Bulletin of Mathematical Biology (2009) 71: 1873–1889
DOI 10.1007/s11538-009-9428-4

ORIGINAL ARTICLE

The Mutation Process in Colored Coalescent Theory

Jianjun Paul Tian[a,*], Xiao-Song Lin[b]

[a] *Mathematics Department, The College of William and Mary, Williamsburg, VA 23187, USA*
[b] *Department of Mathematics, University of California, Riverside, CA 92521, USA*

Received: 18 December 2007 / Accepted: 7 May 2009 / Published online: 22 May 2009

Abstract The mutation process is introduced into the colored coalescent theory. The mutation process can be viewed as an independent Poisson process running on the colored genealogical random tree generated by the colored coalescent process, with the edge lengths of the random tree serving as the time scale for the mutation process. Moving backward along the colored genealogical tree, the color of vertices may change in two ways, when two vertices coalesce, or when a mutation happens. The rule that governs the coalescent change of color involves a parameter x; the rule that governs the mutation involves a parameter μ. Explicit computations of the expectation of the coalescent time (the first hitting time), and the coalescent probabilities (the first hitting probabilities) are carried out. For example, our calculation shows that when $x = 1/2$, for a sample of n colored individuals, the expected time for the colored coalescent process with the mutation process superimposed to first reach a black MRCA or a white MRCA, respectively, is $3 - 2/n$ with probability $1/2$ for any value of the parameter μ. On the other hand, the expected time for the colored coalescent process with mutation to first reach a MRCA, either black or white, is $2 - 2/n$ for any values of the parameters μ and x, which is the same as that for the standard Kingman coalescent process.

Keywords Coalescent theory · Colored coalescent process · Mutation process · Lumped process · Colored random tree

1. Introduction

Coalescent theory is a retrospective model of population genetics that traces all alleles of a gene in a sample from a population to a single ancestral allele shared by all members of the population. First formulated in the seminal work of Kingman in 1982 (Kingman, 1982a, 1982b), coalescent theory offers various sample-based and highly efficient statistical methods for analyzing molecular data such as DNA sequence samples. For recent reviews as well as extensive references of coalescent theory, see Fu and Li (1999) and Rosenberg and Nordborg (2002). A nice introduction to coalescent theory can be found

*Corresponding author.
E-mail address: jptian@math.wm.edu (Jianjun Paul Tian).

in Nordborg (2001), Hein et al. (2005) and Rice (2004). In order to understand the geographical models of the origin of the human being, we introduced a colored coalescent theory in Tian and Lin (2005). The colored coalescent process generates random colored trees. Here a *coloring* of a tree is to color the vertices of the tree by two colors, black (B) and white (W), such that if two vertices are joint by an edge, they may have different colors only when the vertex closer to the root is a branching point of the tree. Thus, in recovering a random colored tree using colored coalescent process, we may end up at a colored tree with the root colored black or colored white.

In present paper, we consider a color-changing mutation process in the colored coalescent theory. The quantities which we are interested in including the following: the probabilities for the mutation superimposed colored coalescent process to first reach a black or white root, respectively; the mean coalescent time, which is the time elapsed before the coalescent process first reaches a black root or a white root, respectively; etc.

We now describe the colored coalescent process and the mutation process in some details so that one may appreciate its motivation and understand the statement of our main results.

The Wright–Fisher model in population genetics assumes discrete, non-overlapping generations $G_0, G_1, G_2, \ldots$ in which each generation contains a fixed number N of individuals. In a haploid population, each member in G_{i+1} is the child of exactly one member in G_i, but the number of children born to the jth member of G_i is a random variable ν_j satisfying the symmetric multinomial distribution

$$\Pr\{\nu_j = n_j\,;\ j = 1, 2, \ldots, N\} = \frac{N!}{n_1!n_2!\cdots n_N!N^N}.$$

In colored coalescent theory, it is assumed that each individual in a generation has two possible colors B and W. In the next generation, if a member is the only child of its parent, then this child will inherit the color of its parent. But when a parent has more than one child in the next generation, the color of children of that common parent satisfies a binomial distribution. More specifically, for a parent with k children in the next generation, $k > 1$, let b be the number of children with B color and w be the number of children with W color (so that $b + w = k$), we have

$$\begin{aligned}
\Pr\{b = k_1, w = k_2\,;\ \text{the parent has color } B\} &= \binom{k}{k_1} p^{k_1}(1-p)^{k_2}, \\
\Pr\{b = k_1, w = k_2\,;\ \text{the parent has color } W\} &= \binom{k}{k_1} (1-q)^{k_1} q^{k_2}
\end{aligned} \tag{1}$$

where $0 \le p, q \le 1$. Following the same argument as in Kingman (1982a, 1982b), there is a limiting coalescent process for a sample of n colored individuals when $N \to \infty$, which is called the colored coalescent process $Z(t)$. In this limiting coalescent process, one is only allowed to have two individuals in the sample to coalesce. When two colored individuals coalesce, the probability of the color of their common parent can be calculated according to (1). The details can be found in Tian and Lin (2005). The mutation process can be viewed as an independent Poisson process running on the random tree generated by this colored coalescent process, with the edge lengths of the random tree serving as the time scale for the mutation process. In particular, we consider that the mutation is symmetrical, that is, the probability that an individual mutates its color from B to W and the probability that an individual mutates its color from W to B is the same. It is assumed that the probability that an individual mutates its color from one to another in a unit time Δt is $\mu \Delta t$. The mutation processes of different individuals are considered to be independent of each other. Therefore, the probability that k of the n individuals within a sample mutate their colors can be calculated by

$$\binom{n}{k}(\mu \Delta t)^k (1-\mu \Delta t)^{n-k}. \tag{2}$$

The stochastic character of a state of the mutation superimposed colored coalescent process, denoted by $M(t)$, with the parameter x and μ turns out to depend only on the number of individuals in this state colored by one color, say B. If we start the colored coalescent process with a sample of n colored individuals, the initial state can be denoted by a pair of non-negative integers (n_1, n_2), where n_1 is the number of individuals colored by B and n_2 is the number of individuals colored by W. Notice that the states $(0,1)$ and $(1,0)$ are no longer absorbing states. In order to be able to talk about quantities like coalescent time, we will assume that, once the process arrives at $(0,1)$ or $(1,0)$, $M(t)$ will stop. Actually, coalescent time here is a type of first passage time. The following is our main result (see also Theorem 2.1).

Theorem 1.1. *Let $P_{(n_1,n_2)(0,1)}$ and $P_{(n_1,n_2)(1,0)}$ be the probabilities that the coalescent process $M(t)$ first reaches $(0,1)$ and $(1,0)$, respectively, given that it starts at (n_1,n_2), $n_1+n_2=n$. Then we have*

$$P_{(n_1,n_2)(0,1)} = \begin{cases} \frac{1}{2}+\frac{1}{2}(1-2x)^{n-1}H_{n,2}(\mu), & \text{if } n_1 \text{ is even}, \\ \frac{1}{2}-\frac{1}{2}(1-2x)^{n-1}H_{n,2}(\mu), & \text{if } n_1 \text{ is odd}; \end{cases} \tag{3}$$

$$P_{(n_1,n_2)(1,0)} = \begin{cases} \frac{1}{2}-\frac{1}{2}(1-2x)^{n-1}H_{n,2}(\mu), & \text{if } n_1 \text{ is even}, \\ \frac{1}{2}+\frac{1}{2}(1-2x)^{n-1}H_{n,2}(\mu), & \text{if } n_1 \text{ is odd}, \end{cases} \tag{4}$$

where $H_{n,k}(\mu) = \frac{(n-1)(n-2)\cdots k}{(4\mu+n-1)(4\mu+n-2)\cdots(4\mu+k)}$.

