mathcal{H}_j) = 1-e^{-E[L_j^-|\mathcal{H}_j]} \geq E[L_j^-|\mathcal{H}_j] - (E[L_j^-|\mathcal{H}_j])^2.$$

The first inequality in (7.16) follows from this result and (7.19). ☐

7.4 The size of a surviving family

The lemma below bounds the probability that some individual will acquire a jth mutation before time ξ_j and have at least $xe^{sq_j(\tau_j'-\tau_j)}$ descendants alive at time τ_j'. Recall from (4.6) and part 1 of Proposition 4.4 that $e^{sq_j(\tau_j'-\tau_j)}$ is approximately the number of type j individuals that we would expect there to be in the population in the absence of such an early type j mutation. This result is the precise version of (3.5), which is the key to understanding why the Bolthausen-Sznitman coalescent describes the genealogy of the population.

Lemma 7.6. *Fix $j\in I$, and recall the definition of $\mathcal{H}_j$ from subsection 7.2. For sufficiently large N, on $\{\tau_j<\zeta\}$, we have for all $x\in[\delta/2,2/\delta]$,*

$$P(X_j^-(\tau_j'-\tau_j) > xe^{sq_j(\tau_j'-\tau_j)}|\mathcal{H}_j) \geq \frac{1-7\delta}{q_jx} \tag{7.20}$$

and for all $x\in[e^{-b},2/\delta]$,

$$P(X_j^+(\tau_j'-\tau_j) > xe^{sq_j(\tau_j'-\tau_j)}|\mathcal{H}_j) \leq \frac{1+7\delta}{q_jx}. \tag{7.21}$$

Proof. Throughout the proof, we work on the event $\{\tau_j<\zeta\}$. We first prove (7.21). Suppose $x\in[e^{-b},2/\delta]$. If $X_j^+(\tau_j'-\tau_j) > xe^{sq_j(\tau_j'-\tau_j)}$, then either two immigrants in the population have descendants alive at time $\tau_j'-\tau_j$, an event whose probability has already been bounded above in (7.17), or else for some $u\in(0,\xi_j-\tau_j]$, an immigrant arrives at time u and has more than $xe^{sq_j(\tau_j'-\tau_j)}$ descendants at time $\tau_j'-\tau_j$. Note that

$$|(\lambda_j^+-\nu_j^+)-sq_j|(\tau_j'-\tau_j) = s(C_4+1)\cdot\frac{3}{sq_j}\log\left(\frac{1}{sq_j}\right)\to 0 \tag{7.22}$$

as $N\to\infty$ by the reasoning in (6.35). Therefore, for sufficiently large N, we have

$$xe^{sq_j(\tau_j'-\tau_j)} \geq (1-\delta)xe^{(\lambda_j^+-\nu_j^+)(\tau_j'-\tau_j)}. \tag{7.23}$$

Suppose an immigrant arrives at time u, and let $X^+_{j,u}(t)$ be the number of descendants of this immigrant in the population at time t. For $t \geq 0$, let

$$W^+_u(t) := e^{-(\lambda^+_j - \nu^+_j)t} X^+_{j,u}(t+u), \tag{7.24}$$

and let $W^+ := \lim_{t\to\infty} W^+_u(t)$, which exists by (7.3). Equations (7.23) and (7.24) imply that for the immigrant to have more than $xe^{sq_j(\tau'_j - \tau_j)}$ descendants in the population at time $\tau'_j - \tau_j$, if N is sufficiently large we must have

$$W^+_u(\tau'_j - \tau_j - u) \geq (1-\delta)xe^{(\lambda^+_j - \nu^+_j)u}. \tag{7.25}$$

To estimate the probability that this occurs, observe that by Lemma 7.1 and (7.2)

$$\begin{aligned} P\big(|W^+ - W^+_u(\tau'_j - \tau_j - u)| > \delta x e^{(\lambda^+_j - \nu^+_j)u}\big) &\leq \frac{2\lambda^+_j e^{-(\lambda^+_j - \nu^+_j)(\tau'_j - \tau_j - u)}}{\delta^2 x^2 e^{2(\lambda^+_j - \nu^+_j)u}(\lambda^+_j - \nu^+_j)} \\ &\leq \frac{2(1+s(q_j + C_4))e^{-(\lambda^+_j - \nu^+_j)(\tau'_j - \tau_j)}}{\delta^2 x^2 s(q_j + C_4 + 1)}. \end{aligned}$$

Since $e^{-(\lambda^+_j - \nu^+_j)(\tau'_j - \tau_j)} \leq e^{-sq_j(\tau'_j - \tau_j)} = (sq_j)^3$, it follows that for sufficiently large N,

$$P\big(|W^+ - W^+_u(\tau'_j - \tau_j - u)| > \delta x e^{(\lambda^+_j - \nu^+_j)u}\big) \leq \frac{3(sq_j)^2}{\delta^2 x^2}. \tag{7.26}$$

Note that $\lambda^+_j - \nu^+_j \geq sq_j$, and $(1-\delta/2)sq_j \leq (\lambda^+_j - \nu^+_j)/\lambda^+_j \leq (1+\delta)sq_j$ for sufficiently large N. Therefore, by (7.2) and (7.4), for sufficiently large N,

$$\begin{aligned} P\big(W^+ > (1-2\delta)xe^{(\lambda^+_j - \nu^+_j)u}\big) &= \left(\frac{\lambda^+_j - \nu^+_j}{\lambda^+_j}\right) e^{-(1-2\delta)xe^{(\lambda^+_j - \nu^+_j)u}(\lambda^+_j - \nu^+_j)/\lambda^+_j} \\ &\leq (1+\delta)sq_j e^{-(1-3\delta)sq_j x e^{sq_j u}}. \end{aligned} \tag{7.27}$$

The probability of the event in (7.25) is bounded above by the sum of the expressions in (7.26) and (7.27). Thus, combining this result with (7.17), we have

$$\begin{aligned} &P(X^+_j(\tau'_j - \tau_j) > xe^{sq_j(\tau'_j - \tau_j)} | \mathcal{H}_j) \\ &\quad \leq \frac{2e^{2b}}{q_j^2} + \int_0^{\xi_j - \tau_j} (1+\delta)se^{s(q_j + C_4)u}\left((1+\delta)sq_j e^{-(1-3\delta)sq_j x e^{sq_j u}} + \frac{3(sq_j)^2}{\delta^2 x^2}\right) du. \end{aligned} \tag{7.28}$$

Using that $e^{sq_j(\xi_j - \tau_j)} = e^b/(sq_j)$ and that $s(\xi_j - \tau_j) \to 0$ as $N \to \infty$ by (6.35), we have, for sufficiently large N,

$$\int_0^{\xi_j - \tau_j} (1+\delta)se^{s(q_j + C_4)u} \cdot \frac{3(sq_j)^2}{\delta^2 x^2}\, du \leq \frac{4e^b s}{\delta^2 x^2}. \tag{7.29}$$

Also, making the substitution $y = (1-3\delta)sq_j x e^{sq_j u}$, so that $dy/du = sq_j y$, and using again that $s(\xi_j - \tau_j) \to 0$ as $N \to \infty$, for sufficiently large N we have

$$\begin{aligned} &\int_0^{\xi_j - \tau_j} (1+\delta)se^{s(q_j + C_4)u} \cdot (1+\delta)sq_j e^{-(1-3\delta)sq_j x e^{sq_j u}}\, du \\ &\qquad \leq (1+\delta)^2 s^2 q_j e^{C_4 s(\xi_j - \tau_j)} \int_0^{\xi_j - \tau_j} e^{sq_j u} e^{-(1-3\delta)sq_j x e^{sq_j u}}\, du \\ &\qquad = (1+\delta)^2 s^2 q_j e^{C_4 s(\xi_j - \tau_j)} \int_{(1-3\delta)sq_j x}^{(1-3\delta)sq_j x e^{sq_j(\xi_j - \tau_j)}} \frac{e^{-y}}{(1-3\delta)s^2 q_j^2 x}\, dy \\ &\qquad \leq \frac{1+6\delta}{q_j x}. \end{aligned} \tag{7.30}$$

From (7.28), (7.29), and (7.30), we get

$$P(X_j^+(\tau_j' - \tau_j) > xe^{sq_j(\tau_j' - \tau_j)}|\mathcal{H}_j) \leq \frac{1}{q_j x}\bigg(1 + 6\delta + \frac{4e^b sq_j}{\delta^2 x} + \frac{2e^{2b}x}{q_j}\bigg). \tag{7.31}$$

Recall that $(1-2\delta)k_N \leq q_j \leq (e+2\delta)k_N$ on $\{\tau_j < \zeta\}$ by part 3 of Proposition 4.4. Since $k_N \to \infty$ and $sk_N \to 0$ as $N \to \infty$ by assumptions A1 and A3 respectively, the upper bound (7.21) follows from (7.31).

Next, we will suppose $x \in [\delta/2, 2/\delta]$ and show (7.20) by similar arguments. We consider only the individuals colored red in the construction given above. Suppose a red immigrant arrives at time u. Then let $X_{j,u}^-(t)$ denote the number of red descendants of this immigrant at time t, and for $t \geq 0$, let

$$W_u^-(t) := e^{-(\lambda_j^- - \nu_j^-)t} X_{j,u}^-(t+u).$$

Let $W^- := \lim_{t\to\infty} W_u^-(t)$. Because $|(\lambda_j^- - \nu_j^-) - sq_j| \to 0$ as $N \to \infty$ by the reasoning in (7.22), the reasoning that led to (7.25) implies that if

$$W_u^-(\tau_j' - \tau_j - u) \geq (1+\delta)xe^{(\lambda_j^- - \nu_j^-)u} \tag{7.32}$$

and N is large enough, then $X_j^-(\tau_j' - \tau_j) > xe^{sq_j(\tau_j' - \tau_j)}$. Because $s(\tau_j' - \tau_j) \to 0$ as $N \to \infty$ by the reasoning in (6.35), we have

$$e^{-(\lambda_j^- - \nu_j^-)(\tau_j' - \tau_j)} \leq (1+\delta)(sq_j)^3$$

for sufficiently large N. Therefore, by the reasoning leading to (7.26), for sufficiently large N we have

$$P\big(|W^- - W_u^-(\tau_j' - \tau_j - u)| > \delta xe^{(\lambda_j^- - \nu_j^-)u}\big) \leq \frac{3(sq_j)^2}{\delta^2 x^2}. \tag{7.33}$$

