Population Genetics Genealogies Under Selection



Hongyu Jiang
Lady Margaret Hall and Department of Statistics
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Statistics
Trinity Term 2013

Population Genetics Genealogies Under Selection

Hongyu Jiang, Lady Margaret Hall

DPhil thesis, Trinity Term 2013

Abstract

In the presence of selection and mutation, the genealogy of a given sample configuration can be described by two classes of ancestral processes, namely the coalescent-in-a-random-background model of Kaplan et al. (1988) and the dual process with typed lines of Etheridge and Griffiths (2009). These two processes are based on the same forwards population genetics model. However, in the former model, selection is reflected in the ancestral frequencies in the population, while in the latter model, there are branching events that generate virtual ancestral lines. We simulate the dual processes with typed lines and derive the limits of the two ancestral processes under strong selection and under selection-mutation balance to address the question of to what extent the genealogy is distorted. The two ancestral processes generate the same limiting genealogy. In a two-allele population under strong selection, the disfavoured individuals in the sample are instantaneously converted to a random number of favoured individuals, and the limiting genealogy is governed by the usual Kingman's coalescent. Under selection-mutation balance, all disfavoured individuals in the sample are instantaneously converted to the favoured type, and the limiting genealogy is determined by a time-changed Kingman's coalescent. The proofs of these limiting processes are based on the convergence result of Möhle (1998, Lemma 1). The studies of selection-mutation balance are then extended to an additive selection model, where each individual is composed of L diallelic loci. In the corresponding dual process with typed lines, the evolution of the virtual lines on a faster timescale can be approximated by a deterministic process, while the evolution of the real lines is independent of the virtual lines. The structure in the limiting genealogy collapses to Kingman's coalescent. We also let L tend to infinity, and obtain a full description of the limiting genealogy in the background selection model.

Acknowledgements

It has been a pleasure to work under the supervision of Robert Griffiths. I am extremely grateful to him for his guidance, advice, encouragement and patience during my DPhil studies.

I would also like to express my gratitude to Alison Etheridge, Ellen Baake, Christina Goldschmidt and James Martin for taking the time to examine the drafts of this thesis and for providing many insightful comments. I also thank Alison for introducing several illuminating references to me.

I could never have finished this work without the support of my colleagues. Particularly I thank Shidong Wang, for his warm friendship, stimulating conversations and invaluable suggestions.

My time at Oxford has been partially supported by the Department of Statistics through the Teaching Assistance Bursary, for which I am very grateful.

Finally, my deepest gratitude goes to my parents and to my wife for their selfless and endless support.

Contents

1 Introduction **1**

1.1 Structure and main results 6

1.2 Ancestral selection graph 11

1.3 Moran model with selection and mutation 16

1.4 Ancestral dual process with typed lines 22

2 Genealogies under strong selection **29**

2.1 Genealogies in Kingman's coalescent 31

2.2 Simulation of ancestral dual processes with typed lines 33

2.3 Shape of genealogies 40

2.3.1 Coalescence times for samples with random configuration . . . 42

2.3.2 Time to the MRCA and length of the genealogies for samples with given configuration 45

2.3.3 Type distribution of the MRCA 47

2.4 Recursive equations under strong selection 49

3 Ancestral processes under strong selection **67**

3.1 Coalescent in a random background 69

3.2 Ancestral dual process with typed lines 85

3.3 Ancestral dual process with three types 91

4 Genealogies under selection-mutation balance **102**

4.1 Deterministic approximation of the diffusion 103

4.2 Coalescent in a random background 109

4.3 Ancestral dual process with typed lines 118

4.4 Connections to the forwards-in-time process 130

5 Selection-mutation balance in an additive selection model **134**

5.1 Moran model and diffusion approximation 136

5.2 Deterministic trajectories 141

5.2.1 One-locus case 142

5.2.2 L-locus case 146

5.3 Ancestral dual process with typed lines 152

5.4 Genealogies under selection-mutation balance 155

5.4.1 One-locus case 155

5.4.2 L-locus case 165

5.5 Limit as $L \to \infty$ 175

Appendix A Proofs **184**

A.1 Proof of Lemma 2.7 184

A.2 Proof of Lemma 2.8 185

A.3 Proof of Lemma 3.13 188

A.4 Proof of Lemma 3.14 189

Bibliography **192**

List of Figures

1.1 The stationary density of the frequency of the disfavoured type 1 individuals in the diffusion limit 21
1.2 A graphical representation of the Moran model with selection and mutation 24

2.1 A genealogy in Kingman's coalescent 31
2.2 A genealogy in the ancestral dual process with typed lines 38
2.3 The effect of selection and mutation on the time to the MRCA for given samples 46
2.4 The effect of selection and mutation on the MRCA type distribution for given samples 48

4.1 A path of the diffusion with selection and mutation, superimposed on the trajectory of the differential equation 105
4.2 The stable equilibrium point of the differential equation (4.2) 106
4.3 Comparison of the stationary density of the diffusion under selection-mutation balance to the density of a normal distribution 107
4.4 The evolution of the virtual lines in the ancestral dual process with typed lines under selection-mutation balance 127

5.1 The stable equilibrium point of the differential equation (5.7) in the one-locus case 144

5.2 A path of the diffusion in the additive selection model in the two-locus case, superimposed on the trajectory of the deterministic system . . . 149
5.3 The evolution of the virtual lines in the ancestral dual process with typed lines in the additive selection model in the one-locus case . . . 159
5.4 The evolution of the virtual lines in the ancestral dual process with typed lines in the additive selection model in the three-locus case . . 170

List of Tables

2.1 The mean number of branching events in the ancestral dual process with typed lines and in the ancestral selection graph 39

2.2 The effect of selection and mutation on the time to the MRCA and the length of the genealogy for random samples 43

2.3 The effect of selection and mutation on the coalescence times for random samples 44

2.4 The time to the MRCA and the length of the genealogy for given samples under strong selection 62

4.1 The time to the MRCA for given samples under selection-mutation balance 130

Chapter 1

Introduction

Mathematical population genetics is an area of study that is used to quantitatively and qualitatively understand the genealogical structure of populations by employing stochastic processes to model the evolutionary forces of genetic drift (random reproduction), mutation and selection that produce and maintain genetic variation. Consider a haploid (one-parent) and panmictic population (e.g., mitochondrial DNA) of size N. Assume the population is neutral, which means that each individual has the same fitness (or reproductive ability). Let ν be the number of offspring an individual may produce in one generation. Kingman (1982a,b) shows that if $\lim_{N\to\infty} \mathrm{Var}(\nu) = c^2$, $0 < c^2 < \infty$, the random genealogical structure of a sample taken from the population can be approximated by a *coalescent* process in the limit as $N \to \infty$, provided time is measured in units of N/c^2 generations.

Two individuals are said to coalesce when they are related via a common ancestor at some time in the past. Let $\alpha_n(t)$ be the number of distinct ancestors (or ancestral lines) at (scaled) time $t > 0$ in the past, of a sample of size $n \geq 1$ taken from the population at time $t = 0$. The ancestral process $\{\alpha_n(t), t \geq 0\}$ in Kingman's coalescent is a continuous-time Markov process on the state space $S = \{1, 2, \ldots, n\}$, which has

n as its initial state and 1 as its absorbing state. The transition rates

$$q(\alpha, \beta) \coloneqq \lim_{\delta t \to 0} \frac{1}{\delta t} \mathbb{P}[\alpha_n(t + \delta t) = \beta \mid \alpha_n(t) = \alpha],$$

where $\alpha, \beta \in S$ and $\alpha \neq \beta$, are given by

$$q(\alpha, \beta) = \begin{cases} \binom{\alpha}{2}, & \text{if } \beta = \alpha - 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.1}$$

The coalescence rate $\binom{\alpha}{2}$ in equation (1.1) is independent of the population size and the structure of the genealogy. The process $\{\alpha_n(t), t \geq 0\}$ is a death process with initial state n and death rate $\binom{\alpha}{2}$.

The beauty of Kingman's coalescent is the simple tree-like genealogy it generates (see Figure 2.1), and the intuitive probabilistic interpretation due to the assumption of *exchangeability* (as a consequence of panmixia). In the context of coalescent theory, exchangeability means individuals are not labelled in any way, and all individuals are equally likely to be involved in a coalescence event (hence the coalescence rate $\binom{\alpha}{2}$ in equation (1.1)). A genealogy consists of two types of information: the branch lengths of the genealogy, and the topology (who shares a common ancestor with whom). The neutral genealogy in Kingman's coalescent will be discussed in detail in Section 2.1.

Although Kingman's coalescent is derived in the setting of the classic Wright-Fisher model (Fisher, 1930; Wright, 1931) with non-overlapping generations, the coalescent is a robust limiting process for a large class of neutral population genetics models, including the Moran model (Moran, 1958) with overlapping generations, in the sense that properties of Kingman's coalescent are shared by all these models that converge to it in the limit. Möhle (1998) illustrates that even some non-exchangeable population genetics models converge to Kingman's coalescent in the limit of large population size (provided time is scaled appropriately). Kingman's coalescent is also extended to model various biological phenomenon, such as diploid, two sexes, selfing and varying population sizes. Hudson (1991), Nordborg (2007), Durrett (2008), Wakeley (2009)

and Etheridge (2011) provide comprehensive introductions and reviews for these extensions. Griffiths and Marjoram (1996, 1997) propose an analogue of the coalescent, the ancestral recombination graph, to describe genealogies in multi-locus models under recombination. Moreover, Kingman's coalescent has led to numerous computationally intensive algorithms for a wide range of evolutionary problems (see Griffiths, 1989; Griffiths and Tavaré, 1994; Tavaré et al., 1997).

The *structured coalescent* process is introduced by Takahata (1988) and Notohara (1990) for populations subdivided into a discrete collection of demes or different genetic backgrounds, each of which is large and panmictic. Donnelly and Tavaré (1995) and Wilkinson-Herbots (1998) provide review articles. Denote the set of non-negative integers by $\mathbb{Z}_+$, and the set in which all elements are 0 by $\mathbf{0}$. Let $|\boldsymbol{\alpha}| \coloneqq \sum_i \alpha_i$ be the sum of elements in a vector. Write $\boldsymbol{\alpha} > \boldsymbol{\beta}$ when $\alpha_i > \beta_i$ for all i. Consider $d \geq 1$ subpopulations and a sample $\boldsymbol{n} \in \mathbb{Z}_+^d$, where each n_i is the number of individuals sampled from the ith subpopulation. The structured ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ is a continuous-time Markov process with initial state $\boldsymbol{\alpha}(0) = \boldsymbol{n}$ and state space $\{\boldsymbol{\alpha} \colon \boldsymbol{\alpha} \in \mathbb{Z}_+^d, 1 \leq |\boldsymbol{\alpha}| \leq |\boldsymbol{n}|\}$, in which each $\alpha_i(t)$, $1 \leq i \leq d$ counts the number of ancestral lines in the ith subpopulation t time units ago. Let the configuration of ancestral lines at time t be $\boldsymbol{\alpha}$. Each pair of ancestral lines within the same subpopulation can coalesce with the transition $\boldsymbol{\alpha} \to \boldsymbol{\alpha} - \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is a unit vector with a 1 in the ith position and 0 everywhere else; and every ancestral line in the ith population can move to the jth subpopulation with the transition $\boldsymbol{\alpha} \to \boldsymbol{\alpha} - \boldsymbol{e}_i + \boldsymbol{e}_j$, $i \neq j$ by some form of migration.

The genealogy in the structured coalescent for a sample of size $|\boldsymbol{n}| \geq 2$ from $d \geq 1$ subpopulations must contain $|\boldsymbol{n}| - 1$ coalescence events, and a minimum of $d - 1$ migrations between subpopulations, since the coalescence of two ancestral lines cannot occur unless the lines are in the same subpopulation. Because of this, when migration rates are small, genealogies in the structured coalescent are stretched

compared with those in the (unstructured) Kingman's coalescent. When population is subdivided according to allelic types, and movement between subpopulations is caused by mutation, the structured coalescent can be used to study genealogies under selection (see Kaplan et al., 1988; Barton et al., 2004; Barton and Etheridge, 2004), which will be discussed in Section 3.1.

A mutation is a heritable change in the genetic material of an organism. Mutation was first recognised by Gregor Mendel from his experiments on peas. Barton et al. (2007, p. 325) state that mutation is the ultimate source of all genetic variation, and without it there would be no evolution. The mutation rate per nucleotide site per generation for completely linked DNA sequences, is small (10^{-8} is a typical value for advanced organisms, see Drake et al., 1998). However, there is a positive correlation between nucleotide diversity and the recombination rate (Lercher and Hurst, 2002).

Consider a population of fixed size N with discrete generations, in which the mutation probability is u per individual per generation. A continuous time limit can be obtained by measuring time $t \geq 0$ in units of N generations. We assume

$$\lim_{N\to\infty} Nu = \theta/2,$$

such that $\theta/2$ is the mutation rate per individual per unit of time that is rescaled in units of N generations. Denote the largest integer not greater than x by $\lfloor x \rfloor$. The number M of mutations that occur on an ancestral line of length $\lfloor Nt \rfloor$ generations, $t \geq 0$, is binomially distributed with probability mass function

$$\mathbb{P}(M = m) = \binom{\lfloor Nt \rfloor}{m} u^m (1-u)^{\lfloor Nt \rfloor - m}, \quad m = 0, 1, \ldots, \lfloor Nt \rfloor.$$

The moment generating function of M is

$$\mathbb{E}(e^{\lambda M}) = (1 - u + ue^{\lambda})^{\lfloor Nt \rfloor} \to e^{\frac{\theta}{2}t(e^{\lambda}-1)},$$

as $N \to \infty$. Thus in the limit, the distribution of M is Poisson with rate $\theta t/2$.

In Kingman's coalescent, mutations can be superimposed on genealogies such that mutations occur at the points of an independent Poisson process of rate $\theta/2$ on each branch of a genealogy. Conditional on the length W of an ancestral line, the number of mutations that arise on this line has a Poisson distribution with rate $W\theta/2$.

The influence of mutations can be described by two models: the infinitely-many-alleles model (Kimura and Crow, 1964) and the infinitely-many-sites model (Watterson, 1975). The former model assumes that each mutation leads to a novel allelic type never seen before in the population, so the model retains the information of the most recent mutation on each ancestral line; the latter model supposes that each mutation occurs at a nucleotide site that is different from those of previous mutations, hence the model has the details of all mutations on the genealogy. In the infinitely-many-alleles model, the number of distinct alleles observed in a neutral sample follows a distribution specified by the Ewens sampling formula (Ewens, 1972), which appears (somehow surprisingly) in Theorem 2.5 for the shape of genealogies under strong selection.

Natural selection, discovered by Charles Darwin, is another force besides genetic drift and mutation, that drives evolution. Selection occurs when alleles in the population have different fitness. Gillespie (1986, 1991) argues that not all genetic similarity can be attributed to genetic drift and mutation. It is therefore important to examine models with selection. It is, however, complicated to incorporate natural selection into coalescent theory, because the fundamental assumption of exchangeability is invalid, as the fitness of an individual depends on its allelic type.

Nevertheless, Neuhauser and Krone (1997) and Krone and Neuhauser (1997) develop a model for describing the genealogy of a *random* sample in a population under selection and mutation, where all possible genealogies are embedded in an exchangeable process named the *ancestral selection graph*. Baake and Bialowons (2008) provide a review article. In the ancestral selection graph for a two-allele population, besides the coalescence events, an ancestral line can *branch* into an original continuing line

and a newly-created incoming line. However, one of the continuing and incoming lines is *virtual* (see Figure 1.2 for an illustration of how virtual lines arise when we trace a genealogy backwards in time). The virtual lines do not belong to the actual genealogy; they are used to capture the effect that individuals of a certain type are selectively favoured. Including virtual lines is essential for retaining the exchangeability of the coalescent.

1.1 Structure and main results

The structure of this thesis and the main results are as follows. In the remainder of this chapter, we describe in detail the ancestral selection graph in the two-type case for incorporating selection into the coalescent process . We illustrate in Proposition 1.1 that the number of branching events in the ancestral selection graph increases exponentially with the selection coefficient $\sigma > 0$. Thus, the ancestral selection graph is not suitable for our investigation of the genealogies in the strong selection limit. Further studies of the genealogies are primarily based on the *dual process with typed ancestral lines* developed by Etheridge and Griffiths (2009). The Moran model with selection and mutation, its diffusion approximation and the corresponding ancestral dual process with typed lines are also introduced in the present chapter. We also carefully explain, with Figure 1.2, how branching events occur in the ancestral dual process and how virtual lines emerge at branching events.

We use the Moran model to capture the way in which we can establish a duality between and forwards-in-time and backwards-in-time processes. We pass the Moran model to its diffusion approximation, so the behaviour of the population is simpler. We then derive the corresponding dual processes in the diffusion limit, which are the starting point of our studies of strong selection and selection-mutation balance limits. The properties of the genealogies are analysed when the parameters of the

dual processes tend to infinity. Our approach is in contrast to the approach where one starts with a Moran model with selection and mutation parameters being order 1 and then lets the population size $N \to \infty$, because many results for the forwards-in-time and backwards-in-time processes in the diffusion limit are in a simpler form.

In Chapter 2, we start to address the question of to what extent strong selection distorts the shape of genealogies. We show in Theorem 2.1 that in the dual process with typed lines, the real and virtual ancestral lines can be identified immediately as the genealogy is traced. One then only needs to simulate up to the time when the number of real lines reduces to one to extract information on the real genealogy. We derive in Theorem 2.3 that the number of branching events in the dual process with typed lines is finite in the limit as the selection coefficient $\sigma \to \infty$. We observe from simulation results in Tables 2.2 and 2.3 and in Figures 2.3 and 2.4 that, although selection does not have a significant impact on the genealogy of a sample with random type configuration (heuristically a random sample under strong selection only contains the favoured type so selection cannot operate), it does substantially affect the genealogy of a sample with given configuration. The expected height of the genealogy of a given sample under strong selection is smaller than that in Kingman's coalescent when there is more than one disfavoured individual in the sample; the genealogy is given by Kingman's coalescent otherwise. This observation makes intuitive sense by considering the forwards-in-time dynamics. There tends to be only one allele existing in the population under strong selection. When the population happens to contain a considerable proportion of the disfavoured allele (so we can draw a sample containing some individuals of type 1), the disfavoured allele will be eliminated rapidly from the population by the strong selective force. The probability that the most recent common ancestor (the remaining single line on the real genealogy) is the favoured type 2, is approximately 1 under strong selection, even when there are no type 2 individuals in the sample. We then derive explicit formulae for the limiting quantities

(under strong selection) that have been simulated. The formulae in Theorems 2.5, 2.15 and 2.16 are obtained by analysing various recursive equations for the desired expectations and variances, where the Ewens sampling formula plays an important role (see Remark 2.6). These theoretical results are novel and are consistent with the simulations.

In Chapter 3, we study the effect of strong selection at a process level with our dual process with typed lines and the coalescent-in-a-random-background model of Kaplan et al. (1988). Both the dual process with typed lines and the coalescent-in-a-random-background model can describe the genealogy of a given sample configuration. However, these two models are different in that in the dual process with typed lines, the ancestral lines can branch and virtual lines can appear, while in the coalescent-in-a-random-background model, there are no branching events in the ancestral process, and selection is reflected in the ancestral frequencies of alleles. In spite of this difference, the genealogies generated by these two models are the same. However, it is still of interest to study how strong selection affects the genealogies in both models. The ancestral process in the coalescent-in-a-random-background model is time-inhomogeneous, which we approximate by a time-homogeneous process, as the fluctuations in the ancestral frequencies under strong selection are negligible. We then derive in Theorem 3.5 that the genealogy in such an approximated model can be described by a limiting process, in which type 1 individuals in the sample are instantaneously converted to a random number of type 2 individuals according to a distribution that is related to the Ewens sampling formula, and then the resulting sample follows Kingman's coalescent. The limiting theorem is new, and its derivation is based on the time-scaling convergence result of Möhle (1998). However, in the dual process with typed lines, as the state space of the dual process is infinite due to the mechanism of virtual lines, Möhle's lemma cannot be applied exactly. Nevertheless, these limiting results are supported by simulations. The extension of the strong selection limit to the ancestral dual

process with three types is also investigated in the chapter. We assume individuals of type 1 and type 2 are strongly disfavoured, while individuals of type 3 are neutral. We observe that, although on different timescales, the disfavoured individuals in the sample instantaneously coalesce and mutate to the favoured type 3 according to the Ewens sampling formula, and then the resulting sample, which consists of individuals of the same type, evolves as in Kingman's coalescent on the original timescale.

In Chapter 4, we examine the genealogy of a sample taken from a two-allele population in which the mutation coefficient θ_1 (for type 2 individuals to mutate to type 1) and the selection coefficient σ (for eliminating type 1 individuals) are both large and comparable such that the forwards-in-time frequency can be approximated by a deterministic trajectory. We again use the constant frequency assumption in the studies of the limiting processes in this chapter. The limits of the ancestral dual processes with constant frequencies are deduced in Theorems 4.3 and 4.13. These two limiting theorems are new. We observe that in the dual process with typed lines, under selection-mutation balance, the evolution of the real lines is independent of the virtual lines, but not vice versa. However, on a faster timescale, the evolution of the scaled numbers of virtual lines is independent of the real lines. The evolution of the virtual lines on the faster timescale can be approximated by a deterministic process in Theorem 4.11, where the numbers of virtual lines increase from (ancestral) time 0 and then stabilise to constants. Although independently of the virtual lines, the real lines evolve on two timescales, and the limit is a time-changed Kingman's coalescent as derived in Theorem 4.13.

In Chapter 5, we consider a more general model where the type of each individual is described by a sequence of L diallelic loci. Let the two alleles on each locus be labelled 0 and 1. We suppose the reproductive ability of an individual is determined by the number of type 1 loci. Thus, there are $L + 1$ different types in the population. We formulate the Moran model with this additive selection scheme, and obtain the

diffusion approximation and the corresponding ancestral dual process with typed lines. Under a selection-mutation balance that is different from the assumption in Chapter 4 (compare equation (4.1) with equation (5.1)), we derive in Theorem 5.6 that the stable equilibrium point of the forwards-in-time frequencies satisfies a binomial formula. We then examine the real genealogy in this additive selection model. The limiting behaviour of the ancestral dual process can be described with the ideas used in Chapter 4, and the results are summarised in Theorems 5.17 and 5.19. In the final section, the limit under selection-mutation balance is extended to the case where $L \to \infty$. The number of type 0 loci that an individual has is finite in the limit as $L \to \infty$, as derived in Lemma 5.20. The corresponding forwards-in-time stationary frequencies become Poisson. Backwards in time, types in a sample are lost one by one on a faster timescale, and the sample is instantaneously converted to the one in which individuals are not composed of any type 0 locus. The real genealogy then follows a time-changed Kingman's coalescent. These results provide new insight into the studies of background selection model. The genealogies in the background selection model have been studied by various authors; see, e.g, Charlesworth et al. (1993, 1995), Hudson and Kaplan (1994, 1996) and Durrett (2008). Those authors investigate various properties of the genealogy under selection-mutation balance, e.g., the nucleotide diversity and the average time to lose a sampled type. We provide a full description of the genealogies backwards in time under the constant frequency assumption; see Remarks 5.21 and 5.23 for detailed discussions.

In the appendix, we present the algebra used in the derivation of some lemmas in the thesis.

1.2 Ancestral selection graph

The ancestral selection graph of Neuhauser and Krone (1997) and Krone and Neuhauser (1997) is a continuous-time Mpro with a branching-coalescing structure existing in the limit of large population size, which can be seen as an analogue of Kingman's coalescent under weak selection. In the ancestral selection graph, the types in the sample are not specified initially to make the ancestral lines exchangeable. We consider the ancestral process $\{\alpha_n(t), 0 \leq t \leq T_{UA}(n)\}$ for a sample of size $n \geq 1$, which is a continuous-time Markov process that counts the number of ancestral lines at time t in the past, starts with $\alpha_n(0) = n$ and stops at the time

$$T_{UA}(n) \coloneqq \inf\{t \geq 0 \colon \alpha_n(t) = 1\}.$$

The remaining single line is called the ultimate ancestor. If $\alpha_n(t) = \alpha \geq 2$, the following transitions occur:

$$\alpha \to \begin{cases} \alpha + 1 & \text{at rate } \frac{1}{2}\alpha\sigma, \\ \alpha - 1 & \text{at rate } \frac{1}{2}\alpha(\alpha - 1), \end{cases} \tag{1.2}$$

where $\sigma \geq 0$ is the selection coefficient. The state space of the ancestral process is $\mathbb{Z}_+ \setminus \{0\}$. The ancestral process will be absorbed into the ultimate ancestor in finite time almost surely, since the quadratic coalescence rate dominates the linear branching rate in this birth-death-type ancestral process (Etheridge, 2011, p. 73).

In the ancestral selection graph, besides the coalescence events, an ancestral line can branch into an original continuing line and a newly-created incoming line. To extract the *real* genealogy which is always a binary tree, knowledge about the types of ancestral lines is required. Once the entire ancestral selection graph has been built, the type of the ultimate ancestor is chosen according to the distribution of a sample of size one in the stationary population (such a sample can be simulated using the rejection algorithm developed by Donnelly et al., 2001). Mutations are then superimposed on the graph forwards in time as points of an independent Poisson process along

each ancestral line, independently of the coalescence and branching events, since all ancestral lines are exchangeable.

One then goes through the graph forwards in time from the ultimate ancestor to the sample, to assign types to ancestral lines and to determine branching events. At a coalescence event, the two ancestral lines involved have the same type as the ancestral line to which they coalesced. When a mutation is encountered, the type of ancestral line changes. The branching events are resolved by following the rules that incoming lines are kept if they are the favoured type. The real genealogy is then extracted by deleting all virtual lines. The remaining single line on the real genealogy is the *most recent common ancestor* (MRCA) of the sample, which may or may not coincide with the ultimate ancestor. Stephens and Donnelly (2003, Algorithm 3.1) provide an algorithm for simulating the ancestral selection graph. Nevertheless, an actual tree structure of the graph has to be simulated in order to discover the real genealogy.

The ancestral selection graph provides a tool for investigating the properties of a genealogy under selection and mutation. The graph can accommodate populations with more than two types, both infinitely-many-allele and infinitely-many-sites mutation schemes, and frequency dependent selection (Krone and Neuhauser, 1997; Neuhauser and Krone, 1997); the approach is also extended to the general diploid population by Donnelly and Kurtz (1999).

However, as we will see in Proposition 1.1, the number of branching events in the ancestral selection graph increases exponentially with the selection coefficient σ. Thus, simulations and studies of strong selection limit are impractical. For instance, when the mutation coefficient $\theta = 2$ (for mutating to both types) and the sample size $n = 10$, it takes 0.3 seconds to simulate 100 graphs for $\sigma = 1$; the computational time soars to 1.2 hours for $\sigma = 10$. Furthermore, these simulations require a considerable amount of memory on a computer to store the structure of each graph.

We notice that in the transition rates (1.2) of the ancestral process, as $\sigma \to \infty$,

the coalescence rate becomes negligible in comparison with the branching rate. The ancestral process behaves like a birth process with birth rate $\alpha\sigma/2$. Let $K_n(t) \in \{n, n+1, \dots\}$ be the number of descendants at time $t \geq 0$ of a sample of size n in this birth process. Then $P_m(t) \coloneqq \mathbb{P}(K_n(t) = m)$ satisfies

$$P_m(t+\delta t) = P_{m-1}(t)\left(\frac{1}{2}\sigma(m-1)\delta t\right) + P_m(t)\left(1 - \frac{1}{2}\sigma m \delta t\right) + o(\delta t), \tag{1.3}$$

where the probability that more than one birth event occurs in a small time interval of length δt is $o(\delta t)$. Equation (1.3) implies

$$\frac{\mathrm{d}P_m(t)}{\mathrm{d}t} = \frac{1}{2}\sigma(m-1)P_{m-1}(t) - \frac{1}{2}\sigma m P_m(t), \tag{1.4}$$

with initial value $P_n(0) = 1$. We can verify that the unique solution to equation (1.4) is

$$P_m(t) = \binom{m-1}{m-n} e^{-\frac{1}{2}\sigma n t}\left(1 - e^{-\frac{1}{2}\sigma t}\right)^{m-n}, \quad m = n, n+1, \dots.$$

Therefore, for fixed t, $K_n(t)$ has a negative binomial distribution with parameters n and $1 - \exp(-\sigma t/2)$, and

$$\mathbb{E}[K_n(t)] = n\left(1 - e^{-\frac{1}{2}\sigma t}\right)e^{\frac{1}{2}\sigma t} \sim n e^{\frac{1}{2}\sigma t},$$

for large σ.

The following proposition shows that the mean number of branching events until the ultimate ancestor increases exponentially with σ (see also the far-right column of Table 2.1). Similar results have been obtained by Ethier and Griffiths (1990, Theorem 5.1 and Corollary 5.2) in the context of the ancestral recombination graph.

Proposition 1.1 (Branching events in the ancestral selection graph)**.** *Consider the ancestral process with transitions* (1.2) *in the ancestral selection graph for a finite sample of size $n \geq 1$. Let $V(n)$ be the number of branching events until the ultimate*

ancestor. The expectation of $V(n)$ satisfies

$$\mathbb{E}[V(n)] = e^{\sigma} \sum_{l=0}^{n-2} \frac{l!}{\sigma^l} - \sum_{l=0}^{n-2} \frac{1}{\sigma^l} \sum_{k=l}^{n-2} \frac{k!}{(k-l)!}. \tag{1.5}$$

Proof. Let $\psi_n := \mathbb{E}[V(n)]$. By considering the first-step transitions of the Markov process, it follows that

$$(n-1+\sigma)\psi_n = (n-1)\psi_{n-1} + \sigma\psi_{n+1} + \sigma, \qquad n \geq 2, \tag{1.6}$$

with $\psi_1 = 0$. We can verify that a unique solution to the recursive equation (1.6) is

$$\psi_n = \sigma \int_0^1 [1-(1-x)^{n-1}] x^{-1} e^{\sigma x} \, \mathrm{d}x.$$

The boundary condition is clearly satisfied. For $n \geq 2$, we have

$$\begin{aligned}
(n-1+\sigma)\psi_n &= (n-1)\sigma \int_0^1 [1-(1-x)^{n-2}(1-x)] x^{-1} e^{\sigma x} \, \mathrm{d}x \\
&\quad + \sigma^2 \int_0^1 [1-(1-x)^{n-1}] x^{-1} e^{\sigma x} \, \mathrm{d}x \\
&= (n-1)\sigma \int_0^1 [1-(1-x)^{n-2}] x^{-1} e^{\sigma x} \, \mathrm{d}x + \sigma \\
&\quad + \sigma^2 \int_0^1 (1-x)^{n-1} e^{\sigma x} \, \mathrm{d}x + \sigma^2 \int_0^1 [x^{-1} - (1-x)^{n-1} x^{-1}] e^{\sigma x} \, \mathrm{d}x \\
&= (n-1)\psi_{n-1} + \sigma\psi_{n+1} + \sigma.
\end{aligned}$$

To obtain the desired expression for $\mathbb{E}[V(n)] \equiv \psi_n$ as in equation (1.5), we write

$$\begin{aligned}
\mathbb{E}[V(n)] &= \sigma e^{\sigma} \int_0^1 \frac{1-(1-x)^{n-1}}{1-(1-x)} e^{-\sigma(1-x)} \, \mathrm{d}x \\
&= \sigma e^{\sigma} \int_0^1 \sum_{k=0}^{n-2} (1-x)^k e^{-\sigma(1-x)} \, \mathrm{d}x.
\end{aligned}$$

Let $y = \sigma(1-x)$. It follows that

$$\mathbb{E}[V(n)] = e^{\sigma} \int_0^{\sigma} \sum_{k=0}^{n-2} \frac{y^k}{\sigma^k} e^{-y} \, \mathrm{d}y$$

$$= e^{\sigma}\left(\int_0^{\infty} \sum_{k=0}^{n-2} \frac{y^k}{\sigma^k} e^{-y}\, \mathrm{d}y - \int_{\sigma}^{\infty} \sum_{k=0}^{n-2} \frac{y^k}{\sigma^k} e^{-y}\, \mathrm{d}y \right)$$
$$= e^{\sigma}\left(\sum_{k=0}^{n-2} \frac{k!}{\sigma^k} - e^{-\sigma} \sum_{k=0}^{n-2} \frac{k!}{\sigma^k} \sum_{j=0}^{k} \frac{\sigma^j}{j!} \right),$$

where the last equality follows Abramowitz and Stegun (1965, Equation 6.5.29). By setting $l = k - j$, we have

$$\mathbb{E}[V(n)] = e^{\sigma} \sum_{k=0}^{n-2} \frac{k!}{\sigma^k} - \sum_{k=0}^{n-2} \sum_{l=0}^{k} \frac{k!}{(k-l)!} \frac{1}{\sigma^l}$$
$$= e^{\sigma} \sum_{l=0}^{n-2} \frac{l!}{\sigma^l} - \sum_{l=0}^{n-2} \frac{1}{\sigma^l} \sum_{k=l}^{n-2} \frac{k!}{(k-l)!},$$

where the last equality follows from interchanging the order of summation. □

Further investigation of genealogies under selection and mutation in this thesis will be mainly based on an ancestral dual process developed by Etheridge and Griffiths (2009). This *dual process with typed ancestral lines* is similar to the ancestral selection graph but has the following advantages: firstly, this dual process allows us to investigate the genealogy of a sample with given type configuration (cf. the genealogy of a random sample in the ancestral selection graph); secondly, as we will show in Theorem 2.1, although the branching events can still occur in the dual process and introduce virtual lines, the genealogy is *controlled* in the sense that the real and virtual lines can be identified immediately so we know which lines are ancestral to the sample and which are not (cf. the real genealogy is extracted after an ultimate ancestor is reached in the ancestral selection graph). This dual process is certainly more practical for simulations, and for the analysis of whether, and to what extent, selection and mutation affect genealogies of samples with given configuration, particularly when selection and mutation are strong. The shortcoming of the dual process with typed lines is that unless one has convenient expressions for the sampling formulae under the stationary distribution (which is only the case for parent-independent mutation), it is then

numerically intensive to evaluate the sampling formulae in the transition rates of the dual process.

In the remaining part of this chapter, we describe the stochastic models that underline the studies in the remainder of the thesis. In Section 1.3, we present a Moran model with selection and mutation, and its diffusion approximation. In Section 1.4, a graphical representation of the Moran model is plotted in Figure 1.2, which we use to illustrate how branching events and virtual ancestral lines emerge in the ancestral process. The ancestral process with typed lines that is dual to the forwards-in-time diffusion, is derived in Theorem 1.9 with the method of duality in Definition 1.8. However, the strong duality can only be established with respect to the Moran model; the duality in the diffusion limit is weak.

1.3 Moran model with selection and mutation

To avoid ambiguity, we specify a Moran model with purifying selection and parent-independent mutation as follows. We only use the Moran model to introduce the diffusion approximation in the limit of large population size. The subsequent studies of strong selection and strong mutation limits are based on the dual processes in the diffusion limit.

Consider a haploid and panmictic population in which there are N individuals, each of which is either the selectively disfavoured type 1 or the favoured type 2. Each individual of type i produces an offspring of the same type at rate $\lambda_i \geq 0$, $i = 1, 2$. Simultaneously, an individual is chosen uniformly at random from the population to die (so the population size is kept constant); an offspring may replace its own parent. We assume the type 2 individuals have a *fertility* selective advantage by setting $\lambda_2 = 1$ and

$$\lambda_1 = 1 - s, \quad 0 \leq s \leq 1.$$

Hence, selection is purifying, as its effect is to eliminate the disfavoured allele from the population.

Mutations are assumed to occur continuously in time, independently of the reproduction process. Each individual mutates at rate $\mu \geq 0$. When a mutation occurs, regardless of the present type of the individual, an individual has probability p_i of mutating to type i, $i = 1, 2$, and $p_1 + p_2 = 1$. The null-mutations that do not change the type of an individual can be removed in the ancestral dual process (see Theorem 1.9). In the two-type case, a more general parent-dependent mutation model can also be rewritten as an equivalent parent-independent mutation model. The parent-independent mutation is desirable when there are more than two types, because the explicit form of the stationary distribution as that in Proposition 1.3 is then known.

Let $Z_1(t)$ denote the number of type 1 individuals in the Moran model population at time $t \geq 0$. Since the population size is fixed for all time t, the evolution of the population can be described by the continuous-time Markov process $\{Z_1(t), t \geq 0\}$ with state space $S = \{0, 1, \ldots, N\}$. For $z, k \in S$ and $z \neq k$, the transition rates of the process $\{Z_1(t), t \geq 0\}$ are

$$q_N(z,k) = \begin{cases} (1-s)z\frac{N-z}{N} + \mu p_1(N-z), & \text{if } k = z+1, \\ (N-z)\frac{z}{N} + \mu p_2 z, & \text{if } k = z-1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.7}$$

For a bounded function $f\colon S \to \mathbb{R}$, the generator of $\{Z_1(t), t \geq 0\}$ is

$$\begin{aligned} \mathcal{L}_N f(z) &\coloneqq \lim_{\delta t \to 0} \frac{1}{\delta t} \mathbb{E}[f(Z_1(t+\delta t)) - f(Z_1(t)) \mid Z_1(t) = z] \\ &= [f(z+1) - f(z)]q_N(z, z+1) + [f(z-1) - f(z)]q_N(z, z-1). \end{aligned} \tag{1.8}$$

Definition 1.2 (Moran model with selection and mutation)**.** *We call the process* $\{Z_1(t), t \geq 0\}$ *with generator given by equation* (1.8) *the Moran model with selection and mutation.*

When the population in the Moran model with selection and two-way mutation has

evolved indefinitely in time, the process $\{Z_1(t), t \geq 0\}$ will approach an equilibrium such that

$$\lim_{t\to\infty} \mathbb{P}(Z_1(t) = z \mid Z_1(0) = k) = \pi(z), \quad \text{for all } z, k \in S.$$

Let $B(a, b)$ be a beta function defined as

$$B(a, b) \coloneqq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad a > 0, \quad b > 0,$$

where $\Gamma(a)$ is a gamma function such that

$$\Gamma(a) \coloneqq \int_0^\infty e^{-u} u^{a-1} \, \mathrm{d}u, \quad a > 0.$$

In the case of parent-independent mutation, an exact formula is known for the stationary distribution $\pi(z)$ (see Etheridge and Griffiths, 2009).

Proposition 1.3 (Stationary distribution of the Moran model)**.** *The process* $\{Z_1(t), t \geq 0\}$ *has a unique stationary distribution with probability mass function*

$$\pi(z) = C\lambda_1^z \lambda_2^{N-z} \binom{N}{z} \frac{B(N\mu p_1/\lambda_1 + z, N\mu p_2/\lambda_2 + N - z)}{B(N\mu p_1/\lambda_1, N\mu p_2/\lambda_2)}, \quad 0 \leq z \leq N, \qquad (1.9)$$

where C *is the normalising constant such that* $\sum_{z=0}^{N} \pi(z) = 1$.

The process $\{Z_1(t), t \geq 0\}$ is time-reversible with respect to its stationary distribution, as it is readily verified that the above probability mass function (1.9) satisfies the following detailed balance equation:

$$\pi(z) q_N(z, z+1) = \pi(z+1) q_N(z+1, z), \quad \text{for all } 0 \leq z < N.$$

When the forces of genetic drift, selection and mutation are comparable in size, the process $\{X_N(t), t \geq 0\}$, where $X_N(t) \coloneqq Z_1(t)/N$, can be approximated by the diffusion process named after Sewall Wright and Ronald Fisher, in the limit as $N \to \infty$, provided that we speed up time by a factor of $N/2$. Ewens (2004) and Etheridge (2011) provide a comprehensive reference for diffusion processes arising in mathematical population

genetics.

We assume $Ns = \sigma$ and $N\mu p_i = \theta_i$, $i = 1, 2$. Let $\theta = \theta_1 + \theta_2$. The derivation of the ancestral dual process in Theorem 1.9 relies on the following diffusion approximation.