Furthermore, let $T_{(n_1,n_2)(0,1)}$ (or $T_{(n_1,n_2)(1,0)}$) be the time for the coalescent process $M(t)$ to first reach $(0,1)$ (or $(1,0)$), given that it starts at (n_1,n_2), $n_1+n_2=n$. Then the expectations of $T_{(n_1,n_2)(0,1)}$ and $T_{(n_1,n_2)(1,0)}$ are given as follows:

$$E(T_{(n_1,n_2)(0,1)}) = \begin{cases} 1-\frac{2}{n}+\frac{1}{2(\bar{x}x+2\mu)} \\ \quad \times(8\mu+1-(1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is even}, \\ 1-\frac{2}{n}+\frac{1}{2(\bar{x}x+2\mu)} \\ \quad \times(8\mu+1+(1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is odd}; \end{cases} \tag{5}$$

$$E(T_{(n_1,n_2)(1,0)}) = \begin{cases} 1-\frac{2}{n}+\frac{1}{2(\bar{x}x+2\mu)} \\ \quad \times(8\mu+1+(1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is even}, \\ 1-\frac{2}{n}+\frac{1}{2(\bar{x}x+2\mu)} \\ \quad \times(8\mu+1-(1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is odd}. \end{cases} \tag{6}$$

It is illuminating to consider the formulas of this theorem in the case when $x = 1/2$. In this case, the probabilities to first reach $(0, 1)$ and $(1, 0)$, respectively, are both 1/2. The mean time to first reach $(0, 1)$ is $3 - 2/n$, and the mean time to first reach $(1, 0)$ is also $3 - 2/n$. In general, these quantities depends on parameters x and μ. Furthermore, for any value of x, the mean time to reach either $(0, 1)$ or $(1, 0)$ is $2 - 2/n$. This quantity is independent of the mutation rate μ. It is consistent with the assumption that the mutation process is independent of the coalescent process.

In Section 2, we discuss in detail the colored coalescent process with mutation, and also mention some interesting questions related to colored coalescent processes with or without mutation. The technique of lumping turns out to be very important in simplifying computations involved. The lumping technique can be found in Tian and Kannan (2006).

2. Mutation processes

2.1. The model and some of its basic parameters

We start with the colored coalescent process $Z(t)$. So an individual in each generation can have two colors, black (B) and white (W). At the initial stage of the process, each of the n individuals in the current generation is given a color, B or W. The process then runs as the standard Kingman coalescent process. When two individuals coalesce, the color of their common ancestor is determined by (1). In other words, when an individual colored by B and another individual colored by B coalesce, the probability that their common ancestor is colored by B is x and colored by W is $\overline{x} := 1 - x$. Other situations are interpreted similarly. Furthermore, non-coalescent individuals will keep their colors unchanged after a coalescent event. We now consider the situation when the coalescent process $Z(t)$ is superimposed with a mutation process. In the language of population genetics, we assume that in the population from which the sample individuals were drawn, the probability that an individual mutates from one colored (B or W) to another in a unit time Δt is $\mu\Delta t$. Furthermore, the mutation processes of different individuals are independent of each other. Then, in a unit time Δt, the probability that k of the n individuals mutate their color is $\binom{n}{k}(\mu\Delta t)^k(1 - \mu\Delta t)^{n-k}$. Ignoring the terms of higher order in Δt, we assume that, in a unit time Δt, there is only one individual which may mutate its color with probability $\mu\Delta t$.

A typical way to study the colored coalescent process is to consider a death process on the lattice points on the plane. Then mutation processes are transitions of states within each diagonal.

Fix an integer $n > 0$. Consider a proper subset $\mathcal{L}$ of the plane integer lattice:

$$\mathcal{L} = \left\{(k, l) \in \mathbb{Z} \times \mathbb{Z};\ k \geq 0, l \geq 0, 0 < k + l \leq n\right\}.$$

A point $(k, l) \in \mathcal{L}$ represents a colored generation of $k + l$ individuals, with the number of B-colored individuals equal to k, and the number of W-colored individuals equal to l. The colored coalescent process is a death process on $\mathcal{L}$. The superimposed mutation process is a Poisson process on the following subsets, Δ_k, of $\mathcal{L}$:

$$\Delta_k = \left\{(0, k), (1, k - 1), (2, k - 2), \ldots, (k, 0)\right\},$$

for $k = 1, 2, \ldots, n$.

Let us denote by $M(t)$ the colored coalescent process superimposed with the mutation process having rate μ on the same state space $\mathcal{L}$. Notice that the states $(0, 1)$ and $(1, 0)$ are no longer absorbing states for the process $M(t)$. In fact, $M(t)$ has no more absorbing states. In order to be able to talk about quantities like coalescent time, we will assume that once the process $M(t)$ arrives at $(0, 1)$ or $(1, 0)$ it will stop. The infinitesimal generator $Q^\mu = (q^\mu_{\zeta\eta})$ of the process $M(t)$ is given by

$$q^\mu_{\zeta\eta} = \begin{cases} \bar{x}\binom{k}{2}, & \text{if } \eta = (k-2, l+1), \text{ given } \zeta = (k, l), \\ x\binom{k}{2} + xkl, & \text{if } \eta = (k-1, l), \text{ given } \zeta = (k, l), \\ \bar{x}\binom{l}{2} + \bar{x}kl, & \text{if } \eta = (k, l-1), \text{ given } \zeta = (k, l), \\ x\binom{l}{2}, & \text{if } \eta = (k+1, l-2), \text{ given } \zeta = (k, l), \\ k\mu, & \text{if } \eta = (k-1, l+1), \text{ given } \zeta = (k, l), k+l > 1, \\ l\mu, & \text{if } \eta = (k+1, l-1), \text{ given } \zeta = (k, l), k+l > 1\ , \\ -\binom{k+l}{2} - (k+l)\mu, & \text{if } \eta = (k, l), \text{ given } \zeta = (k, l), \\ 0, & \text{otherwise.} \end{cases}$$

For the process $M(t)$, we are concerned with questions like: What is the coalescent probability to $(1, 0)$ or $(0, 1)$, given that the process $M(t)$ starts at a state in Δ_n? What is the coalescent time, its mean and its distribution? We may also ask now what is the sojourn time on each diagonal Δ_k on average? Our approach to these questions will be similar to that in our study of the colored coalescent process $Z(t)$ (Tian and Lin, 2005). So we will be somewhat brief in our discussion.

Let us start with the associated jump chain of the coalescent process $M(t)$.