Note that $\lambda_j^- - \nu_j^- \leq sq_j$, and $(1-\delta)sq_j \leq (\lambda_j^- - \nu_j^-)/\lambda_j^- \leq (1+\delta/2)sq_j$ for sufficiently large N. Therefore, by (7.4)

$$\begin{aligned} P\big(W^- > (1+2\delta)xe^{(\lambda_j^- - \nu_j^-)u}\big) &= \bigg(\frac{\lambda_j^- - \nu_j^-}{\lambda_j^-}\bigg) e^{-(1+2\delta)xe^{(\lambda_j^- - \nu_j^-)u}(\lambda_j^- - \nu_j^-)/\lambda_j^-} \\ &\geq (1-\delta)e^{-(1+3\delta)sq_j xe^{sq_j u}}. \end{aligned} \tag{7.34}$$

By using (7.33) and (7.34) to bound from below the probability in (7.32), we get that for sufficiently large N,

$$\begin{aligned} &P(X_j^-(\tau_j' - \tau_j) > xe^{sq_j(\tau_j' - \tau_j)}|\mathcal{H}_j) \\ &\qquad \geq \int_0^{\xi_j - \tau_j} (1-\delta)se^{s(q_j - C_4)}\bigg((1-\delta)e^{-(1+3\delta)sq_j xe^{sq_j u}} - \frac{3(sq_j)^2}{\delta^2 x^2}\bigg)\, du. \end{aligned} \tag{7.35}$$

Following the reasoning in (7.30), this time using the substitution $y = (1+3\delta)sq_j xe^{sq_j u}$, we get

$$\begin{aligned} &\int_0^{\xi_j - \tau_j} (1-\delta)se^{s(q_j - C_4)} \cdot (1-\delta)e^{-(1+3\delta)sq_j xe^{sq_j u}}\, du \\ &\qquad\qquad \geq \frac{1-6\delta}{q_j x}\int_{(1+3\delta)sq_j x}^{(1+3\delta)sq_j xe^{sq_j(\xi_j - \tau_j)}} e^{-y}\, dy \\ &\qquad\qquad = \frac{1-6\delta}{q_j x}\big(e^{-(1+3\delta)sq_j x} - e^{-(1+3\delta)e^b x}\big) \\ &\qquad\qquad \geq \frac{1-6\delta}{q_j x}\big(1 - (1+3\delta)sq_j x - e^{-e^b x}\big). \end{aligned} \tag{7.36}$$

On $\{\tau_j < \zeta\}$, by part 3 of Proposition 4.4, we have $sq_j x \le 2(e+2\delta)sk_N/\delta \to 0$ as $N \to \infty$. Also, using the definition of b from (4.1), we have $e^{-e^b x} \le e^{-12000T/(\delta\varepsilon)}$. Therefore, using (7.36) to bound the first term in (7.35), and using the reasoning of (7.29) to bound the second term, we obtain (7.20). □

In view of (7.13), Lemmas 7.5 and 7.6 show that the number of early type j individuals is well-approximated up to time τ_j' by a continuous-time branching process. The result below tells us that the number of early type j individuals at time τ_{j+1} is usually determined, to within a small error, by the number of such individuals at time τ_j'.

Lemma 7.7. *For $j \in I$, define the event*

$$A_j := \left\{\left|e^{-sq_j(\tau_j'-\tau_j)}X_j'(\tau_j') - e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv}X_j'(\tau_{j+1})\right| > e^{-b}\right\}. \tag{7.37}$$

Then

$$\lim_{N\to\infty} P\biggl(\Lambda \cap \bigcup_{j\in I} A_j\biggr) = 0.$$

Proof. Let S be the set of individuals at time τ_j' descended from individuals that acquired their jth mutation during the time interval $(\tau_j, \xi_j]$, which means there are $X_j'(\tau_j')$ individuals in the set S. Then, using the notation of Corollary 4.9 with τ_j' in place of κ, we get that for $t \ge \tau_j'$,

$$e^{-\int_{\tau_j'}^{t\wedge\tau_{j+1}\wedge\zeta} G_j(v)\,dv} X_j'(t\wedge\tau_{j+1}\wedge\zeta) = X_j'(\tau_j') + Z_j^S(t), \tag{7.38}$$

where $(Z_j^S(\tau_j'+t), t \ge 0)$ is a mean zero martingale. Therefore, on $\{\tau_{j+1} < \zeta\}$, we have

$$\begin{aligned} e^{-sq_j(\tau_j'-\tau_j)}X_j'(\tau_j') &= e^{-sq_j(\tau_j'-\tau_j)-\int_{\tau_j'}^{\tau_{j+1}} G_j(v)\,dv} X_j'(\tau_{j+1}) - e^{-sq_j(\tau_j'-\tau_j)}Z_j^S(\tau_{j+1}) \\ &= e^{\int_{\tau_j}^{\tau_j'}(G_j(v)-sq_j)\,dv} e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv} X_j'(\tau_{j+1}) - e^{-sq_j(\tau_j'-\tau_j)}Z_j^S(\tau_{j+1}). \end{aligned} \tag{7.39}$$

By (4.5), on $\{\tau_{j+1} < \zeta\}$, we have $e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv}X_j'(\tau_{j+1}) \le C_1$. Also, by part 1 of Proposition 4.4, on $\{\tau_{j+1} < \zeta\}$, we have

$$\int_{\tau_j}^{\tau_j'} |G_j(v) - sq_j|\,dv \le C_3 s(\tau_j' - \tau_j), \tag{7.40}$$

which tends to zero as $N \to \infty$ by the argument in (6.35). Thus, (7.39) implies that for sufficiently large N, on $\{\tau_{j+1} < \zeta\}$, we have

$$\left|e^{-sq_j(\tau_j'-\tau_j)}X_j'(\tau_j') - e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv}X_j'(\tau_{j+1})\right| \le \frac{e^{-b}}{2} + e^{-sq_j(\tau_j'-\tau_j)}|Z_j^S(\tau_{j+1})|. \tag{7.41}$$

It remains to bound $|Z_j^S(\tau_{j+1})|$. By Corollary 4.9 and the argument leading to (4.19),

$$\mathrm{Var}(Z_j^S(\tau_j'+t)|\mathcal{F}_{\tau_j'}) \le 3E\left[\int_{\tau_j'}^{(\tau_j'+t)\wedge\tau_{j+1}\wedge\zeta} e^{-2\int_{\tau_j'}^{u} G_j(v)\,dv} X_j'(u)\,du \,\middle|\, \mathcal{F}_{\tau_j'}\right]. \tag{7.42}$$

Because $G_j(v) \ge s(q_j - C_3)$ for $v \in [\tau_j, \tau_{j+1}\wedge\zeta)$ by part 1 of Proposition 4.4, it follows from equations (7.38) and (7.42), Fubini's Theorem, and the fact that $(Z_j^S(\tau_j'+t), t \ge 0)$

is a mean zero martingale that for sufficiently large N,

$$\begin{aligned}\mathrm{Var}(Z_j^S(\tau_j'+t)|\mathcal{F}_{\tau_j'}) &\leq 3E\bigg[\int_{\tau_j'}^{(\tau_j'+t)\wedge\tau_{j+1}\wedge\zeta} e^{-\int_{\tau_j'}^u G_j(v)\,dv}(X_j'(\tau_j')+Z_j^S(t))\,du\bigg|\mathcal{F}_{\tau_j'}\bigg]\\ &\leq 3E\bigg[\int_{\tau_j'}^{(\tau_j'+t)\wedge\tau_{j+1}\wedge\zeta} e^{-s(q_j-C_3)(u-\tau_j')}(X_j'(\tau_j')+Z_j^S(t))\,du\bigg|\mathcal{F}_{\tau_j'}\bigg]\\ &\leq 3X_j'(\tau_j')\int_{\tau_j'}^{\infty} e^{-s(q_j-C_3)(u-\tau_j')}\,du\\ &\leq \frac{4X_j'(\tau_j')}{sq_j}.\end{aligned}$$

Therefore, by the L^2 Maximum Inequality for martingales,

$$P\bigg(\big|Z_j^S(\tau_{j+1})\big| > \frac{e^{-b}}{2}e^{sq_j(\tau_j'-\tau_j)}\bigg|\mathcal{F}_{\tau_j'}\bigg) \leq \frac{CX_j'(\tau_j')}{sq_je^{-2b}}e^{-2sq_j(\tau_j'-\tau_j)} = CX_j'(\tau_j')(sq_j)^5e^{2b}. \quad (7.43)$$

On $\{\kappa_j > \tau_j'-\tau_j\}$, we have $X_j'(\tau_j') \leq X_j^+(\tau_j'-\tau_j)$ by (7.13). Let $\mathcal{F}^*_{\tau_j'}$ be the σ-field generated by $\mathcal{F}_{\tau_j'}$ and the event $\{\kappa_j > \tau_j'-\tau_j\}$. Since the additional Poisson processes $N_{b,i,j}$, $N_{d,i,j}$, and $N_{m,j}$ and random variables $\beta_{\ell,j}$ are independent of the population process $(\mathbf{X}(t), t\geq 0)$, we have on $\{\kappa_j > \tau_j'-\tau_j\}$,

$$P\bigg(|Z_j^S(\tau_{j+1})| > \frac{e^{-b}}{2}e^{sq_j(\tau_j'-\tau_j)}\bigg|\mathcal{F}^*_{\tau_j'}\bigg) \leq CX_j^+(\tau_j'-\tau_j)(sk_N)^5e^{2b}.$$

Therefore, taking conditional expectations of both sides of (7.43) with respect to $\mathcal{F}_{\tau_j}$ and then using Lemma 7.3 and part 3 of Proposition 4.4, we get

$$\begin{aligned}P\bigg(\{\kappa_j > \tau_j'-\tau_j\}\cap\bigg\{\big|Z_j^S(\tau_{j+1})\big| > \frac{e^{-b}}{2}e^{sq_j(\tau_j'-\tau_j)}\bigg\}\bigg|\mathcal{F}_{\tau_j}\bigg)&\\ \leq CE[X_j^+(\tau_j'-\tau_j)|\mathcal{F}_{\tau_j}](sk_N)^5e^{2b}&\\ \leq \frac{C(sk_N)^2e^{2b}}{k_N}\log\bigg(\frac{1}{sk_N}\bigg).& \quad (7.44)\end{aligned}$$