Proposition 1.4 (Wright-Fisher diffusion)**.** *The process* $\{X_N(Nt/2), t \geq 0\}$ *converges in distribution to a diffusion process* $\{X(t), t \geq 0\}$ *on* $[0, 1]$*, as* $N \to \infty$*. The generator of the diffusion process, evaluated on a twice continuously differentiable function* $f\colon [0, 1] \to \mathbb{R}$*, is*

$$\mathcal{L}f(x) = \frac{1}{2}x(1-x)\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} + \frac{1}{2}[-\sigma x(1-x) + \theta_1 - \theta x]\frac{\mathrm{d}f(x)}{\mathrm{d}x}. \tag{1.10}$$

Proof. Consider the generator $\mathcal{L}_N$ defined in equation (1.8) with scaled parameters. Let $x = z/N$. The effect of speeding up time is to multiply $\mathcal{L}_N$ by a factor of $N/2$. It follows that

$$\begin{aligned}
\lim_{N\to\infty} \frac{N}{2}\mathcal{L}_N f(x) = \lim_{N\to\infty} \Bigg\{ &\left[f\left(x + \frac{1}{N}\right) - f(x)\right] \times \left[\frac{N}{2}\left(1 - \frac{\sigma}{N}\right)Nx(1-x) + \frac{N}{2}\theta_1(1-x)\right] \\
&+ \left[f\left(x - \frac{1}{N}\right) - f(x)\right] \times \left[\frac{N^2}{2}(1-x)x + \frac{N}{2}\theta_2 x\right]\Bigg\} \\
= {} & \frac{1}{2}[-\sigma x(1-x) + \theta_1 - \theta x]\,\frac{\mathrm{d}f(x)}{\mathrm{d}x} + \frac{1}{2}x(1-x)\,\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \\
= {} & \mathcal{L}f(x),
\end{aligned}$$

where the Taylor expansion is applied to the functions $f(x + 1/N)$ and $f(x - 1/N)$ about the point x to obtain the second equality. The convergence of the generator is uniform in x. In addition, the state space $[0, 1]$ is compact. The weak convergence of the process $\{X_N(Nt/2), t \geq 0\}$ to the diffusion then follows Ethier and Kurtz (2005, Theorem 6.5 in Chapter 1 and Theorem 7.8 in Chapter 3). □

Denote the set of non-negative real numbers by $\mathbb{R}_+$. Let $p\colon \mathbb{R}_+ \times [0, 1] \times [0, 1] \to \mathbb{R}_+$ be the transition density function of the diffusion $\{X(t), t \geq 0\}$ such that

$$\mathbb{P}[X(t) \in I \mid X(0) = y] = \int_I p(t, y, x)\,\mathrm{d}x, \quad I \subseteq [0, 1].$$

When a sufficiently long period of time has passed, the diffusion will eventually approach an equilibrium such that $\lim_{t\to\infty} p(t, y, x) = \pi(x)$ exists for all $x, y \in [0, 1]$. The density $\pi(x)$ is a special case of the Wright's formula (Wright, 1931, p. 123), which can be computed by solving the following Kolmogorov forward equation:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}[x(1-x)\pi(x)] - \frac{\mathrm{d}}{\mathrm{d}x}[(-\sigma x(1-x) + \theta_1 - \theta x)\pi(x)] = 0.$$

The construction of the dual process in Theorem 1.9 will depend crucially on such stationary distribution.

Proposition 1.5 (Stationary density of the Wright-Fisher diffusion)**.** *The Wright-Fisher diffusion with generator* (1.10) *has a unique stationary distribution with density*

$$\pi(x) = \frac{x^{\theta_1-1}(1-x)^{\theta_2-1}e^{-\sigma x}}{B(\theta_1, \theta_2)M(\theta_1, \theta, -\sigma)}, \quad 0 < x < 1, \tag{1.11}$$

where

$$M(\theta_1, \theta, -\sigma) = \frac{1}{B(\theta_1, \theta_2)} \int_0^1 u^{\theta_1-1}(1-u)^{\theta_2-1}e^{-\sigma u}\,\mathrm{d}u, \quad \theta_1 > 0, \quad \theta_2 > 0, \tag{1.12}$$

is the confluent hypergeometric function of the first kind.

Remark 1.6 (Confluent hypergeometric function)**.** In the neutral case where $\sigma = 0$, $M(\theta_1, \theta, 0) = 1$. There is also a useful Kummer's transformation

$$M(\theta_1, \theta, -\sigma) = e^{-\sigma}M(\theta_2, \theta, \sigma), \quad \sigma \geq 0.$$

One can refer to Abramowitz and Stegun (1965, Chapter 13) for more properties of the confluent hypergeometric functions.

Figure 1.1 plots the stationary density (1.11) with respect to different values of mutation coefficients θ_1 and θ_2, and selection coefficient σ. The figure provides some understanding of the long-term balance of the competing forces of genetic drift, selection and mutation. When there is no selection ($\sigma = 0$) and mutation is very weak

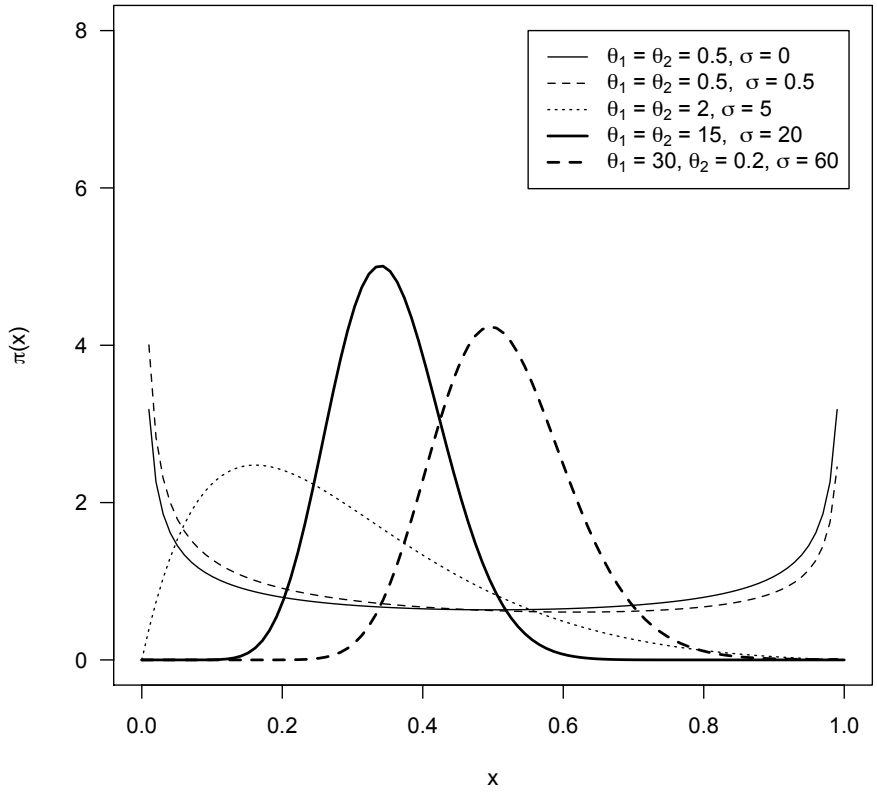


Figure 1.1: The stationary density of the frequency of the disfavoured type 1 individuals in the diffusion limit, as given in equation (1.11), with respect to various strengths of selection and mutation.

($\theta_1 = \theta_2 = 0.5$), genetic drift is the dominant evolutionary force, and the probability mass concentrates around 0 and 1. The (non-bold) dashed and dotted curves illustrate that selection pushes the frequency of the disfavoured allele towards 0. The frequencies of alleles achieve a balance in the population when the strengths of selection and mutation are large and comparable (see the bold solid curve). Moreover, the bold dashed curve shows that the frequencies are also balanced when the values of θ_1 and σ are both large and comparable.

The sampling probability in the following proposition will become useful in the derivation of the ancestral dual process in Theorem 1.9.

Proposition 1.7 (Sampling probability in the diffusion limit)**.** *Consider an* ordered *sample with type configuration* $(n_1, n_2) \in \mathbb{Z}_+^2$ *of size* $n = n_1 + n_2 \geq 1$*. The probability of drawing such a sample from the stationary population governed by the diffusion with generator* (1.10)*, is*

$$\mathscr{H}(n_1, n_2) = \frac{B(n_1 + \theta_1, n_2 + \theta_2) M(n_1 + \theta_1, n + \theta, -\sigma)}{B(\theta_1, \theta_2) M(\theta_1, \theta, -\sigma)}. \tag{1.13}$$

Proof. Let X_1 be the stationary frequency of type 1 individuals in the population governed by the diffusion with generator (1.10). We have

$$\mathscr{H}(n_1, n_2) = \mathbb{E}(X_1^{n_1} X_2^{n_2}),$$

where $X_2 = 1 - X_1$. Since X_1 has the probability density function $\pi(x)$ in equation (1.11), it follows that

$$\begin{aligned} \mathscr{H}(n_1, n_2) &= \int_0^1 x^{n_1} (1-x)^{n_2} \pi(x) \, \mathrm{d}x \\ &= \frac{B(n_1 + \theta_1, n_2 + \theta_2) M(n_1 + \theta_1, n + \theta, -\sigma)}{B(\theta_1, \theta_2) M(\theta_1, \theta, -\sigma)}. \end{aligned}$$

□

The probability of obtaining the corresponding unordered sample is $\binom{n}{n_1} \mathscr{H}(n_1, n_2)$, since there are $\binom{n}{n_1}$ possible orderings of a specific unordered sample, and every ordered sample has the same distribution $\mathscr{H}(n_1, n_2)$.

1.4 Ancestral dual process with typed lines

The algebraic derivation of the ancestral process with typed lines that is dual to the Moran model is provided by Etheridge and Griffiths (2009) for a general case of $d \geq 2$ types under selection and parent-dependent mutation. Etheridge et al. (2010) demonstrate that similar algebra can be applied to the construction of dual processes in the Λ-coalescent limit. However, instead of using the results of Etheridge and

Griffiths (2009) and focusing on two types and parent-independent mutation, and then taking the diffusion limit, we derive the dual process in Theorem 1.9 directly from the diffusion process with generator (1.10). The purpose is to illustrate the method of duality in Definition 1.8 and to present the dual process in the diffusion limit for later usage.

We first demonstrate how an ancestral process arises naturally as a dual to the Moran model of Definition 1.2. The key is to formulate the Moran model as a biased voter model with mutation on a complete graph (see the review by Liggett, 2004). The voter model is introduced into mathematical population genetics by Donnelly (1984) to study the Moran model with mutation. The biased voter model is used to study the Moran model with selection by Krone and Neuhauser (1997) and Etheridge and Griffiths (2009).

A graphical representation of the Moran model is shown in Figure 1.2. The vertical lines represent ancestral lines, time is run from the top of the graph to the bottom, and there is no spatial structure among the lines.

Recall that in the Moran model of Definition 1.2 we set the reproduction rates $\lambda_1 = 1 - s$ and $\lambda_2 = 1$. The Moran model starts forwards in time with the types of the individuals being known. For each ordered pair of lines, arrows are drawn between the lines at the points of a Poisson process with rate λ_2/N to model the reproduction events. The individual at the tail of an arrow is a potential ancestor; it may replace the individual at the head. The selection difference is reflected such that when the potential ancestor is the disfavoured type 1, the individual at the head of the arrow is replaced with probability $\lambda_1/\lambda_2 = 1 - s$ (a thinning procedure is used to produce a Poisson process that gives arrivals at a smaller rate); when the potential ancestor is the favoured type 2, the individual at the head is replaced with probability $\lambda_2/\lambda_2 = 1$. Mutations are superimposed on each ancestral line as the points of an independent Poisson process with rate μ. Given a mutation occurs on a line, the type of the line

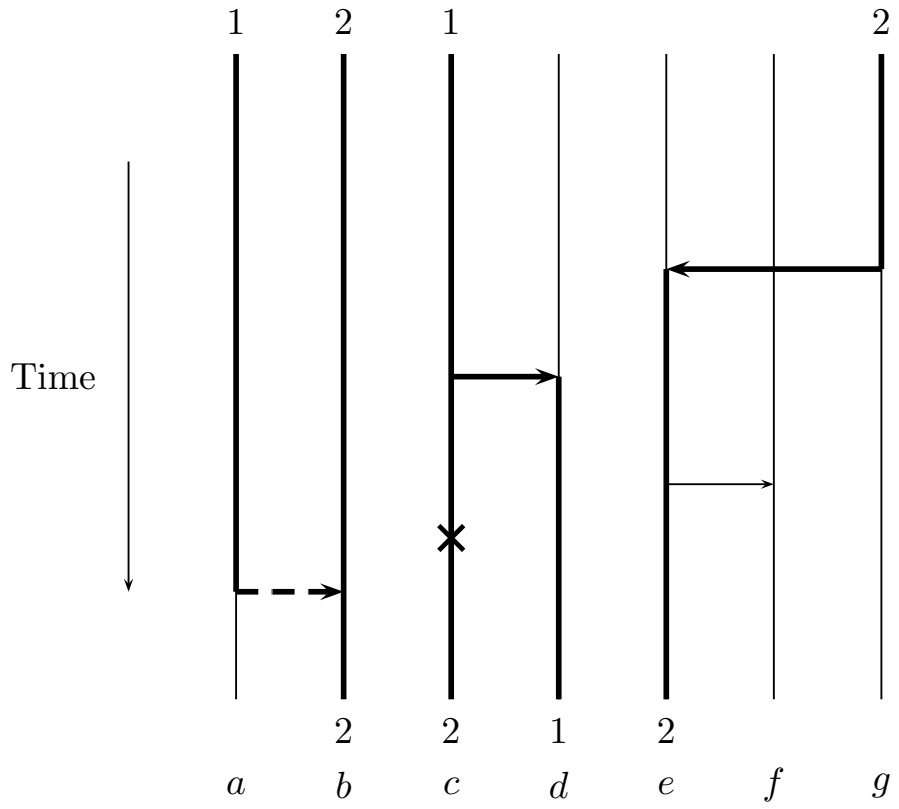


Figure 1.2: A graphical representation of the Moran model with selection and mutation. Forwards in time, line g reproduces and the offspring replaces line e; line c replaces line d; line e replaces line f; a mutation occurs on line c and changes the line to type 2; and line a *fails* to replace line b. Take a sample of the individuals b, c, d and e at the bottom of the graph, whose types are known to be 2, 2, 1 and 2. As the ancestral lines are traced backwards in time in the dual process, line b *branches* and a virtual line a is added, because line a fails to replace line b forwards in time, and line a is not in the sample; line c mutates from type 2 to type 1; the reproduction event from line e to line f does not affect the dual process; lines c and d coalesce, as both lines are in the sample; and line e is replaced by line g, but this event does not appear in the dual process because the event does not change the distribution of the ancestral lines.

changes to i with probability p_i, $i = 1, 2$.

We can recover the genealogy of a sample with given configuration by tracing the ancestral lines backwards in time (see Figure 1.2). In particular, when the potential ancestor *fails* to replace an individual forwards in time, *and* the potential ancestor is outside the genealogy of the sample, this ancestor becomes *virtual*.

The ancestral dual process constructed in Theorem 1.9 is based on the following

method of duality of Ethier and Kurtz (2005, Section 4 in Chapter 4).

Definition 1.8 (Duality)**.** *Consider two Markov processes* $\{X(t), t \geq 0\}$ *and* $\{\alpha(t), t \geq 0\}$ *with state spaces* S_X *and* S_α*, respectively. The two processes are dual to each other with respect to a function* $f\colon S_X \times S_\alpha \to \mathbb{R}$*, if*

$$\mathbb{E}[f(X(t), \alpha) \mid X(0) = x] = \mathbb{E}[f(x, \alpha(t)) \mid \alpha(0) = \alpha], \tag{1.14}$$

for all $x \in S_X$*,* $\alpha \in S_\alpha$ *and* $t \geq 0$*.*

With the power of the duality equation (1.14), we can use the knowledge of the forwards-in-time process to study the backwards-in-time dual process.

Consider a sample of finite size taken from the stationary population governed by the diffusion with generator (1.10). The numbers and types of the ancestral lines at time $t \geq 0$ in the past can be represented by a two-dimensional vector $\boldsymbol{\alpha}(t)$, where $\alpha_i(t)$, $i = 1, 2$ is the number of ancestral lines of type i at time t.

Theorem 1.9 (Dual process with typed lines in the diffusion limit)**.** *For a sample* $\boldsymbol{n} \in \mathbb{Z}_+^2$ *of size* $|\boldsymbol{n}| \geq 1$*, the ancestral dual process* $\{\boldsymbol{\alpha}(t), t \geq 0\}$ *is a two-dimensional Markov process with state space* $\mathbb{Z}_+^2 \setminus \{\mathbf{0}\}$ *and initial state* $\boldsymbol{n}$*. The non-zero entries in the* $\boldsymbol{\alpha}$*th row of the Q-matrix of the dual process are*

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha} + \boldsymbol{e}_1) = \frac{1}{2}\alpha\sigma\frac{(\alpha_1 + \theta_1)}{\alpha + \theta}\frac{\nu(\boldsymbol{\alpha} + \boldsymbol{e}_1 + \boldsymbol{\theta})}{\nu(\boldsymbol{\alpha} + \boldsymbol{\theta})}, \tag{1.15}$$

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_i) = \frac{1}{2}\alpha_i(\alpha_i - 1)\frac{\alpha - 1 + \theta}{\alpha_i - 1 + \theta_i}\frac{\nu(\boldsymbol{\alpha} - \boldsymbol{e}_i + \boldsymbol{\theta})}{\nu(\boldsymbol{\alpha} + \boldsymbol{\theta})}, \tag{1.16}$$

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha} + \boldsymbol{e}_i - \boldsymbol{e}_j) = \frac{1}{2}\alpha_j\theta_j\frac{\alpha_i + \theta_i}{\alpha_j - 1 + \theta_j}\frac{\nu(\boldsymbol{\alpha} + \boldsymbol{e}_i - \boldsymbol{e}_j + \boldsymbol{\theta})}{\nu(\boldsymbol{\alpha} + \boldsymbol{\theta})}, \quad i \neq j, \tag{1.17}$$

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = -\frac{1}{2}[\alpha_1\sigma + \alpha(\alpha - 1) + \alpha_1\theta_2 + \alpha_2\theta_1], \tag{1.18}$$

where $i, j = 1, 2$*,* $\alpha = |\boldsymbol{\alpha}|$*,* $\theta = |\boldsymbol{\theta}|$*, and*

$$\nu(\boldsymbol{\theta}) = \mathbb{E}[e^{-\sigma\xi}] = M(\theta_1, \theta, -\sigma)$$

with ξ being a random variable having the Beta(θ_1, θ_2) distribution and $M(\theta_1, \theta, -\sigma)$ being the confluent hypergeometric function defined in equation (1.12)*.*

Proof. The subtlety of the construction of the dual process with typed lines is to apply the diffusion generator $\mathcal{L}$ in equation (1.10) to a specific test function of the form $x^{\alpha_1}(1-x)^{\alpha_2}$. The quantity $x^{\alpha_1}(1-x)^{\alpha_2}$ can be thought of as the probability of sampling α_1 individuals of type 1 and α_2 individuals of type 2 from the population composed by frequencies x and $1-x$. We obtain

$$\begin{aligned}\mathcal{L}x^{\alpha_1}(1-x)^{\alpha_2} = &\frac{1}{2}\alpha\sigma x^{\alpha_1+1}(1-x)^{\alpha_2} \\ &+ \frac{1}{2}\alpha_1(\alpha_1-1)x^{\alpha_1-1}(1-x)^{\alpha_2} + \frac{1}{2}\alpha_2(\alpha_2-1)x^{\alpha_1}(1-x)^{\alpha_2-1} \\ &+ \frac{1}{2}\alpha_1\theta_1 x^{\alpha_1-1}(1-x)^{\alpha_2+1} + \frac{1}{2}\alpha_2\theta_2 x^{\alpha_1+1}(1-x)^{\alpha_2-1} \\ &- \frac{1}{2}[\alpha_1\sigma + \alpha(\alpha-1) + \alpha_1\theta_2 + \alpha_2\theta_1]x^{\alpha_1}(1-x)^{\alpha_2}. \end{aligned} \tag{1.19}$$

Let X be the stationary frequency with density (1.11), and $\mathscr{H}(\boldsymbol{\alpha})$ be the sampling probability derived in Proposition 1.7. Define

$$g_{\boldsymbol{\alpha}}(x) = \frac{x^{\alpha_1}(1-x)^{\alpha_2}}{\mathbb{E}[X^{\alpha_1}(1-X)^{\alpha_2}]} = \frac{x^{\alpha_1}(1-x)^{\alpha_2}}{\mathscr{H}(\boldsymbol{\alpha})}. \tag{1.20}$$

By applying the diffusion generator $\mathcal{L}$ in equation (1.10) to $g_{\boldsymbol{\alpha}}(x)$ and using equation (1.19), it follows that

$$\begin{aligned}\mathcal{L}g_{\boldsymbol{\alpha}}(x) = &\frac{1}{2}\alpha\sigma\frac{\mathscr{H}(\boldsymbol{\alpha}+\boldsymbol{e}_1)}{\mathscr{H}(\boldsymbol{\alpha})}g_{\boldsymbol{\alpha}+\boldsymbol{e}_1}(x) + \frac{1}{2}\sum_{i=1}^{2}\alpha_i(\alpha_i-1)\frac{\mathscr{H}(\boldsymbol{\alpha}-\boldsymbol{e}_i)}{\mathscr{H}(\boldsymbol{\alpha})}g_{\boldsymbol{\alpha}-\boldsymbol{e}_i}(x) \\ &+ \frac{1}{2}\sum_{i=1}^{2}\sum_{\substack{j=1\\ j\neq i}}^{2}\alpha_j\theta_j\frac{\mathscr{H}(\boldsymbol{\alpha}+\boldsymbol{e}_i-\boldsymbol{e}_j)}{\mathscr{H}(\boldsymbol{\alpha})}g_{\boldsymbol{\alpha}+\boldsymbol{e}_i-\boldsymbol{e}_j}(x) \\ &- \frac{1}{2}[\alpha_1\sigma + \alpha(\alpha-1) + \alpha_1\theta_2 + \alpha_2\theta_1]g_{\boldsymbol{\alpha}}(x) \\ = &[g_{\boldsymbol{\alpha}+\boldsymbol{e}_1}(x) - g_{\boldsymbol{\alpha}}(x)] \times \frac{1}{2}\alpha\sigma\frac{\mathscr{H}(\boldsymbol{\alpha}+\boldsymbol{e}_1)}{\mathscr{H}(\boldsymbol{\alpha})} \\ &+ \sum_{i=1}^{2}[g_{\boldsymbol{\alpha}-\boldsymbol{e}_i}(x) - g_{\boldsymbol{\alpha}}(x)] \times \frac{1}{2}\alpha_i(\alpha_i-1)\frac{\mathscr{H}(\boldsymbol{\alpha}-\boldsymbol{e}_i)}{\mathscr{H}(\boldsymbol{\alpha})}\end{aligned}$$

$$+\sum_{i=1}^{2}\sum_{\substack{j=1\\j\neq i}}^{2}[g_{\boldsymbol{\alpha}+\boldsymbol{e}_i-\boldsymbol{e}_j}(x)-g_{\boldsymbol{\alpha}}(x)]\times\frac{1}{2}\alpha_j\theta_j\frac{\mathscr{H}(\boldsymbol{\alpha}+\boldsymbol{e}_i-\boldsymbol{e}_j)}{\mathscr{H}(\boldsymbol{\alpha})}. \tag{1.21}$$

Now it should be clear that the test function (1.20) is chosen because $\mathbb{E}[g_{\boldsymbol{\alpha}}(X)]=1$ for all choices of $\boldsymbol{\alpha}$, and $\mathbb{E}[\mathcal{L}g_{\boldsymbol{\alpha}}(X)]=0$ when the expectation is taken with respect to the stationary distribution of the diffusion. Hence, by using the method of duality in Definition 1.8, and considering $g_{\cdot}(\cdot)$ as functions of the configuration $\boldsymbol{\alpha}$ of ancestral lines taken from the stationary population, rather than as functions of the frequency x, those terms after $g_{\boldsymbol{\beta}}(x)-g_{\boldsymbol{\alpha}}(x)$ in equation (1.21) can be interpreted as the transition rates of a Markov process.

The transition rates (1.15) to (1.17) then follow a minor algebraic manipulation using equation (1.13) for functions $\mathscr{H}(\cdot)$ and the definition of functions $\nu(\cdot)$.

Etheridge and Griffiths (2009) show that the strong duality can be established with respect to the Moran model through the graphical representation such as that in Figure 1.2. However, in the diffusion limit, the duality is weak. □

It is clear from the transition rates (1.15) to (1.17) that the types are carried at all times by ancestral lines in the dual process, as opposed to the genealogical history being simulated and types assigned later in the ancestral selection graph described in Section 1.2.

The transition $\boldsymbol{\alpha}\to\boldsymbol{\alpha}+\boldsymbol{e}_1$ with rate (1.15) corresponds to a branching event; $\boldsymbol{\alpha}\to\boldsymbol{\alpha}-\boldsymbol{e}_i$ with rate (1.16) corresponds to a coalescence event; and $\boldsymbol{\alpha}\to\boldsymbol{\alpha}+\boldsymbol{e}_i-\boldsymbol{e}_j$, $i\neq j$ with rate (1.17) corresponds to a mutation event. The total transition rate out of the state $\boldsymbol{\alpha}$ is minus the diagonal entry (1.18), i.e.,

$$-q(\boldsymbol{\alpha},\boldsymbol{\alpha})=\lim_{\delta t\to 0}\frac{1}{\delta t}\mathbb{P}[\boldsymbol{\alpha}(t+\delta t)\neq\boldsymbol{\alpha}\mid\boldsymbol{\alpha}(t)=\boldsymbol{\alpha}].$$

Remark 1.10 (Mutations in the dual process)**.** Although in the forwards-in-time process mutations occur independently of the reproduction events, mutations do not

have a simple independence interpretation in the ancestral dual process with typed lines.

Chapter 2

Genealogies under strong selection

The forwards-in-time process whereby a mutant with a strong selective advantage arises in an otherwise neutral population and eventually spreads to the entire population is called a selective sweep, which can have a dramatic effect on levels of variation on the selective locus (and on the neighbouring neutral locus). Selective sweeps and their effect on genetic variation have been investigated (usually with recombination) by a number of authors (see Stephan et al., 1992; Barton, 1998; Durrett and Schweinsberg, 2004; Yu et al., 2012). The spread of the favoured allele in the middle phase of the selective sweep can be approximated by a deterministic process, while the beginning and the end of the sweep are modelled by branching processes. Our intention in this chapter is to provide quantitative and qualitative insights into the shape of the genealogy of a sample under strong selection, using the ancestral dual process with typed lines introduced in Theorem 1.9. We assume the selection coefficient $\sigma \to \infty$ while the mutation coefficients θ_1 and θ_2 remain constant.

For large σ, as we will see in Theorem 2.15, the probability of obtaining an unordered sample (n_1, n_2) is $\theta_{1(n_1)}/\sigma^{n_1} + O(\sigma^{-n_1-1})$, where we write $x_{(n)} = x(x+1)\cdots(x+n-1)$ for $n \in \mathbb{Z}_+$ with $x_{(0)} = 1$. Therefore, forwards in time there tends to be only one allele existing in the population under strong selection. When the population

happens to contain a considerable proportion of the disfavoured allele (so we can draw a sample containing some individuals of type 1), the disfavoured allele will be eliminated rapidly from the population by the strong selective force. We will see in the remainder of this chapter that the genealogies under strong selection exactly reflect this heuristic.

This chapter has the following structure. In Section 2.1, we briefly review the properties of the neutral genealogy generated by Kingman's coalescent. In Section 2.2, we consider the simulation of the ancestral dual process with typed lines. We decompose the ancestral lines into real and virtual ones in Theorem 2.1. It is then unnecessary to simulate beyond the most recent common ancestor (MRCA) in order to extract information on the real genealogy. A realisation of the genealogy is illustrated in Figure 2.2. In Section 2.3, we analyse the properties of the real genealogy for various combinations of selection and mutation coefficients through simulations. We observe the effect of selection and mutation on the coalescence times in Section 2.3.1, on the time to the MRCA and the length of the genealogy in Section 2.3.2, and on the distribution of the types of the MRCA in Section 2.3.3, with particular emphasis on samples with given configuration under strong selection. In Section 2.4, we derive explicit formulae in Theorems 2.3, 2.5, 2.15 and 2.16 and in Corollaries 2.13 and 2.14 for the limiting values of the quantities we simulate, through computations of various recursive equations. These theoretical results are consistent with the simulations. In the generic recursive equations in Lemmas 2.9 and 2.11 and in Corollary 2.10, we also provide the terms of $O(1/\sigma)$, from which the error of our approximation of the limiting expectations and variances can be computed.

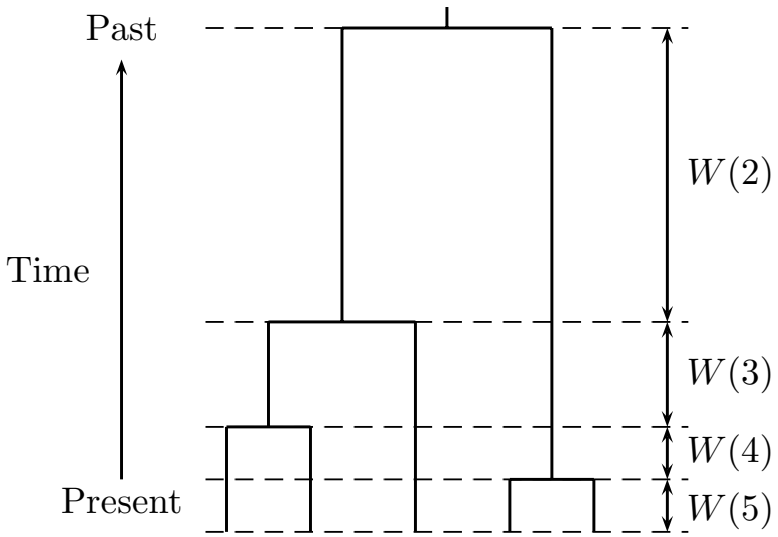


Figure 2.1: A genealogy in Kingman's coalescent for a sample of size $n = 5$.

2.1 Genealogies in Kingman's coalescent

The genealogy in Kingman's coalescent is traced backwards as time recedes into the past. The genealogy can be graphically represented as a rooted binary tree, on which the branches have lengths in units of time. Figure 2.1 plots a genealogy for a sample of size $n = 5$. The tips at the bottom of the genealogy are individuals in the sample. In a coalescence event, when there are i ancestral lines, $i \geq 2$, each of the $\binom{i}{2}$ pairs of ancestral lines is equally likely to merge at rate 1. The number of distinct lines thus follows the ancestral process with transition rates (1.1). The remaining single line on the top of the genealogy is the MRCA of the sample. Let $W(i)$ represent the coalescence time, i.e., the time until the first coalescence event when there are i ancestral lines. $W(i)$ is exponentially distributed with mean $1/\binom{i}{2}$, independently for each i. The genealogy can be simulated by generating $n - 1$ independent and exponentially distributed random time, and independently of these, a random bifurcating topology.

There are two important quantities associated with the shape of a genealogy: the time to the MRCA, i.e., the height of the genealogy, and the total length of all branches on the genealogy. The time to the MRCA satisfies $T_{MRCA}(n) = \sum_{i=2}^{n} W(i)$; and the

length of the genealogy satisfies $L_{MRCA}(n) = \sum_{i=2}^{n} iW(i)$. In Kingman's coalescent, the mean and variance of $T_{MRCA}(n)$ are

$$\mathbb{E}[T_{MRCA}(n)] = 2\left(1 - \frac{1}{n}\right), \tag{2.1}$$

$$\mathrm{Var}[T_{MRCA}(n)] = 8\sum_{i=2}^{n} \frac{1}{i^2} - 4\left(1 - \frac{1}{n}\right)^2; \tag{2.2}$$

and the mean and variance of $L_{MRCA}(n)$ are

$$\mathbb{E}[L_{MRCA}(n)] = 2\sum_{i=1}^{n-1} \frac{1}{i}, \tag{2.3}$$

$$\mathrm{Var}[L_{MRCA}(n)] = 4\sum_{i=1}^{n-1} \frac{1}{i^2}. \tag{2.4}$$

Hudson (1991) derives equations (2.1) and (2.2),while Watterson (1975) deduces equations (2.3) and (2.4).

Even though the mean time for the last two ancestral lines to coalesce is $\mathbb{E}[W(2)] = 1$, we have

$$\lim_{n\to\infty} \mathbb{E}[T_{MRCA}(n)] = 2.$$

Therefore, a large number of coalescence events are likely to occur near the bottom of the genealogy, generating very short ancestral lines. The mean length of the genealogy increases logarithmically without bound, since

$$\mathbb{E}[L_{MRCA}(n)] \sim 2(\log n + C)$$

for large n, where $C \approx 0.58$ is the Euler constant. It is well known (see Tavaré et al., 1997) that

$$\lim_{n\to\infty} \mathrm{Var}[T_{MRCA}(n)] = 4\pi^2/3 - 12 \approx 1.16$$

and

$$\lim_{n\to\infty} \mathrm{Var}[L_{MRCA}(n)] = 2\pi^2/3 \approx 6.58.$$

Since $\text{Var}[T_{MRCA}(2)] = 1$, most of the variation in $T_{MRCA}(n)$ is attributed to the coalescence time of the last two ancestral lines before the MRCA. Hence, the genealogy tends to be dominated by ancestral lines near the MRCA.

In the next section, we will show how the real genealogy in the ancestral dual process with typed lines can be simulated.

2.2 Simulation of ancestral dual processes with typed lines

The ancestral dual process of Theorem 1.9 models the evolution of the real and virtual ancestral lines all together backwards in time. To simulate the real genealogy, it is natural to decompose ancestral lines into real and virtual ones. Although the virtual lines are not part of the real genealogy, we still need to track their numbers and types up to the MRCA, since the transition rates (1.15) to (1.17) of the dual process depend on them. Examination of the transition rates (1.15) to (1.17) reveals that it is only possible to have virtual lines of type 1 emerging at branching events. However, once these virtual lines are introduced into the dual process, subsequent mutations can convert them to type 2. Hence, virtual lines of both types may exist in the dual process. Let $r_i(t)$ be the number of real lines of type i, and $v_i(t)$ be the number of virtual lines of type i, at time $t \geq 0$ in the past. The configuration of real and virtual ancestral lines can be denoted by $(\boldsymbol{r}(t), \boldsymbol{v}(t))$, where $\boldsymbol{r}(t) = (r_1(t), r_2(t))$ and $\boldsymbol{v}(t) = (v_1(t), v_2(t))$.

Let $\pi(x)$ be the stationary density of the diffusion with generator (1.10), as given in equation (1.11). According to Etheridge and Griffiths (2009), by extending the sampling probability of Proposition 1.7, the probability of obtaining an ordered type configuration associated with the decomposition $(\boldsymbol{r}, \boldsymbol{v})$ in the dual process at any fixed

time in the past, is

$$\begin{aligned}\mathscr{H}(\boldsymbol{r},\boldsymbol{v}) &= \int_0^1 x^{r_1+v_1}(1-x)^{r_2+v_2}\pi(x)\,\mathrm{d}x \\ &= \frac{B(r_1+v_1+\theta_1, r_2+v_2+\theta_2)M(r_1+v_1+\theta_1, r+v+\theta, -\sigma)}{B(\theta_1,\theta_2)M(\theta_1,\theta,-\sigma)},\end{aligned} \tag{2.5}$$

where $r = |\boldsymbol{r}|$, $v = |\boldsymbol{v}|$ and $\theta = |\boldsymbol{\theta}|$. The sampling probabilities are stationary; hence, they are independent of time. Etheridge and Griffiths (2009) also argue that the configuration of only real lines in the dual process does not follow the sampling probability (2.5), since the probability is for both real and virtual lines. We have to simulate the stochastic dual process, as in the following theorem, in order to obtain the real genealogy, not just draw a random sample from the population. Including virtual lines is necessary to construct an ancestral process that is Markov; and the virtual lines also implicitly provide information on the composition of the population backwards in time.

Theorem 2.1 (Dual process decomposed into real and virtual lines)**.** *Consider the ancestral dual process decomposed into real and virtual lines,* $\{(\boldsymbol{r}(t),\boldsymbol{v}(t)), t \geq 0\}$*, for a sample* $\boldsymbol{n} \in \mathbb{Z}_+^2$ *of size* $|\boldsymbol{n}| \geq 1$*. The ancestral dual process* $\{(\boldsymbol{r}(t),\boldsymbol{v}(t)), t \geq 0\}$ *is Markov with state space* $\mathbb{Z}_+^4 \setminus \{\boldsymbol{0}\}$ *and initial state* $(\boldsymbol{n},\boldsymbol{0})$*. The non-zero entries in the* $(\boldsymbol{r},\boldsymbol{v})$*th row of the Q-matrix of the dual process are*

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) = \frac{1}{2}\sigma(r+v)\frac{(r_1+v_1+\theta_1)}{r+v+\theta}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_1+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \tag{2.6}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_i,\boldsymbol{v})) = \frac{1}{2}r_i(r_i-1)\frac{r+v-1+\theta}{r_i+v_i-1+\theta_i}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \tag{2.7}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_i)) = \frac{1}{2}(2r_iv_i+v_i(v_i-1))\frac{r+v-1+\theta}{r_i+v_i-1+\theta_i}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \tag{2.8}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_i-\boldsymbol{e}_j,\boldsymbol{v})) = \frac{1}{2}r_j\theta_j\frac{r_i+v_i+\theta_i}{r_j+v_j-1+\theta_j}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \tag{2.9}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j)) = \frac{1}{2}v_j\theta_j\frac{r_i+v_i+\theta_i}{r_j+v_j-1+\theta_j}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \tag{2.10}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v})) = -\frac{1}{2}[(r_1+v_1)\sigma + (r+v)(r+v-1) + \theta_2(r_1+v_1) + \theta_1(r_2+v_2)], \tag{2.11}$$

where $i,j = 1,2$ *and* $i \neq j$, $r = |\boldsymbol{r}|$, $v = |\boldsymbol{v}|$, $\theta = |\boldsymbol{\theta}|$, *and*

$$\nu(\boldsymbol{\theta}) = \mathbb{E}[e^{-\sigma\xi}] = M(\theta_1,\theta,-\sigma)$$

with ξ *being a random variable having the Beta*(θ_1,θ_2) *distribution and* $M(\theta_1,\theta,-\sigma)$ *being the confluent hypergeometric function defined in equation* (1.12).

Remark 2.2. The dual process derived in Theorem 2.1 is closely related to the work of Fearnhead (2002) and Stephens and Donnelly (2003). Those authors construct ancestral selection graphs where the type of individuals in the sample is known. In particular, Fearnhead (2002) studies the history of the common ancestor in a haploid population with parent-independent mutation, and achieves an additional simplification of the dual process by not following the virtual lines that mutate backwards in time (as their type is then a random draw from the stationary distribution and thus does not carry any information about the composition of the population). Stephens and Donnelly (2003) do not take advantage of that additional simplification but consider the general diploid population in the diffusion limit.