The "infinitesimal generator" for each Δ_k is a $(k+1) \times (k+1)$ matrix given as follows:

$$A_{k+1} = \begin{bmatrix} -r_k - k\mu & k\mu & & & & \\ \mu & -r_k - k\mu & (k-1)\mu & & & \\ & 2\mu & -r_k - k\mu & (k-2)\mu & & \\ & & & \ddots & & \\ & & & (k-1)\mu & -r_k - k\mu & \mu \\ & & & & k\mu & -r_k - k\mu \end{bmatrix}$$

for $k = 2, 3, \ldots, n$, and $A_2 = I_2$. Then, the infinitesimal generator Q^μ of $M(t)$ can be written as

$$Q^\mu = \begin{bmatrix} A_{n+1} & B_{n+1,n} & & & & \\ & A_n & B_{n,n-1} & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & A_3 & B_{3,2} \\ & & & & & A_2 \end{bmatrix},$$

where $B_{k+1,k}$ is a $k \times (k-1)$ matrix representing the transition (coalescent event) from the diagonal $\Delta_{k-1} = \{(l_1, l_2) \in \mathcal{L};\, l_1 + l_2 = k-1\}$ to the diagonal $\Delta_{k-2} = \{(l_1, l_2) \in$

$\mathcal{L};\, l_1 + l_2 = k - 2\}$ of $\mathcal{L}$, which is the same as that in the infinitesimal generator Q of the colored coalescent process $Z(t)$. Then the transition matrix for the associated jump chain of $M(t)$ is given by

$$J^{\mu} = \begin{bmatrix} \frac{1}{r_n+n\mu}A_{n+1} + I_{n+1} & \frac{1}{r_n+n\mu}B_{n+1,n} & & & \\ & \frac{1}{r_{n-1}+(n-1)\mu}A_n + I_n & & & \\ & & \ddots & & \\ & & & \ddots & \frac{1}{r_2+2\mu}B_{3,2} \\ & & & & I_2 \end{bmatrix}.$$

Denote

$$J_0^{\mu} = \begin{bmatrix} \frac{1}{r_n+n\mu}A_{n+1} + I_{n+1} & \frac{1}{r_n+n\mu}B_{n+1,n} & & & \\ & \frac{1}{r_{n-1}+(n-1)\mu}A_n + I_n & & & \\ & & \ddots & & \\ & & & \ddots & \frac{1}{r_3+3\mu}B_{4,3} \\ & & & & \frac{1}{r_2+2\mu}A_3 + I_3 \end{bmatrix},$$

then the fundamental matrix of the jump chain is given by $N^{\mu} = (I - J_0^{\mu})^{-1}$. By a direct matrix computation, we get the matrices in the first row of the block matrix N^{μ} as follows:

$$(-1)^{n-k+1}(r_k + k\mu)A_{n+1}^{-1}B_{n+1,n}A_n^{-1}B_{n,n-1}\cdots A_{k+2}^{-1}B_{k+2,k+1}A_{k+1}^{-1}, \tag{7}$$

for $k = n, n-1, \ldots, 2$. We will use these data to calculate the coalescent time and coalescent probability for the process $M(t)$.

For convenience, we start in Δ_n with an initial distribution $\pi = (\pi_0, \pi_1, \ldots, \pi_n)$, $\sum_{i=0}^{n}\pi_i = 1$. If we start at a state (n_1, n_2), just set $\pi_i = \delta_{i,n_1+1}$.

Lemma 2.1. *Let the coalescent process $M(t)$ start in Δ_n with an initial distribution $\pi = (\pi_0, \pi_1, \ldots, \pi_n)$. Denote by $P_{\pi,(0,1)}$ and $P_{\pi,(1,0)}$ the probabilities that $M(t)$ coalesces to* $(0, 1)$ *and* $(1, 0)$, *respectively. Then*

$$(P_{\pi,(0,1)}, P_{\pi,(1,0)}) = (-1)^n \pi A_{n+1}^{-1}B_{n+1,n}A_n^{-1}B_{n,n-1}\cdots A_4^{-1}B_{4,3}A_3^{-1}B_{3,2}.$$

Proof: The absorption probability (first hitting probability) of the Markov process $M(t)$ is the same as that of its jump chain. The latter can be calculated as

$$\pi N^{\mu}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ B_{3,2} \end{bmatrix} = (-1)^n \pi A_{n+1}^{-1}B_{n+1,n}A_n^{-1}B_{n,n-1}\cdots A_4^{-1}B_{4,3}A_3^{-1}B_{3,2}.$$

□

Let $\alpha_{\pi,k} = (a_{\pi(0,k)}, a_{\pi(1,k-1)}, \ldots, a_{\pi(k,0)})$ be the sojourn coefficient vector of Δ_k, that is, each $a_{\pi(k_1,k_2)}$ is the expected number of times that the jump chain J^{μ} will visit the state

(k_1, k_2), $k_1 + k_2 = k$, given that it starts with an initial distribution π on Δ_n. Then we can compute these sojourn coefficient vectors by using the following lemma.

Lemma 2.2. *The sojourn coefficient vector $\alpha_{\pi,k}$ is given by*

$$\alpha_{\pi,k} = (-1)^{n-k+1}(r_k + k\mu)\pi A_{n+1}^{-1} B_{n+1,n} A_n^{-1} B_{n,n-1} \cdots A_{k+2}^{-1} B_{k+2,k+1} A_{k+1}^{-1},$$

where $k = 2, 3, \ldots, n$.

Proof: The $\zeta\eta$-entry of the fundamental matrix N^μ is the expected number of times the jump chain is in the state η, given that it starts at the state ζ, so we get the sojourn coefficient vector $\alpha_{\pi,k}$ by picking up the matrix in the first row of the block matrix N^μ corresponding to states in Δ_k and multiplying it with the initial distribution π. This is what we have in the lemma. $\square$

2.2. *Parity lumping of the coalescent process $M(t)$*

Similar to the approach in the study of colored coalescent process $Z(t)$, we will also consider the parity lumping of the mutation superimposed coalescent process $M(t)$. The calculation for the lumped process $\overline{M}(t)$ would be simpler and we may use the results to study the original process $M(t)$.

In order to lump the process $M(t)$, we partition the state space $\mathcal{L}$ into diagonals Δ_m, $m = 1, 2, \ldots, n$, and then divide each Δ_m into two disjoint subsets, O_m and E_m. A state $(k, l) \in O_m$ when $k + l = m$ and k is odd and $(k, l) \in E_m$ when $k + l = m$ and k is even. Let

$$\overline{\mathcal{L}} = \{O_m, E_m;\ m = 1, 2, \ldots, n\}.$$

We define a new Markov process on $\overline{\mathcal{L}}$ by lumping all states in each O_m or E_m as one state, or simply taking O_m as one state, also E_m as one state. The following lemma guarantees $M(t)$ is lumpable. Let the matrices U and V be matrices for the partition of the state space as in Tian and Kannan (2006). We first have the following lemma, which can be proved by a straightforward computation of products of block matrices.