Using Boole's Inequality and summing over $j\in I$, we now deduce from equations (7.41) and (7.44) and Lemmas 6.2 and 7.4 that

$$P\bigg(\Lambda\cap\bigcup_{j\in I}A_j\bigg) \leq 3Tk_N\cdot\frac{C(sk_N)^2e^{2b}}{k_N}\log\bigg(\frac{1}{sk_N}\bigg),$$

which tends to zero as $N\to\infty$ by assumption A3. □

7.5 The fraction of individuals descended from an early mutation

To determine the genealogy of the population, it will be important to consider the fraction of type j individuals in the population descended from an early type j mutation, as this is an estimate of the fraction of lineages that will coalesce near the time of this mutation. To this end, we let

$$Y_j := \frac{X_j'(\tau_{j+1})}{\lceil s/\mu\rceil}, \quad (7.45)$$

which is the fraction of type j individuals at time τ_{j+1} that are descended from a type j mutation that occurred between times τ_j and ξ_j. Also, define

$$Y_j^- := \frac{(e^{-sq_j(\tau_j'-\tau_j)}X_j^-(\tau_j'-\tau_j)-e^{-b})\vee 0}{((e^{-sq_j(\tau_j'-\tau_j)}X_j^-(\tau_j'-\tau_j)-e^{-b})\vee 0)+1+4\delta}$$

and

$$Y_j^+ := \frac{e^{-sq_j(\tau_j'-\tau_j)}X_j^+(\tau_j'-\tau_j)+e^{-b}}{e^{-sq_j(\tau_j'-\tau_j)}X_j^+(\tau_j'-\tau_j)+e^{-b}+1-4\delta}. \tag{7.46}$$

Lemma 7.8. *Suppose $j \in I$. For sufficiently large N, on $\{\tau_j < \zeta\}$, we have, for all $y \in [\delta, 1-\delta]$,*

$$\frac{(1-y)(1-13\delta)}{q_j y} \leq P(Y_j^- \geq y|\mathcal{H}_j) \leq P(Y_j^+ \geq y|\mathcal{H}_j) \leq \frac{(1-y)(1+13\delta)}{q_j y}. \tag{7.47}$$

Also, defining the event A_j as in (7.37), on $A_j^c \cap \{\tau_{j+1} < \zeta\} \cap \{\kappa_j > \tau_j' - \tau_j\}$ we have

$$Y_j^- \leq Y_j \leq Y_j^+. \tag{7.48}$$

Proof. We first prove (7.47). Suppose $y \in [\delta, 1-\delta]$. The middle inequality in (7.47) is immediate. To prove the third inequality in (7.47), note that $Y_j^+ \geq y$ if and only if

$$e^{-sq_j(\tau_j'-\tau_j)}X_j^+(\tau_j'-\tau_j) \geq \frac{(1+e^{-b}-4\delta)y-e^{-b}}{1-y}. \tag{7.49}$$

Since $e^{-b}/y \leq \delta$ by (4.1), we see that (7.49) implies

$$e^{-sq_j(\tau_j'-\tau_j)}X_j^+(\tau_j'-\tau_j) \geq \frac{(1-5\delta)y}{1-y}.$$

Thus, by Lemma 7.6, for sufficiently large N, on the event $\{\tau_j < \zeta\}$, we have for all $y \in [\delta, 1-\delta]$,

$$P(Y_j^+ \geq y|\mathcal{H}_j) \leq P\left(e^{-sq_j(\tau_j'-\tau_j)}X_j^+(\tau_j'-\tau_j) \geq \frac{(1-5\delta)y}{1-y}\middle|\mathcal{H}_j\right) \leq \frac{(1+7\delta)(1-y)}{(1-5\delta)q_j y},$$

which leads to the third inequality in (7.47). Likewise, note that $Y_j^- \geq y$ if and only if

$$e^{-sq_j(\tau_j'-\tau_j)}X_j^-(\tau_j'-\tau_j) \geq \frac{(1-e^{-b}+4\delta)y+e^{-b}}{1-y},$$

which, since $e^{-b}/y \leq \delta$, will always hold if $e^{-sq_j(\tau_j'-\tau_j)}X_j^+(\tau_j'-\tau_j) \geq (1+5\delta)y/(1-y)$. Therefore, by Lemma 7.6,

$$P(Y_j^- \geq y|\mathcal{H}_j) \geq P\left(e^{-sq_j(\tau_j'-\tau_j)}X_j^+(\tau_j'-\tau_j) \geq \frac{(1+5\delta)y}{1-y}\middle|\mathcal{H}_j\right) \geq \frac{(1-7\delta)(1-y)}{(1+5\delta)q_j y},$$

which implies the first inequality in (7.47). It remains to prove (7.48).

The last statement of part 1 of Proposition 4.1, combined with (4.9), implies that on the event $\{\tau_{j+1} < \zeta\}$, no individual that gets a jth mutation at or before time τ_j has a descendant alive at time τ_{j+1}. In particular, we have $X_j'(\tau_{j+1}) = X_{j,1}(\tau_{j+1})$. Therefore, using also that $X_{j,1}(\tau_{j+1}) + X_{j,2}(\tau_{j+1}) = X_j(\tau_{j+1}) = \lceil s/\mu \rceil$, we get, on $\{\tau_{j+1} < \zeta\}$,

$$\begin{aligned} Y_j &= \frac{X_{j,1}(\tau_{j+1})}{X_{j,1}(\tau_{j+1}) + X_{j,2}(\tau_{j+2})} \\ &= \frac{e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv}X_{j,1}(\tau_{j+1})}{e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv}X_{j,1}(\tau_{j+1}) + e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv}X_{j,2}(\tau_{j+2})}. \end{aligned} \tag{7.50}$$

By (4.6), on $\{\tau_{j+1} < \zeta\}$,

$$1-4\delta \leq e^{-\int_{\tau_j}^{\tau_{j+1}} G_j(v)\,dv}X_{j,2}(\tau_{j+1}) \leq 1+4\delta. \tag{7.51}$$

Combining (7.50), (7.51), and the definition of A_j, we get that on $A_j^c \cap \{\tau_{j+1} < \zeta\}$,

$$\frac{e^{-sq_j(\tau_j'-\tau_j)}X_j'(\tau_j') - e^{-b}}{e^{-sq_j(\tau_j'-\tau_j)}X_j'(\tau_j') - e^{-b} + 1 + 4\delta} \le Y_j \le \frac{e^{-sq_j(\tau_j'-\tau_j)}X_j'(\tau_j') + e^{-b}}{e^{-sq_j(\tau_j'-\tau_j)}X_j'(\tau_j') + e^{-b} + 1 - 4\delta}.$$

Combining this observation with (7.13) and noting that $Y_j \ge 0$, we conclude that (7.48) holds on $A_j^c \cap \{\tau_{j+1} < \zeta\} \cap \{\kappa_j > \tau_j' - \tau_j\}$. ☐

8 Coupling with the Bolthausen-Sznitman coalescent

In this section, we prove Theorem 2.1 by establishing a coupling between the coalescent process $(\Pi_N(u), 0 \le u \le t_0 + 1)$ and the Bolthausen-Sznitman coalescent. Our strategy will involve examining the process at the times τ_j. A very similar idea was used in [12] by Desai, Walczak, and Fisher.

8.1 No coalescence between times τ_L and $a_N T$

Recall from Remark 5.5 that with probability tending to one as $N \to \infty$, no lineages coalesce as they are traced back from time $a_N T$ to time τ_{L+10}. The result below shows that the lineages are also unlikely to coalesce as they are traced back further from time τ_{L+10} to time τ_L, which implies the statement (2.6) from Theorem 2.1. As with Lemmas 5.4 and 6.5, it is sufficient to state the result for the first two lineages.

Lemma 8.1. *We have*

$$\limsup_{N\to\infty} P\big(\Lambda \cap \{T_{1,2} \ge \tau_L\}\big) \le CTe^{-b}. \tag{8.1}$$

In particular, the statement (2.6) holds.

Proof. Let $\ell_1 = U_1(a_N T)$ and $\ell_2 = U_2(a_N T)$. Without loss of generality, suppose $\ell_1 \le \ell_2$. We know from the argument in Remark 5.5 that

$$\lim_{N\to\infty} P\big(\Lambda \cap \{T_{1,2} \ge \tau_{L+10}\}\big) = 0,$$

so we only need to follow these two lineages between times τ_L and τ_{L+10}. By Lemmas 5.2 and 6.4, we know that, outside of an event A such that $\lim_{N\to\infty} P(\Lambda \cap A) = 0$, for $i \in \{1, 2\}$ we have $U_i(\tau_{j+1}) = j$ for $j \in \{L-1, L, \dots, \ell_i\}$ and $U_i(\tau_{j+1}) = \ell_i$ for $j \in \{\ell_i, \dots, L+9\}$. When this occurs, there are only three ways that these lineages could coalesce between times τ_L and τ_{L+10}, in view of the fact that only lineages of the same type can coalesce:

1. We have $\ell_1 = \ell_2$ and $T_{1,2} \ge \tau_{\ell_1+1}$.
2. We have $\ell_1 < \ell_2$ and $\tau_{\ell_1+1} < T_{1,2} < V_{2,\ell_1+1} < \tau_{\ell_1+2}$. That is, as we trace back the ancestral lines, the second lineage gets traced back to a type ℓ_1 individual, then coalesces with the first lineage between times τ_{ℓ_1+1} and τ_{ℓ_1+2}.
3. For some $j \in \{L-1, L, \dots, \ell_1\}$, two type j lineages at time τ_{j+1} are descended from the same type $j-1$ lineage at time τ_j.

Lemma 5.4 bounds the probability of the first possibility above, while Lemma 6.6 bounds the probability of the second possibility. It remains only to consider the third possibility, in which the lineages coalesce between times τ_j and τ_{j+1} for $j \in \{L-1, L, \dots, \ell_i\}$. As noted in the discussion in subsection 6.3, Lemmas 6.5 and 6.6 establish that the probability that such a coalescence event occurs without the ancestor acquiring an early type j mutation is bounded above by CTe^{-b}. Also, because the result of Lemma 7.5 holds even when j is random provided that τ_j is a stopping time, we have

$$P(\Lambda \cap \{X_j^+(\tau_j' - \tau_j) > 0 \text{ for some } j \in \{L-1, L, \dots L+9\}) \le \frac{Ce^b}{k_N},$$

where we have used also part 3 of Proposition 4.4. In view of (7.13) and Lemma 7.4, it follows that the probability that, for some $j \in \{L-1, L, \dots, \ell_i\}$, two type j lineages at time τ_{j+1} are descended from an early type j mutation tends to zero as $N \to \infty$. The result (8.1) now follows from the bounds collected in this paragraph.