Proof of Theorem 2.1. Let $\boldsymbol{\alpha} = \boldsymbol{r} + \boldsymbol{v}$. The dual process decomposed into real and virtual lines is a continuous-time Markov process which remains in a state $(\boldsymbol{r},\boldsymbol{v})$ for an exponentially distributed length of time, with rate $-q(\boldsymbol{\alpha},\boldsymbol{\alpha})$ given in equation (1.18). The process then jumps to a new state according to the following events and transition rates:

1. A line branches, adding a virtual line of type 1 to the genealogy. The corresponding transition is $(\boldsymbol{r},\boldsymbol{v}) \to (\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)$ at rate $q(\boldsymbol{\alpha},\boldsymbol{\alpha}+\boldsymbol{e}_1)$ given in equation (1.15). Since all real and virtual lines are equally likely to branch, we have

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) = q(\boldsymbol{\alpha},\boldsymbol{\alpha}+\boldsymbol{e}_1).$$

2. A pair of lines of type i, $i = 1, 2$ coalesce at rate $q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_i)$ given in equation (1.16). Conditional on this event,

   (a) the coalescence occurs on a pair of real lines of type i, giving a transition $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r} - \boldsymbol{e}_i, \boldsymbol{v})$ as two real lines coalesce to a real line, with probability $\binom{r_i}{2} / \binom{r_i + v_i}{2}$. Since all pairs of real ancestral lines of type i are equally likely to coalesce, we have

   $$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_i, \boldsymbol{v})) = q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_i) \times \frac{\binom{r_i}{2}}{\binom{r_i + v_i}{2}};$$

   (b) the coalescence occurs on a pair of virtual lines of type i, giving a transition $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_i)$ as two virtual lines coalesce to a virtual line, with probability $\binom{v_i}{2} / \binom{r_i + v_i}{2}$;

   (c) the coalescence occurs on a real and a virtual line of type i, giving a transition $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_i)$ as one real and one virtual line coalesce to a real line, with probability $1 - [\binom{r_i}{2} + \binom{v_i}{2}] / \binom{r_i + v_i}{2}$. Since all real and virtual lines of type i are equally likely to be involved, we have

   $$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_i)) = q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_i) \times \frac{\binom{r_i + v_i}{2} - \binom{r_i}{2}}{\binom{r_i + v_i}{2}}.$$

3. A mutation occurs such that a line of type j changes its type to i, $i, j = 1, 2$ and $i \neq j$, at rate $q(\boldsymbol{\alpha}, \boldsymbol{\alpha} + \boldsymbol{e}_i - \boldsymbol{e}_j)$ given in equation (1.17). Conditional on this event,

   (a) a real line of type j mutates to type i, giving a transition $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r} + \boldsymbol{e}_i - \boldsymbol{e}_j, \boldsymbol{v})$ with probability $r_j / (r_j + v_j)$. Since all real ancestral lines of type j are equally likely to mutate, we have

   $$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_i - \boldsymbol{e}_j, \boldsymbol{v})) = q(\boldsymbol{\alpha}, \boldsymbol{\alpha} + \boldsymbol{e}_i - \boldsymbol{e}_j) \times \frac{r_j}{r_j + v_j};$$

   (b) a virtual line of type j mutates to type i, giving a transition $(\boldsymbol{r}, \boldsymbol{v}) \to$

$(\boldsymbol{r}, \boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j)$ with probability $v_j/(r_j+v_j)$. Since all virtual ancestral lines of type j are equally likely to mutate, we have

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j)) = q(\boldsymbol{\alpha}, \boldsymbol{\alpha}+\boldsymbol{e}_i-\boldsymbol{e}_j) \times \frac{v_j}{r_j+v_j}.$$

The theorem is then proved. □

To extract the quantities such as the time to the MRCA and the length of the genealogy, simulation of a tree-like structure is unnecessary, since information about real and virtual lines is already contained in the dual process, and each event occurs on an ancestral line that is picked uniformly at random from the genealogy. The knowledge required is that of the particular transition which occurs, with respect to the resulting configuration of ancestral lines.

The simulation of a genealogy starts from the state $(\boldsymbol{n}, \boldsymbol{0})$ for a given sample $\boldsymbol{n}$. At each state, an event is drawn uniformly from the set of possible events with normalised probabilities that are calculated as their individual transition rates in equations (2.6) to (2.10) divided by the their sum $-q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v}))$ in equation (2.11). The normalised probabilities are the jump probabilities of the Markov process. The configuration of real and virtual lines is then updated accordingly. Inter-jump times are simulated as independent and exponential distributed random variables with rate $-q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v}))$ in equation (2.11).

Although the total number of real and virtual lines does not follow a monotonic process, the numbers of real lines are non-increasing, and will eventually be absorbed into $(0, 1)$ or $(1, 0)$. The simulation stops when a MRCA is reached with the configuration $(\boldsymbol{e}_i, \boldsymbol{v})$, $i = 1, 2$. Figure 2.2 depicts a real genealogy generated by the ancestral dual process with typed lines for a sample containing one individual of type 1 and one of type 2.

Once again, we see the advantages of the dual process with typed lines over the ancestral selection graph. Although the real genealogy in the dual process for a *random*

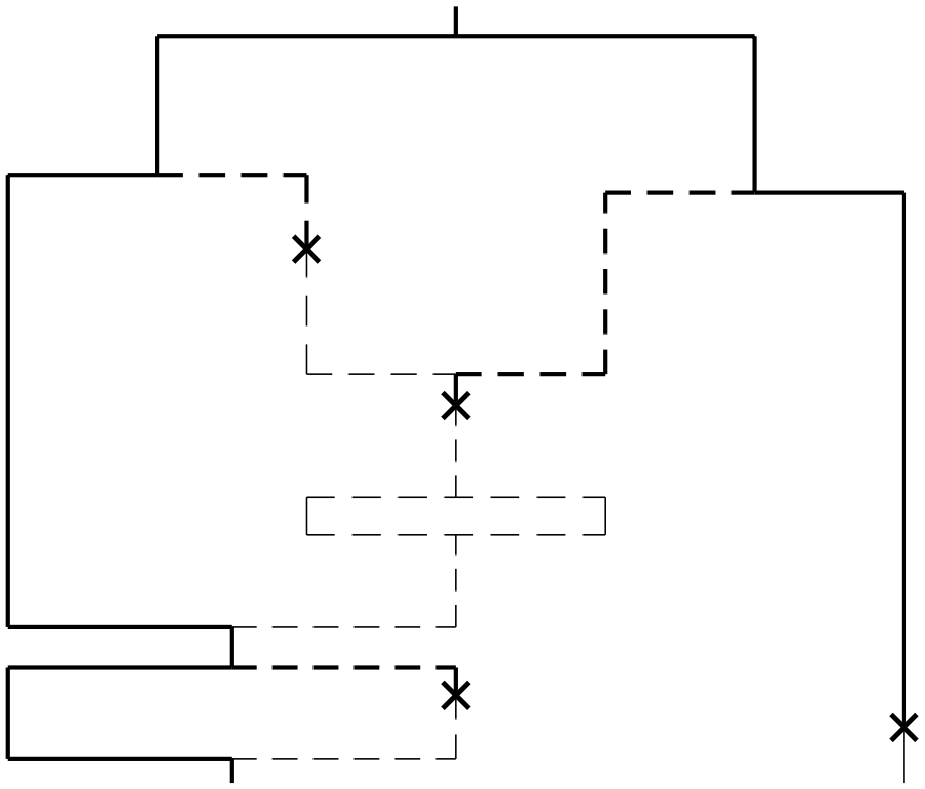


Figure 2.2: A genealogy in the ancestral dual process with typed lines for a sample with configuration $(1, 1)$. The dashed lines represent virtual ancestral lines, the bold lines represent ancestral lines of type 2, and the crosses correspond to mutations. The coefficient values are $\sigma = 2$ and $\theta_1 = \theta_2 = 1.5$.

sample has the same distribution as those embedded in the ancestral selection graph, with the dual process we can study the genealogy of a sample with *given* configuration. Simulation of the dual process is also much more efficient for strong selection and mutation, as the real genealogy can be directly obtained without tracing ancestral lines beyond the MRCA.

Table 2.1 presents a comparison of the number of branching events in the dual process with typed lines and in the ancestral selection graph. For a sample of moderate total size of 5, for $\sigma = 1, 3$ and 5, the number of branching events in dual process with typed lines increases with n_1; while when $\sigma = 1000$, the number of branching events decreases as n_1 increases. For samples of a relatively large size, say $(20, 20)$, $(50, 50)$ and $(100, 100)$ in the dual process, simulations show that the expected numbers of

Table 2.1: The mean number of branching events in the ancestral dual process with typed lines for three sample configurations, and in the ancestral selection graph for random samples of size 2. The mutation coefficients are $\theta_1 = \theta_2 = 1.5$.

σ	$(0,5)$	$(1,4)$	$(2,3)$	$(3,2)$	$(4,1)$	$(5,0)$	ASG
0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0
1	0.8 (0.6)	1.0 (0.9)	1.2 (1.0)	1.3 (1.2)	1.3 (1.2)	1.3 (1.2)	3.2
3	2.5 (1.7)	3.2 (2.3)	3.6 (2.7)	4.1 (3.1)	4.4 (3.4)	4.4 (3.4)	28.5
5	4.0 (2.4)	4.7 (3.1)	5.4 (3.7)	6.0 (4.2)	6.6 (4.8)	6.8 (5.0)	191.7
10^3	6.1 (3.9)	6.1 (3.9)	5.9 (3.8)	5.5 (3.5)	5.0 (3.1)	4.0 (2.4)	-
∞	6.1	6.1	5.9	5.5	5.0	4.0	-

In both the ancestral dual process with typed lines and the ancestral selection graph, we count up to the ultimate ancestor. (The MRCA cannot be determined without reaching the ultimate ancestor in the ancestral selection graph, while it is unnecessary to simulate the ancestral dual process with typed lines beyond the MRCA.) The bracketed values are the mean number of branching events up to the MRCA. These values are obtained from simulations of the ancestral dual process of Theorem 2.1 and the ancestral selection graph. The values in the bottom row are calculated from Theorem 2.3. There are gaps in the table because it is very computationally expensive to simulate those values.

branching events are 7.4, 11.8 and 22.8 up to the MRCA when $\sigma = 1000$. However, in this thesis, we always consider samples of a finite size.

In Proposition 1.1, we have proved that for a random sample of finite size in the ancestral selection graph, the mean number of branching events until the ultimate ancestor is $O(\exp(\sigma))$ for large σ. By contrast, the following theorem shows that in the dual process with typed lines, the mean number of branching events up to the ultimate ancestor is finite as $\sigma \to \infty$. (Although it is unnecessary to simulate up to the ultimate ancestor in the dual process with typed lines.) This result is novel, and the proof of this theorem is postponed to the end of Section 3.2.

Theorem 2.3 (Branching events in the dual process with typed lines)**.** *Consider the dual process* $\{(\boldsymbol{r}(t), \boldsymbol{v}(t)), t \geq 0\}$ *of Theorem 2.1 for a sample* $(n_1, n_2) \in \mathbb{Z}_+^2$ *of size* $n_1 + n_2 \geq 1$*. The ultimate ancestor is reached at the time*

$$T_{UA}(n_1, n_2) \coloneqq \inf\{t \geq 0 \colon |\boldsymbol{r}(t) + \boldsymbol{v}(t)| = 1\}.$$

Let $V(n_1, n_2)$ *be the number of branching events until the ultimate ancestor for the*

sample (n_1, n_2)*. Then,*

$$\lim_{\sigma\to\infty} \mathbb{E}[V(n_1,n_2)] = \sum_{m=0}^{n_1} \frac{\theta_1^m}{\theta_{1(n_1)}} \left|S_{n_1}^{(m)}\right| \left(e^{\theta_1} \sum_{l=0}^{n_2+m-2} \frac{l!}{\theta_1^l} - \sum_{l=0}^{n_2+m-2} \frac{1}{\theta_1^l} \sum_{k=l}^{n_2+m-2} \frac{k!}{(k-l)!} \right), \tag{2.12}$$

where $S_{n_1}^{(m)}$ *are the Stirling numbers of the first kind.*

Remark 2.4 (Stirling numbers)**.** The unsigned Stirling numbers of the first kind, $|S_n^{(m)}|$, arise as coefficients of the rising factorial such that

$$x_{(n)} = \sum_{m=0}^{n} \left|S_n^{(m)}\right| x^m. \tag{2.13}$$

We also have the recursive equation

$$\left|S_n^{(m)}\right| = \left|S_{n-1}^{(m-1)}\right| + (n-1)\left|S_{n-1}^{(m)}\right|, \quad 0 < m < n, \tag{2.14}$$

with boundary values $|S_0^{(0)}| = 1$, and $|S_n^{(0)}| = 0$ for $n > 0$. See Abramowitz and Stegun (1965, Section 24.1) for more properties of the Stirling numbers.

2.3 Shape of genealogies

There is considerable interest in understanding the shape of genealogies under selection and mutation. Coalescent processes and the shape of the genealogies they generate have proved very useful in interpreting and understanding the patterns of variation in population genetics data, as the structure of genetic data usually reflects the underlying genealogy (Donnelly and Tavaré, 1995). Stephens and Donnelly (2003) also claim that many current methods for discovering the genetic basis of common diseases in humans focus on detecting linkage disequilibrium, which is the non-random association of alleles at loci that may or may not be on the same chromosome. The pattern of linkage disequilibrium depends crucially on the shape of genealogies at the loci. The famous statistical tests of neutrality of Fu and Li (1993) are also based on measuring the deviation of branching lengths on a genealogy under purifying selection from the

neutral case.

Kaplan et al. (1988) use the coalescent-in-a-random-background model (which will be discussed in Section 3.1) and Neuhauser and Krone (1997) use the ancestral selection graph (which was introduced in Section 1.2) to both conclude that for a random sample under weak purifying selection, the shape of the genealogy differs only slightly from the neutral expectation. This observation is later advocated by Przeworski et al. (1999). In addition, Kaplan et al. (1988) discover that the genealogy of a sample containing only disfavoured individuals appears much shorter than a neutral one, while the genealogy of a sample containing only favoured individuals looks neutral.

On the one hand, Kaplan et al. (1988) merely consider samples of size 2 for simulations, as the ability of solving diffusion equations (see their Equation 20) on computers was limited. However, Coop and Griffiths (2004) later show that the genealogy of a sample together with the ancestral history of the population in the coalescent-in-a-random-background model can be efficiently simulated by generating a stochastic trajectory of the ancestral frequency backwards in time, and then constructing a genealogy given this trajectory. On the other hand, Neuhauser and Krone (1997) and Przeworski et al. (1999) only simulate the genealogies of random samples with the value of σ being smaller than 6, because of the shortcomings of the ancestral selection graph mentioned in Section 1.2.

Simulation studies of the ancestral dual process with typed lines of Theorem 2.1 have never been done before. It is of interest to us to use these simulations to examine how genealogies depend on the interaction of strong selection and strong mutation, for samples with given configuration, as we are able to simulate genealogies with large selection and mutation coefficients, e.g., $\sigma = 1000$ and $\theta_1 = \theta_2 = 100$.

In the remainder of this section, genealogies of samples with random configuration are investigated in Section 2.3.1, while genealogies of samples with given configuration

are studied in Sections 2.3.2 and 2.3.3. In all cases, simulation results are averaged over 50000 Monte Carlo replicates. It is well known (see Ripley, 2006, p. 120) that the difference between the Monte Carlo expectation and the true one is inversely proportional to the square root of the number of replicates, provided that the Monte Carlo paths are independent and the variance is finite. It is also not feasible to increase the number of replicates indefinitely. Therefore, some numerical instability is inevitable, especially when the values of the selection and mutation coefficients become extreme.

In the following sections, we often compare the genealogy in the dual process with typed lines with that in the neutral structured coalescent and that in Kingman's coalescent. By setting the selection coefficient $\sigma = 0$ in the dual process with typed lines, we obtain the neutral structured coalescent, which is not the same as Kingman's coalescent. The shape of the genealogy in the neutral structured coalescent depends on the mutation rates, while that in Kingman's coalescent does not.

2.3.1 Coalescence times for samples with random configuration

In this section, we study how the coalescence times and the time to the MRCA for samples with random configuration are modified by selection and mutation, with Tables 2.2 and 2.3.

Consider the dual process $\{(\boldsymbol{r}(t), \boldsymbol{v}(t)), t \geq 0\}$ of Theorem 2.1 for a sample with given configuration $(n_1, n_2) \in \mathbb{Z}_+^2$ of size $n_1 + n_2 \geq 1$. Denote the time to the MRCA by

$$T_{MRCA}(n_1, n_2) \coloneqq \inf\{t \geq 0 \colon |\boldsymbol{r}(t)| = 1\}.$$

It follows that the (unconditional) time to the MRCA of a sample with random

Table 2.2: The mean time to the MRCA and the length of the genealogy for random samples of size 10, with respect to various strengths of selection and mutation.

θ_1	σ	$\mathbb{E}[T_{MRCA}(10)]$	$\mathbb{E}[L_{MRCA}(10)]$
10^{-4}	$*$	1.80	5.66
	1.5	1.79	5.63
	#	1.80	5.66
1.5	†	1.80	5.66
	10	1.67	5.32
	10^3	1.79	5.64
10^2	‡	1.80	5.66
	10^3	1.66	5.24
-	-	1.80	5.66

$*$: $\sigma = 0, 0.3$; #: $\sigma = 10, 10^3$; †: $\sigma = 0, 0.3, 1.5$; ‡: $\sigma = 0, 0.3, 1.5, 10$. We set $\theta_1 = \theta_2$. The values in the bottom row are from Kingman's coalescent.

configuration of size n, $T_{MRCA}(n)$, satisfies

$$\mathbb{E}[T_{MRCA}(n)] = \sum_{n_1=0}^{n} \binom{n}{n_1} \mathscr{H}(n_1, n-n_1)\, \mathbb{E}[T_{MRCA}(n_1, n-n_1)],$$

where $\mathscr{H}(n_1, n-n_1)$ is the sampling probability specified in Proposition 1.7. The time to the MRCA is also the height of the real genealogy.

Let $W_n(i)$ be the coalescence time, i.e., the waiting time until the next coalescence event, given that there are currently $i \geq 2$ ancestral lines, on the real genealogy of a sample of size $n \geq 2$. $W_n(i)$ is also the length of the branches on the real genealogy between two consecutive coalescence events. We have $\mathbb{E}[T_{MRCA}(n)] = \sum_{i=2}^{n} \mathbb{E}[W_n(i)]$.

Table 2.2 shows that, for a random sample, the values of $\mathbb{E}[T_{MRCA}(n)]$ are only slightly smaller than the neutral expectation when σ is slightly larger (but not extremely large) compared with θ_1 and θ_2 (see the cases where $\theta_1 = \theta_2 = 1.5$, $\sigma = 10$ and $\theta_1 = \theta_2 = 100$, $\sigma = 1000$). In the case where $\theta_1 = \theta_2 = 100$ and $\sigma = 1000$, the strong selective force is offset by strong mutation force.

Table 2.3 further demonstrates that for moderate strength of mutation ($\theta_1 = \theta_2 = 1.5$), a slightly larger selection coefficient ($\sigma = 10$) mainly affects the coalescence

Table 2.3: The mean coalescence times $\mathbb{E}[W_n(i)]$ for random samples of size 10, with respect to various strengths of selection and mutation.

		Number of ancestral lines, i					
θ_1	σ	2	3	4	6	8	10
10^{-4}	$*$	1.00	0.33	0.17	0.07	0.04	0.02
1.5	†	1.00	0.33	0.17	0.07	0.04	0.02
	10	0.90	0.31	0.16	0.07	0.04	0.02
	10^3	0.99	0.33	0.16	0.07	0.04	0.02
10^2	‡	1.00	0.33	0.17	0.07	0.04	0.02
	10^3	0.92	0.31	0.15	0.06	0.03	0.02
-	-	1.00	0.33	0.17	0.07	0.04	0.02

$*$: $\sigma = 0, 0.3, 1.5, 10, 10^3$; †: $\sigma = 0, 0.3, 1.5$; ‡: $\sigma = 0, 0.3, 1.5, 10$. We set $\theta_1 = \theta_2$. The values in the bottom row are from Kingman's coalescent.

time of the last two ancestral lines before the MRCA, i.e., the value of $\mathbb{E}[W_n(2)]$, and causes the value to decrease, so the whole genealogy has a tendency to be star-shaped. Shorter coalescence times near the MRCA lead to less mutations that occur, thus the variation is also expected to reduce locally. However, when mutation is strong ($\theta_1 = \theta_2 = 100$), a slightly stronger selection coefficient ($\sigma = 1000$) makes almost all branches of the genealogy shorter than those in Kingman's coalescent.

For small values of σ, no drastic effect of selection on the genealogies of random samples can be discovered in Tables 2.2 and 2.3. This observation is consistent with the conclusions of Kaplan et al. (1988), Golding (1997), Neuhauser and Krone (1997) and Przeworski et al. (1999). However, we will see in Sections 2.3.2 and 2.3.3 that for samples with given configuration, selection *does* have a significant impact on the shape of genealogies.

In the limits as $\theta_1, \theta_2 \to 0$, $\theta_1, \theta_2 \to \infty$, or $\sigma \to \infty$, the mean heights and branch lengths of genealogies of random samples in Tables 2.2 and 2.3 are the same as those in Kingman's coalescent. Intuitively, when there is no mutation, or selection is very strong, the population tends to consist of only one allele. Hence, a random sample from such a population contains only one allele, and so the genealogy must be the

same as that in Kingman's coalescent.

2.3.2 Time to the MRCA and length of the genealogies for samples with given configuration

One of the advantages of the ancestral dual process of Theorem 2.1 is that it can generate the genealogy for a given sample configuration. We will observe in this section that genealogies, especially those under strong selection, can be very different from those in Kingman's coalescent, depending on the configuration of samples.

Figure 2.3 plots the values of $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ as a function of the number n_2 of type 2 individuals in a sample of size $n_1 + n_2 = 10$. The balance between selection and mutation distorts genealogies of samples with given configuration. Curves in Figure 2.3 become asymmetric as σ increases. When $\theta_1 = \theta_2 = 10^{-4}$, $\sigma = 1000$ and the values of n_2 are small, the values of $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ can be much smaller than those in the neutral case ($\sigma = 0$). By contrast, the values of $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ approach the neutral expectation as the value of n_2 increases. This observation makes intuitive sense by considering the forwards-in-time dynamics. When there is no mutation, there tends to be only one allele existing in the population. When the population happens to contain a considerable proportion of the disfavoured allele (so we can draw a sample containing some individuals of type 1), the disfavoured allele will be eliminated rapidly from the population by the strong selective force. Thus, the genealogy of our sample viewed backwards in time is much shorter than the neutral one, when the value of n_2 is small.

For $\theta_1 = \theta_2 = 1.5$, $\sigma = 1000$ and $n_2 < 9$, genealogies also tend to be shorter than the neutral expectation, but become longer when $n_2 = 9$ and 10. In fact, the values of $\mathbb{E}[T_{MRCA}(1, 9)]$ and $E[T_{MRCA}(0, 10)]$ are both approximately 1.8, which is the expected height of a genealogy in Kingman's coalescent for a random sample of size 10.

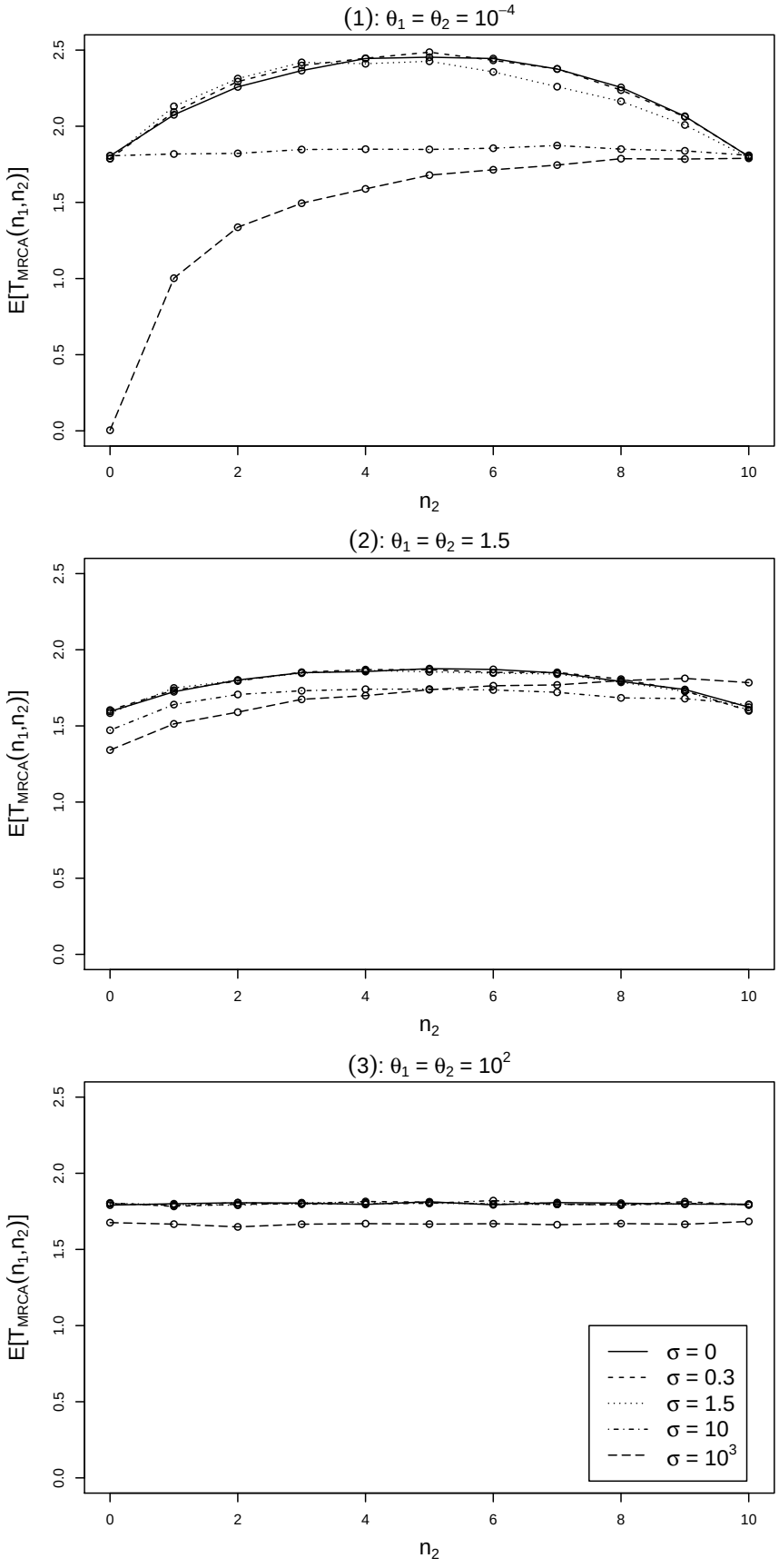


Figure 2.3: The mean time to the MRCA, plotted against the number of type 2 individuals in the sample, with respect to various strengths of selection and mutation.

When mutation is strong ($\theta_1 = \theta_2 = 100$) and selection is relatively weak ($0 < \sigma \leq 10$), no departure from the neutral case can be observed. The selective force is overwhelmed by the strong mutation force. We also notice that, under strong mutation, the height of the genealogy becomes insensitive to initial sample configurations. Intuitively, the reason for the insensitivity is that the types of ancestral lines under strong mutation are assigned according to the stationary distribution of the mutation process such that an ancestral line is type i, $i = 1, 2$ with probability $\theta_i/(\theta_1 + \theta_2)$.

In the neutral case where $\sigma = 0$, for a given sample configuration, the values of $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ tend to be smaller when mutation coefficients become larger. This is because when θ_1 and θ_2 are small, the number of ancestral lines on the genealogy is reduced, due mostly to coalescence events. When all lines of the same type have coalesced, mutation is essential for changing the remaining two ancestral lines to the same type before coalescing to the MRCA, and smaller θ_1 and θ_2 cause this mutation event to take longer.

2.3.3 Type distribution of the MRCA

Consider the ancestral dual process $\{(\boldsymbol{r}(t), \boldsymbol{v}(t)), t \geq 0\}$ of Theorem 2.1 for a sample $(n_1, n_2) \in \mathbb{Z}_+^2$ of size $n_1 + n_2 \geq 1$. Denote the probability that the MRCA is the favoured type 2 by

$$\rho(n_1, n_2) = \mathbb{P}[r_2(T_{MRCA}(n_1, n_2)) = 1]. \tag{2.15}$$

In the neutral case ($\sigma = 0$), the MRCA type distribution of a random sample of size n drawn from the stationary population is equal to the distribution of a random sample of size one.

In this section, we use Monte Carlo simulations of the dual process with typed lines to study the MRCA type distribution for a given sample configuration under selection and mutation. Figure 2.4 plots the values of $\rho(n_1, n_2)$ against the number n_2 of type 2 individuals in a sample of size $n_1 + n_2 = 10$, with respect to different values

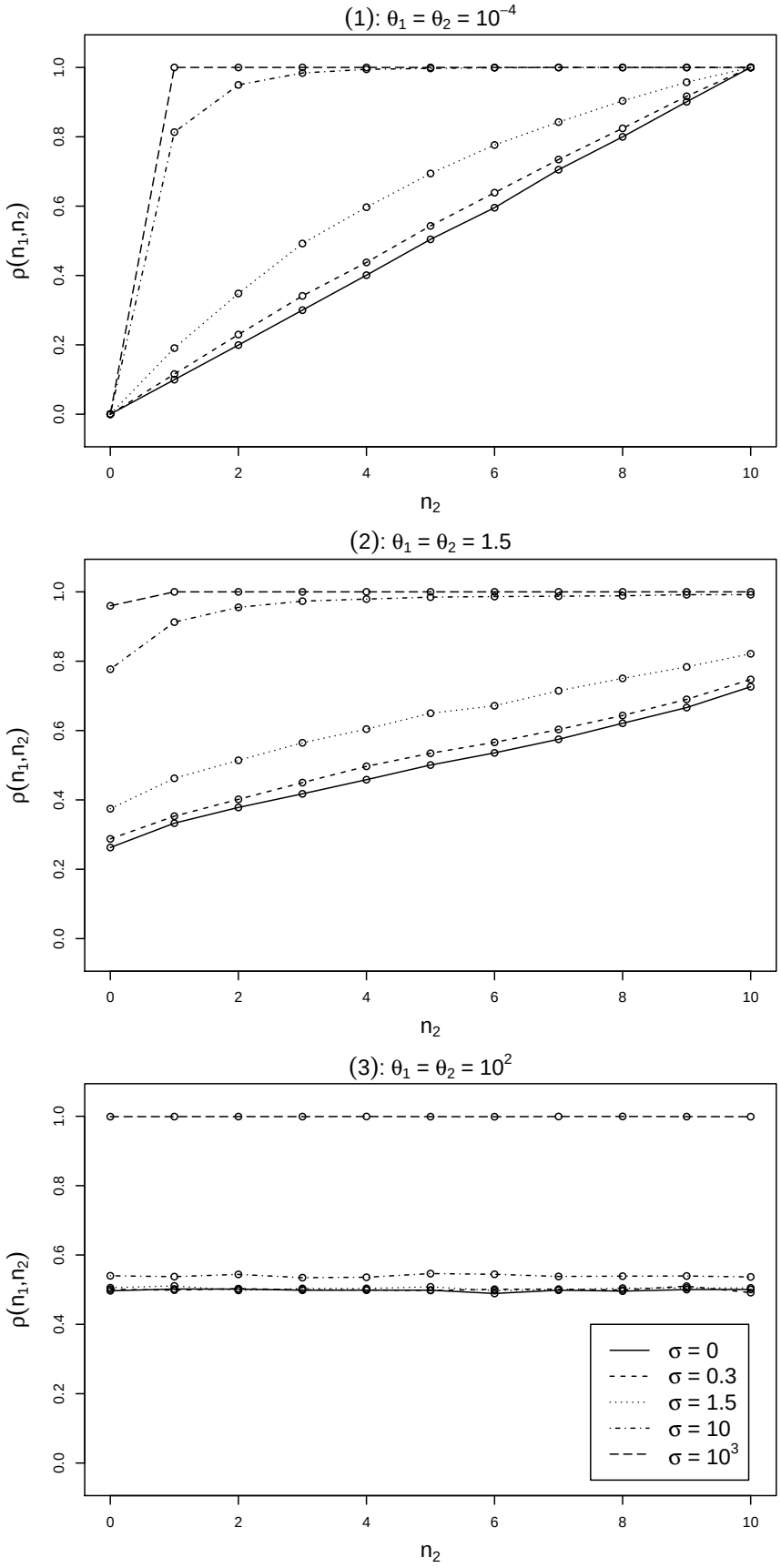


Figure 2.4: The probability that the MRCA is type 2, plotted against the number of type 2 individuals in the sample, with respect to various strengths of selection and mutation.

of θ_1, θ_2 and σ.

Slade (2000) simulates the MRCA type distribution with a modified ancestral selection graph. The author finds that under selection the MRCA type distribution is biased towards the favoured type. This result is also intuitive as the MRCA of the favoured type tends to have more descendants in the population at the present time from which our sample is taken. Figure 2.4 also shows that all probabilities are higher than those in the neutral case, so the probability that the MRCA is type 2 for a random sample must also be higher than the probability of sampling one individual of type 2 randomly from a neutral population.

However, we are particularly interested in observing the effect of strong selection and strong mutation on the MRCA type distribution for a given sample. Figure 2.4 illustrates that in the neutral case or when the values of θ_1, θ_2 and σ are moderate, the probability that the MRCA is type 2 increases almost linearly with the number of type 2 individuals in the sample. Selection makes the distribution skewed towards the favoured allele. When selection is strong ($\sigma = 1000$), the probability is significantly high even when there are no individuals of type 2 in the sample.

The probabilities become insensitive to given sample configurations when the strength of mutation grows, as strong mutation force shuffles the types of the ancestral lines according to the stationary distribution of the mutation process. Since the mutation rates for both types are the same, under moderate strength of selection ($0 \leq \sigma \leq 10$) and strong mutation ($\theta_1 = \theta_2 = 100$), the probability that the MRCA is type 2 is approximately 0.5.

2.4 Recursive equations under strong selection

These limiting phenomena observed from simulations in Sections 2.3.1 to 2.3.3 are not clear by simply looking at the dual process of Theorem 2.1. Some limiting theorems

from formal calculations are needed.

It is not difficult to establish some recursive equations for the expectations we have simulated in Sections 2.3.1 to 2.3.3. For example, by considering the first event that happens in the dual process, $\mathbb{E}[T_{MRCA}(\boldsymbol{r}, \boldsymbol{v})]$ satisfies

$$
\begin{aligned}
&- q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v}))\, \mathbb{E}[T_{MRCA}(\boldsymbol{r}, \boldsymbol{v})] \\
&\quad = 1 + q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1))\, \mathbb{E}[T_{MRCA}(\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1)] \\
&\qquad + \sum_{i=1}^{2} q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_i, \boldsymbol{v}))\, \mathbb{E}[T_{MRCA}(\boldsymbol{r} - \boldsymbol{e}_i, \boldsymbol{v})] \\
&\qquad + \sum_{i=1}^{2} q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_i))\, \mathbb{E}[T_{MRCA}(\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_i)] \\
&\qquad + \sum_{i=1}^{2} \sum_{\substack{j=1 \\ j \neq i}}^{2} q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_i - \boldsymbol{e}_j, \boldsymbol{v}))\, \mathbb{E}[T_{MRCA}(\boldsymbol{r} + \boldsymbol{e}_i - \boldsymbol{e}_j, \boldsymbol{v})] \\
&\qquad + \sum_{i=1}^{2} \sum_{\substack{j=1 \\ j \neq i}}^{2} q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_i - \boldsymbol{e}_j))\, \mathbb{E}[T_{MRCA}(\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_i - \boldsymbol{e}_j)], \qquad (2.16)
\end{aligned}
$$

where $q(\cdot, \cdot)$ are the transition rates given in equations (2.6) to (2.11). The boundary conditions are $\mathbb{E}[T_{MRCA}(\boldsymbol{e}_i, \boldsymbol{v})] = 0$, $i = 1, 2$.

Recursive equations such as the one in equation (2.16) allow us, at least in principle, to compute the relevant quantities. However, such recursive equations are usually too large to solve directly. The main question addressed in this section is how to solve such recursive equations under the assumption of strong selection. This can be approached by asymptotically expanding the transition rates (2.6) to (2.10) for large σ, so that the recursive equations become much more tractable when we ignore terms of $o(1)$, and explicit formulae for the quantities we are interested in can be obtained.

The following theorem for the expectation and variance of the height and the length of the genealogy under strong selection is new, and the proof of it will be presented at the end of this section after we have all the auxiliary lemmas. Note that we assume the mutation coefficients θ_1 and θ_2 are constant but the selection coefficient

σ is very large. The probabilistic interpretation will become clearer from the limiting behaviour of the ancestral process presented in Theorem 3.5 in the next chapter. We will see that backwards in time the disfavoured type 1 individuals in the sample are converted to a random number of type 2 individuals according to the number M of mutations that occur on a faster timescale. The resulting sample that contains only individuals of type 2 then follows the usual Kingman's coalescent.

Theorem 2.5 (Genealogies under strong selection)**.** *Let M be a random variable whose distribution is*

$$\mathbb{P}(M = m \mid n_1) = \frac{\theta_1^m}{\theta_{1(n_1)}} \left| S_{n_1}^{(m)} \right|, \quad 0 \leq m \leq n_1, \tag{2.17}$$

where $S_{n_1}^{(m)}$ are the Stirling numbers of the first kind.

Consider the genealogy generated by the ancestral dual process of Theorem 2.1 for a sample $(n_1, n_2) \in \mathbb{Z}_+^2$ of size $n_1 + n_2 \geq 1$. The mean and variance of the time to the MRCA satisfy

$$\lim_{\sigma \to \infty} \mathbb{E}[T_{MRCA}(n_1, n_2)] = 2 \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1) \left(1 - \frac{1}{n_2 + m} \right), \tag{2.18}$$

$$\lim_{\sigma \to \infty} \mathrm{Var}[T_{MRCA}(n_1, n_2)] = 4 \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1) \left[2 \sum_{i=2}^{n_2+m} \frac{1}{i^2} - \left(1 - \frac{1}{n_2 + m} \right)^2 \right]; \tag{2.19}$$

and the mean and variance of the length of the genealogy satisfy

$$\lim_{\sigma \to \infty} \mathbb{E}[L_{MRCA}(n_1, n_2)] = 2 \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1) \sum_{i=1}^{n_2+m-1} \frac{1}{i}, \tag{2.20}$$

$$\lim_{\sigma \to \infty} \mathrm{Var}[L_{MRCA}(n_1, n_2)] = 4 \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1) \sum_{i=1}^{n_2+m-1} \frac{1}{i^2}. \tag{2.21}$$

Remark 2.6 (Ewens sampling formula)**.**

1. The distribution of M in equation (2.17) is related to the famous Ewens sampling formula discovered by Ewens (1972), which itself is a more detailed formula

for the distribution of the number of times the alleles are represented in a random neutral sample in the infinitely-many-alleles mutation model where every mutation leads to a new allele. M can be interpreted as the number of distinct alleles in a neutral sample of size n_1 in the infinitely-many-alleles mutation model. M can also be understood as the number of mutations that occur until the number of ancestral lines is absorbed from n_1 into 0 in a process, where, when there are $i \geq 2$ ancestral lines, a line is lost through a coalescence event at rate $i(i-1)/2$ or a mutation event at rate $i\theta_1/2$ (see Tavaré, 2004, Section 3.6).

2. Some useful properties of the Stirling numbers are listed in Remark 2.4. In particular, equation (2.13) implies that

$$\sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1) = 1, \quad n_1 \geq 0.$$

Before proceeding to the proof of Theorem 2.5, we present some supporting lemmas. The main ingredient in the proof of Theorem 2.5 is the asymptotic expansions of the transition rates (2.6) to (2.10) for large σ, based on the following lemma.

Lemma 2.7 (Confluent hypergeometric function for large σ)**.** *Consider the confluent hypergeometric function of the first kind, $M(\theta_1, \theta, -\sigma)$, where $\theta = \theta_1 + \theta_2$, $\theta_1, \theta_2 > 0$ and $\sigma > 0$, as defined in equation* (1.12)*. For large σ,*

$$M(\theta_1, \theta, -\sigma) = \frac{\Gamma(\theta)}{\Gamma(\theta_2)\sigma^{\theta_1}} \left[1 + \frac{\theta_1(1-\theta_2)}{\sigma} + \frac{\theta_1(\theta_1+1)(1-\theta_2)(2-\theta_2)}{2\sigma^2} + O(\sigma^{-3}) \right]. \tag{2.22}$$

Proof. See Appendix A.1. □

By applying equation (2.22), we obtain the asymptotic expansions of the transition rates of the ancestral dual process of Theorem 2.1 in the following lemma.

Lemma 2.8. *Consider the non-zero entries* (2.6) *to* (2.11) *in the* $(\boldsymbol{r}, \boldsymbol{v})$*th row of the Q-matrix of the dual process of Theorem 2.1. For large* σ*, these entries can be rewritten as*

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) &= \frac{1}{2}(r+v)(r_1+v_1+\theta_1)[1+\sigma^{-1}(1-r_2-v_2-\theta_2)]+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{r_1(r_1-1)}{2(r_1+v_1-1+\theta_1)}[\sigma+r_2+v_2-1+\theta_2\\
&\qquad +\sigma^{-1}(r_1+v_1+\theta_1)(r_2+v_2-1+\theta_2)]+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1}{2}r_2(r_2-1)[1+\sigma^{-1}(r_1+v_1+\theta_1)]+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1)) &= \frac{v_1(2r_1+v_1-1)}{2(r_1+v_1-1+\theta_1)}[\sigma+r_2+v_2-1+\theta_2\\
&\qquad +\sigma^{-1}(r_1+v_1+\theta_1)(r_2+v_2-1+\theta_2)]+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_2)) &= \frac{1}{2}v_2(2r_2+v_2-1)[1+\sigma^{-1}(r_1+v_1+\theta_1)]+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1}{2}\sigma^{-1}r_2\theta_2(r_1+v_1+\theta_1)+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_2-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{r_1\theta_1}{2(r_1+v_1-1+\theta_1)}[\sigma+r_2+v_2+\theta_2-r_1-v_1-\theta_1\\
&\qquad +\sigma^{-1}(r_1+v_1+\theta_1)(r_2+v_2-1+\theta_2)]+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2)) &= \frac{1}{2\sigma}v_2\theta_2(r_1+v_1+\theta_1)+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_2-\boldsymbol{e}_1)) &= \frac{v_1\theta_1}{2(r_1+v_1-1+\theta_1)}[\sigma+r_2+v_2+\theta_2-r_1-v_1-\theta_1\\
&\qquad +\sigma^{-1}(r_1+v_1+\theta_1)(r_2+v_2-1+\theta_2)]+O(\sigma^{-2}),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v})) &= -\frac{1}{2}[(r_1+v_1)\sigma+(r+v)(r+v-1)+(r_1+v_1)\theta_2+(r_2+v_2)\theta_1].
\end{aligned}
$$

Proof. See Appendix A.2. □

We want to evaluate the expectation and variance of the height and the length of a genealogy in the ancestral dual process under strong selection, as in Theorem 2.5. Calculations that are required rely on the recursive equations in the following lemma.