Lemma 2.3. *The coalescent process $M(t)$ is lumpable, that is, the infinitesimal generator of $M(t)$ satisfies $VUQ^\mu V = Q^\mu V$.*

Next, we consider the jump chain of the lumped coalescent process $\overline{M}(t)$.

Lemma 2.4. *The jumping chain of $\overline{M}(t)$ has the following transition matrix*:

$$\overline{J^\mu} = \begin{bmatrix} W_{n+1} & Y_{n+1} & & & \\ & W_n & Y_n & & \\ & & \ddots & \ddots & \\ & & & W_3 & Y_3 \\ & & & & I_2 \end{bmatrix},$$

where

$$W_{k+1} = \begin{bmatrix} 0 & \frac{k\mu}{r_k+k\mu} \\ \frac{k\mu}{r_k+k\mu} & 0 \end{bmatrix} \quad \textit{and} \quad Y_{k+1} = \begin{bmatrix} \frac{\overline{x}r_k}{r_k+k\mu} & \frac{xr_k}{r_k+k\mu} \\ \frac{xr_k}{r_k+k\mu} & \frac{\overline{x}r_k}{r_k+k\mu} \end{bmatrix}$$

for $k = 2, \ldots, n$.

Droping the last two rows and columns of $\overline{J}^{\mu}$, we get a matrix $\overline{J}^{\mu}_0$. The fundamental matrix of the jump chain of $\overline{M}(t)$ is $\overline{N}^{\mu} = (I - \overline{J}^{\mu}_0)^{-1}$. Set $X_{k+1} = I_2 - W_{k+1}$. The $(1, n-k+1)$th block entry of $\overline{N}^{\mu}$ is given by

$$X_{n+1}^{-1} Y_{n+1} X_n^{-1} Y_n \cdots X_{k+2}^{-1} Y_{k+2} X_{k+1}^{-1}.$$

Let us compute the above product of 2×2 matrices. First, we have

$$\begin{aligned} X_{k+1}^{-1} Y_{k+1} &= \frac{(r_k + k\mu)^2}{r_k(r_k + k\mu)} \begin{pmatrix} 1 & \frac{k\mu}{r_k+k\mu} \\ \frac{k\mu}{r_k+k\mu} & 1 \end{pmatrix} \begin{pmatrix} \overline{x} & x \\ x & \overline{x} \end{pmatrix} \frac{r_k}{r_k + k\mu} \\ &= \frac{1}{r_k + 2k\mu} \begin{pmatrix} r_k + k\mu & k\mu \\ k\mu & r_k + k\mu \end{pmatrix} \begin{pmatrix} \overline{x} & x \\ x & \overline{x} \end{pmatrix} \\ &= \frac{1}{r_k + 2k\mu} \begin{pmatrix} \overline{x} r_k + k\mu & r_k + k\mu \\ r_k + k\mu & \overline{x} r_k + k\mu \end{pmatrix} \\ &= \frac{1}{r_k + 2k\mu} \left(r_k C + k\mu D \right), \end{aligned}$$

where $C = \left(\begin{smallmatrix} \overline{x} & x \\ x & \overline{x} \end{smallmatrix}\right)$ and $D = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. It is easy to see that $CD = D = DC$, and if C^{-1} exists, $C^{-1}D = DC^{-1} = D$.

The product of $X_{k+1}^{-1} Y_{k+1}$, for $k = 2, \ldots, n$ can be calculated as follows:

$$\begin{aligned} & X_{n+1}^{-1} Y_{n+1} \cdots X_{k+2}^{-1} Y_{k+2} X_{k+1}^{-1} \\ &= \frac{1}{r_n + 2n\mu} \cdot \frac{1}{r_{n-1} + 2(n-1)\mu} \cdots \frac{1}{r_{k+1} + 2(k+1)\mu} \\ &\quad \times (r_n C + n\mu D) \cdots \big(r_{k+1} C + (k+1)\mu D\big) X_{k+1}^{-1} \\ &= \left(\prod_{i=k+1}^{n} \frac{i\mu}{r_i + 2i\mu} \cdot \left(\frac{r_i}{i\mu} C + D \right) \right) X_{k+1}^{-1} \\ &= \mu^{n-k} \frac{n!}{k!} C^{n-k} \left(\prod_{i=k+1}^{n} \frac{1}{r_i + 2i\mu} \cdot \left(\frac{r_i}{i\mu} I + D \right) \right) X_{k+1}^{-1}. \end{aligned}$$

Denote

$$G_{n,k}(z) = \left(z + \frac{r_n}{n\mu}\right)\left(z + \frac{r_{n-1}}{(n-1)\mu}\right)\cdots\left(z + \frac{r_{k+1}}{(k+1)\mu}\right)$$
$$= \left(z + \frac{n-1}{2\mu}\right)\left(z + \frac{n-2}{2\mu}\right)\cdots\left(z + \frac{k}{2\mu}\right).$$

We set

$$H_{n,k} = \frac{G_{n,k-1}(0)}{G_{n,k-1}(2)}.$$

In fact, $H_{n,k} = H_{n,k}(\mu)$ is a function of μ. We have a special function

$$H_{n,k} = H_{n,k}(\mu) = \frac{(n-1)(n-2)\cdots k}{(4\mu+n-1)(4\mu+n-2)\cdots(4\mu+k)}.$$

The product can be expressed in terms of this special function, which is the following lemma.

Lemma 2.5. *The* $(1, n-k+1)$*th block entry of the fundamental matrix of the jump chain of* $\overline{M}(t)$ *is given by*

$$X_{n+1}^{-1}Y_{n+1}X_n^{-1}Y_n \cdot X_{k+2}^{-1}Y_{k+2}X_{k+1}^{-1}$$
$$= \frac{r_k + k\mu}{2r_k}\begin{bmatrix} 1+(1-2x)^{n-k}H_{n,k} & 1-(1-2x)^{n-k}H_{n,k} \\ 1-(1-2x)^{n-k}H_{n,k} & 1+(1-2x)^{n-k}H_{n,k} \end{bmatrix}.$$

Proof: This formula is proved by a matrix product computation. We start with the computation of $G_{n,k}(D)$:

$$G_{n,k}(D) = \left(D + \frac{r_n}{n\mu}I\right)\left(D + \frac{r_{n-1}}{(n-1)\mu}I\right)\cdots\left(D + \frac{r_{k+1}}{(k+1)\mu}I\right)$$
$$= D^n + \left(\frac{r_n}{n\mu} + \frac{r_{n-1}}{(n-1)\mu}\cdots\frac{r_{k+1}}{(k+1)\mu}\right)D^{n-1} + \cdots$$
$$+ \left(\frac{r_n}{n\mu}\cdots\frac{r_{k+2}}{(k+2)\mu} + \cdots + \frac{r_{n-1}}{(n-1)\mu}\cdots\frac{r_{k+1}}{(k+1)\mu}\right)D$$
$$+ \left(\frac{r_n}{n\mu}\cdots\frac{r_{k+1}}{(k+1)\mu}\right)I$$
$$= \left[2^{n-1} + \left(\frac{r_n}{n\mu} + \frac{r_{n-1}}{(n-1)\mu} + \cdots + \frac{r_{k+1}}{(k+1)\mu}\right)2^{n-2} + \cdots\right.$$
$$\left.+ \left(\frac{r_n}{n\mu}\cdots\frac{r_{k+2}}{(k+2)\mu} + \cdots + \frac{r_{n-1}}{(n-1)\mu}\cdots\frac{r_{k+1}}{(k+1)\mu}\right)\right]D$$
$$+ \left(\frac{r_n}{n\mu}\cdots\frac{r_{k+1}}{(k+1)\mu}\right)I$$