Finally, since $\tau_L < a_N(T-1)$ on Λ by (5.4) and (4.9), the statement (2.6) follows from (8.1), (4.1), and the fact that $\varepsilon > 0$ and $\delta > 0$ are arbitrary. □

8.2 Representing the early type j mutations by a point process

Fix $j \in I$. Recall from the discussion before Lemma 6.5 that the individuals sampled at time $a_N T$ are typically descended from type j individuals at time τ_{j+1}, and these lineages will typically coalesce only if they are traced back to one individual that acquires its jth mutation before time ξ_j. We construct in this subsection a point process that encodes these coalescence events.

Let Λ_j be the event that Λ occurs and that $U_i(\tau_j) = j+1$ for all $i \in \{1, \dots, n\}$. Suppose we condition on the event Λ_j, the random variables $Y_\ell = X'_\ell(\tau_{\ell+1})/\lceil s/\mu \rceil$ and τ_ℓ for $\ell \in I$, and the partitions $\Pi_N(T - \tau_\ell/a_N)$ for $\ell \in I$ with $\ell \geq j+1$. Denote the blocks of $\Pi_N(T - \tau_\ell/a_N)$ by $B_{\ell,1}, \dots, B_{\ell,n_\ell}$, where we rank the blocks in order by their smallest element. By the definition of Λ_j, the n_{j+1} individuals in the population at time τ_{j+1} that are ancestors of individuals in the sample are all among the $\lceil s/\mu \rceil$ type j individuals in the population at time τ_{j+1}. However, by the symmetry in the process, all $\lceil s/\mu \rceil(\lceil s/\mu \rceil - 1)\dots(\lceil s/\mu \rceil - n_{j+1} + 1)$ possible choices of n_{j+1} individuals out of these $\lceil s/\mu \rceil$ are equally likely to be the ancestors of the individuals in the sample corresponding to the integers in the blocks $B_{j+1,1}, \dots, B_{j+1,n_{j+1}}$ respectively. Also, $X'_j(\tau_{j+1})$ of the $\lceil s/\mu \rceil$ type j individuals at time τ_{j+1} are descended from an individual that got an early type j mutation between times τ_j and ξ_j. We call these type j individuals good.

We now construct some uniformly distributed random variables $Z_{i,j}$ for $i \in \{1, \dots, n\}$ and $j \in I$. Begin by defining random variables $Z^*_{i,j}$ for $i \in \{1, \dots, n\}$ and $j \in I$ which are uniformly distributed on $[0,1]$ and independent of the population process $(\mathbf{X}(t), t \geq 0)$ and of one another. If $j \geq L+1$, then let $Z_{i,j} = Z^*_{i,j}$. Likewise, if either Λ_j does not occur or $n_{j+1} < i \leq n$, then let $Z_{i,j} = Z^*_{i,j}$. Now suppose Λ_j occurs. For $i \in \{1, \dots, n_{j+1}\}$, we call the (i,j) ancestor the individual at time τ_{j+1} that is the ancestor of the individuals in the sample whose label is in the block $B_{j+1,i}$. Let $K_0 = 0$, and for $i \in \{1, \dots, n_{j+1} - 1\}$, let K_i be the number of integers $h \in \{1, \dots, i\}$ such that the (h,j) ancestor is good. Then, conditioning on K_{i-1} in addition to the event Λ_j, the random variables Y_ℓ and τ_ℓ for $\ell \in I$, and the partitions $\Pi_N(T - \tau_\ell/a_N)$ for $\ell \in I$ with $\ell \geq j+1$, the probability that the (i,j) ancestor is good is

$$P_{i,j} := \frac{X'_j(\tau_{j+1}) - K_{i-1}}{\lceil s/\mu \rceil - (i-1)}.$$

Let $Z_{i,j} := Z^*_{i,j} P_{i,j}$ if the (i,j) ancestor is good, and let $Z_{i,j} := P_{i,j} + Z^*_{i,j}(1 - P_{i,j})$ otherwise. Note that $Z_{i,j}$ has a uniform distribution on $[0,1]$, and the (i,j) ancestor is good if and only if $Z_{i,j} \leq P_{i,j}$. Also, the random variables $Z_{i,j}$ are jointly independent of the random variables Y_ℓ and the stopping times τ_ℓ for $\ell \in I$.

Let Φ_N be the point process on $[0, t_0 + 1] \times [0,1]^{n+1}$ consisting of all of the points

$$\left(T - \frac{\tau_j}{a_N}, Y_j, Z_{1,j}, \dots, Z_{n,j} \right)$$

such that $j \in I$, $j \leq L$, and $Y_j > 0$. We use the point process Φ_N to construct a coalescent process $(\Pi^*_N(u), 0 \leq u \leq t_0 + 1)$ as follows. Let $\Pi^*_N(0) = \{\{1\}, \dots, \{n\}\}$. For $u \in (0, t_0 + 1]$, suppose $(u, y, z_1, \dots, z_n)$ is a point of Φ_N and $\Pi^*_N(u-) = \pi$, where π is a partition of $\{1, \dots, n\}$ whose blocks, ordered by their smallest elements, are $B_1, \dots, B_\ell$.

Then $\Pi_N^*(u)$ is obtained from $\Pi_N^*(u-)$ by merging together all of the blocks B_i for which $z_i \leq y$. The result below relates the coalescent processes $(\Pi_N(u), 0 \leq u \leq t_0+1)$ and $(\Pi_N^*(u), 0 \leq u \leq t_0+1)$.

Lemma 8.2. *We have*

$$\liminf_{N \to \infty} P\left(\bigcap_{j \in I} \left\{\Pi_N\left(T - \frac{\tau_j}{a_N}\right) = \Pi_N^*\left(T - \frac{\tau_j}{a_N}\right)\right\}\right) \geq 1 - Cn^2\varepsilon. \tag{8.2}$$

Proof. We claim that the event in (8.2) could fail to hold in the following ways:

1. Either $\Pi_N(T - \tau_L/a_N) \neq \{\{1\}, \dots, \{n\}\}$ or $\Pi_N^*(T - \tau_L/a_N) \neq \{\{1\}, \dots, \{n\}\}$.
2. The event Λ_j could fail to hold for some $j \in I$ with $j \leq L$.
3. For some $j \in I$, either the event A_j' defined in the statement of Lemma 6.5 or the event A_j^* defined in the statement of Lemma 6.6 occurs.
4. For some $j \in I$, two or more individuals at time τ_j have descendants that got a jth mutation before time ξ_j and then have type j descendants in the population at time τ_{j+1}.
5. For some $j \in I$ with $j \leq L$ and $Y_j > 0$, and some $i \in \{1, \dots, n\}$, the random variable $Z_{i,j}$ is between $P_{i,j}$ and Y_j.

To see that these are the only possibilities, recall from the discussion at the beginning of subsection 6.3 that if Λ_ℓ occurs for all $\ell \in I$ with $\ell \leq L$, then unless A_j' or A_j^* occurs, the only way that lineages can coalesce between times τ_j and τ_{j+1} is for two or more lineages at time τ_{j+1} to be traced back to one individual that acquires its jth mutation before time ξ_j. Unless the fourth event listed above occurs, the only way this can happen is for a group of lineages at time τ_{j+1} to get traced back to the same individual that acquires its jth mutation before time ξ_j. In this case, suppose $\Pi_N(T - \tau_{j+1}/a_N) = \Pi_N^*(T - \tau_{j+1}/a_N) = \pi_{j+1}$, and $B_{j+1,1}, \dots, B_{j+1,n_{j+1}}$ are the blocks of π_{j+1}, ranked in order by their smallest elements. By the construction described at the beginning of this subsection, we obtain $\Pi_N(T - \tau_j/a_N)$ by merging the blocks $B_{j+1,i}$ for which $Z_{i,j} \leq P_{i,j}$. We obtain $\Pi_N^*(T - \tau_j/a_N)$ by merging the blocks $B_{j+1,i}$ for which $Z_{i,j} \leq Y_{i,j}$. Therefore, we can only have $\Pi_N(T - \tau_j/a_N) \neq \Pi_N^*(T - \tau_j/a_N)$ if the fifth event listed above occurs.

We thus need to bound the probabilities of the five events listed above. Recall that $P(\Lambda^c) < 2\varepsilon$ by (4.11). By construction, $(T - \tau_j/a_N, Y_j, Z_{1,j}, \dots, Z_{n,j})$ will only be a point of Φ_N if $j \leq L$, and $\tau_L < a_N(T-1)$ on Λ by (5.4) and (4.9). It follows that $\Pi_N^*(T - \tau_L/a_N) = \{\{1\}, \dots, \{n\}\}$ on Λ. Also, by Lemma 8.1, the probability that Λ occurs and $\Pi_N(T - \tau_L/a_N) \neq \{\{1\}, \dots, \{n\}\}$ is at most $Cn^2Te^{-b} \leq Cn^2\varepsilon$ in view of (4.1). By Lemma 6.4, the probability that Λ occurs and the second event above occurs tends to zero as $N \to \infty$. Lemmas 6.5 and 6.6 show that the probability that Λ occurs and the third event above occurs is at most $Cn^2Te^{-b} \leq Cn^2\varepsilon$. The probability that Λ occurs and the fourth event above occurs tends to zero as $N \to \infty$ by (7.17) along with (7.13), Lemma 7.4, and part 3 of Proposition 4.4.

It remains to bound the probability of the fifth event above. For sufficiently large N,

$$|P_{i,j} - Y_j| = \left|\frac{(i-1)X_j'(\tau_{j+1}) - K_{i-1}\lceil s/\mu\rceil}{\lceil s/\mu\rceil(\lceil s/\mu\rceil - (i-1))}\right| \leq \frac{n\lceil s/\mu\rceil}{\lceil s/\mu\rceil(\lceil s/\mu\rceil - (i-1))} \leq \frac{2n\mu}{s}.$$

Because $Z_{i,j}$ has a uniform distribution on $[0,1]$ and is independent of Y_j, the probability that $Z_{i,j}$ is between $P_{i,j}$ and Y_j is at most $2n\mu/s$. Therefore, using Lemma 6.2, the probability that this occurs for some $\ell \in \{1, \dots, n\}$ and $j \in I$ is at most $6n^2Tk_N\mu/s$, which tends to zero as $N \to \infty$ by (2.4) and assumption A2. The lemma follows. $\square$

8.3 A Poisson point process derived from Φ_N

In this subsection, we modify the point process Φ_N to obtain a Poisson point process Φ from which we can construct a Bolthausen-Sznitman coalescent via the technique outlined in subsection 3.1. The random variables $Z_{j,1}, \dots, Z_{j,n}$ are already independent and uniformly distributed on $[0,1]$, and they will remain unchanged. However, we will define new random variables Y_j^* that are coupled with the original random variables Y_j as well as new times T_j^*.