Lemma 2.9. *Let β and γ be functions defined on $\mathbb{Z}_+^4 \setminus \{\mathbf{0}\}$. Suppose β and γ satisfy*

$$\begin{aligned}\gamma(r_1,r_2,v_1,v_2) = \frac{1}{-q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}))} \times \{&\beta(r_1,r_2,v_1,v_2)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1))\gamma(r_1,r_2,v_1+1,v_2)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v}))\gamma(r_1-1,r_2,v_1,v_2)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v}))\gamma(r_1,r_2-1,v_1,v_2)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1))\gamma(r_1,r_2,v_1-1,v_2)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_2))\gamma(r_1,r_2,v_1,v_2-1)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{v}))\gamma(r_1+1,r_2-1,v_1,v_2)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_2-\boldsymbol{e}_1,\boldsymbol{v}))\gamma(r_1-1,r_2+1,v_1,v_2)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2))\gamma(r_1,r_2,v_1+1,v_2-1)\\ &+ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_2-\boldsymbol{e}_1))\gamma(r_1,r_2,v_1-1,v_2+1)\},\end{aligned} \tag{2.23}$$

where $q(\cdot,\cdot)$ are the transition rates given in equations (2.6) *to* (2.11)*. The boundary conditions are $\gamma(1,0,v_1,v_2)$ and $\gamma(0,1,v_1,v_2)$.*

If $r_1+v_1>0$, then for large σ,

$$\begin{aligned}\gamma(r_1,r_2,v_1,v_2) = &\frac{r_1(r_1-1)}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1-1,r_2,v_1,v_2)\\ &+ \frac{v_1(2r_1+v_1-1)}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1,r_2,v_1-1,v_2)\\ &+ \frac{r_1\theta_1}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1-1,r_2+1,v_1,v_2)\\ &+ \frac{v_1\theta_1}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1,r_2,v_1-1,v_2+1)\\ &+ \frac{\phi_1}{\sigma(r_1+v_1)} + O(\sigma^{-2}),\end{aligned} \tag{2.24}$$

where the boundary conditions remain the same, and

$$\begin{aligned}\phi_1 = &2\beta(r_1,r_2,v_1,v_2) + \gamma(r_1,r_2,v_1+1,v_2)(r+v)(r_1+v_1+\theta_1)\\ &+ \gamma(r_1,r_2-1,v_1,v_2)r_2(r_2-1) + \gamma(r_1,r_2,v_1,v_2-1)v_2(2r_2+v_2-1)\end{aligned}$$

$$
\begin{aligned}
&+ [(r_1+v_1)(r_1+v_1-1+\theta_1)]^{-1} \\
&\quad \times \{\gamma(r_1-1,r_2,v_1,v_2) r_1(r_1-1)[(r_1+v_1)(r_2+v_2-1+\theta_2)+\phi_2] \\
&\qquad + \gamma(r_1,r_2,v_1-1,v_2) v_1(2r_1+v_1-1)[(r_1+v_1)(r_2+v_2-1+\theta_2)+\phi_2] \\
&\qquad + \gamma(r_1-1,r_2+1,v_1,v_2) r_1\theta_1[(r_1+v_1)(r_2+v_2+\theta_2-r_1-v_1-\theta_1)-\phi_2] \\
&\qquad + \gamma(r_1,r_2,v_1-1,v_2+1) v_1\theta_1[(r_1+v_1)(r_2+v_2+\theta_2-r_1-v_1-\theta_1)-\phi_2]\}
\end{aligned} \tag{2.25}
$$

with $\phi_2 = (r+v)(r+v-1)+(r_1+v_1)\theta_2+(r_2+v_2)\theta_1$.

Proof. Equation (2.23) arises by considering the first transition when the dual process with transition rates (2.6) to (2.10) is at the state (r_1, r_2, v_1, v_2). Equation (2.24) is derived by substituting the transition rates of the form in Lemma 2.8 into equation (2.23). By rearranging terms, it follows that

$$
\begin{aligned}
\gamma(r_1,r_2,v_1,v_2) = &\frac{r_1(r_1-1)}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1-1,r_2,v_1,v_2) \\
&+ \frac{v_1(2r_1+v_1-1)}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1,r_2,v_1-1,v_2) \\
&+ \frac{r_1\theta_1}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1-1,r_2+1,v_1,v_2) \\
&+ \frac{v_1\theta_1}{(r_1+v_1)(r_1+v_1-1+\theta_1)}\gamma(r_1,r_2,v_1-1,v_2+1) \\
&+ \frac{\phi_1}{\sigma(r_1+v_1)} + O(\sigma^{-2}),
\end{aligned}
$$

where ϕ_1 is given in equation (2.25). □

The next corollary follows immediately from Lemma 2.9.

Corollary 2.10. *Let* β *and* γ *be functions defined on* $\mathbb{Z}_+^4 \setminus \{\mathbf{0}\}$. *If* β *and* γ *satisfy the recursive equation* (2.23) *with* $r_1 = 0$ *and* $v_1 = 1$, *then for large* σ,

$$
\gamma(0,r_2,1,v_2) = \gamma(0,r_2,0,v_2+1) + \frac{1}{\sigma}\phi + O(\sigma^{-2}), \tag{2.26}
$$

where

$$\begin{aligned}\phi &= 2\beta(0,r_2,1,v_2)+\gamma(0,r_2,2,v_2)(r_2+1+v_2)(1+\theta_1)\\ &\quad+\gamma(0,r_2-1,1,v_2)r_2(r_2-1)+\gamma(0,r_2,1,v_2-1)v_2(2r_2+v_2-1)\\ &\quad+\gamma(0,r_2,0,v_2+1)[(r_2+v_2)(r_2+v_2-\theta_1)-1-\theta_1]. \end{aligned} \tag{2.27}$$

When $r_1+v_1=0$, the assumption in Lemma 2.9 no longer holds. We then need the following lemma. Heuristically, the above corollary and the following lemma suggest that under strong selection, when a line branches, a virtual line of type 2 is added to the genealogy instead of the line of type 1.

Lemma 2.11. *Let β and γ be functions defined on $\mathbb{Z}_+^4\setminus\{\mathbf{0}\}$. If β and γ satisfy the recursive equation* (2.23) *with $r_1=0$ and $v_1=0$, then for large σ,*

$$\begin{aligned}\gamma(0,r_2,0,v_2) &= \frac{2\beta(0,r_2,0,v_2)}{(r_2+v_2)(r_2+v_2-1+\theta_1)}+\frac{(r_2+v_2)\theta_1}{(r_2+v_2)(r_2+v_2-1+\theta_1)}\gamma(0,r_2,0,v_2+1)\\ &\quad+\frac{r_2(r_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)}\gamma(0,r_2-1,0,v_2)\\ &\quad+\frac{v_2(2r_2+v_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)}\gamma(0,r_2,0,v_2-1)\\ &\quad+\frac{\theta_1\phi}{\sigma(r_2+v_2)(r_2+v_2-1+\theta_1)}+O(\sigma^{-2}),\end{aligned}$$

where

$$\begin{aligned}\phi &= \gamma(0,r_2,0,v_2+1)(r_2+v_2)(1-r_2-v_2-\theta_2)\\ &\quad+\gamma(0,r_2-1,0,v_2)r_2(r_2-1)+\gamma(0,r_2,0,v_2-1)v_2(2r_2+v_2-1)\\ &\quad+(r_2+v_2)\times\{2\beta(0,r_2,1,v_2)+\gamma(0,r_2,2,v_2)(r_2+v_2+1)(1+\theta_1)\\ &\qquad+\gamma(0,r_2-1,1,v_2)r_2(r_2-1)+\gamma(0,r_2,1,v_2-1)v_2(2r_2+v_2-1)\\ &\qquad+\gamma(0,r_2,0,v_2+1)[(r_2+v_2)(r_2+v_2+\theta_1)-1-\theta_1]\}. \end{aligned} \tag{2.28}$$

Proof. When $r_1 = 0$ and $v_1 = 0$, we have

$$\begin{aligned}\gamma(0, r_2, 0, v_2) = \frac{1}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \times \Big\{ & 2\beta(0, r_2, 0, v_2) \\ & + (r_2+v_2)\sigma \frac{\theta_1}{r_2+v_2+\theta} \frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_1+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \gamma(0, r_2, 1, v_2) \\ & + r_2(r_2-1) \frac{r_2+v_2-1+\theta}{r_2+v_2-1+\theta_2} \frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_2+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \gamma(0, r_2-1, 0, v_2) \\ & + v_2(2r_2+v_2-1) \frac{r_2+v_2-1+\theta}{r_2+v_2-1+\theta_2} \frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_2+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \gamma(0, r_2, 0, v_2-1) \\ & + r_2\theta_2 \frac{\theta_1}{r_2+v_2-1+\theta_2} \frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \gamma(1, r_2-1, 0, v_2) \\ & + v_2\theta_2 \frac{\theta_1}{r_2+v_2-1+\theta_2} \frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \gamma(0, r_2, 1, v_2-1) \Big\}.\end{aligned}$$

By substituting the transition rates in Lemma 2.8 into the above equation, we obtain

$$\begin{aligned}\gamma(0, r_2, 0, v_2) = & \frac{2\beta(0, r_2, 0, v_2)}{(r_2+v_2)(r_2+v_2-1+\theta_1)} + \frac{(r_2+v_2)\theta_1}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \gamma(0, r_2, 1, v_2) \\ & + \frac{r_2(r_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \gamma(0, r_2-1, 0, v_2) \\ & + \frac{v_2(2r_2+v_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \gamma(0, r_2, 0, v_2-1) \\ & + \frac{\theta_1\phi_1}{\sigma(r_2+v_2)(r_2+v_2-1+\theta_1)} + O(\sigma^{-2}), \end{aligned} \tag{2.29}$$

where

$$\begin{aligned}\phi_1 = & \gamma(0, r_2, 1, v_2)(r_2+v_2)(1-r_2-v_2-\theta_2) + \gamma(0, r_2-1, 0, v_2) r_2(r_2-1) \\ & + \gamma(0, r_2, 0, v_2-1) v_2(2r_2+v_2-1).\end{aligned}$$

By substituting equation (2.26) in Corollary 2.10 into equation (2.29), it follows that

$$\begin{aligned}\gamma(0, r_2, 0, v_2) = & \frac{2\beta(0, r_2, 0, v_2)}{(r_2+v_2)(r_2+v_2-1+\theta_1)} + \frac{(r_2+v_2)\theta_1}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \gamma(0, r_2, 0, v_2+1) \\ & + \frac{r_2(r_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \gamma(0, r_2-1, 0, v_2) \\ & + \frac{v_2(2r_2+v_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \gamma(0, r_2, 0, v_2-1) \\ & + \frac{\theta_1\phi}{\sigma(r_2+v_2)(r_2+v_2-1+\theta_1)} + O(\sigma^{-2}),\end{aligned}$$

where ϕ is given in equation (2.28). The proof is then completed. □

For the proof of Theorem 2.5, we still need the following Lemma 2.12 due to Krone and Neuhauser (1997, Theorem 3.12). This lemma is intuitive because when there is no mutation, the population consists of only one allele, so selection cannot have an effect on the genealogy of a sample. Our proof of the lemma is also an easier circumscription of Krone and Neuhauser's proof.

Lemma 2.12 (Ancestral process without mutation)**.** *Consider an ancestral process with selection but no mutation for a sample containing $n \geq 1$ individuals of the same type. We suppose, when the configuration of ancestral lines is (r, v), the process has the following transition rates*

$$q((r,v),(r,v+1)) = \frac{1}{2}(r+v)\psi, \tag{2.30}$$

$$q((r,v),(r-1,v)) = \frac{1}{2}r(r-1), \tag{2.31}$$

$$q((r,v),(r,v-1)) = \frac{1}{2}v(2r+v-1), \tag{2.32}$$

where ψ is some positive selection coefficient. The real genealogy in this ancestral process follows Kingman's coalescent for a sample of size n.

Proof. The transitions at the rates (2.30) and (2.32) change the configuration of virtual lines. The rates of these transitions are determined by both the number of real lines and the number of virtual lines. The configuration of real lines changes according to the coalescence events at rate (2.31), which depends only on the number of real lines, and is the one usually found in Kingman's coalescent. The lemma thus follows. □

The following is the proof of the main theorem of this section.

Proof of Theorem 2.5. We prove equation (2.18) for the values of $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ in the limit as $\sigma \to \infty$. The other equations (2.19) to (2.21) can be shown with an analogous argument.

The recursive equation (2.16) implies that $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ satisfies the recursive equation (2.23) with $v_1 = v_2 = 0$ and $\beta = 1$. Hence, by Lemma 2.9, for large σ, $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ satisfies the following recursive equation where no branching events and virtual lines are involved:

$$\begin{aligned}\mathbb{E}[T_{MRCA}(n_1, n_2)] = &\frac{n_1 - 1}{n_1 - 1 + \theta_1} \mathbb{E}[T_{MRCA}(n_1 - 1, n_2)] \\ &+ \frac{\theta_1}{n_1 - 1 + \theta_1} \mathbb{E}[T_{MRCA}(n_1 - 1, n_2 + 1)] + O(\sigma^{-1}), \qquad (2.33)\end{aligned}$$

with boundary conditions $\mathbb{E}[T_{MRCA}(1, 0)] = \mathbb{E}[T_{MRCA}(0, 1)] = 0$.

By iterating equation (2.33), it follows that

$$\begin{aligned}\mathbb{E}[T_{MRCA}(n_1, n_2)] = &\frac{(n_1 - 1)(n_1 - 2)}{(n_1 - 1 + \theta_1)(n_1 - 2 + \theta_1)} \mathbb{E}[T_{MRCA}(n_1 - 2, n_2)] \\ &+ \frac{(n_1 - 1)\theta_1}{(n_1 - 1 + \theta_1)(n_1 - 2 + \theta_1)} \mathbb{E}[T_{MRCA}(n_1 - 2, n_2 + 1)] \\ &+ \frac{\theta_1(n_1 - 2)}{(n_1 - 1 + \theta_1)(n_1 - 2 + \theta_1)} \mathbb{E}[T_{MRCA}(n_1 - 2, n_2 + 1)] \\ &+ \frac{\theta_1^2}{(n_1 - 1 + \theta_1)(n_1 - 2 + \theta_1)} \mathbb{E}[T_{MRCA}(n_1 - 2, n_2 + 2)] + O(\sigma^{-1}).\end{aligned}$$

We can eventually obtain

$$\mathbb{E}[T_{MRCA}(n_1, n_2)] = \sum_{m=1}^{n_1} \mathbb{P}(M = m \mid n_1) \, \mathbb{E}[T_{MRCA}(0, n_2 + m)] + O(\sigma^{-1}), \qquad (2.34)$$

where the distribution of M is analogous to that of M heads in a series of n_1 coin tosses such that

$$\mathbb{P}(M = m \mid n_1) = \mathbb{P}\left(\sum_{i=0}^{n_1} B_i = m \;\middle|\; n_1 \right),$$

in which B_i are independent, but not identically distributed, random variables having Bernoulli distribution with success probability $\theta_1/(\theta_1 + i)$. It is well known (see Tavaré, 2004, Section 3.6) that the distribution of M then follows the classic Ewens sampling formula such that

$$\mathbb{P}(M = m \mid n_1) = \frac{\theta_1^m}{\theta_{1(n_1)}} \left| S_{n_1}^{(m)} \right|, \quad 0 < m \leq n_1,$$

where $S_{n_1}^{(m)}$ are the Stirling numbers of the first kind. Since $|S_0^{(0)}| = 1$, if $n_1 = 0$, $\mathbb{P}(M = 0 \mid n_1) = 1$.

We can also algebraically verify that equation (2.34) is the unique solution to the recursive equation (2.33). The boundary conditions obviously hold. By substituting equation (2.34) into the right-hand side of equation (2.33), we have

$$\begin{aligned}
RHS &= \frac{n_1 - 1}{n_1 - 1 + \theta_1} \sum_{m=1}^{n_1-1} \frac{\theta_1^m}{\theta_{1(n_1-1)}} \left|S_{n_1-1}^{(m)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + m)] \\
&\quad + \frac{\theta_1}{n_1 - 1 + \theta_1} \sum_{m=1}^{n_1-1} \frac{\theta_1^m}{\theta_{1(n_1-1)}} \left|S_{n_1-1}^{(m)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + 1 + m)] + O(\sigma^{-1}) \\
&= \sum_{m=1}^{n_1-1} \frac{\theta_1^m}{\theta_{1(n_1)}} (n_1 - 1) \left|S_{n_1-1}^{(m)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + m)] \\
&\quad + \sum_{m=2}^{n_1} \frac{\theta_1^m}{\theta_{1(n_1)}} \left|S_{n_1-1}^{(m-1)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + m)] + O(\sigma^{-1}) \\
&= \frac{\theta_1}{\theta_{1(n_1)}} (n_1 - 1) \left|S_{n_1-1}^{(1)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + 1)] + \frac{\theta_1^{n_1}}{\theta_{1(n_1)}} \left|S_{n_1-1}^{(n_1-1)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + n_1)] \\
&\quad + \sum_{m=2}^{n_1-1} \frac{\theta_1^m}{\theta_{1(n_1)}} \left|S_{n_1}^{(m)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + m)] + O(\sigma^{-1}) \\
&= \frac{\theta_1}{\theta_{1(n_1)}} \left|S_{n_1}^{(1)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + 1)] + \frac{\theta_1^{n_1}}{\theta_{1(n_1)}} \left|S_{n_1}^{(n_1)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + n_1)] \\
&\quad + \sum_{m=2}^{n_1-1} \frac{\theta_1^m}{\theta_{1(n_1)}} \left|S_{n_1}^{(m)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + m)] + O(\sigma^{-1}) \\
&= \sum_{m=1}^{n_1} \frac{\theta_1^m}{\theta_{1(n_1)}} \left|S_{n_1}^{(m)}\right| \mathbb{E}[T_{MRCA}(0, n_2 + m)] + O(\sigma^{-1}) \\
&= LHS,
\end{aligned}$$

where the third equality follows the recursive identity (2.14) for the Stirling numbers, and the fourth equality holds because

$$(n_1 - 1)\left|S_{n_1-1}^{(1)}\right| = \left|S_{n_1}^{(1)}\right| - \left|S_{n_1-1}^{(0)}\right| = \left|S_{n_1}^{(1)}\right|,$$

and

$$\left|S_{n_1-1}^{(n_1-1)}\right| = \left|S_{n_1}^{(n_1)}\right| = 1.$$

Now consider $\mathbb{E}[T_{MRCA}(0, n_2+m)] \equiv \mathbb{E}[T_{MRCA}(0, n_2+m, 0, 0)]$, which satisfies the recursive equation (2.23) with $r_1 = v_1 = 0$ and $\beta = 1$. By Lemma 2.11, we have, for large σ,

$$\begin{aligned}\mathbb{E}[T_{MRCA}(0, r_2, 0, v_2)] = &\frac{2}{(r_2+v_2)(r_2+v_2-1+\theta_1)} \\ &+ \frac{(r_2+v_2)\theta_1}{(r_2+v_2)(r_2+v_2-1+\theta_1)}\,\mathbb{E}[T_{MRCA}(0, r_2, 0, v_2+1)] \\ &+ \frac{r_2(r_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)}\,\mathbb{E}[T_{MRCA}(0, r_2-1, 0, v_2)] \\ &+ \frac{v_2(2r_2+v_2-1)}{(r_2+v_2)(r_2+v_2-1+\theta_1)}\,\mathbb{E}[T_{MRCA}(0, r_2, 0, v_2-1)] + O(\sigma^{-1}),\end{aligned}$$

where the boundary condition is $\mathbb{E}[T_{MRCA}(0, 1, 0, v_2)] = 0$. In the limit as $\sigma \to \infty$, the distribution of the time to the MRCA is equal to that in a dual process $\{(r_2(t), v_2(t)), t \geq 0\}$ whose initial state is $(n_2+m, 0)$, and the transitions and rates are

$$(r_2, v_2) \to \begin{cases} (r_2, v_2+1) & \text{at rate } \frac{1}{2}(r_2+v_2)\theta_1, \\ (r_2-1, v_2) & \text{at rate } \frac{1}{2}r_2(r_2-1), \\ (r_2, v_2-1) & \text{at rate } \frac{1}{2}v_2(2r_2+v_2-1). \end{cases}$$

Lemma 2.12 states that the distribution of real genealogy in such an ancestral process follows Kingman's coalescent for a sample of size n_2+m. Thus, we have

$$\lim_{\sigma\to\infty} \mathbb{E}[T_{MRCA}(0, n_2+m)] = 2\left(1 - \frac{1}{n_2+m}\right).$$

Equation (2.18) for the limit of $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ as $\sigma \to \infty$, is thus proved by substituting the above equation into equation (2.34). □

To provide additional support to Theorem 2.5, we simulate the height and the length of the genealogies for samples $(10, 0)$, $(5, 5)$ and $(0, 10)$, with respect to different values of σ. The limiting values calculated from equations (2.18) and (2.20) in Theorem 2.5 are listed in the bottom row of Table 2.4, and simulation results are in other rows. The agreement between the simulated and the theoretic results is quite good, especially in view of the fact that as $\sigma \to \infty$, simulations become more and

Table 2.4: The mean time to the MRCA and the length of the genealogy in the ancestral dual process with typed lines, with respect to various sample configurations and increasing values of σ. The mutation coefficients are $\theta_1 = \theta_2 = 2$.

σ	$\mathbb{E}[T_{MRCA}(10,0)]$	$\mathbb{E}[L_{MRCA}(10,0)]$	$\mathbb{E}[T_{MRCA}(5,5)]$	$\mathbb{E}[L_{MRCA}(5,5)]$	$\mathbb{E}[T_{MRCA}(0,10)]$	$\mathbb{E}[L_{MRCA}(0,10)]$
0	1.64	5.00	1.85	5.86	1.63	4.99
10^2	1.43	3.64	1.75	5.21	1.77	5.55
10^3	1.43	3.54	1.75	5.16	1.81	5.67
10^4	1.43	3.54	1.75	5.16	1.81	5.67
∞	1.43	3.52	1.74	5.14	1.80	5.66

The bottom row lists the limiting values calculated from Theorem 2.5. The other values are obtained from simulations of the ancestral dual process of Theorem 2.1.

more unstable due to the evaluation of the confluent hypergeometric function.

The next corollary shows that in the limit as $\sigma \to \infty$, the distributions of the height and the length of the real genealogy of a sample (n_1, n_2) with $n_1 \leq 1$ is the same as those in Kingman's coalescent. In contrast, the genealogy of a sample (n_1, n_2) with $n_1 > 1$ can be very different. This justifies the observation in Figure 2.3 that when $\theta_1 = \theta_2 = 1.5$ and $\sigma = 1000$, both the values of $\mathbb{E}[T_{MRCA}(0, 10)]$ and $\mathbb{E}[T_{MRCA}(1, 9)]$ are approximately 1.8, which is the expected height of the genealogy in Kingman's coalescent for a sample of size 10.

Corollary 2.13. *Consider the dual process of Theorem 2.1 for a sample (n_1, n_2), where $n_1 = 1$, $n_2 \in \mathbb{Z}_+$. Let $n = 1 + n_2$. The mean and variance of the time to the MRCA satisfy*

$$\lim_{\sigma\to\infty} \mathbb{E}[T_{MRCA}(1, n_2)] = 2\left(1 - \frac{1}{n}\right), \tag{2.35}$$

$$\lim_{\sigma\to\infty} \mathrm{Var}[T_{MRCA}(1, n_2)] = 8\sum_{i=2}^{n} \frac{1}{i^2} - 4\left(1 - \frac{1}{n}\right)^2; \tag{2.36}$$

and the mean and variance of the length of the genealogy satisfy

$$\lim_{\sigma\to\infty} \mathbb{E}[L_{MRCA}(1, n_2)] = 2\sum_{i=1}^{n-1} \frac{1}{i}, \tag{2.37}$$

$$\lim_{\sigma\to\infty} \mathrm{Var}[L_{MRCA}(1, n_2)] = 4\sum_{i=1}^{n-1} \frac{1}{i^2}. \tag{2.38}$$

Proof. Since $|S_1^{(0)}| = 0$ and $|S_1^{(1)}| = 1$, we have

$$\mathbb{P}(M = m \mid n_1 = 1) = \begin{cases} 1, & \text{if } m = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Equations (2.35) to (2.38) then follow immediately from the corresponding equations in Theorem 2.5 and the above equation. □

We notice that these limiting values in Theorem 2.5 for a sample (n_1, n_2) cannot be larger than those for a sample of size $n_1 + n_2$ in Kingman's coalescent, since the maximum value of M is n_1. This agrees with the simulations in Sections 2.3.1 and 2.3.2, where we see that genealogies under strong selection can only be compressed when compared with those in Kingman's coalescent. In the next corollary, the upper and lower bounds of $\mathbb{E}[T_{MRCA}(n_1, n_2)]$ in the limit as $\sigma \to \infty$ are deduced. The bounds of the other quantities in Theorem 2.5 can be given analogously.

Corollary 2.14. *For $n_1, n_2 \in \mathbb{Z}_+$ and $n_2 > 0$, we have*

$$2\left(1 - \frac{1}{n_2}\right) \leq \lim_{\sigma \to \infty} \mathbb{E}[T_{MRCA}(n_1, n_2)] \leq 2\left(1 - \frac{1}{n}\right).$$

Proof. We have, as shown in Theorem 2.5,

$$\lim_{\sigma \to \infty} \mathbb{E}[T_{MRCA}(n_1, n_2)] = 2 \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1)\left(1 - \frac{1}{n_2 + m}\right).$$

Since $0 < m \leq n_1$ for $n_1 > 0$, and $m = 0$ if $n_1 = 0$, it follows that

$$1 - \frac{1}{n_2} \leq 1 - \frac{1}{n_2 + m} \leq 1 - \frac{1}{n_1 + n_2},$$

which implies

$$\begin{aligned} 2 \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1)\left(1 - \frac{1}{n_2}\right) \leq \lim_{\sigma \to \infty} \mathbb{E}[T_{MRCA}(n_1, n_2)] \\ \leq 2 \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1)\left(1 - \frac{1}{n}\right). \end{aligned}$$

Therefore, we have

$$2\left(1-\frac{1}{n_2}\right) \leq \lim_{\sigma\to\infty} \mathbb{E}[T_{MRCA}(n_1,n_2)] \leq 2\left(1-\frac{1}{n}\right).$$

□

The following theorem shows that as $\sigma \to \infty$, the distributions of the height and the length of the real genealogy of a random sample are the same as those in Kingman's coalescent. This is consistent with the simulation results in Table 2.2 in the cases where $\theta_1 = \theta_2 = 1.5$ or 10^{-4}, and $\sigma = 1000$.

Theorem 2.15 (Genealogies of random samples under strong selection)**.** *Consider a random sample of size $n \geq 1$ drawn from the stationary population. The mean and variance of time to the MRCA satisfy*

$$\lim_{\sigma\to\infty} \mathbb{E}[T_{MRCA}(n)] = 2\left(1-\frac{1}{n}\right), \tag{2.39}$$

$$\lim_{\sigma\to\infty} \mathrm{Var}[T_{MRCA}(n)] = 8\sum_{i=2}^{n}\frac{1}{i^2} - 4\left(1-\frac{1}{n}\right)^2; \tag{2.40}$$

and the mean and variance of the length of the genealogy satisfy

$$\lim_{\sigma\to\infty} \mathbb{E}[L_{MRCA}(n)] = 2\sum_{i=1}^{n-1}\frac{1}{i}, \tag{2.41}$$

$$\lim_{\sigma\to\infty} \mathrm{Var}[L_{MRCA}(n)] = 4\sum_{i=1}^{n-1}\frac{1}{i^2}. \tag{2.42}$$

Proof. We are going to show that the sampling probability converges to 1 as $\sigma \to \infty$ and then a random sample drawn from the population is neutral.

Recall that we define $\mathscr{H}(n_1,n_2)$ in equation (1.13) to be the probability of obtaining an ordered sample with configuration (n_1,n_2) from the stationary population. The corresponding unordered sampling probability is $\binom{n}{n_1}\mathscr{H}(n_1,n_2)$.

Given equation (2.22), we have, for large σ,

$$\mathscr{H}(n_1,n_2) = \frac{\theta_{1(n_1)}}{\sigma^{n_1}} + O(\sigma^{-n_1-1}).$$

Thus,

$$\lim_{\sigma\to\infty} \mathscr{H}(n_1, n_2) = \begin{cases} 1, & \text{if } n_1 = 0, \\ 0, & \text{if } n_1 > 0. \end{cases}$$

That is, in the limit of $\sigma \to \infty$, a random sample drawn from the population is neutral with probability 1. The theorem then follows immediately. □

Simulations in Section 2.3.3 show that as the value of σ becomes larger, even when there are no type 2 individuals in the sample, the probability that the MRCA is type 2, approaches 1. This result is proved in the next theorem.

Theorem 2.16 (MRCA type distribution under strong selection)**.** *Consider the dual process of Theorem 2.1 for a sample* $(n_1, n_2) \in \mathbb{Z}_+^2$ *of size* $n_1 + n_2 \geq 1$*. Let* $\rho(n_1, n_2)$ *be the probability that the MRCA is type* 2*, as defined in equation* (2.15)*. The probability* $\rho(n_1, n_2)$ *satisfies*

$$\lim_{\sigma\to\infty} \rho(n_1, n_2) = 1.$$

Proof. The probability $\rho(n_1, n_2)$ satisfies the recursive equation (2.23) with $v_1 = v_2 = 0$ and $\beta = 0$. Hence, by Lemma 2.9, for large σ, $\rho(n_1, n_2)$ satisfies the recursive equation

$$\rho(n_1, n_2) = \frac{n_1 - 1}{n_1 - 1 + \theta_1}\rho(n_1 - 1, n_2) + \frac{\theta_1}{n_1 - 1 + \theta_1}\rho(n_1 - 1, n_2 + 1) + O(\sigma^{-1}),$$

with boundary conditions $\rho(1, 0) = 0$ and $\rho(0, 1) = 1$. We have

$$\lim_{\sigma\to\infty} \rho(n_1, n_2) = \frac{n_1 - 1}{n_1 - 1 + \theta_1} \lim_{\sigma\to\infty} \rho(n_1 - 1, n_2) + \frac{\theta_1}{n_1 - 1 + \theta_1} \lim_{\sigma\to\infty} \rho(n_1 - 1, n_2 + 1).$$

It follows that

$$\lim_{\sigma\to\infty} \rho(n_1, n_2) = \sum_{m=0}^{n_1} \left[\mathbb{P}(M = m \mid n_1) \lim_{\sigma\to\infty} \rho(0, n_2 + m) \right], \tag{2.43}$$

by using an argument similar to that in the proof of Theorem 2.5, where the distribution of M is given in equation (2.17).

Consider $\rho(0, n_2 + m) \equiv \rho(0, n_2 + m, 0, 0)$ with $n_2 + m > 0$, which satisfies the

recursive equation (2.23) with $r_1 = v_1 = 0$ and $\beta = 0$. By Lemma 2.11, we have

$$\begin{aligned}\lim_{\sigma\to\infty} \rho(0, r_2, 0, v_2) = &\frac{(r_2 + v_2)\theta_1}{(r_2 + v_2)(r_2 + v_2 - 1 + \theta_1)} \lim_{\sigma\to\infty} \rho(0, r_2, 0, v_2 + 1) \\ &+ \frac{r_2(r_2 - 1)}{(r_2 + v_2)(r_2 + v_2 - 1 + \theta_1)} \lim_{\sigma\to\infty} \rho(0, r_2 - 1, 0, v_2) \\ &+ \frac{v_2(2r_2 + v_2 - 1)}{(r_2 + v_2)(r_2 + v_2 - 1 + \theta_1)} \lim_{\sigma\to\infty} \rho(0, r_2, 0, v_2 - 1),\end{aligned}$$

where the boundary condition is $\rho(0, 1, 0, v_2) = 1$.

It is clear from the above recursive equation that no lines of type 1 can be added to the genealogy. The MRCA can only be type 2. Therefore,

$$\lim_{\sigma\to\infty} \rho(0, n_2 + m) = 1,$$

for any $n_2 + m > 0$. By substituting the above equation into equation (2.43), we have

$$\lim_{\sigma\to\infty} \rho(n_1, n_2) = \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1) = 1.$$

The theorem is thus proved. □

Chapter 3

Ancestral processes under strong selection

In the previous chapter, we have observed some interesting aspects of the genealogy under strong selection from simulations in Section 2.3. Then, in Theorems 2.5, 2.15 and 2.16, we have derived explicit formulae for those quantities simulated in the limit as the selection coefficient $\sigma \to \infty$.

In this chapter, we analyse the ancestral processes under strong selection at a process level. We consider two classes of ancestral processes that can describe the genealogy of a *given* sample configuration, namely the coalescent-in-a-random-background model, which will be introduced in Proposition 3.2, and the dual process with typed lines of Theorem 2.1. Although the two ancestral processes are based on the same forwards-in-time model for gene frequency evolution, the two processes are distinct in that in the coalescent-in-a-random-background model, there are no branching events in the ancestral process, and selection is reflected in the ancestral frequencies of alleles, while in the dual process with typed lines, the ancestral lines can branch and virtual lines can appear.

The genealogies generated by the coalescent-in-a-random-background model and

the dual process with typed lines are the same by construction. The genealogies are also the same under strong selection, as we will see in this chapter. However, it is of interest to see how the limiting arguments work in both models. The proofs of the limiting theorems are based on the idea of separation of timescales in the ancestral processes, which dates back to the investigation of the coalescent process with selfing (Nordborg and Donnelly, 1997), strong migration limit in the structured coalescent process (Nordborg and Krone, 2002) and infinite demes limit of the ancestral recombination graph in a subdivided population (Lessard and Wakeley, 2004).

Möhle (1998) provides a formal convergence result for the ancestral processes arising in population genetics where the timescales of events are drastically different in some limit. Consider a matrix $\boldsymbol{A} = (a_{ij})$. The following lemma of Möhle (1998, Lemma 1) is a generalisation of a well-known matrix equation $\lim_{k\to\infty}(\boldsymbol{I} + \boldsymbol{A}/k)^k = \exp(\boldsymbol{A})$. We define the norm $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ which is the maximum of the absolute row sums of the matrix. For a stochastic matrix $\boldsymbol{A}$, we have $\|A\|_\infty = 1$.

Lemma 3.1 (Möhle's lemma)**.** *Let $t, K \geq 0$ be fixed and let $\{c_k, k \in \mathbb{N}\}$ be a sequence of positive real numbers with $\lim_{k\to\infty} c_k = 0$. Further let A be a matrix with $\|A\|_\infty = 1$ such that $\boldsymbol{P} \coloneqq \lim_{m\to\infty} \boldsymbol{A}^m$ exists. Then*

$$\lim_{k\to\infty} \sup_{\|B\|_\infty \leq K} \|(\boldsymbol{A} + c_k\boldsymbol{B})^{[t/c_k]} - (\boldsymbol{P} + c_k\boldsymbol{B})^{[t/c_k]}\|_\infty = 0.$$

If $\{\boldsymbol{B}_k, k \in \mathbb{N}\}$ is a matrix sequence such that $\boldsymbol{G} \coloneqq \lim_{k\to\infty} \boldsymbol{P}\boldsymbol{B}_k\boldsymbol{P}$ exists, then

$$\lim_{k\to\infty} (\boldsymbol{A} + c_k\boldsymbol{B}_k)^{[t/c_k]} = \boldsymbol{P} - \boldsymbol{I} + e^{\boldsymbol{G}t}$$

for all $t > 0$.

We apply the above Möhle's lemma to the proofs of the limiting ancestral processes. The idea is that for a large selection coefficient σ, the transition rates of the ancestral processes can be $o(1)$, $O(1)$ or $O(\sigma)$. The events with transition rates of $O(\sigma)$ occur instantaneously such that between the events with transition rates of $O(1)$, the genealogy

is converted according to the equilibrium of these instantaneous events. The ancestral process in the coalescent-in-a-random-background model is time-inhomogeneous, which we approximate by a time-homogeneous process, as the fluctuations in the ancestral frequencies under strong selection are negligible. In the dual process with typed lines, as the state space of the dual process is infinite due to the mechanism of virtual lines, Möhle's lemma cannot be applied exactly. Nevertheless, these limiting results are supported by simulations.

Studies of ancestral processes with a branching-coalescing structure under strong selection have been rare. Wakeley (2008) uses the separation of timescales to describe a modified ancestral selection graph in the limit of strong purifying selection, and Wakeley and Sargsyan (2009) use a similar approach to explore the limit under strong balancing selection. Their outcomes are mainly heuristic and are based on the ancestral selection graph, which is different from the two models we consider.

The remaining part of this chapter is divided into three sections. In Section 3.1, we examine the ancestral process in the coalescent-in-a-random-background model under strong selection. In Section 3.2, we show the limiting ancestral dual process with typed lines. In Section 3.3, we investigate the limiting behaviour of the real genealogy in the ancestral dual process with three types under strong selection. The description of the limiting genealogy in Section 3.3 is heuristic, because, although the methods in Sections 3.1 and 3.2 still apply, there are a few different timescales involved in the three-type case, which make formal analysis difficult.

3.1 Coalescent in a random background

The coalescent analysis of genealogies under selection originates from the work of Kaplan et al. (1988). Their approach is to subdivide the population into genetic backgrounds according to allelic types, within which there is no selection, and then

approximate the ancestral frequencies of alleles in the population by some deterministic process, and use the structured coalescent discussed at the beginning of Chapter 1 to investigate the effect of selection on the height and the length of the genealogy. This approach is numerically practical for examining strong selection of arbitrary form, since only the ancestral frequencies need to be altered, not the structure of the ancestral process. However, when there is weak selection, the fluctuations in ancestral frequencies can no longer be ignored. Darden et al. (1989) develop a numerical scheme to calculate the coalescence times in the genealogy in the case where ancestral frequencies are stochastic. Barton et al. (2004) prove the convergence of a Markov process describing the genealogy of a Moran model as the population size $N \to \infty$, and the existence of the ancestral process in the following Proposition 3.2. Etheridge (2011, Section 5.5) provides a comprehensive guide to this model.

We now introduce the coalescent-in-a-random-background model. We will present its limit under strong selection in Theorem 3.5. There is no concept of virtual lines in the coalescent-in-a-random-background model; all ancestral lines belong to the real genealogy. For a sample $\boldsymbol{n} \in \mathbb{Z}_+^2$ of size $|\boldsymbol{n}| \geq 1$, the distribution of its genealogy can be described by an ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$, where $\alpha_i(t)$, $i = 1, 2$ is the number of ancestral lines of type i at time t in the past. The ancestral process has the state space $S = \{\boldsymbol{\alpha} \colon \boldsymbol{\alpha} \in \mathbb{Z}_+^2, 1 \leq |\boldsymbol{\alpha}| \leq |\boldsymbol{n}|\}$, and the initial state $\boldsymbol{n}$. The following proposition is due to Barton et al. (2004).