$$
\begin{aligned}
&= \frac{1}{2}\Bigg[2^n + \left(\frac{r_n}{n\mu} + \cdots + \frac{r_{k+1}}{(k+1)\mu}\right)2^{n-1} + \cdots \\
&\quad + \left(\frac{r_n}{n\mu}\cdots\frac{r_{k+2}}{(k+2)\mu} + \cdots + \frac{r_{n-1}}{(n-1)\mu}\cdots\frac{r_{k+1}}{(k+1)\mu}\right)2\Bigg]D \\
&\quad + \left(\frac{r_n}{n\mu}\cdots\frac{r_{k+1}}{(k+1)\mu}\right)I \\
&= \frac{1}{2}\big(G_{n,k}(2) - G_{n,k}(0)\big)D + G_{n,k}(0)I \\
&= \begin{bmatrix} \frac{G_{n,k}(2)+G_{n,k}(0)}{2} & \frac{G_{n,k}(2)-G_{n,k}(0)}{2} \\ \frac{G_{n,k}(2)-G_{n,k}(0)}{2} & \frac{G_{n,k}(2)+G_{n,k}(0)}{2} \end{bmatrix}.
\end{aligned}
$$

So, we have

$$
\begin{aligned}
C^{n-k}G_{n,k}(D) &= \begin{bmatrix} \frac{1}{2}+\frac{1}{2}(1-2x)^{n-k} & \frac{1}{2}-\frac{1}{2}(1-2x)^{n-k} \\ \frac{1}{2}-\frac{1}{2}(1-2x)^{n-k} & \frac{1}{2}+\frac{1}{2}(1-2x)^{n-k} \end{bmatrix} \\
&\quad \times \begin{bmatrix} \frac{G_{n,k}(2)+G_{n,k}(0)}{2} & \frac{G_{n,k}(2)-G_{n,k}(0)}{2} \\ \frac{G_{n,k}(2)-G_{n,k}(0)}{2} & \frac{G_{n,k}(2)+G_{n,k}(0)}{2} \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} G_{n,k}(2)+(1-2x)^{n-k}G_{n,k}(0) & G_{n,k}(2)-(1-2x)^{n-k}G_{n,k}(0) \\ G_{n,k}(2)-(1-2x)^{n-k}G_{n,k}(0) & G_{n,k}(2)+(1-2x)^{n-k}G_{n,k}(0) \end{bmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
&C^{n-k}G_{n,k}(D)X_{k+1}^{-1} \\
&\quad = \frac{1}{2}\,\frac{r_k+k\mu}{r_k(r_k+2k\mu)} \\
&\quad\quad \times \begin{bmatrix} (r_k+2k\mu)G_{n,k}(2)+r_k(1-2x)^{n-k}G(0) & (r_k+2k\mu)G_{n,k}(2)-r_k(1-2x)^{n-k}G(0) \\ (r_k+2k\mu)G_{n,k}(2)-r_k(1-2x)^{n-k}G(0) & (r_k+2k\mu)G_{n,k}(2)+r_k(1-2x)^{n-k}G(0) \end{bmatrix}.
\end{aligned}
$$

Now, we get

$$
\begin{aligned}
&X_{n+1}^{-1}Y_{n+1}X_n^{-1}Y_n\cdots X_{k+2}^{-1}Y_{k+2}X_{k+1}^{-1} \\
&\quad = \left(\prod_{i=k+1}^{n}\frac{1}{r_i+2i\mu}\right)\mu^{n-k}\,\frac{n!}{k!}\,\frac{1}{2}\,\frac{r_k+k\mu}{r_k(r_k+2k\mu)} \\
&\quad\quad \times \begin{bmatrix} (r_k+2k\mu)G_{n,k}(2)+r_k(1-2x)^{n-k}G(0) & (r_k+2k\mu)G_{n,k}(2)-r_k(1-2x)^{n-k}G(0) \\ (r_k+2k\mu)G_{n,k}(2)-r_k(1-2x)^{n-k}G(0) & (r_k+2k\mu)G_{n,k}(2)+r_k(1-2x)^{n-k}G(0) \end{bmatrix} \\
&\quad = \left(\prod_{i=k}^{n}\frac{1}{r_i+2i\mu}\right)\mu^{n-k+1}\,\frac{n!}{k!}\,\frac{1}{2}\,\frac{r_k+k\mu}{r_k} \\
&\quad\quad \times \begin{bmatrix} G_{n,k-1}(2)+(1-2x)^{n-k}G_{n,k-1}(0) & G_{n,k-1}(2)-(1-2x)^{n-k}G_{n,k-1}(0) \\ G_{n,k-1}(2)-(1-2x)^{n-k}G_{n,k-1}(0) & G_{n,k-1}(2)+(1-2x)^{n-k}G_{n,k-1}(0) \end{bmatrix}
\end{aligned}
$$

$$= \frac{r_k + k\mu}{2r_k} \begin{bmatrix} 1 + (1-2x)^{n-k} \frac{G_{n,k-1}(0)}{G_{n,k-1}(2)} & 1 - (1-2x)^{n-k} \frac{G_{n,k-1}(0)}{G_{n,k-1}(2)} \\ 1 - (1-2x)^{n-k} \frac{G_{n,k-1}(0)}{G_{n,k-1}(2)} & 1 + (1-2x)^{n-k} \frac{G_{n,k-1}(0)}{G_{n,k-1}(2)} \end{bmatrix}$$

$$= \frac{r_k + k\mu}{2r_k} \begin{bmatrix} 1 + (1-2x)^{n-k} H_{n,k} & 1 - (1-2x)^{n-k} H_{n,k} \\ 1 - (1-2x)^{n-k} H_{n,k} & 1 + (1-2x)^{n-k} H_{n,k} \end{bmatrix}. \qquad \square$$

Let the lumped coalescent process $\overline{M}(t)$ start in $\{E_n, O_n\}$ with a distribution $\pi = (\pi_E, \pi_O)$. The sojourn coefficients at $\{E_k, O_k\}$ are denoted by (α_k, β_k). We have

$$(\alpha_k, \beta_k) = (\pi_E, \pi_O) \begin{bmatrix} 1 + (1-2x)^{n-k} H_{n,k}(\mu) & 1 - (1-2x)^{n-k} H_{n,k}(\mu) \\ 1 - (1-2x)^{n-k} H_{n,k}(\mu) & 1 + (1-2x)^{n-k} H_{n,k}(\mu) \end{bmatrix} \frac{r_k + k\mu}{2r_k}.$$

Therefore,

$$\begin{aligned} \alpha_k &= \frac{r_k + k\mu}{2r_k} + (\pi_E - \pi_O) \frac{r_k + k\mu}{2r_k} (1-2x)^{n-k} H_{n,k}(\mu), \\ \beta_k &= \frac{r_k + k\mu}{2r_k} + (\pi_O - \pi_E) \frac{r_k + k\mu}{2r_k} (1-2x)^{n-k} H_{n,k}(\mu). \end{aligned} \tag{8}$$

Let $P_{\pi,E}$, $P_{\pi,O}$ be the coalescent probabilities to E_1 and O_1, respectively, given that the process starts with the distribution $\pi = (\pi_E, \pi_O)$.