For $j \in I$, let Z_j be a random variable having the uniform distribution on $[0,1]$ that is independent of the population process. Recall the definition of the σ-field $\mathcal{H}_j$ from subsection 7.2. Define the random function

$$H_j(y,z) := P(Y_j^+ < y|\mathcal{H}_j) + zP(Y_j^+ = y|\mathcal{H}_j), \quad \text{for all } y,z \in [0,1].$$

Also, let $F_j(y) := P(Y_j^+ \leq y|\mathcal{H}_j) = H_j(y,1)$, and for $x \in [0,1]$, let $F_j^{-1}(x) := \sup\{y : F_j(y) \leq x\}$. Then it is easy to see that almost surely

$$Y_j^+ = F_j^{-1}(H_j(Y_j^+, Z_j)). \tag{8.3}$$

Note that if $0 < x < 1$, then there is a random integer $K(x)$ such that

$$P\biggl(Y_j^+ \leq \frac{K(x)}{\lceil s/\mu \rceil}\bigg|\mathcal{H}_j\biggr) \leq x < P\biggl(Y_j^+ \leq \frac{K(x)+1}{\lceil s/\mu \rceil}\bigg|\mathcal{H}_j\biggr).$$

Then

$$\begin{aligned} P(H_j(Y_j^+, Z_j) \leq x|\mathcal{H}_j) &= P\biggl(Y_j^+ \leq \frac{K(x)}{\lceil s/\mu \rceil}\bigg|\mathcal{H}_j\biggr) + P\biggl(Y_j^+ = \frac{K(x)+1}{\lceil s/\mu \rceil}\bigg|\mathcal{H}_j\biggr) \\ &\qquad \times P\biggl(Z_j \leq \frac{x - P(Y_j^+ \leq K(x)/\lceil s/\mu \rceil|\mathcal{H}_j)}{P(Y_j^+ = (K(x)+1)/\lceil s/\mu \rceil|\mathcal{H}_j)}\biggr) \\ &= x. \end{aligned}$$

Therefore, the conditional distribution of $H_j(Y_j^+, Z_j)$ given $\mathcal{H}_j$ is uniform on $[0,1]$. For $x \geq 0$, let

$$K_j(x) := \begin{cases} e^{-(\tau_{j+1}^* - \tau_j^*)(1-x)/a_N x} & \text{if } \varepsilon \leq x \leq 1 \\ e^{-(\tau_{j+1}^* - \tau_j^*)(1-\varepsilon)/a_N \varepsilon} & \text{if } 0 \leq x < \varepsilon \\ 0 & \text{if } x < 0 \end{cases}$$

For $x \in [0,1]$, let $K_j^{-1}(x) = \sup\{y : K_j(y) \leq x\}$. Also, let

$$Y_j^* := K_j^{-1}(H_j(Y_j^+, Z_j)). \tag{8.4}$$

Then for all $x \geq 0$, we have

$$P(Y_j^* \leq x|\mathcal{H}_j) = K_j(x). \tag{8.5}$$

Note that Y_j^* never takes a value between 0 and ε, so if $Y_j^* > 0$, then $Y_j^* \geq \varepsilon$.

We now continue with the construction of Φ. For all $j \in I$, independently of the population process $(\mathbf{X}(t), t \geq 0)$ and of all other auxiliary random variables introduced up to this point, let T_j^* be uniformly distributed on $[T - \tau_{j+1}^*/a_N, T - \tau_j^*/a_N]$, and let Φ_j' be a Poisson point process on $[T - \tau_{j+1}^*/a_N, T - \tau_j^*/a_N] \times [0,1]^{n+1}$ with intensity

$$du \times x^{-2}\, dx \times dz_1 \times \cdots \times dz_n.$$

For all j such that $T_j^* \in [1, t_0+1]$ and $Y_j^* > 0$, the point process Φ will include the point $(T_j^*, Y_j^*, Z_{j,1}, \dots, Z_{j,n})$. Also, for all j such that $Y_j^* > 0$, the point process Φ will include all points of Φ_j' whose first coordinate is in $[1, t_0+1]$ and whose second coordinate is in the interval (ε, Y_j^*). Finally, Φ will include all points of Φ_j' whose first coordinate is in $[1, t_0+1]$ and whose second coordinate is less than ε.

Lemma 8.3. *The point process Φ defined above is a Poisson point process on $[1, t_0+1] \times [0,1]^{n+1}$ with intensity*

$$du \times x^{-2}\, dx \times dz_1 \times \cdots \times dz_n. \tag{8.6}$$

Proof. We separately consider, for each j, the restriction of Φ to points whose first coordinate is in the interval $[T - \tau^*_{j+1}/a_N, T - \tau^*_j/a_N]$. For a Poisson point process with intensity (8.6), the expected number of points in the region $[T - \tau^*_{j+1}/a_N, T - \tau^*_j/a_N] \times [x, 1] \times [0,1]^n$ is

$$\left(\frac{\tau^*_{j+1} - \tau^*_j}{a_N}\right) \int_x^1 y^{-2}\, dy = \frac{(\tau^*_{j+1} - \tau^*_j)(1-x)}{a_N x}.$$

Therefore, from (8.5), we see that if $x \geq \varepsilon$, then $P(Y^*_j \geq x | \mathcal{H}_j)$ is the probability that there are no points in this region. Using also that T^*_j is uniformly distributed on $[T - \tau^*_{j+1}/a_N, T - \tau^*_j/a_N]$ and that the random variables $Z_{j,1}, \dots, Z_{j,n}$ are uniformly distributed on $[0,1]^n$, it follows that

$$(T^*_j, Y^*_j, Z_{j,1}, \dots, Z_{j,n})$$

has the same distribution as the point whose second coordinate is the largest among points of a Poisson process with intensity (8.6) restricted to $[T - \tau^*_{j+1}/a_N, T - \tau^*_j/a_N] \times [\varepsilon, 1] \times [0,1]^n$. Furthermore, conditional on the event that such a Poisson process has a point whose second coordinate is y and no point whose second coordinate is larger than y, the distribution of the restriction of the Poisson process to $[T - \tau^*_{j+1}/a_N, T - \tau^*_j/a_N] \times [\varepsilon, y) \times [0,1]^n$ is that of a Poisson process with intensity (8.6). It thus follows from the construction of Φ that the restriction of Φ to $[T - \tau^*_{j+1}/a_N, T - \tau^*_j/a_N]$ has intensity given by (8.6).

Finally, because of the conditioning on $\mathcal{H}_j$ in (8.5), the random variables Y^*_j for $j \in I$ are independent. Because the Poisson processes Φ'_j are independent, it follows that the restrictions of Φ to the intervals $[T - \tau^*_{j+1}/a_N, T - \tau^*_j/a_N]$ are independent. The lemma now follows from the superposition theorem for Poisson processes. □

The next step is to use the Poisson point process Φ to construct a coalescent process $(\Pi(u), 0 \leq u \leq t_0+1)$. Let $\Pi(u) = \{\{1\}, \dots, \{n\}\}$ for $u \in [0,1]$. For $u \in (1, t_0+1]$, suppose $(u, y, z_1, \dots, z_n)$ is a point of Φ and $\Pi(u-) = \pi$, where π is a partition of $\{1, \dots, n\}$ into the blocks $B_1, \dots, B_\ell$, ordered by their smallest element. Then $\Pi(u)$ is obtained from $\Pi(u-)$ by merging together all of the blocks B_i for which $z_i \leq y$. As discussed in subsection 3.1, this construction is well-defined, and the process $(\Pi(1+u), 0 \leq u \leq t_0)$ obeys the law of the Bolthausen-Sznitman coalescent.

8.4 Comparing Y_j and Y^*_j

The goal in this subsection is to prove two lemmas that establish that, with high probability, the random variables Y_j and Y^*_j are close. Lemma 8.6 bounds the probability that either Y_j or Y^*_j is greater than ε, but the other is not. Lemma 8.7 bounds the probability that the difference between Y_j and Y^*_j is more than ε^2. We will need a couple of preliminary estimates.

Lemma 8.4. *For $j \in I$, let*

$$A''_j := \left\{ \left| \frac{q_j}{k_N} - q\left(\frac{\tau_j}{a_N}\right) \right| > \delta \right\}.$$

Then

$$\lim_{N \to \infty} P\left(\Lambda \cap \bigcup_{j \in I} A''_j\right) = 0.$$

Proof. Lemma 6.2 and part 1 of Proposition 4.1 imply that on Λ, the fittest individual in the population at time τ_j must have either j or $j-1$ mutations. It therefore follows from (4.2) and (4.12), along with the fact that $\tau_j > a_N + 2a_N/k_N$ for all $j \in I$ by Lemma 6.2, that $Q(\tau_j)$ must either equal q_j or $q_j - 1$ on Λ for all $j \in I$. Let $S = [1 + (T - (t_0 + 2))/2, T]$, which is a compact subset of $(1, \infty)$. It follows from Proposition 4.5 that

$$\sup_{t \in S} \left| \frac{Q(a_N t)}{k_N} - q(t) \right| \to_p 0,$$

where $\to_p$ denotes convergence in probability as $N \to \infty$. By (6.17) and Lemma 6.2, on Λ we have $\tau_j/a_N \in S$ for all $j \in I$. Therefore,

$$\sup_{j \in I} \left| \frac{q_j}{k_N} - q\left(\frac{\tau_j}{a_N}\right) \right| \mathbb{1}_\Lambda \to_p 0,$$

which implies the lemma. □

Lemma 8.5. *There is a positive constant C such that if $\varepsilon \le y \le 1$ and $j \in I$, then on the event $\{\tau_j < \zeta\} \cap (A_j'')^c \in \mathcal{H}_j$, we have for sufficiently large N,*

$$\frac{(1-y)(1-C\delta T)}{q_j y} \le P(Y_j^* \ge y | \mathcal{H}_j) \le \frac{(1-y)(1+C\delta T)}{q_j y}.$$