Proposition 3.2 (Coalescent in a random background)**.** *Let* $\{\rho(t), t \geq 0\}$ *on* $[0, 1]$ *be the time-reversal of the forwards-in-time diffusion process with generator* (1.10)*.* $\rho(t)$ *is the ancestral frequency of type* 1 *individuals at time* t *in the past.*

When the ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ *is at a state* $\boldsymbol{\alpha}$ *at time* t*, conditional on* $\rho(t)$*, the transition rates are*

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_1) = \binom{\alpha_1}{2} \frac{1}{\rho(t)}, \tag{3.1}$$

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_2) = \binom{\alpha_2}{2} \frac{1}{1-\rho(t)}, \tag{3.2}$$

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_1 + \boldsymbol{e}_2) = \frac{1}{2}\alpha_1\theta_1 \frac{(1-\rho(t))}{\rho(t)}, \tag{3.3}$$

$$q(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_2 + \boldsymbol{e}_1) = \frac{1}{2}\alpha_2\theta_2 \frac{\rho(t)}{1-\rho(t)}, \tag{3.4}$$

where θ_1 and θ_2 are the mutation coefficients.

Barton et al. (2004, Proposition 4.3) prove that as long as θ_1 and θ_2 are positive, the ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ is well defined as the value of $\rho(t)$ tends to 0 or 1, in the sense that the lines ancestral to the sample will have jumped to the other background before the frequency of one background actually hits zero. The ancestral process is a time-inhomogeneous Markov process, as the probability of a given transition in the time interval $(t, t + \delta t)$ depends on t. The transition rates (3.1) and (3.2) correspond to the coalescence events that occur at rates that are inversely proportional to the ancestral frequencies; the transition rates (3.3) and (3.4) correspond to mutation events that occur at rates that are scaled by the ratios of the ancestral frequencies. There are no branching events in the ancestral process. The effect of selection is reflected in the path of $\{\rho(t), t \geq 0\}$.

Recall the diffusion process $\{X(t), t \geq 0\}$ with generator (1.10), where $X(t)$ is the forwards-in-time frequency of type 1 individuals at time t. By Feller's boundary classification for one-dimensional diffusions (see Durrett, 2008, Example 7.33), if the mutation coefficient $\theta_2 \geq 0.5$, the point 0 is an entrance boundary such that the diffusion $\{X(t), t \geq 0\}$ started from a point in $(0, 1)$ can never reach 0, but the diffusion started from 0 is well-defined. The following lemma shows that the diffusion $\{X(t), t \geq 0\}$ can be approximated by a deterministic process, and the errors follow an Ornstein-Uhlenbeck process.

Lemma 3.3. *Consider the diffusion process $\{X(t), t \geq 0\}$ with generator (1.10) and*

initial value $X(0) = \zeta \in (0,1)$. *Let*

$$x(t) = \frac{\zeta}{\zeta + (1-\zeta)e^{t/2}}, \quad t \geq 0. \tag{3.5}$$

Define $Z_\sigma(t) = \sqrt{\sigma}(X(t/\sigma) - x(t))$. *As* $\sigma \to \infty$ *under strong selection, the real-valued process* $\{Z_\sigma(t), t \geq 0\}$ *converges in finite-dimensional distribution to an Ornstein-Uhlenbeck process* $\{Z(t), t \geq 0\}$ *that satisfies the stochastic differential equation*

$$\mathrm{d}Z(t) = -\frac{1}{2}(1-2x(t))Z(t)\,\mathrm{d}t + \sqrt{x(t)(1-x(t))}\,\mathrm{d}W(t),$$

where $\{W(t), t \geq 0\}$ *is a standard Brownian motion. The process* $\{Z(t), t \geq 0\}$ *approaches* 0 *with probability* 1 *as* $t \to \infty$.

Proof. The dynamics of the diffusion $\{X(t), t \geq 0\}$ with generator (1.10) can be represented by the stochastic differential equation

$$\mathrm{d}X(t) = \frac{1}{2}[-\sigma X(t)(1-X(t)) + \theta_1 - \theta X(t)]\,\mathrm{d}t + \sqrt{X(t)(1-X(t))}\,\mathrm{d}W(t),$$

where $\{W(t), t \geq 0\}$ is a standard Brownian motion. Write $\tilde{X}(t) = X(t/\sigma)$. It follows immediately that

$$\mathrm{d}\tilde{X}(t) = \frac{1}{2\sigma}[-\sigma\tilde{X}(t)(1-\tilde{X}(t)) + \theta_1 - \theta\tilde{X}(t)]\,\mathrm{d}t + \sqrt{\frac{1}{\sigma}\tilde{X}(t)(1-\tilde{X}(t))}\,\mathrm{d}W(t).$$

Define a real-valued process $\{Z_\sigma(t), t \geq 0\}$ where $Z_\sigma(t) \coloneqq \sqrt{\sigma}(\tilde{X}(t) - x(t))$ with $x(t)$ given in equation (3.5). We have $Z_\sigma(0) = 0$. It follows from Itō's formula that

$$\begin{aligned}
\mathrm{d}Z_\sigma(t) &= \sqrt{\sigma}\,\mathrm{d}\tilde{X}(t) - \sqrt{\sigma}\,\mathrm{d}x(t) \\
&= \frac{\sqrt{\sigma}}{2}\left[-\tilde{X}(t)(1-\tilde{X}(t)) + \frac{\theta_1 - \theta\tilde{X}(t)}{\sigma} + x(t)(1-x(t))\right]\mathrm{d}t + \sqrt{\tilde{X}(t)(1-\tilde{X}(t))}\,\mathrm{d}W(t) \\
&= \frac{\sqrt{\sigma}}{2}\left[-\left(\frac{1}{\sqrt{\sigma}}Z_\sigma(t) + x(t)\right)\left(1 - \frac{1}{\sqrt{\sigma}}Z_\sigma(t) - x(t)\right) + \frac{\theta_1}{\sigma} - \frac{\theta}{\sigma}\left(\frac{1}{\sqrt{\sigma}}Z_\sigma(t) + x(t)\right)\right. \\
&\qquad \left. + x(t)(1-x(t))\right]\mathrm{d}t
\end{aligned}$$

$$+\sqrt{\left(\frac{1}{\sqrt{\sigma}}Z_\sigma(t)+x(t)\right)\left(1-\frac{1}{\sqrt{\sigma}}Z_\sigma(t)-x(t)\right)}\,\mathrm{d}W(t)$$
$$\coloneqq m_\sigma(t,Z_\sigma(t))\,\mathrm{d}t+\sqrt{\nu_\sigma(t,Z_\sigma(t))}\,\mathrm{d}W(t).$$

The generator of the process $\{Z_{\theta_1}(t), t\geq 0\}$ is thus

$$\mathcal{L}_\sigma f(z)=m_\sigma(t,z)\,\frac{\mathrm{d}f(z)}{\mathrm{d}z}+\frac{1}{2}\nu_\sigma(t,z)\,\frac{\mathrm{d}^2f(z)}{\mathrm{d}z^2}\,,$$

for a twice continuously differentiable function $f\colon \mathbb{R}\to\mathbb{R}$. We have

$$\lim_{\sigma\to\infty}\mathcal{L}_\sigma f(z)=m(t,z)\,\frac{\mathrm{d}f(z)}{\mathrm{d}z}+\frac{1}{2}\nu(t,z)\,\frac{\mathrm{d}^2f(z)}{\mathrm{d}z^2}\,,$$

where

$$m(t,z)=\lim_{\sigma\to\infty}m_\sigma(t,z)=-\frac{1}{2}(1-2x(t))z,$$

and

$$\nu(t,z)=\lim_{\theta_1\to\infty}\nu_{\theta_1}(t,z)=x(t)(1-x(t)).$$

Therefore, as $\sigma\to\infty$ under strong selection, the process $\{Z_\sigma(t), t\geq 0\}$ converges in finite-dimensional distribution to an Ornstein-Uhlenbeck process $\{Z(t), t\geq 0\}$ that satisfies the stochastic differential equation

$$\mathrm{d}Z(t)=-\frac{1}{2}(1-2x(t))Z(t)\,\mathrm{d}t+\sqrt{x(t)(1-x(t))}\,\mathrm{d}W(t).$$

To deduce the stationary distribution of the process $\{Z(t), t\geq 0\}$, we apply Itō's formula to

$$Z(t)e^{\int_0^t\frac{1}{2}(1-2x(u))\,\mathrm{d}u}.$$

The solution of the resulting stochastic differential equation gives

$$Z(t)=e^{-\int_0^t\frac{1}{2}(1-2x(u))\,\mathrm{d}u}\times\int_0^t\sqrt{x(u)(1-x(u))}e^{\int_0^u\frac{1}{2}(1-2x(v))\,\mathrm{d}v}\,\mathrm{d}W(u).$$

Therefore, the distribution of $Z(t)$ is normal with mean 0 and variance

$$\mathrm{Var}(Z(t)) = e^{-\int_0^t (1-2x(u))\,\mathrm{d}u} \times \int_0^t x(u)(1-x(u))e^{\int_0^u (1-2x(v))\,\mathrm{d}v}\,\mathrm{d}u.$$

We use L'Hôpital's rule to obtain

$$\lim_{t\to\infty} \mathrm{Var}(Z(t)) = \lim_{t\to\infty} \frac{x(t)(1-x(t))}{[1-2x(t)]} = 0.$$

The proof of the lemma is completed. □

Furthermore, the following lemma shows that once the diffusion process becomes stationary, the frequency of individuals of type 1 is $O(\sigma^{-1})$ for large σ.

Lemma 3.4. *Let X be the stationary frequency of type* 1 *individuals in the population governed by the diffusion with generator* (1.10)*. Then σX converges in probability to* 1 *as $\sigma \to \infty$.*

Proof. Given the density of X in equation (1.11) and using equation (2.22) for the confluent hypergeometric functions $M(\cdot,\cdot,\cdot)$, the mean of X satisfies

$$\begin{aligned}\mathbb{E}(X) &= \frac{B(\theta_1+1,\theta_2)M(\theta_1+1,\theta+1,-\sigma)}{B(\theta_1,\theta_2)M(\theta_1,\theta,-\sigma)} \\ &= \frac{\theta_1}{\sigma} + O(\sigma^{-2}),\end{aligned}$$

for large σ; and the variance of X satisfies

$$\begin{aligned}\mathrm{Var}(X) &= \frac{B(\theta_1+2,\theta_2)M(\theta_1+2,\theta+2,-\sigma)}{B(\theta_1,\theta_2)M(\theta_1,\theta,-\sigma)} - (\mathbb{E}(X))^2 \\ &= \frac{\theta_1}{\sigma^2} + O(\sigma^{-3}).\end{aligned}$$

By Chebyshev's inequality, for any $\beta_\sigma \to \infty$ as $\sigma \to \infty$,

$$\mathbb{P}\left(|\sigma X - \theta_1| \geq \beta_\sigma\right) \leq \frac{\sigma^2 \,\mathrm{Var}(X)}{\beta_\sigma^2} \to 0,$$

as $\sigma \to \infty$. The lemma then follows. □

In the remainder of this section, we assume the ancestral frequency $\rho(t) = 1/\sigma$ for all t in coalescent-in-a-random-background model. This approximation was also considered by Kaplan et al. (1988) as the ancestral frequencies of alleles under strong selection are tightly regulated. Thus, the ancestral process of Proposition 3.2 becomes a time-homogeneous Markov process, while the state space and initial state remain the same. By ignoring terms of $o(1)$, the non-zero entries in the $\boldsymbol{\alpha}$th row of the Q-matrix of the simplified ancestral process are

$$q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_1) = \frac{1}{2}\alpha_1(\alpha_1 - 1)\sigma, \tag{3.6}$$

$$q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_2) = \frac{1}{2}\alpha_2(\alpha_2 - 1), \tag{3.7}$$

$$q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_1 + \boldsymbol{e}_2) = \frac{1}{2}\alpha_1\theta_1(\sigma - 1), \tag{3.8}$$

$$q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = -\frac{1}{2}[\alpha_1(\alpha_1 - 1)\sigma + \alpha_2(\alpha_2 - 1) + \alpha_1\theta_1(\sigma - 1)]. \tag{3.9}$$

Denote the ancestral process with such a Q-matrix by $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$. The initial state of $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ is the sample $\boldsymbol{n}$, and the state space is $S = \{\boldsymbol{\alpha} \colon \boldsymbol{\alpha} \in \mathbb{Z}_+^2, 1 \leq |\boldsymbol{\alpha}| \leq |\boldsymbol{n}|\}$.

The next theorem describes the limit of our simplified ancestral process $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ as $\sigma \to \infty$. Heuristically, the coalescence events and mutation events (to type 2) for type 1 individuals occur in the genetic background with a proportion $1/\sigma$ of the total population. Thus, these events happen on a much faster timescale; and in the equilibrium, all type 1 individuals in the sample are lost due to the coalescences and mutations. The limiting genealogy then follows Kingman's coalescent for a sample containing $M + n_2$ individuals of the same type, where M is the number of mutations which have occurred on the faster timescale.

Theorem 3.5. *For a sample $(n_1, n_2) \in \mathbb{Z}_+^2$ of size $n_1 + n_2 \geq 1$, as $\sigma \to \infty$, the ancestral process $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ with transition rates (3.6) to (3.8) converges in finite-dimensional distribution to a process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ with the same initial state and*

state space.

There is an instantaneous jump at $t = 0$ in the limiting process such that

$$\boldsymbol{\alpha}(0+) = (0, M + n_2),$$

where M is distributed as the number of alleles in the Ewens sampling formula with mutation coefficient θ_1, as in equation (2.17). For $t > 0$, the transition rates of the limiting process are

$$q((0, \alpha_2), (0, \alpha_2 - 1)) = \frac{1}{2}\alpha_2(\alpha_2 - 1), \quad \alpha_2 \geq 2. \tag{3.10}$$

Remark 3.6 (Connections to previous work)**.** Hartl and Campbell (1982) and Sawyer (1983) consider a Wright-Fisher model with infinitely-many-alleles mutation in which there are two classes of alleles, and each class consists of different alleles with the same fitness. The authors derive that the type configuration of a sample containing only disfavoured alleles follows the Ewens sampling formula, provided that selection is sufficiently strong such that the equilibrium frequency of the disfavoured alleles is $o(1)$ (cf. our assumption of the ancestral frequency of type 1 being $1/\sigma$). In our model, by Theorem 3.5, backwards in time, the disfavoured type 1 individuals in the sample are converted to a random number of type 2 individuals according to the number M of mutations that occur on a faster timescale, where M is determined by the Ewens sampling formula.

We now prove Theorem 3.5. It is well known that the exponential of $\boldsymbol{D}$, which is an $n \times n$ matrix, is defined as

$$e^{\boldsymbol{D}} = \sum_{k=0}^{\infty} \frac{1}{k!}\boldsymbol{D}^k,$$

where the series converges for any square matrix $\boldsymbol{D}$ (Bellman, 1997, Theorem 2 in

Chapter 10). The limits of matrices are defined component-wise such that

$$\left(\lim_{\sigma\to\infty} \boldsymbol{D}_\sigma\right)_{ij} \coloneqq \lim_{\sigma\to\infty} (\boldsymbol{D}_\sigma)_{ij},$$

provided all the limits on the right-hand side exist.

Proof of Theorem 3.5. Consider the ancestral process $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ with transition rates (3.6) to (3.8) for a sample (n_1, n_2) of finite size $n = n_1 + n_2 \geq 1$. Let S be the state space of the ancestral process, which is the set of all possible configurations of ancestral lines. The order of the states in S is arbitrary but fixed. We let S have three mutually disjoint subsets such that $S = \cup_{i=1}^{3} S_i$. Each state in S_1 contains only lines of type 2, and $S_1 = \{(0, n), (0, n-1), \ldots, (0, 1)\}$; each state in S_2 consists of only lines of type 1, and $S_2 = \{(n, 0), (n-1, 0), \ldots, (1, 0)\}$; and S_3 is broken up into levels that are determined by the total number of ancestral lines, and $S_3 = \{(1, n-1), (2, n-2), \ldots, (n-1, 1), (1, n-2), (2, n-3), \ldots, (n-2, 1), \ldots, (1, 1)\}$. The size of the state space is $|S| = \sum_{i=1}^{3} |S_i| = 2n + \sum_{i=1}^{n-1} i = n(n+3)/2$. The absorbing states are $(0, 1)$ and $(1, 0)$. See Example 3.8 for a demonstration of the structure of the matrices in this proof.

Let $\boldsymbol{Q}_\sigma$ be the Q-matrix of the ancestral process. The entries of $\boldsymbol{Q}_\sigma$ are specified in equations (3.6) to (3.9). $\boldsymbol{Q}_\sigma$ can be written in the block form as

$$\boldsymbol{Q}_\sigma = \begin{pmatrix} \boldsymbol{Q}_{11} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} & \boldsymbol{Q}_{23} \\ \boldsymbol{Q}_{31} & \mathbf{0} & \boldsymbol{Q}_{33} \end{pmatrix},$$

where each $\boldsymbol{Q}_{ij}$ contains the transition rates from the states in S_i to S_j, and $\mathbf{0}$ represents the zero matrix of which all the entries are 0.

The matrix $\boldsymbol{Q}_\sigma$ can be decomposed such that $\boldsymbol{Q}_\sigma = \sigma \boldsymbol{A} + \boldsymbol{B}$, where

$$\boldsymbol{A} = \lim_{\sigma\to\infty} \boldsymbol{Q}_\sigma / \sigma,$$

and

$$\boldsymbol{B} = \lim_{\sigma\to\infty}(\boldsymbol{Q}_\sigma - \sigma\boldsymbol{A}).$$

The entries a_{ij} of $\boldsymbol{A}$, where $i = (\alpha_1, \alpha_2) \in S, j \in S$, are

$$a_{ij} = \begin{cases} \frac{1}{2}\alpha_1(\alpha_1 - 1), & \text{if } j = (\alpha_1 - 1, \alpha_2), \\ \frac{1}{2}\alpha_1\theta_1, & \text{if } j = (\alpha_1 - 1, \alpha_2 + 1), \\ -\frac{1}{2}\alpha_1(\alpha_1 - 1 + \theta_1), & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

In the block form, we have

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} \\ \boldsymbol{A}_{31} & \boldsymbol{0} & \boldsymbol{A}_{33} \end{pmatrix}.$$

The matrix $\boldsymbol{A}$ is a Q-matrix. There exists a Markov process with the corresponding Q-matrix $\boldsymbol{A}$, in which the states in S_1 are absorbing states and all other states are transient states.

The entries b_{ij} of $\boldsymbol{B}$, where $i = (\alpha_1, \alpha_2) \in S, j \in S$, are

$$b_{ij} = \begin{cases} \frac{1}{2}\alpha_2(\alpha_2 - 1), & \text{if } j = (\alpha_1, \alpha_2 - 1), \\ -\frac{1}{2}\alpha_1\theta_1, & \text{if } j = (\alpha_1 - 1, \alpha_2 + 1), \\ -\frac{1}{2}[\alpha_2(\alpha_2 - 1) - \alpha_1\theta_1], & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

In the block form, we have

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \boldsymbol{B}_{23} \\ \boldsymbol{B}_{31} & \boldsymbol{0} & \boldsymbol{B}_{33} \end{pmatrix}.$$

The transition matrix of the ancestral process $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ over a period of time t in the past, is $\boldsymbol{\Pi}_\sigma(t) = \exp(\boldsymbol{Q}_\sigma t)$. It follows algebraically that

$$\begin{aligned} \lim_{\sigma\to\infty}\boldsymbol{\Pi}_\sigma(t) &= \lim_{\sigma\to\infty} e^{(\boldsymbol{A}+\sigma^{-1}\boldsymbol{B})\sigma t} \\ &= \lim_{\sigma\to\infty}\left(e^{\boldsymbol{A}} + e^{\boldsymbol{A}}\sum_{k=1}^{\infty}\frac{\boldsymbol{B}^k}{\sigma^k k!}\right)^{\sigma t} \end{aligned}$$

$$= \lim_{\sigma\to\infty} \left(e^{\boldsymbol{A}} + \frac{1}{\sigma} e^{\boldsymbol{A}} \boldsymbol{B} + O(\sigma^{-2}) \right)^{\sigma t}.$$

According to Lemma 3.1, we have

$$\lim_{\sigma\to\infty} \boldsymbol{\Pi}_\sigma(t) = \boldsymbol{P} e^{\boldsymbol{G}t} := \boldsymbol{\Pi}(t), \tag{3.12}$$

where

$$\boldsymbol{P} = \lim_{m\to\infty} e^{\boldsymbol{A}m} \tag{3.13}$$

is the result of letting the process with Q-matrix $\boldsymbol{A}$ run to its conclusion; and

$$\boldsymbol{G} = \boldsymbol{P} e^{\boldsymbol{A}} \boldsymbol{B} \boldsymbol{P} = \boldsymbol{P} \left(\sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k}{k!} \right) \boldsymbol{B} \boldsymbol{P} = \boldsymbol{P} \boldsymbol{B} \boldsymbol{P}, \tag{3.14}$$

where the last equality follows from the fact that $\boldsymbol{P}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{P} = \boldsymbol{0}$.

It follows from equation (3.12) that $\boldsymbol{P} = \lim_{t\to 0} \boldsymbol{\Pi}(t)$. Hence, the limiting ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ with transition matrix $\boldsymbol{\Pi}(t)$ jumps instantaneously at time $t = 0$ from a state i to a state j, with probabilities specified by the entries of $\boldsymbol{P}$. By the definition of $\boldsymbol{P}$ in equation (3.13), the instantaneous jump is a result of the coalescence and mutation events occurring at the rates in $\boldsymbol{A}$ on a faster timescale where one unit of time corresponds to σ units of time on the original timescale.

It follows from equation (3.13) that $\boldsymbol{P}$ is a projection matrix such that $\boldsymbol{P}^2 = \boldsymbol{P}$. Thus, we have $\boldsymbol{\Pi}(t) = \boldsymbol{P} e^{\boldsymbol{G}t} = \boldsymbol{P} - \boldsymbol{I} + e^{\boldsymbol{G}t}$. The matrix $\boldsymbol{G}$ in equation (3.14) then satisfies

$$\boldsymbol{G} = \lim_{t\to 0} \frac{1}{t} [\boldsymbol{\Pi}(t) - \boldsymbol{\Pi}(0+)].$$

$\boldsymbol{G}$ is the Q-matrix of the limiting process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ and includes both the process described by $\boldsymbol{B}$ and the instantaneous jumps represented by $\boldsymbol{P}$.

We investigate the entries of $\boldsymbol{P}$. The matrix $\boldsymbol{P}$ can be written as

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{P}_{21} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{P}_{31} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix},$$

where $\boldsymbol{I} \coloneqq (\delta_{ij})$ is the identity matrix with $\delta_{ij} = \mathbb{1}_{\{i=j\}}$ being the Kronecker delta. The entries of $\boldsymbol{P}_{i1}$, $i = 2, 3$, are the probabilities of absorption from the states in S_i into the states in S_1. These entries are equal to the probabilities that a state (α_1, α_2) becomes absorbed into $(0, M+\alpha_2)$ after a series of coalescence events among ancestral lines of type 1 and M mutation events from type 1 to type 2 (on the faster timescale). The distribution of M is well known (see Tavaré, 2004, Section 3.6) to follow the Ewens sampling formula in equation (2.17). Hence, the entries of $\boldsymbol{P}$ are

$$p_{ij} = \begin{cases} 1, & \text{if } i \in S_1, j = i, \\ \frac{\theta_1^m}{\theta_{1(\alpha_1)}} \left| S_{\alpha_1}^{(m)} \right|, & \text{if } i = (\alpha_1, \alpha_2) \in S_2 \cup S_3, j = (0, m+\alpha_2) \in S_1, 0 < m \leq \alpha_1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

We can also algebraically verify that the entries p_{ij} given in the above equation (3.15) solve the equation $\boldsymbol{AP} = \boldsymbol{0}$, i.e.,

$$\sum_{l \in S} a_{il} p_{lj} = 0, \quad i, j \in S, \tag{3.16}$$

as follows. Equation (3.16) holds for any $i \in S_1$, since by equation (3.11), we have $a_{il} = 0$ for any $i \in S_1$ and $l \in S$. Equation (3.16) also holds for any $j \in S_2 \cup S_3$, since by equation (3.15), we have $p_{lj} = 0$ for any $l \in S$ and $j \in S_2 \cup S_3$. For $i = (\alpha_1, \alpha_2) \in S_2 \cup S_3$ and $j = (0, m+\alpha_2) \in S_1$ with $0 < m \leq \alpha_1$, we have

$$\begin{aligned} \sum_{l \in S} a_{il} p_{lj} &= \frac{1}{2}\alpha_1(\alpha_1 - 1) \times \frac{\theta_1^m}{\theta_{1(\alpha_1-1)}} \left| S_{\alpha_1-1}^{(m)} \right| + \frac{1}{2}\alpha_1\theta_1 \times \frac{\theta_1^{m-1}}{\theta_{1(\alpha_1-1)}} \left| S_{\alpha_1-1}^{(m-1)} \right| \\ &\quad - \frac{1}{2}\alpha_1(\alpha_1 + \theta_1 - 1) \times \frac{\theta_1^m}{\theta_{1(\alpha_1)}} \left| S_{\alpha_1}^{(m)} \right| \\ &= \frac{1}{2}\alpha_1 \frac{\theta_1^m}{\theta_{1(\alpha_1-1)}} \left[(\alpha_1 - 1) \left| S_{\alpha_1-1}^{(m)} \right| + \left| S_{\alpha_1-1}^{(m-1)} \right| - \left| S_{\alpha_1}^{(m)} \right| \right] \\ &= 0, \end{aligned}$$

where we use the recursive identity (2.14) for the Stirling numbers in the last equality. The proof of equation (3.15) is completed.

The limiting ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ moves instantaneously at time 0 according to $\boldsymbol{P}$, from a sample $(n_1, n_2) \in S$ to a state $(0, m + n_2) \in S_1$. After entering such a state, the limiting process behaves like a Markov chain with initial state $(0, m + \alpha_2)$ and Q-matrix $\boldsymbol{G}$ running on the original timescale. The following equation (3.17) indicates that the genealogy follows Kingman's coalescent.

The matrix $\boldsymbol{G} = \boldsymbol{PBP}$ has the block form

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{B}_{11} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{P}_{21}\boldsymbol{B}_{11} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{P}_{31}\boldsymbol{B}_{11} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Let g_{ij} be the entries of the matrix $\boldsymbol{G}$. We have $g_{ij} = \sum_{l,k \in S} p_{il} b_{lk} p_{kj}$. For a state $i = (0, \alpha_2) \in S_1$, since $p_{ij} = \mathbb{1}_{\{j=i\}}$ for $j \in S$, we have

$$g_{ij} = b_{ij} = \begin{cases} \frac{1}{2}\alpha_2(\alpha_2 - 1), & \text{if } j = (0, \alpha_2 - 1) \in S_1, \\ -\frac{1}{2}\alpha_2(\alpha_2 - 1), & \text{if } j = (0, \alpha_2) \in S_1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.17}$$

For a state $i = (\alpha_1, \alpha_2) \in S_2 \cup S_3$ and $j = (0, m + \alpha_2) \in S_1$ with $0 < m \leq \alpha_1$, we have

$$g_{ij} = \frac{\theta_1^{m+1} |S_{\alpha_1}^{(m+1)}|}{\theta_{1(\alpha_1)}} \times \frac{1}{2}(m + 1 + \alpha_2)(m + \alpha_2) - \frac{\theta_1^{m} |S_{\alpha_1}^{(m)}|}{\theta_{1(\alpha_1)}} \times \frac{1}{2}(m + \alpha_2)(m + \alpha_2 - 1),$$

since for a state $l = (0, \alpha_2) \in S_1$, we have $b_{lk} = \alpha_2(\alpha_2 - 1)/2$, if $k = (0, \alpha_2 - 1) \in S_1$; $b_{lk} = -\alpha_2(\alpha_2 - 1)/2$, if $k = l$; and $b_{lk} = 0$, otherwise. The other entries of the matrix $\boldsymbol{G}$ are all zero. The proof is then completed. □

In the following corollary, we specify the entries of the limiting transition matrix $\boldsymbol{\Pi}(t)$ in equation (3.12) in the proof of Theorem 3.5. We write

$$x_{[n]} = x(x - 1) \cdots (x - n + 1), \quad n \in \mathbb{Z}_+,$$

for the falling factorial, with $x_{[0]} = 1$.

Corollary 3.7. *Consider the ancestral process $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ with transition rates (3.6) to (3.8) for a sample $(n_1, n_2) \in \mathbb{Z}_+^2$ of size $n_1 + n_2 \geq 1$. For fixed $t > 0$, the*

transition probability of the ancestral process satisfies

$$\lim_{\sigma\to\infty} \mathbb{P}[\boldsymbol{\alpha}_\sigma(t) = (\alpha_1, \alpha_2) \mid \boldsymbol{\alpha}_\sigma(0) = (n_1, n_2)] = \begin{cases} \sum_{m=0}^{n_1} \frac{\theta_1^m}{\theta_{1(n_1)}} \left| S_{n_1}^{(m)} \right| d_{m+n_2,\alpha_2}(t), & \text{if } \alpha_1 = 0, \\ 0, & \text{otherwise,} \end{cases} \tag{3.18}$$

where $d_{ij}(t)$ is defined as

$$d_{ij}(t) = \begin{cases} \sum_{l=j}^{i} e^{-\binom{l}{2}t} \frac{(2l-1)(-1)^{l-j} j_{(l-1)} i_{[l]}}{j!(l-j)! i_{(l)}}, & \text{if } 2 \le j \le i, \\ 1 - \sum_{l=2}^{i} e^{-\binom{l}{2}t} \frac{(2l-1)(-1)^{l} i_{[l]}}{i_{(l)}}, & \text{if } j = 1, j \le i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.19}$$

Proof. We have seen in the proof of Theorem 3.5 that the transition matrix $\boldsymbol{\Pi}_\sigma(t)$ of the ancestral process $\{\boldsymbol{\alpha}_\sigma(t), t \ge 0\}$ on the state space S satisfies

$$\lim_{\sigma\to\infty} \boldsymbol{\Pi}_\sigma(t) = \boldsymbol{\Pi}(t) = \boldsymbol{P} e^{\boldsymbol{G}t}, \qquad t > 0,$$

where the matrix $\boldsymbol{P}$ represents the instantaneous jumps, and $\boldsymbol{G} = \boldsymbol{PBP}$ is the Q-matrix of the limiting process. Equation (3.17) shows that once the limiting ancestral process enters the state $(0, m + n_2)$, the process follows Kingman's coalescent. By matrix multiplication, the entries of $\boldsymbol{\Pi}(t)$ are

$$\pi_{ij}(t) = \sum_{k\in S} p_{ik} \left(e^{\boldsymbol{G}t}\right)_{kj}.$$

The entries of $\boldsymbol{P}$ are given in equation (3.15), and the transition probabilities $d_{ij}(t)$ in equation (3.19) are those of Kingman's coalescent derived by Griffiths (1980). Equation (3.18) then follows. □

In the following Example 3.8, we will illustrate the structure of the matrices in the proofs of Theorem 3.5 and Corollary 3.7. Before proceeding to the example, we introduce a standard technique for calculating the transition matrix of a continuous-time Markov process with a given Q-matrix. Let $\boldsymbol{A}$ be the Q-matrix of a Markov process, and $\boldsymbol{P}(t) = \exp(\boldsymbol{A}t), t \ge 0$ be the corresponding transition matrix. The

entries of $\boldsymbol{P}(t)$ can be calculated by considering the resolvent $R(\lambda)$, i.e., the Laplace transform of $\boldsymbol{P}(t)$. We have

$$R(\lambda) \coloneqq \int_0^\infty e^{-\lambda t}\boldsymbol{P}(t)\,\mathrm{d}t = \int_0^\infty e^{-(\lambda\boldsymbol{I}-\boldsymbol{A})t}\,\mathrm{d}t = (\lambda\boldsymbol{I}-\boldsymbol{A})^{-1},$$

where the integrals over matrix-valued functions are understood component-wise, and the last equality is well known (see Ethier and Kurtz, 2005, Proposition 2.1 in Chapter 1). The entries of $\boldsymbol{P}(t)$ can then be recovered by inverting the Laplace transform.

Example 3.8. Consider the ancestral process $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ with transition rates (3.6) to (3.8) for a sample (n_1, n_2) of size $n_1 + n_2 = 3$. The state space is

$$S = \{(0,3),(0,2),(0,1),(3,0),(2,0),(1,0),(1,2),(2,1),(1,1)\}.$$

According to the entries specified in equations (3.6) to (3.9), the Q-matrix $\boldsymbol{Q}_\sigma$ of $\{\boldsymbol{\alpha}_\sigma(t), t \geq 0\}$ is

$$\begin{pmatrix}
-3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -3\sigma-\frac{3}{2}\theta_1(\sigma-1) & 3\sigma & 0 & 0 & \frac{3}{2}\theta_1(\sigma-1) & 0 \\
0 & -1 & 1 & 0 & -\sigma-\theta_1(\sigma-1) & \sigma & 0 & 0 & \theta_1(\sigma-1) \\
0 & 0 & \frac{1}{2}\theta_1(\sigma-1) & 0 & 0 & -\frac{1}{2}\theta_1(\sigma-1) & 0 & 0 & 0 \\
\frac{1}{2}\theta_1(\sigma-1) & 0 & 0 & 0 & 0 & 0 & -1-\frac{1}{2}\theta_1(\sigma-1) & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & \theta_1(\sigma-1) & -\sigma-\theta_1(\sigma-1) & \sigma \\
0 & \frac{1}{2}\theta_1(\sigma-1) & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2}\theta_1(\sigma-1)
\end{pmatrix},$$

in which, for example, the first row shows that when the ancestral process is at the state $(0,3)$, the process jumps after an exponential waiting time of mean $1/3$ to a state $(0,2)$ with probability 1.

The Q-matrix can be separated into $\sigma\boldsymbol{A}$ and $\boldsymbol{B}$ with respect to the order of the

terms in each entry. The terms of $O(\sigma)$ are collected in

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3-\frac{3}{2}\theta_1 & 3 & 0 & 0 & \frac{3}{2}\theta_1 & 0 \\ 0 & 0 & 0 & 0 & -1-\theta_1 & 1 & 0 & 0 & \theta_1 \\ 0 & 0 & \frac{1}{2}\theta_1 & 0 & 0 & -\frac{1}{2}\theta_1 & 0 & 0 & 0 \\ \frac{1}{2}\theta_1 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2}\theta_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_1 & -1-\theta_1 & 1 \\ 0 & \frac{1}{2}\theta_1 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2}\theta_1 \end{pmatrix};$$

and the terms of $O(1)$ are collected in

$$\boldsymbol{B} = \begin{pmatrix} -3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{3}{2}\theta_1 & 0 & 0 & 0 & -\frac{3}{2}\theta_1 & 0 \\ 0 & -1 & 1 & 0 & \theta_1 & 0 & 0 & 0 & -\theta_1 \\ 0 & 0 & -\frac{1}{2}\theta_1 & 0 & 0 & \frac{1}{2}\theta_1 & 0 & 0 & 0 \\ -\frac{1}{2}\theta_1 & 0 & 0 & 0 & 0 & 0 & -1+\frac{1}{2}\theta_1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\theta_1 & \theta_1 & 0 \\ 0 & -\frac{1}{2}\theta_1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}\theta_1 \end{pmatrix}.$$

Then, as given in the proof of Theorem 3.5, the limiting ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ undergoes an instantaneous jump at time $t = 0$ according to

$$\boldsymbol{P} = \lim_{m\to\infty} e^{\boldsymbol{A}m} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\theta_1^2}{(\theta_1+1)(\theta_1+2)} & \frac{3\theta_1}{(\theta_1+1)(\theta_1+2)} & \frac{2}{(\theta_1+1)(\theta_1+2)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\theta_1}{\theta_1+1} & \frac{1}{\theta_1+1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\theta_1}{\theta_1+1} & \frac{1}{\theta_1+1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The entries of $\boldsymbol{P}(m) = \exp(\boldsymbol{A}m)$ are calculated by considering the resolvent introduced before the example, and $\boldsymbol{P}$ is obtained by letting $m \to \infty$ in $\boldsymbol{P}(m)$.

The Q-matrix of the limiting ancestral process is

$$\boldsymbol{G} = \boldsymbol{PBP} = \begin{pmatrix} -3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{3\theta_1^2}{(\theta_1+1)(\theta_1+2)} & \frac{3\theta_1^2-3\theta_1}{(\theta_1+1)(\theta_1+2)} & \frac{3\theta_1}{(\theta_1+1)(\theta_1+2)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\theta_1}{\theta_1+1} & \frac{\theta_1}{\theta_1+1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{3\theta_1}{\theta_1+1} & \frac{3\theta_1-1}{\theta_1+1} & \frac{1}{\theta_1+1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The transition matrix of the limiting ancestral process is

$$\boldsymbol{P}e^{\boldsymbol{G}t} = \begin{pmatrix} e^{-3t} & \frac{3}{2}e^{-t} - \frac{3}{2}e^{-3t} & 1 + \frac{1}{2}e^{-3t} - \frac{3}{2}e^{-t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-t} & 1 - e^{-t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\theta_1^2 e^{-3t}}{(\theta_1+1)(\theta_1+2)} & \frac{3(\theta_1+2)\theta_1 e^{-t} - 3\theta_1^2 e^{-3t}}{2(\theta_1+1)(\theta_1+2)} & 1 + \frac{\theta_1^2 e^{-3t} - 3\theta_1(\theta_1+2)e^{-t}}{2(\theta_1+1)(\theta_1+2)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\theta_1 e^{-t}}{\theta_1+1} & 1 - \frac{\theta_1 e^{-t}}{\theta_1+1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{-3t} & \frac{3}{2}e^{-t} - \frac{3}{2}e^{-3t} & 1 + \frac{1}{2}e^{-3t} - \frac{3}{2}e^{-t} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\theta_1 e^{-3t}}{\theta_1+1} & \frac{-3\theta_1 e^{-3t} + (3\theta_1+2)e^{-t}}{2(\theta_1+1)} & 1 + \frac{\theta_1 e^{-3t} - (3\theta_1+2)e^{-t}}{2(\theta_1+1)} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-t} & 1 - e^{-t} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where the entries are exactly those derived in equation (3.18) of Corollary 3.7. We have also verified Theorem 3.5.

3.2 Ancestral dual process with typed lines

In this section, we study the limiting behaviour of the ancestral dual process with typed lines under strong selection. In such a dual process, the ancestral lines can branch and virtual lines can appear. As a consequence, the state space of the dual process with typed lines is not fixed, so Möhle's lemma does not apply. It is not easy to prove a generalisation of Möhle's lemma for infinite-dimensional matrices, so in the remainder of this section, our arguments are heuristic and we show some interesting aspects of the calculations, provided that Möhle's lemma holds for infinite-dimensional matrices. In

fact, we will see in Remark 3.10 that the limiting ancestral process in this case is still a Markov process with a branching-coalescing structure, but in which there are only individuals of type 1, thus the real genealogy is neutral. Given such a limiting process, the time to the MRCA, the length of the genealogy, and the type distribution of the MRCA in the limit as $\sigma \to \infty$, follow immediately. The expectations and variances of these quantities are consistent with those derived in Theorems 2.5 and 2.16 in the previous chapter by considering recursive equations, as well as the limiting process in the coalescent-in-a-random-background model in Theorem 3.5.

We do not distinguish the real and virtual lines for now, and consider the non-zero entries (1.15) to (1.18) in the $\boldsymbol{\alpha}$th row of the Q-matrix of the dual process with typed lines of Theorem 1.9. The following lemma can be shown with a similar argument to that in the proof of Lemma 2.8.

Lemma 3.9. *Consider the non-zero entries* (1.15) *to* (1.18) *in the* $\boldsymbol{\alpha}$*th row of the* Q*-matrix of the dual process of Theorem 1.9. For large* σ*, these entries can be rewritten as*

$$\begin{aligned}
q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} + \boldsymbol{e}_1) &= \frac{1}{2}\alpha(\alpha_1 + \theta_1) + O(\sigma^{-1}),\\
q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_1) &= \frac{1}{2}\alpha_1(\alpha_1 - 1)\frac{\sigma + \alpha_2 - 1 + \theta_2}{\alpha_1 - 1 + \theta_1} + O(\sigma^{-1}),\\
q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_2) &= \frac{1}{2}\alpha_2(\alpha_2 - 1) + O(\sigma^{-1}),\\
q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} + \boldsymbol{e}_1 - \boldsymbol{e}_2) &= O(\sigma^{-1}),\\
q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha} + \boldsymbol{e}_2 - \boldsymbol{e}_1) &= \frac{1}{2}\alpha_1\theta_1\frac{\sigma + \alpha_2 + \theta_2 - \alpha_1 - \theta_1}{\alpha_1 - 1 + \theta_1} + O(\sigma^{-1}),\\
q_\sigma(\boldsymbol{\alpha}, \boldsymbol{\alpha}) &= -\frac{1}{2}[\alpha_1\sigma + \alpha(\alpha - 1) + \alpha_1\theta_2 + \alpha_2\theta_1]. \qquad (3.20)
\end{aligned}$$

Let $\boldsymbol{Q}_\sigma$ be the Q-matrix of the ancestral dual process with transition rates (3.20). We then follow the same approach as that in the proof of Theorem 3.5 to separate the terms in the entries of $\boldsymbol{Q}_\sigma$ by their orders, and write $\boldsymbol{Q}_\sigma = \sigma\boldsymbol{A} + \boldsymbol{B} + O(\sigma^{-1})$.