Lemma 2.6. *The coalescent probabilities are given by*

$$(P_{\pi,E}, P_{\pi,O}) = (\pi_E, \pi_O) X_{n+1}^{-1} Y_{n+1} X_n^{-1} Y_n \cdots X_3^{-1} Y_3 \begin{pmatrix} \frac{\overline{x}}{1+2\mu} & \frac{x}{1+2\mu} \\ \frac{x}{1+2\mu} & \frac{\overline{x}}{1+2\mu} \end{pmatrix}.$$

In particular, let $P_{E,E}$ be the probability of first reaching E_1, given that the process starts at E_n, and let other quantities $P_{E,O}$, $P_{O,E}$, $P_{O,O}$ be defined similarly. Then we have

$$\begin{aligned} P_{E,E} &= \frac{1}{2} + \frac{1}{2}(1-2x)^{n-1} H_{n,2}(\mu), & P_{E,O} &= \frac{1}{2} - \frac{1}{2}(1-2x)^{n-1} H_{n,2}(\mu), \\ P_{O,E} &= \frac{1}{2} - \frac{1}{2}(1-2x)^{n-1} H_{n,2}(\mu), & P_{O,O} &= \frac{1}{2} + \frac{1}{2}(1-2x)^{n-1} H_{n,2}(\mu). \end{aligned} \tag{9}$$

We now calculate the expectation of the coalescent time for the lumped coalescent process $\overline{M(t)}$. Let τ_k be a random variable distributed exponentially with the mean $(r_k + k\mu)^{-1}$. Recall that the sojourn coefficients α_k (respectively, β_k) is the expected number of times that the jump chain of $\overline{M(t)}$ visits the state E_k (respectively, O_k), given that it starts with an initial distribution $\pi = (\pi_E, \pi_O)$ on the initial states $\{E_n, O_n\}$. To calculate the expectation of the coalescent time $T_{\pi,E}$ of first reaching the state E_1 for the process $\overline{M(t)}$, given that it starts with the initial distribution π, we also need to define a random variable $\widetilde{T}_{\pi,E}$ so that $E(T_{\pi,E}) = E(\widetilde{T}_{\pi,E})$. With the same idea as in the study of coalescent process $Z(t)$ (Tian and Lin, 2005), we define a random variable

$$\widetilde{T}_{\pi,E} = \sum_{k=3}^{n} \alpha_k \tau_k + \sum_{k=3}^{n} \beta_k \tau_k + \widehat{\alpha}_2 \frac{1+2\mu}{\overline{x}+2\mu} \tau_2 + \widehat{\beta}_2 \frac{1+2\mu}{x+2\mu} \tau_2,$$

where $\widehat{\alpha}_2$ and $\widehat{\beta}_2$ are sojourn coefficients for the conditional process which is obtained by deleting the state O_1. Since this does not change $\overline{M(t)}$ too much, we still can use previous formula to compute those coefficients. Actually, we have

$$(\widehat{\alpha}_2, \widehat{\beta}_2) = (\pi_E, \pi_O) X_{n+1}^{-1} Y_{n+1} \cdots X_4^{-1} Y_4 \begin{bmatrix} 1 & -\frac{2\mu}{\overline{x}+2\mu} \\ -\frac{2\mu}{x+2\mu} & 1 \end{bmatrix}^{-1}.$$

This gives us

$$\widehat{\alpha}_2 = \frac{[4\mu + x + (\pi_E - \pi_O)x(1-2x)^{n-2} H_{n,2}](\overline{x} + 2\mu)}{2(x\overline{x} + 2\mu)},$$

$$\widehat{\beta}_2 = \frac{[4\mu + \overline{x} + (\pi_O - \pi_E)\overline{x}(1-2x)^{n-2} H_{n,2}](x + 2\mu)}{2(x\overline{x} + 2\mu)}$$

and

$$\widetilde{T}_{\pi,E} = \sum_{k=3}^{n} \frac{r_k + k\mu}{r_k} \tau_k + \frac{1+2\mu}{2(\overline{x}x + 2\mu)} \big[8\mu + 1 + (\pi_O - \pi_E)(1-2x)^{n-1} H_{n,2}\big] \tau_2.$$

Similarly, we have the random variable $\widetilde{T}_{\pi,O}$ which has the same expectation as the coalescent time $T_{\pi,O}$. It is given as follows:

$$\widetilde{T}_{\pi,O} = \sum_{k=3}^{n} \frac{r_k + k\mu}{r_k} \tau_k + \frac{1+2\mu}{2(\overline{x}x + 2\mu)} \big[8\mu + 1 + (\pi_E - \pi_O)(1-2x)^{n-1} H_{n,2}\big] \tau_2.$$

Using the Feller relation (see Syski, 1992), we get the following lemma.

Lemma 2.7. *The expectations of the coalescent times to E_1 and O_1, respectively, are*

$$E(T_{\pi,E}) = 1 - \frac{2}{n} + \frac{1}{2(\overline{x}x + 2\mu)} \big[8\mu + 1 + (\pi_O - \pi_E)(1-2x)^{n-1} H_{n,2}(\mu)\big]$$

and

$$E(T_{\pi,O}) = 1 - \frac{2}{n} + \frac{1}{2(\overline{x}x + 2\mu)} \big[8\mu + 1 + (\pi_E - \pi_O)(1-2x)^{n-1} H_{n,2}(\mu)\big].$$

For the coalescent process $M(t)$ (or equivalently, the lumped coalescent event $\overline{M}(t)$), let us also define ρ_k to be the parity of a state after the kth coalescent event. That is, ρ_k is a random variable which takes the value 0 if the process is in E_k and takes the value 1 if the process is in O_k in after the kth coalescent event. Then, the distribution of ρ_k depends on that of the starting state, i.e. the distribution of ρ_0. Suppose the initial distribution of the parity is $\pi = (\pi_E, \pi_O)$. When we compute the probability distribution of ρ_k, that is, the probabilities that the lumped coalescent process $\overline{M}(t)$ or its jump chain first reaches the states E_{n-k} and O_{n-k} after k coalescent events, we can view these two states as absorbing states and consider a new Markov process or its jump chain with the state space $\{E_1, O_1, E_2, O_2, \ldots, E_{n-k}, O_{n-k}\}$. Use the same method as in the study of the process $Z(t)$, the absorption probabilities, technically called first hitting probabilities, can be calculated by multiplying the sojourn coefficient vector $(\alpha_{n-k+1}, \beta_{n-k+1})$ with the transition matrix from $\{E_{n-k+1}, O_{n-k+1}\}$ to $\{E_{n-k}, O_{n-k}\}$ of the jump chain $\overline{J}^{\mu}$. Denote by $\Psi_{\pi,k}$ the probability distribution of the parity ρ_k. Then, we have

$$\begin{aligned}\Psi_{\pi,k} &= (\alpha_{n-k+1}, \beta_{n-k+1}) \begin{bmatrix} \frac{\bar{x} r_{n-k+1}}{r_{n-k+1}+(n-k+1)\mu} & \frac{x r_{n-k+1}}{r_{n-1}+(n-k+1)\mu} \\ \frac{x r_{n-k+1}}{r_{n-k+1}+(n-k+1)\mu} & \frac{\bar{x} r_{n-k+1}}{r_{n-k+1}+(n-k+1)\mu} \end{bmatrix} \\ &= \frac{1}{2}\big(1 + (\pi_E - \pi_O)(1-2x)^k H_{n,n-k}(\mu), 1 + (\pi_O - \pi_E)(1-2x)^k H_{n,n-k}(\mu)\big).\end{aligned}$$