Proof. By (6.1),

$$\left| \frac{\tau_{j+1}^* - \tau_j^*}{a_N} - \frac{1}{q_j} \right| = \left| \frac{1}{k_N q(\tau_j^*/a_N)} - \frac{1}{q_j} \right| = \frac{1}{k_N} \left| \frac{1}{q(\tau_j^*/a_N)} - \frac{k_N}{q_j} \right|. \tag{8.7}$$

Also, by (4.15) and (6.4), we have on $\{\tau_j < \zeta\} \cap (A_j'')^c$,

$$\left| \frac{q_j}{k_N} - q\left(\frac{\tau_j^*}{a_N}\right) \right| \le \delta + \left| q\left(\frac{\tau_j}{a_N}\right) - q\left(\frac{\tau_j^*}{a_N}\right) \right| \le \delta + 10e\delta T. \tag{8.8}$$

Therefore, using (8.7) and (8.8) along with the facts that $q(\tau_j^*/a_N) \ge 1$ by Proposition 4.6 and that $q_j/k_N \ge 1 - 2\delta$ on $\{\tau_j < \zeta\}$ by part 3 of Proposition 4.4, we get that on $\{\tau_j < \zeta\} \cap (A_j'')^c$,

$$\left| \frac{\tau_{j+1}^* - \tau_j^*}{a_N} - \frac{1}{q_j} \right| \le \frac{C\delta T}{k_N}.$$

Because $|(1 - e^{-x}) - x| \le x^2/2$ for $x \ge 0$ and (6.1) holds, it follows that when $\varepsilon \le y \le 1$, we have for sufficiently large N, on $\{\tau_j < \zeta\} \cap (A_j'')^c$,

$$\left| (1 - K_j(y)) - \frac{(1-y)}{q_j y} \right| \le \frac{1}{2} \left(\frac{(\tau_{j+1}^* - \tau_j^*)(1-y)}{a_N y} \right)^2 + \frac{1-y}{y} \left| \frac{\tau_{j+1}^* - \tau_j^*}{a_N} - \frac{1}{q_j} \right| \le \frac{1-y}{y} \cdot \frac{C\delta T}{k_N}.$$

Because $q_j \le (e + 2\delta)k_N$ on $\{\tau_j < \zeta\}$ by part 3 of Proposition 4.4, the result follows. □

Lemma 8.6. *Letting $\triangle$ denote the symmetric difference between two events, for sufficiently large N we have*

$$P\left(\Lambda \cap \bigcup_{j \in I} \left(\{Y_j \ge \varepsilon\} \triangle \{Y_j^* \ge \varepsilon\} \right) \right) \le \frac{C\delta T^2}{\varepsilon}.$$

Proof. By Lemmas 7.8 and 8.5 and part 3 of Proposition 4.4,

$$\left| P(Y_j^+ \ge \varepsilon | \mathcal{H}_j) - P\big(Y_j^* \ge \varepsilon | \mathcal{H}_j\big) \right| \le \frac{C\delta T}{\varepsilon k_N}$$

for sufficiently large N on $\{\tau_j < \zeta\} \cap (A''_j)^c$, and the same result holds with Y_j^- in place of Y_j^+. Because $Y_j^- \leq Y_j^+$, and the random variables Y_j^+ and Y_j^* are monotone functions of the same uniformly distributed random variable by (8.3) and (8.4), it follows that

$$P(\{Y_j^+ \geq \varepsilon\} \triangle \{Y_j^* \geq \varepsilon\} | \mathcal{H}_j) \leq \frac{C\delta T}{\varepsilon k_N}$$

on $\{\tau_j < \zeta\} \cap (A''_j)^c$, and the same result holds with Y_j^- in place of Y_j^+. Let

$$\Psi_j := A_j^c \cap \{\tau_{j+1} < \zeta\} \cap \{\kappa_j > \tau'_j - \tau_j\}.$$

By (7.48), we have

$$(\{Y_j^- \geq \varepsilon\} \cap \Psi_j) \subset (\{Y_j \geq \varepsilon\} \cap \Psi_j) \subset (\{Y_j^+ \geq \varepsilon\} \cap \Psi_j).$$

It follows that on $\{\tau_j < \zeta\} \cap (A''_j)^c$, we have

$$P\big((\{Y_j \geq \varepsilon\} \triangle \{Y_j^* \geq \varepsilon\}) \cap \Psi_j \big| \mathcal{H}_j\big) \leq \frac{C\delta T}{\varepsilon k_N}.$$

The result follows by taking expectations, summing over $j \in I$, and using Lemmas 7.4, 7.7, and 8.4, along with the fact that the cardinality of I is at most $3Tk_N$ by Lemma 6.2. ☐

Lemma 8.7. *There is a positive constant C^*, not depending on ε, δ, or T, such that for sufficiently large N, we have*

$$P\bigg(\Lambda \cap \bigcup_{j \in I} \big(\{|Y_j - Y_j^*| > C^* \varepsilon^2\} \cap \{Y_j \geq \varepsilon\} \cap \{Y_j^* \geq \varepsilon\}\big)\bigg) \leq \frac{C\delta T \log(1/\varepsilon)}{\varepsilon^2}.$$

Proof. We first compare Y_j^* to Y_j^+. In view of (8.3) and (8.4), we need to compare the functions F_j^{-1} and K_j^{-1}. Suppose $z \in (0,1)$. If $F_j^{-1}(1-z) \in [\delta, 1-\delta]$, then (7.47) implies that on $\{\tau_j < \zeta\}$, we have

$$\frac{1 - 13\delta}{q_j z + 1 - 13\delta} \leq F_j^{-1}(1-z) \leq \frac{1 + 13\delta}{q_j z + 1 + 13\delta}.$$

Likewise, Lemma 8.5 implies that if $K_j^{-1}(1-z) \geq \varepsilon$, then on $\{\tau_j < \zeta\} \cap (A''_j)^c$, we have

$$\frac{1 - C\delta}{q_j z + 1 - C\delta} \leq K_j^{-1}(1-z) \leq \frac{1 + C\delta}{q_j z + 1 + C\delta}.$$

It follows that on the event $\{\tau_j < \zeta\} \cap (A''_j)^c$, if $F_j^{-1}(1-z) \in [\delta, 1-\delta]$ and $K_j^{-1}(1-z) \in [\varepsilon, 1]$, then

$$|F_j^{-1}(1-z) - K_j^{-1}(1-z)| \leq C\delta. \tag{8.9}$$

Because F_j^{-1} and K_j^{-1} are increasing functions taking their values in $[0,1]$, and $\delta < \varepsilon$ by (5.1), we see that (8.9) holds on $\{\tau_j < \zeta\} \cap (A''_j)^c$ as long as $F_j^{-1}(1-z) \in [\varepsilon, 1]$ and $K_j^{-1}(1-z) \in [\varepsilon, 1]$. Since $\delta < \varepsilon^2$ by (5.1), it follows that there is a positive constant C^* such that on $\{\tau_j < \zeta\} \cap (A''_j)^c$, we have

$$|Y_j^+ - Y_j^*| \mathbb{1}_{\{Y_j^* \geq \varepsilon\}} \mathbb{1}_{\{Y_j^+ \geq \varepsilon\}} \leq (C^* - 1)\varepsilon^2. \tag{8.10}$$

It remains to control the difference between Y_j^+ and Y_j. By (7.47), on $\{\tau_j < \zeta\}$,

$$\begin{aligned}
&E[Y_j^+ \mathbb{1}_{\{Y_j^+ \geq \varepsilon\}} - Y_j^- \mathbb{1}_{\{Y_j^- \geq \varepsilon\}} | \mathcal{H}_j] \\
&\qquad = \int_0^1 \big(P(Y_j^+ \mathbb{1}_{\{Y_j^+ \geq \varepsilon\}} \geq y | \mathcal{H}_j) - P(Y_j^- \mathbb{1}_{\{Y_j^- \geq \varepsilon\}} \geq y | \mathcal{H}_j) \big)\, dy \\
&\qquad = \int_0^\varepsilon \big(P(Y_j^+ \geq \varepsilon | \mathcal{H}_j) - P(Y_j^- \geq \varepsilon | \mathcal{H}_j) \big)\, dy \\
&\qquad\qquad + \int_\varepsilon^1 \big(P(Y_j^+ \geq y | \mathcal{H}_j) - P(Y_j^- \geq y | \mathcal{H}_j) \big)\, dy \\
&\qquad \leq \varepsilon \cdot \frac{(1-\varepsilon)C\delta}{q_j \varepsilon} + \int_\varepsilon^{1-\delta} \frac{C\delta(1-y)}{q_j y}\, dy + \delta \cdot \frac{\delta(1+C\delta)}{q_j(1-\delta)} \\
&\qquad \leq \frac{C\delta \log(1/\varepsilon)}{q_j}.
\end{aligned}$$

Let $\Psi_j = A_j^c \cap \{\tau_{j+1} < \zeta\} \cap \{\kappa_j > \tau_j' - \tau_j\}$. Because $Y_j^- \leq Y_j \leq Y_j^+$ on Ψ_j, by (7.48),

$$E\big[(Y_j^+ \mathbb{1}_{\{Y_j^+ \geq \varepsilon\}} - Y_j \mathbb{1}_{\{Y_j \geq \varepsilon\}}) \mathbb{1}_{\Psi_j} | \mathcal{H}_j\big] \leq \frac{C\delta \log(1/\varepsilon)}{q_j}.$$

Now Markov's Inequality implies that

$$P\big(\{|Y_j \mathbb{1}_{\{Y_j \geq \varepsilon\}} - Y_j^+ \mathbb{1}_{\{Y_j^+ \geq \varepsilon\}}| > \varepsilon^2\} \cap \Psi_j \big| \mathcal{H}_j\big) \leq \frac{C\delta \log(1/\varepsilon)}{q_j \varepsilon^2}.$$

Combining this result with (8.10) and part 3 of Lemma 4.4 gives, for sufficiently large N,

$$P\big(\{|Y_j - Y_j^*| > C^* \varepsilon^2\} \cap \{Y_j \geq \varepsilon\} \cap \{Y_j^* \geq \varepsilon\} \cap \Psi_j \cap (A_j'')^c \cap \Lambda\big) \leq \frac{C\delta \log(1/\varepsilon)}{k_N \varepsilon^2}.$$

The result follows by summing over j and using Lemmas 7.4, 7.7, and 8.4. □

8.5 Small coalescence events

Lemma 8.8 below shows that it is unlikely that lineages will coalesce between times τ_j and τ_{j+1} if $Y_j \leq \varepsilon$.