According to equations (3.20), the entries a_{ij} of $\boldsymbol{A}$, where $i = (\alpha_1, \alpha_2)$, are

$$a_{ij} = \begin{cases} \frac{1}{2}\frac{\alpha_1(\alpha_1-1)}{\alpha_1-1+\theta_1}, & \text{if } j = (\alpha_1 - 1, \alpha_2), \\ \frac{1}{2}\frac{\alpha_1\theta_1}{\alpha_1-1+\theta_1}, & \text{if } j = (\alpha_1 - 1, \alpha_2 + 1), \\ -\frac{1}{2}\alpha_1, & \text{if } j = i, \\ 0, & \text{otherwise;} \end{cases}$$

the entries b_{ij} of $\boldsymbol{B}$, where $i = (\alpha_1, \alpha_2)$, are

$$b_{ij} = \begin{cases} \frac{1}{2}\alpha(\alpha_1 + \theta_1), & \text{if } j = (\alpha_1 + 1, \alpha_2), \\ \frac{1}{2}\alpha_1(\alpha_1 - 1)\frac{\alpha_2-1+\theta_2}{\alpha_1-1+\theta_1}, & \text{if } j = (\alpha_1 - 1, \alpha_2), \\ \frac{1}{2}\alpha_2(\alpha_2 - 1), & \text{if } j = (\alpha_1, \alpha_2 - 1), \\ \frac{1}{2}\alpha_1\theta_1\frac{\alpha_2+\theta_2-\alpha_1-\theta_1}{\alpha_1-1+\theta_1}, & \text{if } j = (\alpha_1 - 1, \alpha_2 + 1), \\ -\frac{1}{2}\big[\alpha(\alpha_2 - 1) + \alpha_1(\theta_2 - \theta_1) + \alpha(\alpha_1 + \theta_1)\big], & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.21}$$

Since $\boldsymbol{A}$ is a Q-matrix, there exists a Markov process with the corresponding Q-matrix $\boldsymbol{A}$. In this Markov process, the states in S_1 are absorbing states and all other states are transient states.

Let $\boldsymbol{\Pi}_\sigma(t) = \exp(\boldsymbol{Q}_\sigma t)$ be the transition matrix of the dual process with transition rates (3.20) over a period of time t in the past. We have

$$\boldsymbol{\Pi}_\sigma(t) = \exp\left[\left(\boldsymbol{A} + \frac{\boldsymbol{B}}{\sigma} + O(\sigma^{-2})\right)\sigma t\right] = \left[e^{\boldsymbol{A}} + \frac{1}{\sigma}e^{\boldsymbol{A}}\boldsymbol{B} + O(\sigma^{-2})\right]^{\sigma t}.$$

If Möhle's lemma holds for infinite-dimensional matrices, we would have

$$\lim_{\sigma\to\infty} \boldsymbol{\Pi}_\sigma(t) = \boldsymbol{P}e^{\boldsymbol{G}t} := \boldsymbol{\Pi}(t),$$

where

$$\boldsymbol{P} = \lim_{m\to\infty} e^{\boldsymbol{A}m} \tag{3.22}$$

represents the equilibrium of the Markov process with Q-matrix $\boldsymbol{A}$; and $\boldsymbol{G} = \boldsymbol{P}e^{\boldsymbol{A}}\boldsymbol{B}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{B}\boldsymbol{P}$, where the second equality follows from the fact that $\boldsymbol{P}\exp(\boldsymbol{A}) = \boldsymbol{P}$ because $\boldsymbol{P}\boldsymbol{A} = \boldsymbol{0}$.

Since $\boldsymbol{P} = \lim_{t\to 0}\boldsymbol{\Pi}(t)$, the limiting ancestral process with transition matrix $\boldsymbol{\Pi}(t)$

jumps instantaneously at time $t = 0$ from the sample (n_1, n_2) to a state $(0, m+n_2)$, with probabilities specified by the entries of $\boldsymbol{P}$. By the definition of $\boldsymbol{P}$ in equation (3.22), the instantaneous jump is a result of the coalescence events and M mutation events that occurred at the rates in $\boldsymbol{A}$ on a faster timescale where one unit of time corresponds to σ units of time on the original timescale. The distribution of M is well known (see Tavaré, 2004, Section 3.6) to be the Ewens sampling formula given in equation (2.17).

The entries of the matrix $\boldsymbol{P}$ are

$$p_{ij} = \begin{cases} 1, & \text{if } i = (0, \alpha_2), j = i, \\ \frac{\theta_1^m}{\theta_{1(\alpha_1)}} \left| S_{\alpha_1}^{(m)} \right|, & \text{if } i = (\alpha_1, \alpha_2), j = (0, m + \alpha_2) \text{ with } 0 < m \leq \alpha_1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.23}$$

We can verify that the entries p_{ij} given in the above equation (3.23) satisfy $\sum_l a_{il} p_{lj} = 0$ for all i, j, as follows. The equation holds for any $i = (0, \alpha_2)$, since then $a_{il} = 0$ for any l. The equation also holds for any $j = (\alpha_1, \alpha_2)$, since then $p_{lj} = 0$ for any l. For $i = (\alpha_1, \alpha_2)$ and $j = (0, m + \alpha_2)$ fixed with $0 < m \leq \alpha_1$, we have

$$\begin{aligned} \sum_l a_{il} p_{lj} &= \frac{1}{2} \frac{\alpha_1(\alpha_1 - 1)}{\alpha_1 - 1 + \theta_1} \times \frac{\theta_1^m}{\theta_{1(\alpha_1 - 1)}} \left| S_{\alpha_1 - 1}^{(m)} \right| + \frac{1}{2} \frac{\alpha_1 \theta_1}{\alpha_1 - 1 + \theta_1} \times \frac{\theta_1^{m-1}}{\theta_{1(\alpha_1 - 1)}} \left| S_{\alpha_1 - 1}^{(m-1)} \right| \\ &\quad - \frac{1}{2} \alpha_1 \times \frac{\theta_1^m}{\theta_{1(\alpha_1)}} \left| S_{\alpha_1}^{(m)} \right| \\ &= \frac{1}{2} \alpha_1 \frac{\theta_1^m}{\theta_{1(\alpha_1)}} \left[(\alpha_1 - 1) \left| S_{\alpha_1 - 1}^{(m)} \right| + \left| S_{\alpha_1 - 1}^{(m-1)} \right| - \left| S_{\alpha_1}^{(m)} \right| \right] \\ &= 0, \end{aligned}$$

where we use the recursive identity (2.14) for the Stirling numbers to obtain the last equality. The proof of equation (3.23) is completed.

The limiting ancestral process with transition matrix $\boldsymbol{\Pi}(t)$ moves instantaneously at time 0 according to $\boldsymbol{P}$, from the sample (n_1, n_2) to a state $(0, m + n_2)$. At time $t > 0$, the limiting process behaves like a Markov chain with initial state $(0, m + n_2)$ and Q-matrix $\boldsymbol{G}$. Let g_{ij} be the entries of the matrix $\boldsymbol{G}$. We have $g_{ij} = \sum_{l,k} p_{il} b_{lk} p_{kj}$. First consider the state $i = (0, \alpha_2)$. We have $p_{il} = \mathbb{1}_{\{l=i\}}$; and given $l = (0, \alpha_2)$, $b_{lk} = 0$

unless $k = (0, \alpha_2 - 1)$, $(1, \alpha_2)$ or l. It then follows that

$$g_{ij} = \begin{cases} \frac{1}{2}\alpha_2\theta_1 \frac{\theta_1}{\theta_{1(1)}} \left| S_1^{(1)} \right| = \frac{1}{2}\alpha_2\theta_1, & \text{if } j = (0, \alpha_2 + 1), \\ \frac{1}{2}\alpha_2(\alpha_2 - 1), & \text{if } j = (0, \alpha_2 - 1), \\ -\frac{1}{2}\alpha_2(\alpha_2 - 1 + \theta_1), & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.24}$$

Then consider the states $i = (\alpha_1, \alpha_2)$ and $j = (0, m + \alpha_2)$ with $0 < m \leq \alpha_1$ fixed. We have

$$\begin{aligned} g_{ij} = {} & \frac{\theta_1^{m+1}}{\theta_{1(\alpha_1)}} \left| S_{\alpha_1}^{(m+1)} \right| \times \frac{1}{2}(m + 1 + \alpha_2)(m + \alpha_2) + \frac{\theta_1^{m-1}}{\theta_{1(\alpha_1)}} \left| S_{\alpha_1}^{(m-1)} \right| \times \frac{1}{2}(m - 1 + \alpha_2)\theta_1 \\ & - \frac{\theta_1^m}{\theta_{1(\alpha_1)}} \left| S_{\alpha_1}^{(m)} \right| \times \frac{1}{2}[(m + \alpha_2)(m + \alpha_2 - 1) + (m + \alpha_2)\theta_1]. \end{aligned}$$

The other entries of the matrix $\boldsymbol{G}$ are all zero.

Therefore, according to equation (3.24), once a sample (n_1, n_2) is converted (on the faster timescale) to $(0, m+n_2)$ at time 0+, the transitions and rates of the limiting ancestral dual process on the original timescale are

$$(0, \alpha_2) \to \begin{cases} (0, \alpha_2 + 1) & \text{at rate } \frac{1}{2}\alpha_2\theta_1, \\ (0, \alpha_2 - 1) & \text{at rate } \frac{1}{2}\alpha_2(\alpha_2 - 1). \end{cases}$$

Remark 3.10. If Möhle's lemma holds for infinite-dimensional matrices, we would arrive at a limiting process for the ancestral dual process with typed lines under strong selection with the above arguments.

There is an instantaneous jump at $t = 0$ in the limiting process such that

$$\boldsymbol{\alpha}(0+) = (0, M + n_2),$$

where M is distributed as the number of distinct alleles in the Ewens sampling formula with mutation coefficient θ_1, as in equation (2.17). For $t > 0$, the transitions and rates of the limiting process are

$$(0, \alpha_2) \to \begin{cases} (0, \alpha_2 + 1), & \text{at rate } \frac{1}{2}\alpha_2\theta_1, \\ (0, \alpha_2 - 1), & \text{at rate } \frac{1}{2}\alpha_2(\alpha_2 - 1), \end{cases} \tag{3.25}$$

for $\alpha_2 \geq 2$.

Remark 3.11. The limiting ancestral process with transitions (3.25) can be interpreted as an ancestral process that can branch with coefficient θ_1, and coalesce at the usual coalescence rate , but which cannot mutate. By Lemma 2.12, the real genealogy in such an ancestral process is determined by Kingman's coalescent.

Given the limiting dual process in Remark 3.10, the time to the MRCA, the length of the genealogy, and the type distribution of the MRCA in the limit as $\sigma \to \infty$, follow immediately. It is clear that the expectation and variance of these quantities are the same as those deduced in Theorems 2.5 and 2.16 by considering recursive equations, as well as those from the limiting process in the coalescent-in-a-random-background model in Theorem 3.5.

In addition, it is clear from the limiting dual process that the mean and variance of the coalescence time for a sample (n_1, n_2) satisfy

$$\begin{aligned}\lim_{\sigma\to\infty} \mathbb{E}[W_{(n_1,n_2)}(i)] &= \mathbb{E}(\text{coalescence time in Kingman's coalescent}) \\ &\quad \times \mathbb{P}(M + n_2 \text{ is greater than or equal to } i \text{ in the limiting dual process}) \\ &= \frac{2}{i(i-1)} \sum_{m=\max\{1,i-n_2\}}^{n_1} \mathbb{P}(M = m \mid n_1),\end{aligned}$$

for $2 \leq i \leq n$.

We complete this section by proving Theorem 2.3 for $\mathbb{E}[V(n_1, n_2)]$ in the limit as $\sigma \to \infty$, i.e., the mean number of branching events to the ultimate ancestor in the dual process of Theorem 2.1 under strong selection. In Theorem 2.3, we claim that $\lim_{\sigma\to\infty} \mathbb{E}[V(n_1, n_2)]$ is finite (cf. the number of branching events grows exponentially with σ in the ancestral selection graph).

Proof of Theorem 2.3. Given the limiting dual process in Remark 3.10, it is clear that

$$\lim_{\sigma\to\infty} \mathbb{E}[V(n_1, n_2)] = \sum_{m=0}^{n_1} \mathbb{P}(M = m \mid n_1)\, \mathbb{E}[V(n_2 + m)], \tag{3.26}$$

where the distribution of M is given in equation (2.17), and $V(n_2+m)$ is the number of branching events in an ancestral process that starts with a sample containing n_2+m individuals of the same type, and has the following transitions and rates:

$$\alpha \to \begin{cases} \alpha+1, & \text{at rate } \frac{1}{2}\alpha\theta_1, \\ \alpha-1, & \text{at rate } \binom{\alpha}{2}. \end{cases}$$

These transitions and rates are similar to those of the ancestral selection graph in (1.2), expect that the branching rate is determined by θ_1 instead of σ. Thus, by Proposition 1.1, we have

$$\mathbb{E}[V(n_2+m)] = e^{\theta_1} \sum_{l=0}^{n_2+m-2} \frac{l!}{\theta_1^l} - \sum_{l=0}^{n_2+m-2} \frac{1}{\theta_1^l} \sum_{k=l}^{n_2+m-2} \frac{k!}{(k-l)!}. \tag{3.27}$$

The theorem is then proved by substituting equation (3.27) into equation (3.26). □

3.3 Ancestral dual process with three types

It is natural to extend the studies of the strong selection limit in the previous sections to the population in which there are individuals of three types. The aim of this section is to investigate the limiting behaviour of the ancestral dual process with typed lines in the three-type case, in which individuals of type 1 and type 2 have a strong selective disadvantage and individuals of type 3 are neutral. Although the methods in Sections 3.1 and 3.2 still apply, our description of the genealogy in this section is heuristic, because there are a few different timescales involved, which make formal analysis difficult.

From an algebraic derivation similar to that in the two-type case in Theorem 1.9, we can obtain an ancestral process with typed lines that is dual to the forwards-in-time diffusion in the three-types case (see Etheridge and Griffiths, 2009). Then, by an argument similar to that in Theorem 2.1, we can decompose the ancestral lines in the dual process into real and virtual ones. The configuration of decomposed ancestral

lines at time t in the past can be represented by $(\boldsymbol{r}(t), \boldsymbol{v}(t))$, where the elements of the three-dimensional vectors $\boldsymbol{r}(t)$ and $\boldsymbol{v}(t)$ are, respectively, the numbers of real and virtual ancestral lines of corresponding types. We have the following proposition for the ancestral dual process with typed lines in the three-type case with selection coefficients σ_1 and σ_2 and mutation coefficients θ_1, θ_2 and θ_3.

Proposition 3.12 (Ancestral dual process with three types in the diffusion limit)**.** *For a sample $\boldsymbol{n} \in \mathbb{Z}_+^3$ of size $|\boldsymbol{n}| \geq 1$, the dual process $\{(\boldsymbol{r}(t), \boldsymbol{v}(t)), t \geq 0\}$ is a Markov process with state space $\mathbb{Z}_+^6 \setminus \{\boldsymbol{0}\}$ and initial state $(\boldsymbol{n}, \boldsymbol{0})$. The non-zero entries in the $(\boldsymbol{r}, \boldsymbol{v})$th row of the Q-matrix of the dual process are*

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_i)) &= \frac{1}{2}(r+v)\sigma_i \frac{(r_i+v_i+\theta_i)}{r+v+\theta}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \quad i \neq 3,\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_i,\boldsymbol{v})) &= \frac{1}{2}r_i(r_i-1)\frac{r+v-1+\theta}{r_i+v_i-1+\theta_i}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_i)) &= \frac{1}{2}v_i(2r_i+v_i-1)\frac{r+v-1+\theta}{r_i+v_i-1+\theta_i}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_i-\boldsymbol{e}_j,\boldsymbol{v})) &= \frac{1}{2}r_j\theta_j\frac{r_i+v_i+\theta_i}{r_j+v_j-1+\theta_j}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \quad i \neq j,\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j)) &= \frac{1}{2}v_j\theta_j\frac{r_i+v_i+\theta_i}{r_j+v_j-1+\theta_j}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}, \quad i \neq j,\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v})) &= -\frac{1}{2}\Bigg[\sum_{i=1}^{2}(r_i+v_i)\sigma_i + (r+v)(r+v-1) + \sum_{i=1}^{3}\sum_{\substack{j=1\\ j\neq i}}^{3}(r_i+v_i)\theta_j\Bigg],
\end{aligned}
\tag{3.28}
$$

where $i, j = 1, 2, 3$, $r = |\boldsymbol{r}|$, $v = |\boldsymbol{v}|$, $\theta = |\boldsymbol{\theta}|$, and

$$
\nu(\boldsymbol{\theta}) = \mathbb{E}[\exp(-\sigma_1\xi_1 - \sigma_2\xi_2)], \tag{3.29}
$$

with (ξ_1, ξ_2) having the Dirichlet($\boldsymbol{\theta}$) distribution with density

$$
\mathcal{D}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)}x_1^{\theta_1-1}x_2^{\theta_2-1}(1-x_1-x_2)^{\theta_3-1},
$$

for $x_1 > 0$, $x_2 > 0$ and $x_1 + x_2 < 1$.

For strong selection, we assume $\sigma_1 \to \infty$, $\sigma_2 \to \infty$. For the sake of arguments, we further distinguish two cases, namely where the disfavoured types have a strong selective disadvantage of the same order such that

$$\lim_{\sigma_2 \to \infty} \sigma_1/\sigma_2 = \gamma \tag{3.30}$$

with $0 < \gamma < \infty$ fixed; and where the disfavoured types have a strong selective disadvantage of different orders such that

$$\lim_{\sigma_2 \to \infty} \sigma_1/\sigma_2 = \infty \quad \text{and} \quad \lim_{\sigma_2 \to \infty} \sigma_1/\sigma_2^2 = 0,$$

i.e.,

$$\sigma_1 \gg \sigma_2 \gg \frac{\sigma_1}{\sigma_2} \gg 1. \tag{3.31}$$

We will see in Remark 3.15, the limiting genealogies in these two cases are the same.

The first task in this section is to asymptotically expand the transition rates of the ancestral dual process of Proposition 3.12 under our assumption of strong selection. In the two-type case, the function $\nu(\boldsymbol{\theta})$ in equation (3.29) is equal to the confluent hypergeometric function of the first kind. However, in the three-type case, the following lemma is needed.

Lemma 3.13. *Consider the function $\nu(\boldsymbol{\theta})$ defined in equation* (3.29)*. For large σ_1 and σ_2, we have*

$$\nu(\boldsymbol{\theta}) = \frac{\Gamma(\theta)}{\Gamma(\theta_3)\sigma_1^{\theta_1}\sigma_2^{\theta_2}} \Bigg[1 + \frac{(1-\theta_3)\theta_1}{\sigma_1} + \frac{(1-\theta_3)\theta_2}{\sigma_2} + \frac{(1-\theta_3)(2-\theta_3)}{2} \left(\frac{\theta_1(1+\theta_1)}{\sigma_1^2} + \frac{2\theta_1\theta_2}{\sigma_1\sigma_2} + \frac{\theta_2(1+\theta_2)}{\sigma_2^2} \right) + o(\sigma_1^{-2}) \Bigg].$$

Proof. See Appendix A.3. □

We then have the following lemma for the transition rates in (3.28) under our assumption (3.30) of strong selection.

Lemma 3.14. *Suppose the disfavoured types have a strong selective disadvantage of the same order such that $\sigma_1 \to \infty$, $\sigma_2 \to \infty$ and*

$$\lim_{\sigma_2 \to \infty} \sigma_1/\sigma_2 = \gamma,$$

with $0 < \gamma < \infty$ fixed. Ignoring the terms of $o(1)$, the non-zero entries (3.28) in the $(\boldsymbol{r}, \boldsymbol{v})$th row of the Q-matrix of the ancestral process become

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1)) = \frac{1}{2}(r + v)(r_1 + v_1 + \theta_1), \tag{3.32}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_2)) = \frac{1}{2}(r + v)(r_2 + v_2 + \theta_2), \tag{3.33}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_1, \boldsymbol{v})) = \frac{1}{2}r_1(r_1 - 1)\frac{\gamma\sigma_2 + r_3 + v_3 - 1 + \theta_3}{r_1 + v_1 - 1 + \theta_1}, \tag{3.34}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_2, \boldsymbol{v})) = \frac{1}{2}r_2(r_2 - 1)\frac{\sigma_2 + r_3 + v_3 - 1 + \theta_3}{r_2 + v_2 - 1 + \theta_2}, \tag{3.35}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_3, \boldsymbol{v})) = \frac{1}{2}r_3(r_3 - 1), \tag{3.36}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_1)) = \frac{1}{2}v_1(2r_1 + v_1 - 1)\frac{\gamma\sigma_2 + r_3 + v_3 - 1 + \theta_3}{r_1 + v_1 - 1 + \theta_1}, \tag{3.37}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_2)) = \frac{1}{2}v_2(2r_2 + v_2 - 1)\frac{\sigma_2 + r_3 + v_3 - 1 + \theta_3}{r_2 + v_2 - 1 + \theta_2}, \tag{3.38}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_3)) = \frac{1}{2}v_3(2r_3 + v_3 - 1), \tag{3.39}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_1 - \boldsymbol{e}_2, \boldsymbol{v})) = \frac{1}{2}r_2\theta_2\frac{r_1 + v_1 + \theta_1}{\gamma(r_2 + v_2 - 1 + \theta_2)}, \tag{3.40}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_2 - \boldsymbol{e}_1, \boldsymbol{v})) = \frac{1}{2}r_1\theta_1\frac{\gamma(r_2 + v_2 + \theta_2)}{(r_1 + v_1 - 1 + \theta_1)}, \tag{3.41}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_3 - \boldsymbol{e}_1, \boldsymbol{v})) = \frac{1}{2}r_1\theta_1\frac{\gamma\sigma_2 + r_3 + v_3 + \theta_3 - r_1 - v_1 - \theta_1 - \gamma(r_2 + v_2 + \theta_2)}{r_1 + v_1 - 1 + \theta_1}, \tag{3.42}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_3 - \boldsymbol{e}_2, \boldsymbol{v})) = \frac{1}{2}r_2\theta_2\frac{\sigma_2 + r_3 + v_3 + \theta_3 - r_2 - v_2 - \theta_2 - (r_1 + v_1 + \theta_1)/\gamma}{r_2 + v_2 - 1 + \theta_2}, \tag{3.43}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1 - \boldsymbol{e}_2)) = \frac{1}{2}v_2\theta_2\frac{r_1 + v_1 + \theta_1}{\gamma(r_2 + v_2 - 1 + \theta_2)}, \tag{3.44}$$

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_2 - \boldsymbol{e}_1)) = \frac{1}{2}v_1\theta_1\frac{\gamma(r_2 + v_2 + \theta_2)}{(r_1 + v_1 - 1 + \theta_1)}, \tag{3.45}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_1)) = \frac{1}{2}v_1\theta_1\frac{\gamma\sigma_2 + r_3 + v_3 + \theta_3 - r_1 - v_1 - \theta_1 - \gamma(r_2+v_2+\theta_2)}{r_1+v_1-1+\theta_1}, \tag{3.46}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_2)) = \frac{1}{2}v_2\theta_2\frac{\sigma_2 + r_3 + v_3 + \theta_3 - r_2 - v_2 - \theta_2 - (r_1+v_1+\theta_1)/\gamma}{r_2+v_2-1+\theta_2}, \tag{3.47}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v})) = -\frac{1}{2}\Bigg[\sum_{i=1}^{2}(r_i+v_i)\sigma_i + (r+v)(r+v-1) + \sum_{i=1}^{3}\sum_{\substack{j=1\\j\neq i}}^{3}(r_i+v_i)\theta_j\Bigg]. \tag{3.48}$$

Proof. See Appendix A.4. □

It is not difficult to check that the sum of the above non-zero entries in the $(\boldsymbol{r},\boldsymbol{v})$th row is zero. The rates (3.32) and (3.33) correspond to branching events; when an ancestral line branches, a virtual line of either type 1 or type 2 can be added to the genealogy. The rates (3.34) to (3.39) correspond to coalescence events; ancestral lines of type 3 with a selective advantage coalesce at the rate of lower order than that for lines of type 1 and 2. The rates (3.40) to (3.47) correspond to mutations; ancestral lines of type 3 cannot mutate to type 1 or 2, and virtual lines of type 3 may also appear in the dual process as a consequence of mutations of virtual lines of type 1 and 2.

According to the idea in the previous sections, we can separate the terms in the transition rates (3.32) to (3.47) with respect to their orders. The transitions with rates of $O(\sigma_2)$ are

$$\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{1}{2}r_1(r_1-1)\frac{\gamma\sigma_2}{r_1+v_1-1+\theta_1},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1}{2}r_2(r_2-1)\frac{\sigma_2}{r_2+v_2-1+\theta_2},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1)) &= \frac{1}{2}v_1(2r_1+v_1-1)\frac{\gamma\sigma_2}{r_1+v_1-1+\theta_1},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_2)) &= \frac{1}{2}v_2(2r_2+v_2-1)\frac{\sigma_2}{r_2+v_2-1+\theta_2},
\end{aligned}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_1,\boldsymbol{v})) = \frac{1}{2}r_1\theta_1\frac{\gamma\sigma_2}{r_1+v_1-1+\theta_1},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_2,\boldsymbol{v})) = \frac{1}{2}r_2\theta_2\frac{\sigma_2}{r_2+v_2-1+\theta_2},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_1)) = \frac{1}{2}v_1\theta_1\frac{\gamma\sigma_2}{r_1+v_1-1+\theta_1},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_1)) = \frac{1}{2}v_2\theta_2\frac{\sigma_2}{r_2+v_2-1+\theta_2};$$

and the transitions with rates of $O(1)$ are

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) = \frac{1}{2}(r+v)(r_1+v_1+\theta_1),$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_2)) = \frac{1}{2}(r+v)(r_2+v_2+\theta_2),$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v})) = \frac{1}{2}r_1(r_1-1)\frac{r_3+v_3-1+\theta_3}{r_1+v_1-1+\theta_1},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v})) = \frac{1}{2}r_2(r_2-1)\frac{r_3+v_3-1+\theta_3}{r_2+v_2-1+\theta_2},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_3,\boldsymbol{v})) = \frac{1}{2}r_3(r_3-1),$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1)) = \frac{1}{2}v_1(2r_1+v_1-1)\frac{r_3+v_3-1+\theta_3}{r_1+v_1-1+\theta_1},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_2)) = \frac{1}{2}v_2(2r_2+v_2-1)\frac{r_3+v_3-1+\theta_3}{r_2+v_2-1+\theta_2},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_3)) = \frac{1}{2}v_3(2r_3+v_3-1),$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{v})) = \frac{1}{2}r_2\theta_2\frac{r_1+v_1+\theta_1}{\gamma(r_2+v_2-1+\theta_2)},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_2-\boldsymbol{e}_1,\boldsymbol{v})) = \frac{1}{2}r_1\theta_1\frac{\gamma(r_2+v_2+\theta_2)}{r_1+v_1-1+\theta_1},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_1,\boldsymbol{v})) = \frac{1}{2}r_1\theta_1\frac{r_3+v_3+\theta_3-r_1-v_1-\theta_1-\gamma(r_2+v_2+\theta_2)}{r_1+v_1-1+\theta_1},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_2,\boldsymbol{v})) = \frac{1}{2}r_2\theta_2\frac{r_3+v_3+\theta_3-r_2-v_2-\theta_2-(r_1+v_1+\theta_1)/\gamma}{r_2+v_2-1+\theta_2},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2)) = \frac{1}{2}v_2\theta_2\frac{r_1+v_1+\theta_1}{\gamma(r_2+v_2-1+\theta_2)},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_2-\boldsymbol{e}_1)) = \frac{1}{2}v_1\theta_1\frac{\gamma(r_2+v_2+\theta_2)}{r_1+v_1-1+\theta_1},$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_1)) = \frac{1}{2}v_1\theta_1\frac{r_3+v_3+\theta_3-r_1-v_1-\theta_1-\gamma(r_2+v_2+\theta_2)}{r_1+v_1-1+\theta_1},$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_2))=\frac{1}{2}v_2\theta_2\frac{r_3+v_3+\theta_3-r_2-v_2-\theta_2-(r_1+v_1+\theta_1)/\gamma}{r_2+v_2-1+\theta_2}. \tag{3.49}$$

The events with transition rates of $O(\sigma_2)$ occur much more quickly on a faster timescale when compared with the events with transition rates of $O(1)$ on the original timescale. The number of these faster events tends to infinity before another event occurs on the original timescale. Thus, between the slower events on the original timescale, the genealogy is converted instantaneously according to the equilibrium of the faster events.

Consider a sample drawing from the stationary population, with the configuration $\boldsymbol{n}>0$ so all three types are presented. We call the instantaneous jump of the limiting dual process at time $t=0$ the first phase, and the evolution on the original timescale at time $t>0$ the second phase. Let time be rescaled such that one unit of time on the faster timescale corresponds to σ_2 units of time on the original timescale. At time 0, transitions with rates of $O(\sigma_2)$ occur on the faster timescale before those with rates of $O(1)$. The branching events in particular, cannot occur on this faster timescale. Hence, no virtual lines will be added to the genealogy in the first phase. The transition rates of the events on the faster timescale are

$$\begin{aligned}
q(\boldsymbol{r},\boldsymbol{r}-\boldsymbol{e}_1)&=\frac{1}{2}\frac{r_1(r_1-1)\gamma}{r_1-1+\theta_1},\\
q(\boldsymbol{r},\boldsymbol{r}-\boldsymbol{e}_2)&=\frac{1}{2}\frac{r_2(r_2-1)}{r_2-1+\theta_2},\\
q(\boldsymbol{r},\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_1)&=\frac{1}{2}\frac{r_1\theta_1\gamma}{r_1-1+\theta_1},\\
q(\boldsymbol{r},\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_2)&=\frac{1}{2}\frac{r_2\theta_2}{r_2-1+\theta_2}.
\end{aligned} \tag{3.50}$$

The ancestral lines of type 1 and type 2 with a selective disadvantage can coalesce with the lines of the same type or mutate to type 3. Furthermore, there is no interaction between type 1 and type 2 by considering the transition rates (3.50). The genealogies of the individuals of type 1 and type 2 in the sample are independent in the first

phase, and each genealogy follows Kingman's coalescent in which there are coalescence events and mutation events to type 3. The sample $\boldsymbol{n}$ is converted to $M_1 + M_2 + n_3$ real lines of type 3 by the end of the first phase on the faster timescale, where M_i, $i = 1, 2$ is the number of mutation events occurred, in which ancestral lines of type i mutate to type 3. As already given in equation (2.17), the distribution of M_i, $i = 1, 2$, satisfies the Ewens sampling formula

$$\mathbb{P}(M_i = m_i \mid n_i) = \frac{\theta_i^{m_i}}{\theta_{i(n_i)}} \left| S_{n_i}^{(m_i)} \right|, \quad 0 < m_i \leq n_i, \tag{3.51}$$

where $S_{n_i}^{(m_i)}$ are the Stirling numbers of the first kind (see Remark 2.4).

The duration of the first phase is $O(\sigma_2^{-1})$ for a finite sample, which is negligible on the original timescale in limit as $\sigma_2 \to \infty$. After the instantaneous jump in the first phase, at the beginning of the second phase at time 0+, the configuration of ancestral lines is $((M_1 + M_2 + n_3)\boldsymbol{e}_3, \mathbf{0})$. The ancestral dual process then undergoes the transitions with rates of $O(1)$. The transition rates of possible events at time $t = 0+$ on the original timescale are

$$\begin{aligned}
q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1)) &= \frac{1}{2} r_3 \theta_1, \\
q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_2)) &= \frac{1}{2} r_3 \theta_2, \\
q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_3, \boldsymbol{v})) &= \frac{1}{2} r_3 (r_3 - 1).
\end{aligned}$$

The possible next events are the coalescences among ancestral lines of type 3, or an ancestral line branching into a virtual line of either type 1 or type 2. If a branching event occurs, and a virtual line of type 1 is added to the genealogy (the argument is similar if a virtual line of type 2 is added), the possible next transitions are

$$\begin{aligned}
q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1)) &= \frac{1}{2} (r_3 + 1)(\theta_1 + 1), \\
q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_2)) &= \frac{1}{2} (r_3 + 1)\theta_2, \\
q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_3, \boldsymbol{v})) &= \frac{1}{2} r_3 (r_3 - 1),
\end{aligned}$$

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_1)) = \frac{1}{2}[\gamma\sigma_2+\theta_3+r_3-\theta_1-1-\theta_2\gamma],$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_2-\boldsymbol{e}_1)) = \frac{1}{2}\theta_2\gamma.$$

These transition rates are again composed of terms of $O(\sigma_2)$ and $O(1)$. The newly-added virtual line of type 1 will mutate to type 3 at rate $\gamma\sigma_2/2$ in a time that is negligible on the original timescale in the limit as $\sigma_2 \to \infty$. So whenever a branching event occurs, a virtual line of type 3 is added to the genealogy on the original timescale. The transition rates of the events that are observable on the original timescale are

$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3)) = \frac{1}{2}(r_3+v_3)(\theta_1+\theta_2),$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_3,\boldsymbol{v})) = \frac{1}{2}r_3(r_3-1),$$
$$q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_3)) = \frac{1}{2}v_3(2r_3+v_3-1).$$

We are only interested in the real genealogy in the dual process, and the above rates show that the real lines can only coalesce at rate $r_3(r_3-1)/2$. The evolution of the real lines is independent of the virtual lines. Therefore, after the instantaneous jump at time $t=0$, the limiting ancestral dual process starts with the configuration $((M_1+M_2+n_3)\boldsymbol{e}_3,\boldsymbol{0})$ at time $t=0+$, and evolves according to Kingman's coalescent with coalescence rate $r_3(r_3-1)/2$ on the original timescale.

Now we turn to the other case in which the individuals of types 1 and 2 are strongly disfavoured, and the individuals of type 3 are neutral, i.e., we assume

$$\sigma_1 \gg \sigma_2 \gg \frac{\sigma_1}{\sigma_2} \gg 1.$$

The transition rates of the dual process under strong selections of different orders can be rewritten in the same way as those in Lemma 3.14, except that we cannot write σ_1 as $\gamma\sigma_2$ since σ_1 and σ_2 are of a different order now. In this case the type 2 lines cannot mutate to type 1 anymore, as the type 2 lines are much more selectively favoured than the type 1 lines.

The terms in the transition rates now have three different orders. The transitions with rates of $O(\sigma_1)$ are

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{1}{2}r_1(r_1-1)\frac{\sigma_1}{r_1+v_1-1+\theta_1},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1)) &= \frac{1}{2}v_1(2r_1+v_1-1)\frac{\sigma_1}{r_1+v_1-1+\theta_1},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{1}{2}r_1\theta_1\frac{\sigma_1}{r_1+v_1-1+\theta_1},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_1)) &= \frac{1}{2}v_1\theta_1\frac{\sigma_1}{r_1+v_1-1+\theta_1};
\end{aligned}
$$

the transitions with rates of $O(\sigma_2)$ are

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1}{2}r_2(r_2-1)\frac{\sigma_2}{r_2+v_2-1+\theta_2},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_2)) &= \frac{1}{2}v_2(2r_2+v_2-1)\frac{\sigma_2}{r_2+v_2-1+\theta_2},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1}{2}r_2\theta_2\frac{\sigma_2}{r_2+v_2-1+\theta_2},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_2)) &= \frac{1}{2}v_2\theta_2\frac{\sigma_2}{r_2+v_2-1+\theta_2};
\end{aligned}
$$

and the transitions with rates of $O(1)$ are the coalescences with the same rates in (3.49) and mutations with the following rates

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{1}{2}r_1\theta_1\frac{r_3+v_3+\theta_3-r_1-v_1-\theta_1}{r_1+v_1-1+\theta_1},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_1)) &= \frac{1}{2}v_1\theta_1\frac{r_3+v_3+\theta_3-r_1-v_1-\theta_1}{r_1+v_1-1+\theta_1},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_3-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1}{2}r_2\theta_2\frac{r_3+v_3+\theta_3-r_2-v_2-\theta_2}{r_2+v_2-1+\theta_2},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_2)) &= \frac{1}{2}v_2\theta_2\frac{r_3+v_3+\theta_3-r_2-v_2-\theta_2}{r_2+v_2-1+\theta_2}.
\end{aligned}
$$

Thus, according to the same argument, on the σ_1-timescale (where one unit time corresponds to σ_1 units of time on the original timescale), ancestral lines of type 1 can coalesce and mutate to type 3. The sample (n_1, n_2, n_3) is converted to $(0, n_2, M_1 + n_3)$. On the σ_2-timescale, ancestral lines of type 2 can coalesce and mutate to type 3. The sample is further converted to $(0, 0, M_1 + M_2 + n_3)$. These transitions all happen before

the transitions with rates of $O(1)$. The limiting dual process then starts on the original timescale with configuration $((M_1+M_2+n_3)\boldsymbol{e}_3, \boldsymbol{0})$ at time 0+. As soon as a branching event occurs, the newly-created virtual line of type 1 or type 2 instantaneously mutates to type 3. Thus, the transition rates of the events that are observable on the original timescale are

$$\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_3)) &= \frac{1}{2}(r_3+v_3)(\theta_1+\theta_2),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_3,\boldsymbol{v})) &= \frac{1}{2}r_3(r_3-1),\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_3)) &= \frac{1}{2}v_3(2r_3+v_3-1).
\end{aligned}$$

Although the evolution of the virtual lines depends on the number of real lines, the evolution of the real lines does not depend on the virtual lines. The real lines coalesce in the limiting dual process at the rate $r_3(r_3-1)/2$.

Remark 3.15. Consider the ancestral dual process of Proposition 3.12 for a sample $\boldsymbol{n} \in \mathbb{Z}_+^3$ of size $|\boldsymbol{n}| \geq 1$. Suppose $\sigma_1 \to \infty$ and $\sigma_2 \to \infty$. The real genealogy in the limiting ancestral dual process is determined by Kingman's coalescent for a sample which contains only $M_1 + M_2 + n_3$ individuals of type 3. The distribution of M_i, $i = 1, 2$, is

$$\mathbb{P}(M_i = m_i \mid n_i) = \frac{\theta_i^{m_i}}{\theta_{i(n_i)}}\left|S_{n_i}^{(m_i)}\right|, \quad 0 < m_i \leq n_i,$$

where $S_{n_i}^{(m_i)}$ are the Stirling numbers of the first kind.

Chapter 4

Genealogies under selection-mutation balance

In this chapter, we study the limiting behaviour of the ancestral processes when selection and mutation are balanced so that the forwards-in-time frequencies are constants in equilibrium. Recall that in the population governed by the diffusion with generator (1.10), selection operates in a way to eliminate type 1 individuals, but mutation converts type 2 individuals to type 1 at rate $\theta_1/2$ per individual per unit of time. Suppose the selection coefficient $\sigma \to \infty$, the mutation coefficient $\theta_1 \to \infty$ and

$$\lim_{\theta_1 \to \infty} \theta_1/\sigma = \tau, \tag{4.1}$$

where $0 < \tau < \infty$ is fixed. Note that the mutation coefficient θ_2 is assumed to be a constant. This is the case usually referred to as unidirectional mutation in the literature. We expect the frequencies of the two alleles to reach a balance by the interaction of selection and mutation. The bold dashed curve in Figure 1.1 on page 21 demonstrates that when $\theta_1 = 30$, $\sigma = 60$ and $\theta_2 = 0.2$, the stationary density of the diffusion process concentrates around 0.5.