In particular,

$$\begin{aligned} \Pr(\rho_k = 0 \mid \rho_0 = 0) &= \frac{1}{2} + \frac{1}{2}(1-2x)^k H_{n,n-k}(\mu), \\ \Pr(\rho_k = 1 \mid \rho_0 = 0) &= \frac{1}{2} - \frac{1}{2}(1-2x)^k H_{n,n-k}(\mu), \\ \Pr(\rho_k = 0 \mid \rho_0 = 1) &= \frac{1}{2} - \frac{1}{2}(1-2x)^k H_{n,n-k}(\mu), \\ \Pr(\rho_k = 1 \mid \rho_0 = 1) &= \frac{1}{2} + \frac{1}{2}(1-2x)^k H_{n,n-k}(\mu). \end{aligned} \tag{10}$$

2.3. Back to the coalescent process with mutation $M(t)$

For the parity lumping of the coalescent process $\overline{M}(t)$, we also have the commutative diagram

$$\begin{array}{ccc} M(t) & \xrightarrow{\text{jump chain}} & J^{\mu} \\ {\scriptstyle\text{lumping}}\big\downarrow & & \big\downarrow{\scriptstyle\text{lumping}} \\ \overline{M}(t) & \xrightarrow{\text{jump chain}} & \overline{J}^{\mu} \end{array}$$

as in the study of the process $Z(t)$. Thus, we can get information about $M(t)$ from the knowledge of the lumped coalescent process $\overline{M}(t)$, particularly, we can recover certain parameters for $M(t)$ from the jump chain $\overline{J}^{\mu}$.

We have the fundamental matrix N^{μ} of the jump chain J^{μ} and the fundamental matrix $\overline{N}^{\mu}$ of the jump chain $\overline{J}^{\mu}$. Let U_0 and V_0 be matrices derived from U and V by dropping the last two rows and columns which correspond to the states $(0,1)$ and $(1,0)$. It is easy to check that N^{μ} is lumpable by U_0 and V_0, that is, $V_0 U_0 N^{\mu} V_0 = N^{\mu} V_0$. Then we have $\overline{N}^{\mu} = U_0 N^{\mu} V_0$. The following is the main theorem of this article.

Theorem 2.1. *Let $P_{(n_1,n_2)(0,1)}$ and $P_{(n_1,n_2)(1,0)}$ be the probabilities that the coalescent process $M(t)$ first reaches $(0,1)$ and $(1,0)$, respectively, given that it starts at (n_1, n_2), $n_1 + n_2 = n$. Then we have*

$$P_{(n_1,n_2)(0,1)} = \begin{cases} \frac{1}{2} + \frac{1}{2}(1-2x)^{n-1} H_{n,2}(\mu), & \text{if } n_1 \text{ is even}, \\ \frac{1}{2} - \frac{1}{2}(1-2x)^{n-1} H_{n,2}(\mu), & \text{if } n_1 \text{ is odd}; \end{cases} \tag{11}$$

$$P_{(n_1,n_2)(1,0)} = \begin{cases} \frac{1}{2} - \frac{1}{2}(1-2x)^{n-1}H_{n,2}(\mu), & \text{if } n_1 \text{ is even}, \\ \frac{1}{2} + \frac{1}{2}(1-2x)^{n-1}H_{n,2}(\mu), & \text{if } n_1 \text{ is odd}. \end{cases} \tag{12}$$

Furthermore, let $T_{(n_1,n_2)(0,1)}$ (*or* $T_{(n_1,n_2)(1,0)}$) *be the time for the coalescent process* $M(t)$ *to first reach* $(0,1)$ (*or* $(1,0)$), *given that it starts at* (n_1,n_2), $n_1+n_2=n$. *Then the expectations of* $T_{(n_1,n_2)(0,1)}$ *and* $T_{(n_1,n_2)(1,0)}$ *are given as follows*:

$$E(T_{(n_1,n_2)(0,1)}) = \begin{cases} 1 - \frac{2}{n} + \frac{1}{2(\overline{x}x+2\mu)} \\ \quad \times(8\mu + 1 - (1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is even}, \\ 1 - \frac{2}{n} + \frac{1}{2(\overline{x}x+2\mu)} \\ \quad \times(8\mu + 1 + (1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is odd}; \end{cases} \tag{13}$$

$$E(T_{(n_1,n_2)(1,0)}) = \begin{cases} 1 - \frac{2}{n} + \frac{1}{2(\overline{x}x+2\mu)} \\ \quad \times(8\mu + 1 + (1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is even}, \\ 1 - \frac{2}{n} + \frac{1}{2(\overline{x}x+2\mu)} \\ \quad \times(8\mu + 1 - (1-2x)^{n-1}H_{n,2}(\mu)), & \text{if } n_1 \text{ is odd}. \end{cases} \tag{14}$$

Proof: The absorption probabilities of the jump chains J^μ and $\overline{J}^\mu$ are calculated from the matrices

$$N^\mu \begin{bmatrix} 0 \\ \vdots \\ 0 \\ C \end{bmatrix} \quad \text{and} \quad \overline{N}^\mu \begin{bmatrix} 0 \\ \vdots \\ 0 \\ C \end{bmatrix} = U_0 N^\mu V_0 \begin{bmatrix} 0 \\ \vdots \\ 0 \\ C \end{bmatrix} = U_0 N^\mu \begin{bmatrix} 0 \\ \vdots \\ 0 \\ C \end{bmatrix},$$

respectively. So $P_{(n_1,n_2)(0,1)}$, for $n_1+n_2=n$ and n_1 even, is equal to $P_{E,E}$. By (9), we get the desired value for $P_{(n_1,n_2)(0,1)}$ in this case. All other cases can be obtained similarly.