Lemma 8.8. *For sufficiently large N, we have*

$$P\bigg(\Lambda \cap \bigcup_{j \in I} \bigg(\bigg\{ \Pi_N^*\Big(T - \frac{\tau_j}{a_N}\Big) \neq \Pi_N^*\Big(T - \frac{\tau_{j+1}}{a_N}\Big) \bigg\} \cap \{Y_j \leq \varepsilon\} \bigg)\bigg) \leq CTn^2 \varepsilon.$$

Proof. Suppose $j \in I$. Let $\Psi_j = A_j^c \cap \{\tau_{j+1} < \zeta\} \cap \{\kappa_j > \tau_j' - \tau_j\}$, where A_j is the event defined in Lemma 7.7 and κ_j is defined in (7.14). Define the σ-field $\mathcal{H}_j$ as in subsection 7.2. Let $\mathcal{G}_j$ be the σ-field generated by the σ-field $\mathcal{H}_j$, the random variable Y_j defined in (7.45), and the event Ψ_j. Conditional on $\mathcal{G}_j$, the probability that at least two of the random variables $Z_{1,j}, \dots, Z_{n,j}$ are less than or equal to Y_j is at most $\binom{n}{2} Y_j^2$. Therefore, on $\{\tau_j < \zeta\}$, we have

$$P\bigg(\bigg\{ \Pi_N^*\Big(T - \frac{\tau_j}{a_N}\Big) \neq \Pi_N^*\Big(T - \frac{\tau_{j+1}}{a_N}\Big) \bigg\} \cap \{Y_j \leq \varepsilon\} \cap \Psi_j \bigg| \mathcal{G}_j \bigg) \leq \binom{n}{2} Y_j^2 \mathbb{1}_{\{Y_j \leq \varepsilon\}} \mathbb{1}_{\Psi_j}.$$

Now take conditional expectations of both sides with respect to $\mathcal{H}_j$ to get that on $\{\tau_j < \zeta\}$,

$$\begin{aligned}
&P\bigg(\bigg\{ \Pi_N^*\Big(T - \frac{\tau_j}{a_N}\Big) \neq \Pi_N^*\Big(T - \frac{\tau_{j+1}}{a_N}\Big) \bigg\} \cap \{Y_j \leq \varepsilon\} \cap \Psi_j \bigg| \mathcal{H}_j \bigg) \\
&\qquad\qquad\qquad \leq \binom{n}{2} E\big[Y_j^2 \mathbb{1}_{\{Y_j \leq \varepsilon\}} \mathbb{1}_{\Psi_j} \big| \mathcal{H}_j\big]. \qquad (8.11)
\end{aligned}$$

Recall that for any nonnegative random variable X, we have $E[X^2] = \int_0^\infty 2xP(X \geq x)\,dx$. Therefore, on $\{\tau_j < \zeta\}$,

$$\begin{aligned} E\big[Y_j^2 \mathbb{1}_{\{Y_j \leq \varepsilon\}} \mathbb{1}_{\Psi_j} \big| \mathcal{H}_j\big] &= \int_0^\infty 2xP\big(Y_j \mathbb{1}_{\{Y_j \leq \varepsilon\}} \mathbb{1}_{\Psi_j} > x \big| \mathcal{H}_j\big)\,dx \\ &\leq \int_0^\varepsilon 2xP\big(Y_j \mathbb{1}_{\Psi_j} > x \big| \mathcal{H}_j\big)\,dx. \end{aligned} \tag{8.12}$$

Recall from (7.48) that $Y_j \leq Y_j^+$ on Ψ_j. Also, from (7.13), we see that on Ψ_j, if $Y_j > 0$ then $X_j^+(\tau_j' - \tau_j) > 0$, and on $\{\tau_j < \zeta\}$, we have $q_j \geq (1-2\delta)k_N$ by part 3 of Proposition 4.4. Therefore, by Lemma 7.5,

$$P(Y_j \mathbb{1}_{\Psi_j} > 0 | \mathcal{H}_j) \leq \frac{Ce^b}{k_N}. \tag{8.13}$$

Also, on Ψ_j, if $Y_j > x$ and $3e^{-b} \leq x \leq \varepsilon$, it follows from (7.46) that if ε is sufficiently small, then

$$e^{-sq_j(\tau_j' - \tau_j)} X_j^+(\tau_j' - \tau_j) \geq \frac{(e^{-b} + 1 - 4\delta)x - e^{-b}}{1-x} \geq \frac{x}{2}.$$

Therefore, Lemma 7.6 implies that if ε is sufficiently small and N is sufficiently large, and if $3e^{-b} \leq x \leq \varepsilon$, then

$$P(Y_j \mathbb{1}_{\Psi_j} > x | \mathcal{H}_j) \leq \frac{C}{k_N x}. \tag{8.14}$$

Dividing the integral on the right-hand side of (8.12) into two pieces and using (8.13) to estimate the first piece and (8.14) to estimate the second piece, we get

$$\begin{aligned} E\big[Y_j^2 \mathbb{1}_{\{Y_j \leq \varepsilon\}} \mathbb{1}_{\Psi_j} \big| \mathcal{H}_j\big] &\leq \int_0^{3e^{-b}} 2x \cdot \frac{Ce^b}{k_N}\,dx + \int_{3e^{-b}}^\varepsilon 2x \cdot \frac{C}{k_N x}\,dx \\ &\leq \frac{Ce^{-b}}{k_N} + \frac{C\varepsilon}{k_N} \\ &\leq \frac{C\varepsilon}{k_N}. \end{aligned} \tag{8.15}$$

Using (8.15) to bound the right-hand side of (8.11) and then taking expectations, we get

$$P\bigg(\bigg\{\Pi_N^*\bigg(T - \frac{\tau_j}{a_N}\bigg) \neq \Pi_N^*\bigg(T - \frac{\tau_{j+1}}{a_N}\bigg)\bigg\} \cap \{Y_j \leq \varepsilon\} \cap \Psi_j\bigg) \leq \frac{Cn^2\varepsilon}{k_N}. \tag{8.16}$$

The result now follows by summing over j and using Lemmas 7.4 and 7.7. □

8.6 Completion of the coupling argument

Fix a positive integer d and times $0 < t_1 < \dots < t_d \leq t_0$. Recall that equation (2.6) was established as part of Lemma 8.1. Therefore, to prove Theorem 2.1, we need to show that the joint distribution of $(\Pi_N(1+t_1), \dots, \Pi_N(1+t_d))$ converges as $N \to \infty$ to the joint distribution of $(\Pi(1+t_1), \dots, \Pi(1+t_d))$, where $(\Pi(u), 0 \leq u \leq t_0 + 1)$ is the coalescent process derived from the Poisson point process Φ at the end of subsection 8.3.

Proof of Theorem 2.1. The key to the proof will be to show that with high probability, we have

$$\Pi\bigg(T - \frac{\tau_j^*}{a_N}\bigg) = \Pi_N^*\bigg(T - \frac{\tau_j}{a_N}\bigg) \quad \text{for all } j \in I \text{ with } j \leq L. \tag{8.17}$$

Recall that the coalescent process Π_N^* was constructed from the point process Φ_N in the same way that Π was constructed from Φ. Therefore, we simply need to compare the two constructions. If (8.17) fails to hold, then one of the following must occur:

1. Either $\Pi_N^*(T - \tau_L/a_N) \neq \{\{1\}, \dots, \{n\}\}$ or $\Pi(T - \tau_L^*/a_N) \neq \{\{1\}, \dots, \{n\}\}$.
2. For some $j \in I$, we have either $Y_j \geq \varepsilon$ and $Y_j^* < \varepsilon$, or $Y_j < \varepsilon$ and $Y_j^* \geq \varepsilon$.
3. For some $j \in I$, we have $\Pi_N^*(T - \tau_j/a_N) \neq \Pi_N^*(T - \tau_{j+1}/a_N)$ and $Y_j < \varepsilon$.
4. For some $u \in [1, t_0 + 1]$, we have $\Pi(u) \neq \Pi(u-)$ but u does not equal T_j^* for any j.
5. For some $j \in I$ with $j \leq L$, we have $Y_j \geq \varepsilon$, $Y_j^* \geq \varepsilon$, and $\Pi_N^*((T - \tau_j/a_N)-) = \Pi(T_j^*-)$, but $\Pi_N^*(T - \tau_j/a_N) \neq \Pi(T_j^*)$.

We now bound the probabilities of these five events. As for the first event, note that (8.1) and (4.11) imply that $P(\Pi_N(T - \tau_L/a_N) \neq \{\{1\}, \dots, \{n\}\}) \leq C\varepsilon + CTn^2 e^{-b}$. Combining this result with Lemma 8.2 and (4.1) gives

$$P\bigg(\Pi_N^*\bigg(T - \frac{\tau_L}{a_N}\bigg) \neq \{\{1\}, \dots, \{n\}\}\bigg) \leq Cn^2\varepsilon.$$

By (5.4), we have $T - \tau_L/a_N \leq 1 + 3/k_N$, so (6.4) implies $T - \tau_L^*/a_N \leq 1 + 3/k_N + 10\delta T$. Because each pair of lineages in the Bolthausen-Sznitman coalescent merges at rate 1, it follows that for sufficiently large N,

$$P\bigg(\Pi\bigg(T - \frac{\tau_L^*}{a_N}\bigg) \neq \{\{1\}, \dots, \{n\}\}\bigg) \leq \binom{n}{2}\bigg(\frac{3}{k_N} + 10\delta T\bigg) \leq Cn^2\delta T.$$

It follows from Lemma 8.6, along with (4.11) and the fact that $\delta < \varepsilon^2$ by (5.1), that the probability that the second of the five events above occurs is at most $C\varepsilon T^2$. Likewise, it follows from Lemma 8.8 and (4.11) that the probability that the third of the five events occurs is bounded above by $CTn^2\varepsilon$.