In the remainder of this chapter, the term "under selection-mutation balance"

means under the assumption (4.1). In Section 4.1, we review the classic results regarding the dynamics of the forwards-in-time allelic frequencies under selection-mutation balance. The two backgrounds are approximately constant in this case. We again use the constant frequency assumption in the studies of the limiting processes in this chapter. In Section 4.2, we derive the limit in the coalescent-in-a-random-background model as $\theta_1 \to \infty$ under selection-mutation balance in Theorem 4.3. In Section 4.3, we deduce the limit in the ancestral dual process with typed lines in Theorem 4.13. These limiting processes are novel and are additionally supported by simulation results in Table 4.1. There are no branching events in the ancestral process in the coalescent-in-a-random-background model, but branching events can occur in the ancestral dual process with typed lines, and virtual ancestral lines can appear. We discover in Theorem 4.11 that in the ancestral dual process with typed lines, as $\theta_1 \to \infty$ under selection-mutation balance, the numbers of virtual lines increase and stabilise to a level of $O(\theta_1)$ on a much faster timescale. In Section 4.4, we discuss the connections of the ancestral dual process with typed lines under selection-mutation balance to the forwards-in-time deterministic process with Theorem 4.16.

4.1 Deterministic approximation of the diffusion

In this section, we consider the Wright-Fisher diffusion with selection and mutation, which is specified by the generator (1.10). The diffusion can be approximated deterministically as $\theta_1 \to \infty$ under selection-mutation balance, as follows. The stable equilibrium point of the deterministic trajectory is derived in equation (4.4). We also study the fluctuations about the deterministic trajectory in Lemma 4.2. These results are known in the literature, but are a prerequisite for the investigation of the limit of the ancestral processes in the remainder of this chapter.

Consider the diffusion process $\{X(t), t \geq 0\}$ with generator (1.10) and initial

value $X(0) = \zeta$ for a constant $\zeta \in (0,1)$. The generator of the time-changed process $\{X(t/\theta_1), t \geq 0\}$, evaluated on a twice continuously differentiable function $f\colon [0,1] \to \mathbb{R}$, satisfies

$$\begin{aligned}\lim_{\theta_1\to\infty} \frac{1}{\theta_1}\mathcal{L}_{\theta_1} f(x) = \lim_{\theta_1\to\infty} \frac{1}{2}\Bigg\{ \Bigg[-\frac{\sigma}{\theta_1}x(1-x) + 1 - \left(1+\frac{\theta_2}{\theta_1}\right)x \Bigg] \frac{\mathrm{d}f(x)}{\mathrm{d}x} \\ + \frac{1}{\theta_1}x(1-x)\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \Bigg\} \\ = \frac{1}{2}(1-x)(1-\tau^{-1}x)\frac{\mathrm{d}f(x)}{\mathrm{d}x},\end{aligned}$$

under selection-mutation balance (4.1). Thus the time-changed diffusion can be approximated by $\{x(t), t \geq 0\}$ where $x(t)$ is the solution to the equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{1}{2}(1-x(t))(1-\tau^{-1}x(t)), \tag{4.2}$$

with initial value $x(0) = \zeta$.

Figure 4.1 shows a path of the diffusion with generator (1.10), superimposed on the trajectory of the differential equation (4.2). Note that in the simulation we use the Moran model with a large population size to approximate the diffusion process.

The solution to the differential equation (4.2) is

$$x(t) = \begin{cases} \tau + \frac{(1-\tau)(\tau-\zeta)}{\tau-\zeta-(1-\zeta)e^{\frac{1-\tau}{2\tau}t}}, & \text{if } \tau \neq 1, \\ 1 - \frac{2(1-\zeta)}{2+(1-\zeta)t}, & \text{if } \tau = 1. \end{cases} \tag{4.3}$$

Remark 4.1 (Stable equilibrium under selection-mutation balance)**.** For any initial value $x(0) = \zeta \in (0,1)$, we have

$$\lim_{t\to\infty} x(t) = \begin{cases} \tau, & \text{if } 0 < \tau < 1, \\ 1, & \text{if } \tau \geq 1, \end{cases} \tag{4.4}$$

which is the stable equilibrium point of the differential equation (4.2).

Figure 4.2 plots the trajectories of the differential equation (4.2) with $\tau = 0.3$ and different initial values. Recall that we assume $\theta_1, \sigma \to \infty$ and $\theta_1/\sigma \to \tau$. If $\tau \geq 1$, type 1 individuals will become fixed in the population as $t \to \infty$; and if $0 < \tau < 1$,

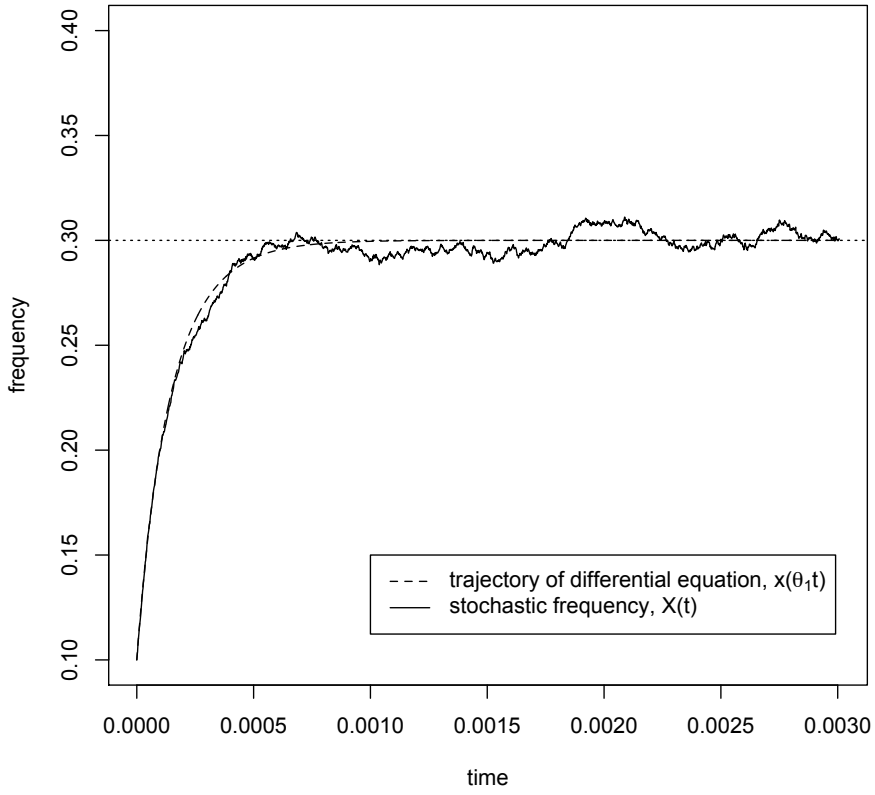


Figure 4.1: A path of the diffusion with generator (1.10), superimposed on the trajectory of the differential equation (4.2). In the simulation the diffusion is approximated by the Moran model of Definition 1.2 with population size $N = 20000$. The other coefficient values are $\theta_1 = 5000$, $\tau = 0.3$ and $\theta_2 = 2$.

the frequency stabilises to a fixed proportion τ of the total population. The latter case is more interesting, so in the remaining part of this chapter, we always assume $0 < \tau < 1$ unless otherwise stated.

The limiting results for the ancestral processes in the remainder of this chapter rely on the deterministic approximation and the stable equilibrium point in equation (4.4). However, before the discussion of the ancestral processes, we investigate the deviation from the deterministic trajectory in the following lemma due to Norman (1974, Theorem 1).

Lemma 4.2 (Fluctuations about the deterministic trajectory)**.** *Consider the diffusion*

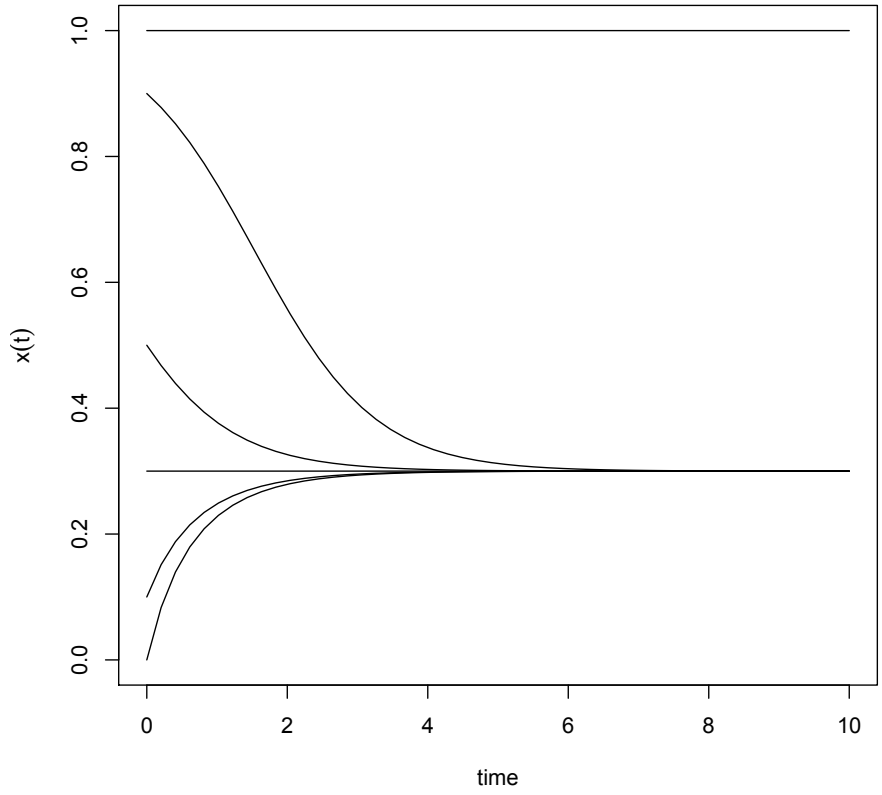


Figure 4.2: The solution to the differential equation (4.2) when $\tau = 0.3$, with respect to various initial values in $\{0, 0.1, 0.3, 0.5, 0.9, 1\}$.

process $\{X(t), t \geq 0\}$ *with generator* (1.10) *and initial value* $X(0) = \zeta \in (0, 1)$. *Let*

$$x(t) = \tau + \frac{(1-\tau)(\tau-\zeta)}{\tau - \zeta - (1-\zeta)e^{\frac{1-\tau}{2\tau}t}}, \quad t \geq 0, \tag{4.5}$$

as in equation (4.3).

Define $Z_{\theta_1}(t) = \sqrt{\theta_1}(X(t/\theta_1) - x(t))$. *As* $\theta_1 \to \infty$ *under selection-mutation balance, the real-valued process* $\{Z_{\theta_1}(t), t \geq 0\}$ *converges in finite-dimensional distribution to an Ornstein-Uhlenbeck process* $\{Z(t), t \geq 0\}$ *that satisfies the stochastic differential equation*

$$\mathrm{d}Z(t) = -\frac{1}{2\tau}(1 + \tau - 2x(t))Z(t)\,\mathrm{d}t + \sqrt{x(t)(1-x(t))}\,\mathrm{d}W(t),$$

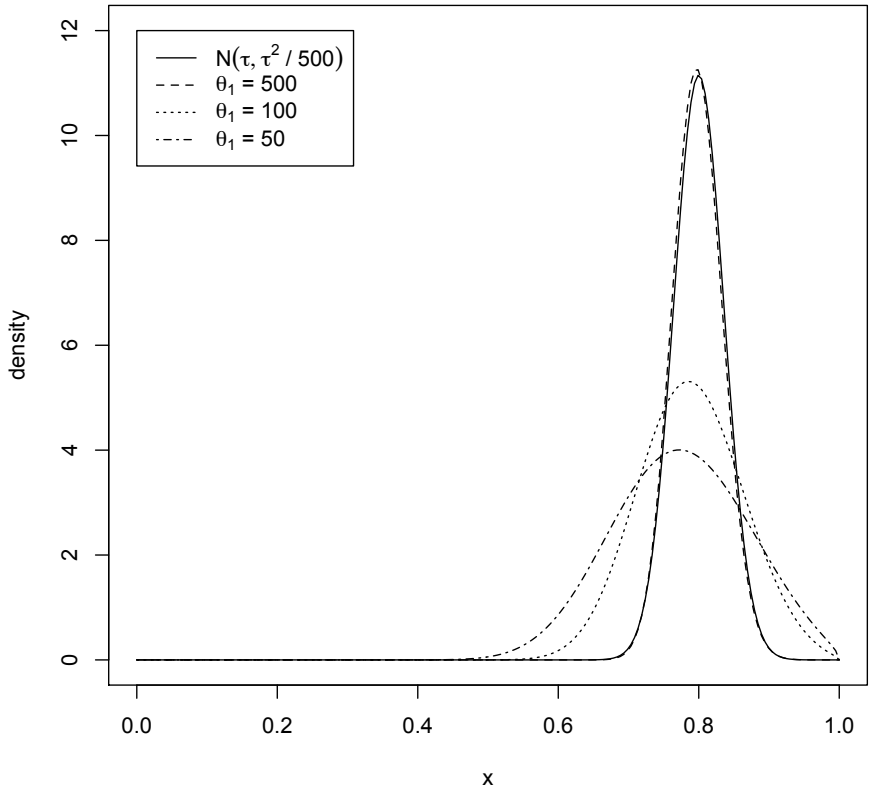


Figure 4.3: The comparison of the stationary density (1.11) of the diffusion with generator (1.10) to the density of $N(\tau, \tau^2/\theta_1)$ distribution, with respect to various values of θ_1. The other coefficient values are $\tau = 0.8$ and $\theta_2 = 1.5$.

where $\{W(t), t \geq 0\}$ *is a standard Brownian motion. The process* $\{Z(t), t \geq 0\}$ *has a unique stationary distribution* $N(0, \tau^2)$.

See Figure 4.3 for a comparison of the stationary density (1.11) of the diffusion to the density of $N(\tau, \tau^2/\theta_1)$ distribution, with respect to various strengths of selection and mutation.

Proof of Lemma 4.2. The dynamics of the diffusion $\{X(t), t \geq 0\}$ with generator (1.10) can be represented by the stochastic differential equation

$$\mathrm{d}X(t) = \frac{1}{2}[-\sigma X(t)(1 - X(t)) + \theta_1 - \theta X(t)]\,\mathrm{d}t + \sqrt{X(t)(1 - X(t))}\,\mathrm{d}W(t),$$

where $\{W(t), t \geq 0\}$ is a standard Brownian motion. Write $\tilde{X}(t) = X(t/\theta_1)$. It follows immediately that

$$\mathrm{d}\tilde{X}(t) = \frac{1}{2\theta_1}[-\sigma\tilde{X}(t)(1-\tilde{X}(t)) + \theta_1 - \theta\tilde{X}(t)]\,\mathrm{d}t + \sqrt{\frac{1}{\theta_1}\tilde{X}(t)(1-\tilde{X}(t))}\,\mathrm{d}W(t).$$

Define a real-valued process $\{Z_{\theta_1}(t), t \geq 0\}$ where $Z_{\theta_1}(t) \coloneqq \sqrt{\theta_1}(\tilde{X}(t) - x(t))$ with $x(t)$ given in equation (4.5). We have $Z_{\theta_1}(0) = 0$. It follows from Itō's formula that

$$\begin{aligned}
\mathrm{d}Z_{\theta_1}(t) &= \sqrt{\theta_1}\,\mathrm{d}\tilde{X}(t) - \sqrt{\theta_1}\,\mathrm{d}x(t) \\
&= \frac{\sqrt{\theta_1}}{2}\bigg[-\frac{\sigma}{\theta_1}\tilde{X}(t)(1-\tilde{X}(t)) + 1 - \bigg(1+\frac{\theta_2}{\theta_1}\bigg)\tilde{X}(t) \\
&\qquad + \frac{1}{\tau}x(t)(1-x(t)) - 1 + x(t)\bigg]\,\mathrm{d}t + \sqrt{\tilde{X}(t)(1-\tilde{X}(t))}\,\mathrm{d}W(t) \\
&= \frac{\sqrt{\theta_1}}{2}\bigg[-\frac{\sigma}{\theta_1}\bigg(\frac{1}{\sqrt{\theta_1}}Z_{\theta_1}(t) + x(t)\bigg)\bigg(1 - \frac{1}{\sqrt{\theta_1}}Z_{\theta_1}(t) - x(t)\bigg) \\
&\qquad - \bigg(1+\frac{\theta_2}{\theta_1}\bigg)\bigg(\frac{1}{\sqrt{\theta_1}}Z_{\theta_1}(t) + x(t)\bigg) + \frac{1}{\tau}x(t)(1-x(t)) + x(t)\bigg]\,\mathrm{d}t \\
&\quad + \sqrt{\bigg(\frac{1}{\sqrt{\theta_1}}Z_{\theta_1}(t) + x(t)\bigg)\bigg(1 - \frac{1}{\sqrt{\theta_1}}Z_{\theta_1}(t) - x(t)\bigg)}\,\mathrm{d}W(t) \\
&\coloneqq m_{\theta_1}(t, Z_{\theta_1}(t))\,\mathrm{d}t + \sqrt{\nu_{\theta_1}(t, Z_{\theta_1}(t))}\,\mathrm{d}W(t).
\end{aligned}$$

The generator of the process $\{Z_{\theta_1}(t), t \geq 0\}$ is thus

$$\mathcal{L}_{\theta_1} f(z) = m_{\theta_1}(t,z)\,\frac{\mathrm{d}f(z)}{\mathrm{d}z} + \frac{1}{2}\nu_{\theta_1}(t,z)\,\frac{\mathrm{d}^2 f(z)}{\mathrm{d}z^2}\,,$$

for a twice continuously differentiable function $f\colon \mathbb{R} \to \mathbb{R}$. We have

$$\lim_{\theta_1\to\infty} \mathcal{L}_{\theta_1} f(z) = m(t,z)\,\frac{\mathrm{d}f(z)}{\mathrm{d}z} + \frac{1}{2}\nu(t,z)\,\frac{\mathrm{d}^2 f(z)}{\mathrm{d}z^2}\,,$$

where

$$m(t,z) = \lim_{\theta_1\to\infty} m_{\theta_1}(t,z) = -\frac{1}{2\tau}(1+\tau-2x(t))z,$$

and

$$\nu(t,z) = \lim_{\theta_1\to\infty} \nu_{\theta_1}(t,z) = x(t)(1-x(t)).$$

Therefore, as $\theta_1 \to \infty$ under selection-mutation balance, the process $\{Z_{\theta_1}(t), t \geq 0\}$ converges in finite-dimensional distribution to an Ornstein-Uhlenbeck process $\{Z(t), t \geq 0\}$ that satisfies the stochastic differential equation

$$\mathrm{d}Z(t) = -\frac{1}{2\tau}(1+\tau-2x(t))Z(t)\,\mathrm{d}t + \sqrt{x(t)(1-x(t))}\,\mathrm{d}W(t).$$

To deduce the stationary distribution of the process $\{Z(t), t \geq 0\}$, we apply Itō's formula to

$$Z(t)e^{\int_0^t \frac{1}{2\tau}(1+\tau-2x(u))\,\mathrm{d}u}.$$

The solution of the resulting stochastic differential equation gives

$$Z(t) = e^{-\int_0^t \frac{1}{2\tau}(1+\tau-2x(u))\,\mathrm{d}u} \times \int_0^t \sqrt{x(u)(1-x(u))}e^{\int_0^u \frac{1}{2\tau}(1+\tau-2x(v))\,\mathrm{d}v}\,\mathrm{d}W(u).$$

Therefore, the distribution of $Z(t)$ is normal with mean 0 and variance

$$\mathrm{Var}(Z(t)) = e^{-\int_0^t \frac{1}{\tau}(1+\tau-2x(u))\,\mathrm{d}u} \times \int_0^t x(u)(1-x(u))e^{\int_0^u \frac{1}{\tau}(1+\tau-2x(v))\,\mathrm{d}v}\,\mathrm{d}u.$$

We use L'Hôpital's rule to obtain

$$\lim_{t\to\infty} \mathrm{Var}(Z(t)) = \lim_{t\to\infty} \frac{x(t)(1-x(t))}{[1+\tau-2x(t)]/\tau} = \tau^2.$$

The proof of the lemma is completed. □

4.2 Coalescent in a random background

We now investigate the ancestral processes under the assumption (4.1) of selection-mutation balance. In this section, we focus on the limit of the coalescent-in-a-random-background model of Proposition 3.2, while in Section 4.3, we will study the limit of the ancestral dual process with typed lines of Theorem 2.1. Recall that these two models for the genealogy of a given sample configuration under selection and mutation are different in that in the coalescent-in-a-random-background model, there are no

branching events in the ancestral process, and selection is reflected in the ancestral frequencies of alleles. In the dual process with typed lines on the other hand, the ancestral lines can branch and virtual lines can appear.

The deterministic approximation of the diffusion with generator equation (1.10) by equation (4.2) and equation (4.4) show that, as $\theta_1 \to \infty$ under selection-mutation balance, the forwards-in-time frequencies in the population rapidly approach the stable equilibrium point $(\tau, 1-\tau)$. Consider the coalescent-in-a-random-background model introduced in Proposition 3.2. We approximate the time-inhomogeneous ancestral process of Proposition 3.2 by a time-homogeneous process, in which we assume the ancestral frequency of type 1 individuals $\rho(t) = \tau$ for all $t \geq 0$ in the past. Therefore, for a sample $\boldsymbol{n} \in \mathbb{Z}_+^2$ of size $|\boldsymbol{n}| \geq 1$, we consider the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$, which is a Markov process with state space $\{\boldsymbol{\alpha} \colon \boldsymbol{\alpha} \in \mathbb{Z}_+^2, 1 \leq |\boldsymbol{\alpha}| \leq |\boldsymbol{n}|\}$ and initial state $\boldsymbol{n}$. When the ancestral process is at a state $\boldsymbol{\alpha}$, the transition rates are

$$\begin{aligned}
q_{\theta_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_1) &= \frac{1}{2\tau}\alpha_1(\alpha_1 - 1), \\
q_{\theta_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_2) &= \frac{1}{2(1-\tau)}\alpha_2(\alpha_2 - 1), \\
q_{\theta_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_1 + \boldsymbol{e}_2) &= \frac{1-\tau}{2\tau}\alpha_1\theta_1, \\
q_{\theta_1}(\boldsymbol{\alpha}, \boldsymbol{\alpha} - \boldsymbol{e}_2 + \boldsymbol{e}_1) &= \frac{\tau}{2(1-\tau)}\alpha_2\theta_2.
\end{aligned} \tag{4.6}$$

Kaplan et al. (1988) use the above time-homogeneous ancestral process to study how the shape of genealogies is altered by different strengths of selection and mutation. A similar assumption of constant frequencies has been used broadly in studies of the background selection model (see Remark 5.23 in Chapter 5). The novelty of our work is that we give a formal description of the limiting ancestral process as $\theta_1 \to \infty$ under selection-mutation balance in the next theorem.

Heuristically, under selection-mutation balance the population is subdivided into two genetic backgrounds, where the proportion of the type 1 population is τ. Since we

assume $\theta_1 \to \infty$, the mutations of type 1 individuals to type 2 occur at rate of $O(\theta_1)$. These events happen much faster than the other events with rates of $O(1)$. Thus, at time 0, all type 1 individuals in the sample instantaneously mutate to type 2, and the resulting sample undergoes Kingman's coalescent in the genetic background of the type 2 population, which takes a proportion $1-\tau$ of the total population. In contrast, in Section 3.1 in the previous chapter where we assume the population under strong selection is subdivided into backgrounds of proportion $1/\sigma$ and $1-1/\sigma$, individuals of type 1 may also coalesce as well as mutate on the faster timescale. Therefore type 1 individuals in the sample are converted to a random number of type 2 individuals on the faster timescale, and the genealogy of the resulting sample is determined by the usual Kingman's coalescent.

Theorem 4.3. *Consider the ancestral process* $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ *with transition rates* (4.6) *for a sample* $\boldsymbol{n} \in \mathbb{Z}_+^2$ *of size* $n = |\boldsymbol{n}| \geq 1$. *As* $\theta_1 \to \infty$, *the process* $\{\boldsymbol{\alpha}_{\theta_1}(t) \colon t \geq 0\}$ *converges in finite-dimensional distribution to a process* $\{\boldsymbol{\alpha}(t), t \geq 0\}$ *with the same initial state and state space.*

There is an instantaneous jump at $t = 0$ *in the limiting process such that*

$$\boldsymbol{\alpha}(0+) = (0, n).$$

For $t > 0$, *the transition rates of the limiting process are*

$$q((0,\alpha_2),(0,\alpha_2-1)) = \frac{1}{2(1-\tau)}\alpha_2(\alpha_2-1), \quad \alpha_2 \geq 2. \tag{4.7}$$

Before proceeding to the proof of Theorem 4.3, we deduce the limit of the expectation and variance of the height and the length of the genealogy under selection-mutation balance, as an immediate consequence of Theorem 4.3. In Kingman's coalescent with the usual coalescence rate

$$\frac{1}{2}\alpha(\alpha-1), \quad \alpha \geq 2,$$

the mean and variance of the time to the MRCA and the length of the genealogy

are given in equations (2.1) to (2.4). By Theorem 4.3, the genealogy in the limiting ancestral process has the same distribution as that in a time-changed Kingman's coalescent.

Remark 4.4 (Genealogies under selection-mutation balance)**.** Consider the genealogy generated by the ancestral process with transition rates (4.6) for a sample $\boldsymbol{n} \in \mathbb{Z}_+^2$ of size $n \geq 1$. The mean and variance of the time to the MRCA satisfy

$$\lim_{\theta_1 \to \infty} \mathbb{E}[T_{MRCA}(\boldsymbol{n})] = 2(1-\tau)\left(1 - \frac{1}{n}\right), \tag{4.8}$$

$$\lim_{\theta_1 \to \infty} \mathrm{Var}[T_{MRCA}(\boldsymbol{n})] = 4(1-\tau)\left[2\sum_{i=2}^{n} \frac{1}{i^2} - \left(1 - \frac{1}{n}\right)^2\right]; \tag{4.9}$$

and the mean and variance of the length of the genealogy satisfy

$$\lim_{\theta_1 \to \infty} \mathbb{E}[L_{MRCA}(\boldsymbol{n})] = 2(1-\tau)\sum_{i=1}^{n-1} \frac{1}{i}, \tag{4.10}$$

$$\lim_{\theta_1 \to \infty} \mathrm{Var}[L_{MRCA}(\boldsymbol{n})] = 4(1-\tau)\sum_{i=1}^{n-1} \frac{1}{i^2}. \tag{4.11}$$

We now prove Theorem 4.3. The main ingredient in the proof is to separate the timescales and apply Lemma 3.1.

Proof of Theorem 4.3. Consider the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ with transition rates (4.6) for a sample (n_1, n_2) of size $n = n_1 + n_2 \geq 1$. Let S be the state space, which is the set of all possible configurations of ancestral lines. The order of the states in S is arbitrary but fixed. We let S have three mutually disjoint subsets such that $S = \cup_{i=1}^{3} S_i$. The states are ordered in a way that each state in S_1 contains only lines of type 2, so $S_1 = \{(0, n), (0, n-1), \ldots, (0, 1)\}$; each state in S_2 consists of only lines of type 1, thus $S_2 = \{(n, 0), (n-1, 0), \ldots, (1, 0)\}$; and $S_3 = \{(1, n-1), (2, n-2), \ldots, (n-1, 1), (1, n-2), (2, n-3), \ldots, (n-2, 1), \ldots, (1, 1)\}$. The size of the state space is $|S| = 2n + \sum_{i=1}^{n-1} i = n(n+3)/2$. The absorbing states are $(0, 1)$ and $(1, 0)$. See Example 4.6 for a demonstration of the structure of the matrices in this proof.

The Q-matrix $\boldsymbol{Q}_{\theta_1}$ of the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ is $(n(n+3)/2) \times (n(n+3)/2)$-dimensional of which the non-zero off-diagonal entries are specified in equations (4.6), and the diagonal entries are minus the sum of off-diagonal entries. $\boldsymbol{Q}_{\theta_1}$ can be written in the block form as

$$\boldsymbol{Q}_{\theta_1} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} & \boldsymbol{Q}_{13} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} & \boldsymbol{Q}_{23} \\ \boldsymbol{Q}_{31} & \boldsymbol{Q}_{32} & \boldsymbol{Q}_{33} \end{pmatrix},$$

where each $\boldsymbol{Q}_{ij}$ contains the rates at which transitions from the states in S_i to the ones in S_j occur.

The matrix $\boldsymbol{Q}_{\theta_1}$ can be decomposition as $\boldsymbol{Q}_{\theta_1} = \theta_1 \boldsymbol{A} + \boldsymbol{B}$, where $\boldsymbol{A} = \lim_{\theta_1 \to \infty} \boldsymbol{Q}_{\theta_1}/\theta_1$ and $\boldsymbol{B} = \lim_{\theta_1 \to \infty}(\boldsymbol{Q}_{\theta_1} - \theta_1 \boldsymbol{A})$. For $i = (\alpha_1, \alpha_2) \in S$ and $j \in S$, the entries a_{ij} of $\boldsymbol{A}$ are

$$a_{ij} = \begin{cases} \frac{1-\tau}{2\tau}\alpha_1, & \text{if } j = (\alpha_1 - 1, \alpha_2 + 1), \\ -\frac{1-\tau}{2\tau}\alpha_1, & \text{if } j = i, \\ 0, & \text{otherwise;} \end{cases}$$

and the entries b_{ij} of $\boldsymbol{B}$ are

$$b_{ij} = \begin{cases} \frac{1}{2\tau}\alpha_1(\alpha_1 - 1), & \text{if } j = (\alpha_1 - 1, \alpha_2), \\ \frac{1}{2(1-\tau)}\alpha_2(\alpha_2 - 1), & \text{if } j = (\alpha_1, \alpha_2 - 1), \\ \frac{\tau}{2(1-\tau)}\alpha_2\theta_2, & \text{if } j = (\alpha_1 + 1, \alpha_2 - 1), \\ -\left[\frac{1}{2\tau}\alpha_1(\alpha_1 - 1) + \frac{1}{2(1-\tau)}\alpha_2(\alpha_2 - 1 + \tau\theta_2)\right], & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

In the block form, we have

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} \\ \boldsymbol{A}_{31} & \boldsymbol{0} & \boldsymbol{A}_{33} \end{pmatrix},$$

and

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{B}_{13} \\ \boldsymbol{0} & \boldsymbol{B}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_{32} & \boldsymbol{B}_{33} \end{pmatrix}.$$

Since $\boldsymbol{A}$ is a Q-matrix, there exists a Markov process with corresponding Q-matrix $\boldsymbol{A}$. This Markov process has the states in S_1 as absorbing states, and all other states

as transient states.

Let $\boldsymbol{\Pi}_{\theta_1}(t)$ be the transition matrix of the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ over a period of time t in the past. We have $\boldsymbol{\Pi}_{\theta_1}(t) = \exp(\boldsymbol{Q}_{\theta_1} t)$. According to Lemma 3.1 and the algebra similar to that in the proof of Theorem 3.5, we have

$$\lim_{\theta_1 \to \infty} \boldsymbol{\Pi}_{\theta_1}(t) = \boldsymbol{P} e^{\boldsymbol{G}t} := \boldsymbol{\Pi}(t), \tag{4.12}$$

where $\boldsymbol{P} = \lim_{m \to \infty} e^{\boldsymbol{A}m}$ and $\boldsymbol{G} = \boldsymbol{PBP}$. $\boldsymbol{\Pi}(t)$ is the transition matrix of the limiting ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$.

Since $\boldsymbol{P} = \lim_{t \to 0} \boldsymbol{\Pi}(t)$, the limiting ancestral process jumps instantaneously at time $t = 0$ from a state $i \in S$ to a state $j \in S_1$, with probability p_{ij} (entries of $\boldsymbol{P}$), i.e., according to the mutation events occurring at the rates in $\boldsymbol{A}$ on a faster timescale where one unit of time corresponds to θ_1 units of time on the original timescale. $\boldsymbol{P}$ has the block form

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{P}_{31} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

The entries of $\boldsymbol{P}_{i1}$, $i = 2, 3$ are the probabilities of absorption from the states in S_i into the states in S_1, which are equal to the probabilities that the state $(\alpha_1, \alpha_2) \in S_i$ becomes absorbed into $(0, \alpha_1 + \alpha_2) \in S_1$ after a series of mutation events from type 1 to type 2 (on the faster timescale). Thus, the entries of the matrix $\boldsymbol{P}$ are

$$p_{ij} = \begin{cases} 1, & \text{if } i = (\alpha_1, \alpha_2) \in S, j = (0, \alpha_1 + \alpha_2) \in S_1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.13}$$

The limiting ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ with transition matrix $\boldsymbol{\Pi}(t)$ moves instantaneously at time 0 according to $\boldsymbol{P}$, from the sample $(n_1, n_2) \in S$ to a state $(0, n_1 + n_2) \in S_1$. After entering such a state in S_1, the limiting process behaves like a Markov process (on the original timescale) with initial state $(0, n_1 + n_2)$ and Q-matrix $\boldsymbol{G}$. The following equation (4.14) shows that the limiting ancestral process behaves like the death process of a time-changed Kingman's coalescent.

The matrix $\boldsymbol{G}$ has the block form

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{B}_{11} + \boldsymbol{B}_{12} + \boldsymbol{B}_{13}\boldsymbol{P}_{31} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{B}_{11} + \boldsymbol{B}_{12} + \boldsymbol{B}_{13}\boldsymbol{P}_{31} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{P}_{31}\boldsymbol{B}_{11} + \boldsymbol{P}_{31}\boldsymbol{B}_{12} + \boldsymbol{P}_{31}\boldsymbol{B}_{13}\boldsymbol{P}_{31} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

The entries g_{ij} of $\boldsymbol{G}$ satisfy $g_{ij} = \sum_{l,k\in S} p_{il} b_{lk} p_{kj}$. For a state $i = (0, \alpha_2) \in S_1$, since $p_{il} = \mathbb{1}_{\{l=i\}}$, we have

$$g_{ij} = \begin{cases} \frac{1}{2(1-\tau)}\alpha_2(\alpha_2 - 1), & \text{if } j = (0, \alpha_2 - 1) \in S_1, \\ \frac{\tau}{2(1-\tau)}\alpha_2\theta_2 - \frac{1}{2(1-\tau)}\alpha_2(\alpha_2 - 1 + \tau\theta_2) = -\frac{1}{2(1-\tau)}\alpha_2(\alpha_2 - 1), & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{4.14}$$

For a state $i = (\alpha_1, \alpha_2) \in S_2 \cup S_3$, p_{il} is non-zero only when $j = (0, \alpha_1 + \alpha_2) \in S_1$. Thus,

$$g_{ij} = \begin{cases} \frac{1}{2(1-\tau)}(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_2 - 1), & \text{if } j = (0, \alpha_1 + \alpha_2 - 1) \in S_1, \\ -\frac{1}{2(1-\tau)}(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_2 - 1), & \text{if } j = (0, \alpha_1 + \alpha_2) \in S_1, \\ 0, & \text{otherwise.} \end{cases}$$

The proof is then completed. □

The following corollary specifies the entries of the transition matrix $\boldsymbol{\Pi}(t)$ of the limiting ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ of Theorem 4.3.

Corollary 4.5. *Consider the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ with transition rates (4.6) for a sample $(n_1, n_2) \in \mathbb{Z}_+^2$ of size $n_1 + n_2 \geq 1$. For fixed $t > 0$, the transition probability of the ancestral process satisfies*

$$\lim_{\theta_1 \to \infty} \mathbb{P}[\boldsymbol{\alpha}_{\theta_1}(t) = (\alpha_1, \alpha_2) \mid \boldsymbol{\alpha}_{\theta_1}(0) = (n_1, n_2)] = \begin{cases} d_{n_1+n_2,\alpha_2}(t), & \textit{if } \alpha_1 = 0, \\ 0, & \textit{otherwise,} \end{cases} \tag{4.15}$$

where $d_{ij}(t)$ is given by

$$d_{ij}(t) = \begin{cases} \sum_{l=j}^{i} e^{-\binom{l}{2}\frac{t}{1-\tau}} \frac{(2l-1)(-1)^{l-j} j_{(l-1)} i_{[l]}}{j!(l-j)! i_{(l)}}, & \textit{if } 2 \leq j \leq i, \\ 1 - \sum_{l=2}^{i} e^{-\binom{l}{2}\frac{t}{1-\tau}} \frac{(2l-1)(-1)^{l} i_{[l]}}{i_{(l)}}, & \textit{if } j = 1, j \leq i, \\ 0, & \textit{otherwise.} \end{cases} \tag{4.16}$$

Proof. We have seen in the proof of Theorem 4.3 that the transition matrix $\boldsymbol{\Pi}_{\theta_1}(t)$ of the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ satisfies

$$\lim_{\theta_1 \to \infty} \boldsymbol{\Pi}_{\theta_1}(t) = \boldsymbol{\Pi}(t) = \boldsymbol{P} e^{\boldsymbol{G}t}, \qquad t > 0,$$

where the matrix $\boldsymbol{P}$ represents the instantaneous jumps, and $\boldsymbol{G} = \boldsymbol{PBP}$ is the Q-matrix of the limiting process with state space S. Therefore, by matrix multiplication, the entries of $\boldsymbol{\Pi}(t)$ are

$$\pi_{ij}(t) = \sum_{k \in S} p_{ik} \big(e^{\boldsymbol{G}t}\big)_{kj}.$$

The entries of $\boldsymbol{P}$ are given in equation (4.13); and $d_{ij}(t)$ in equation (4.16) are the entries of the transition matrix of the death process of a time-changed Kingman's coalescent, which was derived by Griffiths (1980). The corollary then follows. □

Example 4.6. We illustrate the structure of the matrices in the proof of Theorem 4.3 and Corollary 4.5. Consider a sample (n_1, n_2) of size $n = n_1 + n_2 = 3$. The state space of the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ is

$$S = \{(0,3), (0,2), (0,1), (3,0), (2,0), (1,0), (1,2), (2,1), (1,1)\}.$$

The Q-matrix $\boldsymbol{Q}_{\theta_1}$ of the ancestral process $\{\boldsymbol{\alpha}_{\theta_1}(t), t \geq 0\}$ is

$$\begin{pmatrix}
\frac{-3(1+\tau\theta_2/2)}{1-\tau} & \frac{3}{1-\tau} & 0 & 0 & 0 & 0 & \frac{3\tau\theta_2}{2(1-\tau)} & 0 & 0 \\
0 & \frac{-1-\tau\theta_2}{1-\tau} & \frac{1}{1-\tau} & 0 & 0 & 0 & 0 & 0 & \frac{\tau\theta_2}{1-\tau} \\
0 & 0 & \frac{-\tau\theta_2}{2(1-\tau)} & 0 & 0 & \frac{\tau\theta_2}{2(1-\tau)} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{-3(1+(1-\tau)\theta_1)}{2\tau} & \frac{3}{2\tau} & 0 & 0 & \frac{3(1-\tau)\theta_1}{2\tau} & 0 \\
0 & 0 & 0 & 0 & \frac{-1-(1-\tau)\theta_1}{\tau} & \frac{1}{\tau} & 0 & 0 & \frac{(1-\tau)\theta_1}{\tau} \\
0 & 0 & \frac{(1-\tau)\theta_1}{2\tau} & 0 & 0 & -\frac{(1-\tau)\theta_1}{2\tau} & 0 & 0 & 0 \\
\frac{(1-\tau)\theta_1}{2\tau} & 0 & 0 & 0 & 0 & 0 & -\frac{1+\tau\theta_2}{1-\tau} - \frac{(1-\tau)\theta_1}{2\tau} & \frac{\tau\theta_2}{1-\tau} & \frac{1}{1-\tau} \\
0 & 0 & 0 & \frac{\tau\theta_2}{2(1-\tau)} & 0 & 0 & \frac{(1-\tau)\theta_1}{\tau} & -\frac{1+(1-\tau)\theta_1}{\tau} - \frac{\tau\theta_2}{2(1-\tau)} & \frac{1}{\tau} \\
0 & \frac{(1-\tau)\theta_1}{2\tau} & 0 & 0 & \frac{\tau\theta_2}{2(1-\tau)} & 0 & 0 & 0 & -\frac{(1-\tau)\theta_1}{2\tau} - \frac{\tau\theta_2}{2(1-\tau)}
\end{pmatrix}.$$

The Q-matrix can be decomposed according to the order of the terms in each entry,

such that $\boldsymbol{Q}_{\theta_1} = \theta_1 \boldsymbol{A} + \boldsymbol{B}$. We have

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{3(1-\tau)}{2\tau} & 0 & 0 & 0 & \frac{3(1-\tau)}{2\tau} & 0 \\ 0 & 0 & 0 & 0 & -\frac{1-\tau}{\tau} & 0 & 0 & 0 & \frac{1-\tau}{\tau} \\ 0 & 0 & \frac{1-\tau}{2\tau} & 0 & 0 & -\frac{1-\tau}{2\tau} & 0 & 0 & 0 \\ \frac{1-\tau}{2\tau} & 0 & 0 & 0 & 0 & 0 & -\frac{1-\tau}{2\tau} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-\tau}{\tau} & -\frac{1-\tau}{\tau} & 0 \\ 0 & \frac{1-\tau}{2\tau} & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1-\tau}{2\tau} \end{pmatrix}$$

and

$$\boldsymbol{B} = \begin{pmatrix} \frac{-3(1+\tau\theta_2/2)}{1-\tau} & \frac{3}{1-\tau} & 0 & 0 & 0 & 0 & \frac{3\tau\theta_2}{2(1-\tau)} & 0 & 0 \\ 0 & \frac{-1-\tau\theta_2}{1-\tau} & \frac{1}{1-\tau} & 0 & 0 & 0 & 0 & 0 & \frac{\tau\theta_2}{1-\tau} \\ 0 & 0 & \frac{-\tau\theta_2}{2(1-\tau)} & 0 & 0 & \frac{\tau\theta_2}{2(1-\tau)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-3}{2\tau} & \frac{3}{2\tau} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{\tau} & \frac{1}{\tau} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1-\tau\theta_2}{1-\tau} & \frac{\tau\theta_2}{1-\tau} & \frac{1}{1-\tau} \\ 0 & 0 & 0 & \frac{\tau\theta_2}{2(1-\tau)} & 0 & 0 & 0 & -\frac{1}{\tau} - \frac{\tau\theta_2}{2(1-\tau)} & \frac{1}{\tau} \\ 0 & 0 & 0 & 0 & \frac{\tau\theta_2}{2(1-\tau)} & 0 & 0 & 0 & \frac{-\tau\theta_2}{2(1-\tau)} \end{pmatrix}.$$

As shown in the proof of Theorem 4.3, the limiting ancestral process $\{\boldsymbol{\alpha}(t), t \geq 0\}$ undergoes instantaneous jumps according to

$$\boldsymbol{P} = \lim_{m \to \infty} e^{\boldsymbol{A}m} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The Q-matrix of the limiting ancestral process is

$$\boldsymbol{G} = \boldsymbol{PBP} = \begin{pmatrix} -\frac{3}{1-\tau} & \frac{3}{1-\tau} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{1-\tau} & \frac{1}{1-\tau} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{3}{1-\tau} & \frac{3}{1-\tau} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{1-\tau} & \frac{1}{1-\tau} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{3}{1-\tau} & \frac{3}{1-\tau} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{3}{1-\tau} & \frac{3}{1-\tau} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{1-\tau} & \frac{1}{1-\tau} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The transition matrix of the limiting ancestral process is $\boldsymbol{\Pi}(t) = \boldsymbol{P}e^{\boldsymbol{G}t}$, where

$$\boldsymbol{P}e^{\boldsymbol{G}t} = \begin{pmatrix} e^{-\frac{3t}{1-\tau}} & \frac{3}{2}e^{-\frac{t}{1-\tau}} - \frac{3}{2}e^{-\frac{3t}{1-\tau}} & 1 - \frac{3}{2}e^{-\frac{t}{1-\tau}} + \frac{1}{2}e^{-\frac{3t}{1-\tau}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-\frac{t}{1-\tau}} & 1 - e^{-\frac{t}{1-\tau}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{-\frac{3t}{1-\tau}} & \frac{3}{2}e^{-\frac{t}{1-\tau}} - \frac{3}{2}e^{-\frac{3t}{1-\tau}} & 1 - \frac{3}{2}e^{-\frac{t}{1-\tau}} + \frac{1}{2}e^{-\frac{3t}{1-\tau}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-\frac{t}{1-\tau}} & 1 - e^{-\frac{t}{1-\tau}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{-\frac{3t}{1-\tau}} & \frac{3}{2}e^{-\frac{t}{1-\tau}} - \frac{3}{2}e^{-\frac{3t}{1-\tau}} & 1 - \frac{3}{2}e^{-\frac{t}{1-\tau}} + \frac{1}{2}e^{-\frac{3t}{1-\tau}} & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{-\frac{3t}{1-\tau}} & \frac{3}{2}e^{-\frac{t}{1-\tau}} - \frac{3}{2}e^{-\frac{3t}{1-\tau}} & 1 - \frac{3}{2}e^{-\frac{t}{1-\tau}} + \frac{1}{2}e^{-\frac{3t}{1-\tau}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-\frac{t}{1-\tau}} & 1 - e^{-\frac{t}{1-\tau}} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We have verified Corollary 4.5, and the fact that the limiting ancestral process under selection-mutation balance behaves like the death process of the time-changed Kingman's coalescent, as stated in Theorem 4.3.