To compute the coalescent time, we first denote by $a_{(n_1,n_2)(k_1,k_2)}$ the sojourn coefficient of the jump chain J^μ, which is the expect number of times the jump chain J^μ visits the state (k_1,k_2), given that it starts at the state (n_1,n_2), $n_1+n_2=n$. Since the jump chain J^μ is lumpable and its lumping is the jump chain $\overline{J}^\mu$, we have

$$\begin{aligned} \sum_{k_1+k_2=k,\, k_1 \text{ even}} a_{(n_1,n_2)(k_1,k_2)} &= \alpha_k, \\ \sum_{k_1+k_2=k,\, k_1 \text{ odd}} a_{(n_1,n_2)(k_1,k_2)} &= \beta_k, \end{aligned} \tag{15}$$

where α_k and β_k are the sojourn coefficients of $\overline{J}^\mu$ corresponding to the expected number of times $\overline{J}^\mu$ visits E_k and O_k, respectively, given that it starts with the distribution

$$\pi = (\pi_E, \pi_O) = \begin{cases} (1,0), & \text{if } n_1 \text{ is even}, \\ (0,1), & \text{if } n_1 \text{ is odd}. \end{cases} \tag{16}$$

Thus, we have

$$\begin{aligned}
\widetilde{T}_{(n_1,n_2)(0,1)} &= \sum_{k=3}^{n} \Bigg(\sum_{k_1+k_2=k} a_{(n_1,n_2)(k_1,k_2)} \Bigg) \tau_k \\
&\quad + a_{(n_1,n_2)(0,2)} \frac{1+2\mu}{\overline{x}+2\mu} \tau_2 + a_{(n_1,n_2)(2,0)} \frac{1+2\mu}{\overline{x}+2\mu} \tau_2 + a_{(n_1,n_2)(1,1)} \frac{1+2\mu}{x+2\mu} \tau_2 \\
&= \sum_{k=3}^{n} \alpha_k \tau_k + \sum_{k=3}^{n} \beta_k \tau_k + \widehat{\alpha}_2 \frac{1+2\mu}{\overline{x}+2\mu} \tau_2 + \widehat{\beta}_2 \frac{1+2\mu}{x+2\mu} \tau_2 \\
&= \widetilde{T}_{\pi,E},
\end{aligned}$$

where π is the distribution given by (16). So Lemma 2.7 gives us the desired expectation of $T_{(n_1,n_2)(0,1)}$. The calculation of $E(T_{(n_1,n_2)(1,0)})$ is exactly the same. $\square$

Corollary 2.1. *The random variable*

$$T = \sum_{k=2}^{n} \alpha_k \tau_k + \sum_{k=2}^{n} \beta_k \tau_k$$

has the same expectation as the random variable of the time that the process $M(t)$ first reaches either E_1 or O_1, and which is $E(T) = 2 - 2/n$.

The fact that $E(T)$ is independent of the mutation rate μ is consistent with the assumption that the mutation process is independent of the coalescent process. In general, by (13) and (14), $E(T_{(n_1,n_2)(0,1)})$ and $E(T_{(n_1,n_2)(1,0)})$ depend on μ. It is interesting to note that when $x = 1/2$, we have $E(T_{(n_1,n_2)(0,1)}) = E(T_{(n_1,n_2)(1,0)}) = 3 - 2/n$. In population genetics, one is interested in the total branch length of the random genealogical tree defined as

$$T_{\text{tot}} = \sum_{k=2}^{n} k\, \omega_k,$$

where $E(\omega_k) = r_k^{-1}$. This is because the number of mutations that are expected to occur on a random genealogical tree is proportional to

$$E(T_{\text{tot}}) = \sum_{k=2}^{n-1} \frac{2}{k} \approx 2(\gamma + \log n)$$

($\gamma \approx 0.577216$ is the Euler constant), which has important consequences for estimating the mutation rate, as well as for inferences that depend on estimates of the mutation rate. In the case when $x = 1/2$, the expected total branch length of the random genealogical tree with the root E_1 (or O_1) can be calculated as

$$E\Bigg(\sum_{k=3}^{n} k\, \omega_k + 4\omega_2 \Bigg) \approx 2(1 + \gamma + \log n).$$

2.4. Interesting questions

It is worthwhile to point out that although these colored coalescent processes are purely mathematical, they come from a variant of the most studied population models, the neutral Wright–Fisher model. On the other hand, we think that these coalescent processes are of mathematical interest in their own right. For example, the mean time for the colored coalescent process with or without mutation to reach either a black root or a white root is the same as the mean time for the Kingman coalescent process to reach a MRCA. We also see the mean time for the colored coalescent process with or without mutation to reach a black root, and the mean time to reach a white root. These three quantities apparently are not related with each other in a simple way. Such a phenomenon can be compared with a similar situation of two independent Poisson processes: At an airport, one needs to wait for a taxi for T_1 minutes and for a bus for T_2 minutes on average. If the average time one needs to wait for either a taxi or a bus is T, then we have the simple relation that

$$\frac{1}{T} = \frac{1}{T_1} + \frac{1}{T_2}. \tag{17}$$

For the colored coalescent process, $Z(t)$ or $M(t)$, there is no longer such a simple relation among the mean time to reach a black root, the mean time to reach a white root, and the mean time to reach either a black root or a white root. In this setting, we have a Markov process e^{tQ} with two absorbing states, or a persistent cycle with two states. Conditional on only first reaching one of these two states, we get another two Markov processes e^{tQ_1} and e^{tQ_2}. Thus we have the mean time T for the process e^{tQ} to first reach either these two states, and the mean time T_1 (respectively, T_2) for the process e^{tQ_1} (respectively, e^{tQ_2}) to reach its absorbing state. Are these three quantities T, T_1 and T_2 related in a certain way? How much (17) is altered for the colored coalescent processes? Can this deviation from (17) in the colored coalescent processes be attributed to the structure of the random trees they generate? These are some of the questions about the colored coalescent processes that we are interested in.

Acknowledgements

We thank Prof. Michael T. Clegg for introducing us to the coalescent theory and its significance in population genetics. The first author would also like to express his greatest regret of the death of Prof. Xiao-Song Lin, a co-author of this article.

References

Fu, Y.-X., Li, W.-H., 1999. Coalescing into the 21th century: an overview and prospects of coalescent theory. Theor. Popul. Biol. 56, 1–10.

Hein, J., Schierup, M.H., Wiuf, C., 2005. Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. Oxford University Press, London.

Kingman, J.F.C., 1982a. The coalescent. Stoch. Process. Their Appl. 13, 235–248.

Kingman, J.F.C., 1982b. On the genealogy of large populations. J. Appl. Probab. 19A, 27–43.

Nordborg, M., 2001. Coalescent theory. In: Balding, D., Bishop, M., Cannings, C. (Eds.), Handbook of Statistical Genetics. Wiley, New York. Chapter 7.

Rice, S.H., 2004. Evolutionary Theory: Mathematical and Conceptual Foundations. Sinauer, Sunderland.
Rosenberg, N.A., Nordborg, M., 2002. Genealogical tress, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. 3, 380–390.
Syski, R., 1992. Passage Times for Markov Chains. IOS Press, Amsterdam.
Tian, J.P., Kannan, D., 2006. Lumpability and commutativity of Markov processes. Stoch. Anal. Appl. 24(3), 685–702.
Tian, J.P., Lin, X.-S., 2005, Colored coalescent theory. Discrete Contin. Dyn. Syst., Sup. Vol., pp. 833–845.