Consider next the fourth event listed above. From the construction, this can only happen either if, for some $j \in I$, there are two points of Φ in $[T - \tau_{j+1}^*/a_N, T - \tau_j^*/a_N] \times [\varepsilon, 1] \times [0,1]^n$, or if there is some point $(u, y, z_1, \dots, z_n)$ in Φ in which $y \leq \varepsilon$ but two of the points $z_1, \dots, z_n$ are less than or equal to y. Recall that if X has the Poisson distribution with mean λ, then $P(X \geq 2) \leq \lambda^2$. Therefore, using also (6.2), the probability that, for some $j \in I$, there are two points of Φ in $[T - \tau_{j+1}^*/a_N, T - \tau_j^*/a_N] \times [\varepsilon, 1] \times [0,1]^n$ is bounded above by

$$\sum_{j \in I} \bigg(\frac{\tau_{j+1}^* - \tau_j^*}{a_N} \cdot \frac{1 - \varepsilon}{\varepsilon}\bigg)^2 \leq \sum_{j \in I} \frac{1}{(\varepsilon k_N)^2} \leq \frac{CT}{\varepsilon^2 k_N},$$

which tends to zero as $N \to \infty$. Note that if y is the second coordinate of a point in Φ, the probability that two of the points $z_1, \dots, z_n$ are less than or equal to y is at most $\binom{n}{2}y^2$. Therefore, the probability that there is a point $(u, y, z_1, \dots, z_n)$ in Φ in which $y \leq \varepsilon$ but two of the points $z_1, \dots, z_n$ are less than or equal to y is bounded above by

$$t_0 \int_0^\varepsilon y^{-2} \cdot \binom{n}{2} y^2 \, dy = \binom{n}{2} t_0 \varepsilon \leq CTn^2\varepsilon.$$

Finally, consider the fifth of the possibilities above, which means that the coalescence at time $T - \tau_j/a_N$ in the process Π_N^* does not match the coalescence that occurs at time T_j^* in the process Π. One way this could happen would be if the time interval $[T - \tau_{j+1}^*/a_N, T - \tau_j^*/a_N]$ is not entirely contained in the interval $[1, t_0 + 1]$. By (6.2) and (6.15), the number of $j \in I$ for which this interval is not contained in $[1, t_0 + 1]$ is at most $C\delta T k_N$. By Lemmas 7.4, 7.7, and 7.8, along with (4.11) and part 3 of Proposition 4.4, the probability that $Y_j > \varepsilon$ for some such j is at most

$$C\delta T k_N \cdot \frac{(1 - \varepsilon)(1 + 13\delta)}{(1 - 2\delta)k_N\varepsilon} + 2\varepsilon \leq \frac{C\delta T}{\varepsilon} + C\varepsilon.$$

The other way that the coalescence at time $T - \tau_j/a_N$ in the process Π_N^* might not match the coalescence that occurs at time T_j^* in the process Π would be if one of the random variables $Z_{j,1}, \dots, Z_{j,n}$ is between Y_j and Y_j^*. By Lemma 8.7, the probability that this happens when $|Y_j - Y_j^*| > \varepsilon^2$ is bounded above by

$$\frac{C\delta T \log(1/\varepsilon)}{\varepsilon^2}.$$

Using Lemmas 7.4, 7.7, and 7.8, we see that the probability that this happens when $|Y_j - Y_j^*| \leq \varepsilon^2$ is at most

$$\sum_{j \in I} \frac{C}{k_N \varepsilon} \cdot n\varepsilon^2 \leq CTn\varepsilon.$$

Combining the bounds obtained above, we see that for sufficiently large N, the probability that (8.17) fails to hold is bounded above by

$$CTn^2\varepsilon + Cn^2\delta T + C\varepsilon T^2 + \frac{C\delta T \log(1/\varepsilon)}{\varepsilon^2}. \tag{8.18}$$

By Lemma 8.2, we can replace Π_N^* by Π_N in (8.17) and conclude that the probability that

$$\Pi\left(T - \frac{\tau_j^*}{a_N}\right) = \Pi_N\left(T - \frac{\tau_j}{a_N}\right) \quad \text{for all } j \in I \text{ with } j \leq L \tag{8.19}$$

fails to hold is also bounded above by the expression in (8.18) for sufficiently large N.

Now suppose that indeed (8.19) holds and Λ occurs. Fix $i \in \{1, \dots, d\}$. Then there exists $j \in I$ such that $T - \tau_{j+1}/a_N \leq t_i < T - \tau_j/a_N$. By (4.9) and (6.4), for sufficiently large N,

$$T - \frac{\tau_j^*}{a_N} \leq t_i + \frac{2}{k_N} + 10\delta T \leq t_i + 11\delta T$$

and

$$T - \frac{\tau_{j+1}^*}{a_N} \geq t_i - \frac{2}{k_N} - 10\delta T \geq t_i - 11\delta T.$$

Thus, as long as $\Pi(t_i - 11\delta T) = \Pi(t_i + 11\delta T)$ and (8.19) holds, we must have $\Pi(t_i) = \Pi_N(t_i)$. However, because each pair of lineages in the Bolthausen-Sznitman coalescent merges at rate one, we have

$$P\big(\Pi(t_i - 11\delta T) \neq \Pi(t_i + 11\delta T)\big) \leq \binom{n}{2} \cdot 22\delta T.$$

Taking the union over $i \in \{1, \dots, d\}$ and using (8.18), it follows that for sufficiently large N,

$$P(\Pi_N(t_i) \neq \Pi(t_i) \text{ for some } i \in \{1, \dots, d\}) \leq CTn^2\varepsilon + Cdn^2\delta T + C\varepsilon T^2 + \frac{C\delta T \log(1/\varepsilon)}{\varepsilon^2}.$$

Since $\delta < \varepsilon^3$ by (5.1) and $\varepsilon > 0$ can be chosen arbitrarily small for any fixed T, the theorem follows. ☐

References

[1] Athreya, K. B. and Ney, P. E.: Branching Processes. *Springer-Verlag*, Berlin, 1972. xi+287 pp. MR-0373040

[2] Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K. W., Velculescu, V. E., Vogelstein, B., and Nowak, M. A.: Genetic progression and the waiting time to cancer. *PLoS Comput. Biol.* **3**, (2007), 2239–2246. MR-2369267

[3] Bérard, J. and Gouéré, J.-B.: Brunet-Derrida behavior of branching-selection particle systems on the line. *Comm. Math. Phys.* **298**, (2010), 323–342. MR-2669438

[4] Berestycki, J., Berestycki, N., and Schweinsberg, J.: The genealogy of branching Brownian motion with absorption. *Ann. Probab.* **41**, (2013), 527–618. MR-3077519

[5] Berestycki, J., Berestycki, N., and Schweinsberg, J.: Critical branching Brownian motion with absorption: particle configurations. *Ann. Inst. H. Poincaré Probab. Statist.* **51**, (2015), 1215–1250. MR-3414446

[6] Bolthausen, E. and Sznitman, A.-S.: On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197**, (1998), 247–276. MR-1652734

[7] Brunet, É. and Derrida, B.: Shift in the velocity of a front due to a cutoff. *Phys. Rev. E* **56**, (1997), 2597–2604. MR-1473413

[8] Brunet, É., Derrida, B., Mueller, A. H., and Munier, S.: Noisy traveling waves: effect of selection on genealogies. *Europhys. Lett.* **76**, (2006), 1–7. MR-2299937

[9] Brunet, É., Derrida, B., Mueller, A. H., and Munier, S.: Effect of selection on ancestry: an exactly soluble case and its phenomenological generalization. *Phys. Rev. E* **76**, (2007), 041104. MR-2365627

[10] Brunet, É., Rouzine, I. M., and Wilke, C. O.: The stochastic edge in adaptive evolution. *Genetics* **179**, (2008), 603–620.

[11] Desai, M. M. and Fisher, D. S.: Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**, (2007), 1759–1798.

[12] Desai, M. M., Walczak, A. M., and Fisher, D. S.: Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics* **193**, (2013), 565–585.

[13] Durrett, R., Foo, J., Leder, K., Mayberry, J., and Michor, F.: Intratumor heterogeneity in evolutionary models of tumor progression. *Genetics* **188**, (2011), 461–477.

[14] Durrett, R. and Mayberry, J.: Traveling waves of selective sweeps. *Ann. Appl. Probab.* **21**, (2011), 699–744. MR-2807971

[15] Durrett, R. and Moseley, S.: Evolution of resistance and progression to disease during clonal expansion of cancer. *Theo. Pop. Biol.* **77**, (2010), 42–48.

[16] Durrett, R. and Schweinsberg, J.: A coalescent model for the effect of advantageous mutations on the genealogy of a population. *Stochastic Process. Appl.* **115**, (2005), 1628–1657. MR-2165337

[17] Kingman, J. F. C.: The coalescent. *Stochastic Process. Appl.* **13**, (1982), 235–248. MR-0671034

[18] Leviyang, S.: The coalescence of intrahost HIV lineages under symmetric CTL attack. *Bull. Math. Biol.* **74**, (2012), 1818–1856. MR-2949979

[19] Maillard, P.: Speed and fluctuations of N-particle branching Brownian motion with spatial selection. *Probab. Theory Relat. Fields* **166**, (2016), 1061–1173. MR-3568046

[20] Moran, P. A. P.: Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54**, (1958), 60–71. MR-0127989

[21] Mueller, C., Mytnik, L., and Quastel, J.: Effect of noise on front propagation in reaction-diffusion equations of KPP type. *Invent. Math.* **184**, (2011), 405–453. MR-2793860

[22] Neher, R. A. and Hallatschek, O.: Genealogies of rapidly adapting populations. *Proc. Natl. Acad. Sci.* **110**, (2013), 437–442.

[23] Pitman, J.: Coalescents with multiple collisions. *Ann. Probab.* **27**, (1999), 1870–1902. MR-1742892

[24] Pitman, J. and Yor, M.: The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, (1997), 855–900. MR-1434129

[25] Rouzine, I. M., Brunet, É., and Wilke, C. O.: The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theor. Pop. Biol* **73**, (2008), 24–46.

[26] Rouzine, I. M., Wakeley, J., and Coffin, J. M.: The solitary wave of asexual evolution. *Proc. Natl. Acad. Sci.* **100**, (2003), 587–592.

[27] Sagitov, S.: The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, (1999), 1116–1125. MR-1742154

[28] Schweinsberg, J.: Rigorous results for a population model with selection I: evolution of the fitness distribution. *Electron. J. Probab.* **22**, (2017), 1–94.

[29] Yu, F., Etheridge, A., and Cuthbertson, C.: Asymptotic behavior of the rate of adaptation. *Ann. Appl. Probab.* **20**, (2010), 978–1004. MR-2680555

Acknowledgments. The author thanks two referees for helpful comments.