4.3 Ancestral dual process with typed lines

In this section, we examine the limiting behaviour of the ancestral dual process with typed lines of Theorem 2.1 under selection-mutation balance. This investigation is new, and the main results are in Theorems 4.11 and 4.13.

The starting point is to asymptotically expand the confluent hypergeometric functions appearing in the transition rates (2.6) to (2.10) of the ancestral dual process with typed lines. The confluent hypergeometric function $M(r_1 + v_1 + \theta_1, r + v + \theta, -\sigma) \equiv$

$\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})$, where $r=|\boldsymbol{r}|$, $v=|\boldsymbol{v}|$, $\theta=|\boldsymbol{\theta}|$ and $\sigma>0$, is defined in equation (1.12). The following lemma is due to López and Pagola (2010).

Lemma 4.7 (Confluent hypergeometric function under selection-mutation balance)**.** *For large θ_1 and σ such that $\lim_{\theta_1\to\infty}\theta_1/\sigma=\tau$, and $r_1,r_2,v_1,v_2\sim O(1)$, the confluent hypergeometric function has the following asymptotic expansion:*

$$\begin{aligned}
&M(r_1+v_1+\theta_1, r+v+\theta, -\sigma)\\
&\quad\sim \frac{\Gamma(r+v+\theta)(1-\gamma)^{r_2+v_2+\theta_2-1}}{\Gamma(r_1+v_1+\theta_1)\Gamma(r_2+v_2+\theta_2)}\left(\frac{\gamma}{e}\right)^{\gamma\sigma}\sqrt{\frac{2\gamma\pi}{\sigma}}\\
&\qquad\times\left[1+\frac{1}{\sigma}\left(\frac{(2-\gamma(r_2+v_2+\theta_2))(1-r_2-v_2-\theta_2)}{2(1-\gamma)^2}+\frac{1}{12\gamma}\right)+O(\sigma^{-2})\right],
\end{aligned}$$

where $\gamma=\tau+(r_1+v_1-1)/\sigma$.

Note that a condition for the asymptotic expansion in the above Lemma 4.7 to hold is $r_1,r_2,v_1,v_2\sim O(1)$. The numbers of real ancestral lines are always of a constant order, but the numbers of virtual lines could potentially be of a higher order. The asymptotic expansions of the transition rates (2.6) to (2.10) are obtained in the following Lemma 4.8, for v_1,v_2 being of a constant order. The situation where v_1,v_2 are of a higher order is dealt with in Remark 4.9.

Lemma 4.8. *Consider the transition rates* (2.6) *to* (2.10) *of the ancestral dual process with typed lines. Suppose $v_1,v_2\sim O(1)$, for large θ_1 under selection-mutation balance, these transition rates can be rewritten as*

$$\begin{aligned}
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) &= \frac{1}{2}(r+v)\theta_1+\frac{(r+v)[r_1+v_1-\tau(r+v-1+\theta_2)]}{2(1-\tau)}+O(\theta_1^{-1}),\\
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{1}{2\tau}r_1(r_1-1)+O(\theta_1^{-1}),\\
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1}{2(1-\tau)}r_2(r_2-1)+O(\theta_1^{-1}),\\
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1)) &= \frac{1}{2\tau}v_1(2r_1+v_1-1)+O(\theta_1^{-1}),\\
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_2)) &= \frac{1}{2(1-\tau)}v_2(2r_2+v_2-1)+O(\theta_1^{-1}),
\end{aligned}$$

$$
\begin{aligned}
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{\tau}{2(1-\tau)} r_2\theta_2 + O(\theta_1^{-1}),\\
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_2-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{1-\tau}{2\tau} r_1\theta_1 + \frac{r_1[\tau(r+v-1+\theta_2)-r_1-v_1]}{2\tau(1-\tau)} + O(\theta_1^{-1}),\\
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2)) &= \frac{\tau}{2(1-\tau)} v_2\theta_2 + O(\theta_1^{-1}),\\
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_2-\boldsymbol{e}_1)) &= \frac{1-\tau}{2\tau} v_1\theta_1 + \frac{v_1[\tau(r+v-1+\theta_2)-r_1-v_1]}{2\tau(1-\tau)} + O(\theta_1^{-1}).
\end{aligned}
\tag{4.17}
$$

Proof. We substitute the asymptotic formula for the confluent hypergeometric functions in Lemma 4.7 into the transition rates (2.6) to (2.10) of the ancestral dual process with typed lines to obtain the asymptotic expansions of the rates. □

Remark 4.9. We notice that the transition rates (2.6) to (2.10) of the ancestral dual process of Theorem 2.1 can be interpreted, using Bayes' rule, as the unconditional rate of an event multiplied by the sampling probability of ancestral lines given the event, then divided by the sampling probability before the event. For example, the branching rate (2.6) can be expressed in the form

$$
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) = \frac{1}{2}(r+v)\sigma\frac{\mathscr{H}(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},
$$

where $\mathscr{H}(\cdot,\cdot)$ are the sampling probabilities given in equation (2.5). Let X_1 and $X_2 \equiv 1-X_1$ be the stationary frequencies of the two alleles in the population governed by the diffusion with generator (1.10). X_1 and X_2 are asymptotically τ and $1-\tau$, respectively, as $\theta_1 \to \infty$ under selection-mutation balance, as argued in Lemma 4.2. Thus, we have

$$
\mathscr{H}(\boldsymbol{r},\boldsymbol{v}) = \mathbb{E}(X_1^{r_1+v_1}X_2^{r_2+v_2}) \sim \tau^{r_1+v_1}(1-\tau)^{r_2+v_2}.
$$

Therefore, the branching rate satisfies

$$
q_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) \sim \frac{1}{2}(r+v)\sigma\frac{\tau^{r_1+v_1+1}(1-\tau)^{r_2+v_2}}{\tau^{r_1+v_1}(1-\tau)^{r_2+v_2}}
$$

$$= \frac{1}{2}(r+v)\sigma\tau$$
$$\sim \frac{1}{2}(r+v)\theta_1, \tag{4.18}$$

since $\lim_{\theta_1\to\infty} \theta_1/\sigma = \tau$. This asymptotic form of the branching rate agrees with our algebraic result in equations (4.17). In fact, the asymptotic rates in (4.17) are all in the form of equation (4.18) plus a lower-order term representing additional randomness.

Therefore, for v_1, v_2 being of any order, the transition rates should still be in the form of equations (4.18) plus a lower-order term representing additional randomness. Thus, we consider the process $\{(\boldsymbol{r}_{\theta_1}(t), \boldsymbol{v}_{\theta_1}(t)), t \geq 0\}$ with the following transition rates, as an approximation to the ancestral process with the full transition rates under selection-mutation balance:

$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) = \frac{1}{2}(r+v)\theta_1, \tag{4.19}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v})) = \frac{1}{2\tau}r_1(r_1-1), \tag{4.20}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v})) = \frac{1}{2(1-\tau)}r_2(r_2-1), \tag{4.21}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1)) = \frac{1}{2\tau}v_1(2r_1+v_1-1), \tag{4.22}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_2)) = \frac{1}{2(1-\tau)}v_2(2r_2+v_2-1), \tag{4.23}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{v})) = \frac{\tau}{2(1-\tau)}r_2\theta_2, \tag{4.24}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_2-\boldsymbol{e}_1,\boldsymbol{v})) = \frac{1-\tau}{2\tau}r_1\theta_1, \tag{4.25}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2)) = \frac{\tau}{2(1-\tau)}v_2\theta_2, \tag{4.26}$$
$$\tilde{q}_{\theta_1}((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_2-\boldsymbol{e}_1)) = \frac{1-\tau}{2\tau}v_1\theta_1, \tag{4.27}$$

where $r = r_1 + r_2$ and $v = v_1 + v_2$.

The transition rates (4.19) to (4.27) are composed of terms of $O(1)$ and ones of a higher order which is $O(\theta_1)$ when v_1 and v_2 are $O(1)$, but if the orders of v_1 and v_2 become $O(\theta_1)$, the orders of some terms in the rates can potentially become

$O(\theta_1^2)$. Nevertheless, we can still use the idea of separation of timescales to study the behaviour of the limiting ancestral dual process, which is a combination of a process running on a faster timescale and one running on the original timescale.

The rates (4.19), (4.22), (4.23), (4.26) and (4.27) of the transitions $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1)$, $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_1)$, $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_2)$, $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1 - \boldsymbol{e}_2)$ and $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_2 - \boldsymbol{e}_1)$ that change the configuration of virtual ancestral lines, are dependent on both the numbers of virtual and the numbers of real lines. However, the rates (4.20), (4.21), (4.24) and (4.25) of the transitions $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r} - \boldsymbol{e}_1, \boldsymbol{v})$, $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r} - \boldsymbol{e}_2, \boldsymbol{v})$, $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r} + \boldsymbol{e}_1 - \boldsymbol{e}_2, \boldsymbol{v})$ and $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r} + \boldsymbol{e}_2 - \boldsymbol{e}_1, \boldsymbol{v})$ that change the configuration of real ancestral lines, are *independent* of the numbers of virtual lines.

Remark 4.10. The real genealogy in the ancestral process with transition rates (4.19) to (4.27) is governed by the marginal process of the real ancestral lines, which is independent of the numbers of virtual lines. We will see in Theorem 4.11 that on a faster timescale, the evolution of the virtual lines is also independent on the real lines.

We now give a description of the evolution of the virtual lines. The limiting behaviour of the real genealogy will be derived in Theorem 4.13. The numbers of virtual lines are both zero at time 0 when a sample is taken. The possible events occurring on the faster timescale where one unit of time corresponds to θ_1 units of time on the original timescale, are: the branching events, the mutation events for real lines of type 1 to type 2, and the mutation events for virtual lines of type 1 to type 2, with the following rates (on the faster timescale):

$$q_{\theta_1}((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1)) = \frac{1}{2}(r + v), \tag{4.28}$$

$$q_{\theta_1}((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_2 - \boldsymbol{e}_1, \boldsymbol{v})) = \frac{1 - \tau}{2\tau} r_1, \tag{4.29}$$

$$q_{\theta_1}((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_2 - \boldsymbol{e}_1)) = \frac{1}{2} v_1 \frac{1 - \tau}{\tau}. \tag{4.30}$$

The mutation events do not change the total number of real lines nor the total

number of virtual lines. Thus, the branching rate (4.28) is not affected by these mutation events. The total number of virtual lines follows a birth process with birth rate $(r+v)/2$. The rate (4.30) at which the virtual lines of type 1 mutate to type 2 is of the same order as the branching rate, so the numbers of virtual lines of both types will keep growing in such a process.

Consider the ancestral dual process $\{(\boldsymbol{r}_{\theta_1}(t), \boldsymbol{v}_{\theta_1}(t)), t \geq 0\}$ with transition rates (4.19) to (4.27). For a sample $\boldsymbol{n}$, we have $\boldsymbol{0} \leq \boldsymbol{r}(t) \leq \boldsymbol{n}$ for all $t \geq 0$; however, $\boldsymbol{v}(t)$ is in $\mathbb{Z}_+^2$. Define $\boldsymbol{\kappa}_{\theta_1}(t) = \boldsymbol{v}_{\theta_1}(t)/\theta_1$. Consider the process $\{(\boldsymbol{r}_{\theta_1}(t), \boldsymbol{\kappa}_{\theta_1}(t)), t \geq 0\}$ with state space $S = \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\} \times \mathbb{R}_+ \times \mathbb{R}_+$.

Theorem 4.11 (Virtual lines under selection-mutation balance)**.** *Assume $\boldsymbol{\kappa}_{\theta_1}(0) \neq \boldsymbol{0}$. As $\theta_1 \to \infty$ under selection-mutation balance, the process $\{(\boldsymbol{r}_{\theta_1}(t/\theta_1), \boldsymbol{\kappa}_{\theta_1}(t/\theta_1)), t \geq 0\}$ converges in finite-dimensional distribution to a process $\{(\boldsymbol{r}(t), \boldsymbol{\kappa}(t)), t \geq 0\}$ with state space S running on the faster timescale where one unit of time corresponds to θ_1 units of time on the original timescale.*

Suppose $h(\boldsymbol{r}, \boldsymbol{\kappa})\colon S \to \mathbb{R}$ is continuously differentiable with respect to $\boldsymbol{\kappa}$. The process $\{(\boldsymbol{r}(t), \boldsymbol{\kappa}(t)), t \geq 0\}$ is specified by the following generator:

$$\begin{aligned}\mathcal{L}h(\boldsymbol{r}, \boldsymbol{\kappa}) = {} & [h(\boldsymbol{r} - \boldsymbol{e}_1 + \boldsymbol{e}_2, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{1-\tau}{2\tau} r_1 \\ & + \left[\frac{1}{2}(\kappa_1 + \kappa_2) - \frac{1}{2\tau}\kappa_1^2 - \frac{1-\tau}{2\tau}\kappa_1\right] \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_1} \\ & + \left[\frac{1-\tau}{2\tau}\kappa_1 - \frac{1}{2(1-\tau)}\kappa_2^2\right] \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_2}\,. \end{aligned} \tag{4.31}$$

Proof. Consider the process $\{(\boldsymbol{r}_{\theta_1}(t), \boldsymbol{\kappa}_{\theta_1}(t)), t \geq 0\}$ that has the transition rates (4.19) to (4.27) and the generator $\mathcal{L}_{\theta_1}$. Suppose $h(\boldsymbol{r}, \boldsymbol{\kappa})\colon S \to \mathbb{R}$ is twice continuously differentiable with respect to $\boldsymbol{\kappa}$. For $\kappa_1 = v_1/\theta_1$ and $\kappa_2 = v_2/\theta_1$, we have

$$\begin{aligned}\mathcal{L}_{\theta_1}h(\boldsymbol{r}, \boldsymbol{\kappa}) = {} & [h(\boldsymbol{r} - \boldsymbol{e}_1, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{1}{2\tau} r_1(r_1 - 1) \\ & + [h(\boldsymbol{r} - \boldsymbol{e}_2, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{1}{2(1-\tau)} r_2(r_2 - 1)\end{aligned}$$

$$
\begin{aligned}
&+ [h(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{\tau}{2(1-\tau)}r_2\theta_2\\
&+ [h(\boldsymbol{r}-\boldsymbol{e}_1+\boldsymbol{e}_2,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1-\tau}{2\tau}r_1\theta_1\\
&+ [h(\boldsymbol{r},\boldsymbol{\kappa}+\boldsymbol{e}_1/\theta_1)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2}\left(\frac{r_1}{\theta_1}+\frac{r_2}{\theta_1}+\kappa_1+\kappa_2\right)\theta_1^2\\
&+ [h(\boldsymbol{r},\boldsymbol{\kappa}-\boldsymbol{e}_1/\theta_1)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2}\kappa_1\left(\frac{2r_1}{\theta_1}+\kappa_1-\frac{1}{\theta_1}\right)\frac{1}{\tau}\theta_1^2\\
&+ [h(\boldsymbol{r},\boldsymbol{\kappa}-\boldsymbol{e}_2/\theta_1)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2}\kappa_2\left(\frac{2r_2}{\theta_1}+\kappa_2-\frac{1}{\theta_1}\right)\frac{1}{1-\tau}\theta_1^2\\
&+ [h(\boldsymbol{r},\boldsymbol{\kappa}+\boldsymbol{e}_1/\theta_1-\boldsymbol{e}_2/\theta_1)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2}\kappa_2\theta_2\frac{\tau}{1-\tau}\theta_1\\
&+ [h(\boldsymbol{r},\boldsymbol{\kappa}-\boldsymbol{e}_1/\theta_1+\boldsymbol{e}_2/\theta_1)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2}\kappa_1\frac{1-\tau}{\tau}\theta_1^2\\
=\ &[h(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2\tau}r_1(r_1-1)\\
&+ [h(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2(1-\tau)}r_2(r_2-1)\\
&+ [h(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{\tau}{2(1-\tau)}r_2\theta_2\\
&+ [h(\boldsymbol{r}-\boldsymbol{e}_1+\boldsymbol{e}_2,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1-\tau}{2\tau}r_1\theta_1\\
&+ \left[\frac{1}{\theta_1}\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1}+\frac{1}{2\theta_1^2}\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1^2}+O(\theta_1^{-3})\right]\times\frac{1}{2}\left(\frac{r_1}{\theta_1}+\frac{r_2}{\theta_1}+\kappa_1+\kappa_2\right)\theta_1^2\\
&+ \left[-\frac{1}{\theta_1}\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1}+\frac{1}{2\theta_1^2}\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1^2}+O(\theta_1^{-3})\right]\times\frac{1}{2}\kappa_1\left(\frac{2r_1}{\theta_1}+\kappa_1-\frac{1}{\theta_1}\right)\frac{1}{\tau}\theta_1^2\\
&+ \left[-\frac{1}{\theta_1}\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2}+\frac{1}{2\theta_1^2}\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2^2}+O(\theta_1^{-3})\right]\times\frac{1}{2}\kappa_2\left(\frac{2r_2}{\theta_1}+\kappa_2-\frac{1}{\theta_1}\right)\frac{1}{1-\tau}\theta_1^2\\
&+ \left[\frac{1}{\theta_1}\left(\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1}-\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2}\right)+\frac{1}{2\theta_1^2}\left(\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1^2}+\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2^2}-2\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1\partial\kappa_2}\right)\right.\\
&\qquad\left.+O(\theta_1^{-3})\right]\times\frac{1}{2}\kappa_2\theta_2\frac{\tau}{1-\tau}\theta_1\\
&+ \left[\frac{1}{\theta_1}\left(\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2}-\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1}\right)+\frac{1}{2\theta_1^2}\left(\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1^2}+\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2^2}-2\frac{\partial^2 h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1\partial\kappa_2}\right)\right.\\
&\qquad\left.+O(\theta_1^{-3})\right]\times\frac{1}{2}\kappa_1\frac{1-\tau}{\tau}\theta_1^2.
\end{aligned}
$$

It follows that the generator of the time-changed process $\{(\boldsymbol{r}_{\theta_1}(t/\theta_1),\boldsymbol{\kappa}_{\theta_1}(t/\theta_1)),t\geq 0\}$

satisfies

$$\begin{aligned}\lim_{\theta_1\to\infty}\frac{1}{\theta_1}\mathcal{L}_{\theta_1}h(\boldsymbol{r},\boldsymbol{\kappa}) &= [h(\boldsymbol{r}-\boldsymbol{e}_1+\boldsymbol{e}_2,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1-\tau}{2\tau}r_1\\ &\quad+\frac{1}{2}(\kappa_1+\kappa_2)\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1}-\frac{1}{2\tau}\kappa_1^2\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1}-\frac{1}{2(1-\tau)}\kappa_2^2\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2}\\ &\quad+\frac{1-\tau}{2\tau}\kappa_1\left(\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_2}-\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_1}\right)\\ &= \mathcal{L}h(\boldsymbol{r},\boldsymbol{\kappa}).\end{aligned}$$

The proof is thus completed. □

Remark 4.12. Given the generator (4.31), on the faster timescale, the Markov jump process of the numbers of real lines $\{\boldsymbol{r}(t), t\geq 0\}$ and the continuous process of the scaled numbers of virtual lines $\{\boldsymbol{\kappa}(t), t\geq 0\}$ are independent.

The scaled numbers of virtual lines $\{\boldsymbol{\kappa}(t), t\geq 0\}$ can be approximated by the trajectory of the following system of equations:

$$\begin{aligned}\frac{\mathrm{d}\kappa_1(t)}{\mathrm{d}t} &= \frac{1}{2}\left[\kappa_1(t)+\kappa_2(t)-\frac{1}{\tau}\kappa_1^2(t)-\frac{1-\tau}{\tau}\kappa_1(t)\right],\\ \frac{\mathrm{d}\kappa_2(t)}{\mathrm{d}t} &= \frac{1}{2}\left[\frac{1-\tau}{\tau}\kappa_1(t)-\frac{1}{1-\tau}\kappa_2^2(t)\right],\end{aligned}\tag{4.32}$$

with some initial value $\boldsymbol{\kappa}(0)\neq\mathbf{0}$.

The possible equilibrium point (κ_1^*,κ_2^*) of the system (4.32), where $\kappa_i^*=\lim_{t\to\infty}\kappa_i(t)$, $i=1,2$, can be derived by setting the left-hand sides of the equations (4.32) to zero. We then have the following non-linear algebraic system:

$$\begin{aligned}0 &= \kappa_1^*+\kappa_2^*-\frac{1}{\tau}\kappa_1^{*2}-\frac{1-\tau}{\tau}\kappa_1^*,\\ 0 &= \frac{1-\tau}{\tau}\kappa_1^*-\frac{1}{1-\tau}\kappa_2^{*2},\end{aligned}$$

which gives two distinct sets of solutions, namely $\kappa_1^*=\kappa_2^*=0$ and

$$\kappa_1^*=\tau,\quad \kappa_2^*=1-\tau.$$

The Jacobian matrix of the system (4.32) is

$$\boldsymbol{J}(\kappa_1(t), \kappa_2(t)) = \begin{pmatrix} \frac{2\tau-1-2\kappa_1(t)}{2\tau} & \frac{1}{2} \\ \frac{1-\tau}{2\tau} & -\frac{\kappa_2(t)}{1-\tau} \end{pmatrix}.$$

The eigenvalues of the matrix $\boldsymbol{J}(\tau, 1-\tau)$ are $-1/2 < 0$ and $-(1+\tau)/(2\tau) < 0$. Then, according to Hale (1980, Theorem 4.2 in Chapter 3), the equilibrium point $(\tau, 1-\tau)$ is asymptotically stable in the sense that whenever $(\kappa_1(0), \kappa_2(0)) \neq (0, 0)$,

$$\lim_{t\to\infty} \big\|(\kappa_1(t), \kappa_2(t)) - (\tau, 1-\tau)\big\| = 0,$$

where for an n-dimensional vector $\boldsymbol{x}$, we consider the Euclidean norm $\|\boldsymbol{x}\| \coloneqq \sqrt{\sum_{i=1}^n x_i^2}$.

Therefore, on the faster timescale, the numbers of virtual lines will initially grow because of branching and mutation events. When the numbers of virtual lines reach the level of $O(\theta_1)$, the rates (4.22) and (4.23) at which the virtual lines coalesce become $O(\theta_1^2)$, so these coalescence events take effect on the faster timescale. The numbers of virtual lines then stabilise to $\tau\theta_1$ and $(1-\tau)\theta_1$. See Figure 4.4 for a plot of the evolution of the virtual lines from simulations.

However, the total number of virtual lines will eventually become one via the events occurring on the original timescale. This follows because if we do not distinguish the types and the real and virtual lines and set the mutation coefficients to zero, the ancestral dual process is just a birth-death process with quadratic death rate and linear birth rate (see Etheridge and Griffiths, 2009, Equation 37). These events for virtual lines on the original timescale have been ignored in our approximation as they are complicated to handle and our interests focus mainly on the real genealogy.

As stated in Remark 4.10, the shape of the real genealogy is independent of the evolution of the virtual lines. The real genealogy is governed by a process where only real ancestral lines are evolved, with the following transition rates:

$$q_{\theta_1}(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{e}_1) = \frac{1}{2\tau} r_1(r_1 - 1),$$

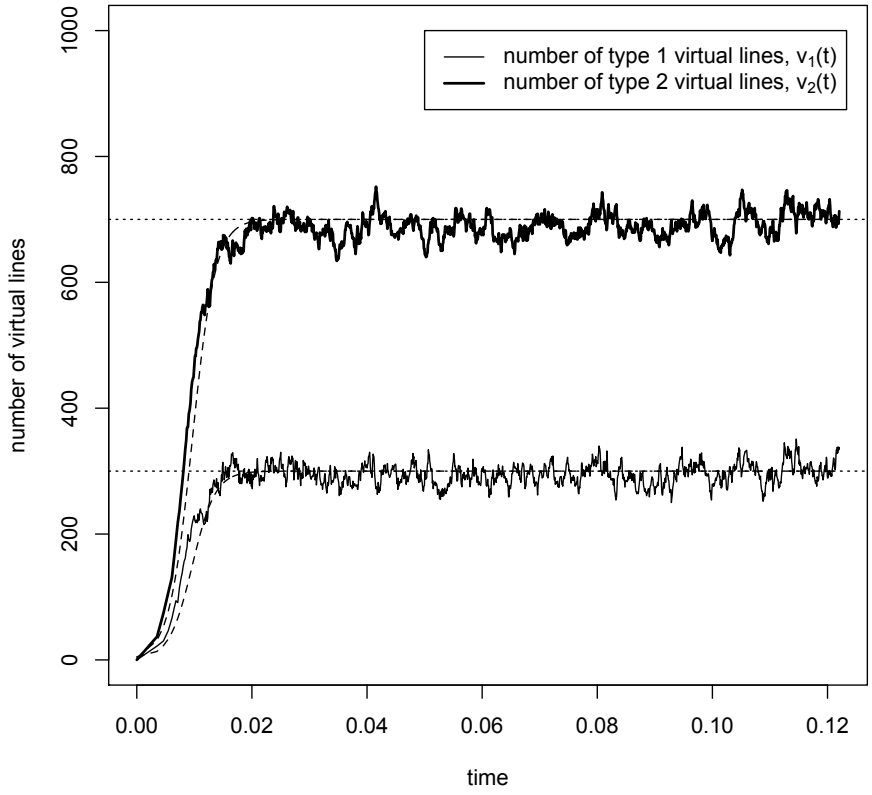


Figure 4.4: The numbers of virtual lines in the ancestral dual process of Theorem 2.1, superimposed on the deterministic trajectory of the system (4.32). Time is measured on the original timescale. The sample is $(5, 3)$, and the coefficient values are $\theta_1 = 1000$, $\tau = 0.3$ and $\theta_2 = 2$. The horizontal dotted lines represent the stable equilibrium point $(\theta_1\tau, \theta_1(1-\tau))$.

$$
\begin{aligned}
q_{\theta_1}(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{e}_2) &= \frac{1}{2(1-\tau)} r_2(r_2 - 1), \\
q_{\theta_1}(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_1 - \boldsymbol{e}_2) &= \frac{\tau}{2(1-\tau)} r_2\theta_2, \\
q_{\theta_1}(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_2 - \boldsymbol{e}_1) &= \frac{1-\tau}{2\tau} r_1\theta_1.
\end{aligned} \tag{4.33}
$$

We provide a heuristic description of the evolution of the real lines here. The formal derivation will be given in the proof of Theorem 4.13. We see that the transitions $\boldsymbol{r} \to \boldsymbol{r} + \boldsymbol{e}_2 - \boldsymbol{e}_1$ occur at a rate of $O(\theta_1)$, therefore much more quickly than the other events with rates of $O(1)$, in the limit as $\theta_1 \to \infty$. The number of these faster mutation

events tends to infinity before another event occurs on the original timescale. Consider a sample $(n_1, n_2) \in \mathbb{Z}_+^2$. The sample is instantaneously converted to $(0, n_1+n_2)$, which has only individuals of type 2.

The transitions on the original timescale starts at time 0+. Since there is no real line of type 1 at time 0+, the coalescence events between a pair of real lines of type 1 cannot occur. The only possible transitions on the original timescale are

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_2, \boldsymbol{v})) = \frac{1}{2(1-\tau)} r_2(r_2 - 1),$$
$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_1 - \boldsymbol{e}_2, \boldsymbol{v})) = \frac{\tau}{2(1-\tau)} r_2\theta_2.$$

However, as soon as a real line of type 2 mutates to type 1 so that r_1 becomes 1, this line of type 1 will mutate back to type 2 on the faster timescale because the corresponding mutation rate is $O(\theta_1)$. Thus, the only observable events on the original timescale are the coalescence events of real lines of type 2 at rate

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_2, \boldsymbol{v})) = \frac{1}{2(1-\tau)} r_2(r_2 - 1).$$

The following theorem is for the real genealogy in the limiting ancestral dual process as $\theta_1 \to \infty$ under selection-mutation balance.

Theorem 4.13 (Real lines under selection-mutation balance)**.** *Consider the ancestral process* $\{\boldsymbol{r}_{\theta_1}(t), t \geq 0\}$ *for a sample* $\boldsymbol{n} \in \mathbb{Z}_+^2$ *of size* $n = |\boldsymbol{n}| \geq 1$*, which has the state space* $\{\boldsymbol{r} \colon \boldsymbol{r} \in \mathbb{Z}_+^2, 1 \leq |\boldsymbol{r}| \leq |\boldsymbol{n}|\}$*, the initial state* $\boldsymbol{n}$ *and the transitions rates* (4.33)*. As* $\theta_1 \to \infty$*, the process* $\{\boldsymbol{r}_{\theta_1}(t), t \geq 0\}$ *converges in finite-dimensional distribution to a process* $\{\boldsymbol{r}(t), t \geq 0\}$ *with the same state space and initial state.*

There is an instantaneous jump at $t = 0$ *in the limiting process such that*

$$\boldsymbol{r}(0+) = (0, n).$$

For $t > 0$, the transition rates of the limiting process are

$$q((0, r_2), (0, r_2 - 1)) = \frac{1}{2(1-\tau)} r_2(r_2 - 1), \quad r_2 \geq 2.$$

Proof. The process $\{\boldsymbol{r}_{\theta_1}(t), t \geq 0\}$ with transition rates (4.33) is the same as the ancestral process with transition rates (4.6), so the present theorem follows the proof of Theorem 4.3. □

Remark 4.14. Although the coalescent-in-a-random-background model and the ancestral dual process with typed lines are different, as mentioned at the beginning of this chapter, the limiting real genealogies in Theorems 4.3 and 4.13 are the same. In fact, the two models generate the same real genealogy in the limit of selection-mutation balance as well as strong selection (see Theorem 3.5 and Remark 3.10).

Given Theorem 4.13, the real genealogy in the limiting ancestral dual process has the same distribution as that in a time-changed Kingman's coalescent. We can immediately write down formulae for the limit of the expectation and variance of the time to the MRCA and the length of the genealogy , as in the following Remark 4.15. We can also derive the limiting values of the other quantities relating to the genealogies, e.g, the coalescence times and the type distribution of the MRCA.

Remark 4.15 (Genealogies under selection-mutation balance)**.** According to Theorem 4.13, as $\theta_1 \to \infty$ under selection-mutation balance (4.1), the mean and variance of the time to the MRCA satisfy equations (4.8) and (4.9), and the mean and variance of the length of the genealogy satisfy equations (4.10) and (4.11).

Table 4.1 lists the values of the mean time to the MRCA obtained from simulations of the ancestral dual process of Theorem 2.1, with respect to various sample configurations and values of θ_1. The simulation results are averaged over 20000 Monte Carlo replicates. It is observed that, as the value of θ_1 increases (the value of σ also increases), the values of the mean time to the MRCA decrease for sample configurations

Table 4.1: The mean time to the MRCA in the ancestral dual process with typed lines, with respect to various sample configurations and increasing values of θ_1. The other coefficient values are $\tau = 0.3$ and $\theta_2 = 0.2$.

θ_1	$\mathbb{E}[T_{MRCA}(2,0)]$	$\mathbb{E}[T_{MRCA}(1,1)]$	$\mathbb{E}[T_{MRCA}(0,2)]$
2	0.97	1.27	0.59
5	0.82	0.87	0.61
10	0.71	0.75	0.65
30	0.69	0.71	0.68
150	0.69	0.70	0.69
∞	0.70	0.70	0.70

The bottom row lists the limiting values calculated from equation (4.8). The other values are obtained from simulations of the ancestral dual process of Theorem 2.1.

$(2,0)$ and $(1,1)$ but increase for sample configuration $(0,2)$, and eventually approach the theoretical limiting value in equation (4.8). These simulation results provide an additional support to the limiting quantities in Remark 4.15 regarding the shape of genealogies under selection-mutation balance, and also to the limiting ancestral dual process in Theorem 4.13.

4.4 Connections to the forwards-in-time process

If, instead of the diffusion process with generator (1.10), we assume, in the first place, the forwards-in-time frequency in the population is governed by the deterministic process satisfying the differential equation (4.2), we can derive the corresponding ancestral dual process as in the following theorem.

Theorem 4.16 (Ancestral process dual to a deterministic process)**.** *Consider a population (under selection-mutation balance) in which the forwards-in-time frequency of type* 1 *individuals is governed by the deterministic process satisfying the differential equation* (4.2)*.*

Let $\{(\boldsymbol{r}(t), \boldsymbol{v}(t)), t \geq 0\}$ *be the ancestral process that is dual to the deterministic process. For a sample* $(n_1, n_2) \in \mathbb{Z}_+^2$ *of size* $n_1 + n_2 \geq 1$*, the ancestral dual process*

is Markov with state space $\mathbb{Z}_+^4 \setminus \{\mathbf{0}\}$ *and initial state* $(\boldsymbol{n}, \mathbf{0})$. *When the ancestral dual process is at a state* $(\boldsymbol{r}, \boldsymbol{v})$, *the transition rates are*

$$\begin{aligned} q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1)) &= \frac{1}{2}(r + v), \\ q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_1 + \boldsymbol{e}_2, \boldsymbol{v})) &= \frac{1 - \tau}{2\tau} r_1, \\ q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_1 + \boldsymbol{e}_2)) &= \frac{1 - \tau}{2\tau} v_1, \end{aligned}$$

where $r = |\boldsymbol{r}|$ *and* $v = |\boldsymbol{v}|$.

Proof. By equation (4.2), the infinitesimal generator of the deterministic process, evaluated on a continuously differentiable function $f\colon [0,1] \to \mathbb{R}$, is

$$\mathcal{L}f(x) = \frac{1}{2}(1 - x)\big(1 - \tau^{-1}x\big)\,\frac{\mathrm{d}f(x)}{\mathrm{d}x}\,.$$

By applying the deterministic generator $\mathcal{L}$ to a test function of the form $x^{\alpha_1}(1-x)^{\alpha_2}$, it follows that

$$\begin{aligned} \mathcal{L}x^{\alpha_1}(1-x)^{\alpha_2} &= \frac{1}{2}\alpha_1 x^{\alpha_1 - 1}(1-x)^{\alpha_2} - \frac{1}{2}\alpha_2 x^{\alpha_1}(1-x)^{\alpha_2 - 1} \\ &\quad + \frac{1}{2\tau}\alpha_2 x^{\alpha_1 + 1}(1-x)^{\alpha_2 - 1} + \frac{1}{2}\alpha_2 x^{\alpha_1}(1-x)^{\alpha_2 - 1}(1 - (1-x)) \\ &\quad - \frac{1}{2\tau}\alpha_1 x^{\alpha_1}(1-x)^{\alpha_2} - \frac{1}{2}\alpha_1 x^{\alpha_1 - 1}(1-x)^{\alpha_2}(1 - (1-x)) \\ &\quad + \frac{1}{2\tau}\alpha_1 x^{\alpha_1 + 1}(1-x)^{\alpha_2} - \frac{1}{2\tau}\alpha_2 x^{\alpha_1 + 1}(1-x)^{\alpha_2 - 1}(1 - (1-x)) \\ &= \frac{1}{2\tau}\alpha x^{\alpha_1 + 1}(1-x)^{\alpha_2} + \frac{1}{2}\alpha_1 \tau x^{\alpha_1 - 1}(1-x)^{\alpha_2 + 1} \\ &\quad - \frac{1}{2}(\alpha_2 + \tau^{-1}\alpha_1)x^{\alpha_1}(1-x)^{\alpha_2}, \end{aligned}$$

were $\alpha = \alpha_1 + \alpha_2$.

Let $x(t)$ be the solution to the differential equation (4.2). Equation (4.4) shows that $x(t) \to \tau$ as $t \to \infty$. Thus, $\tau^{\alpha_1}(1-\tau)^{\alpha_2}$ is the probability of sampling α_1 individuals of type 1 and α_2 individuals of type 2 from the stationary population governed by the deterministic process.

By applying the deterministic generator $\mathcal{L}$ to the function

$$g_{\boldsymbol{\alpha}}(x) = \frac{x^{\alpha_1}(1-x)^{\alpha_2}}{\tau^{\alpha_1}(1-\tau)^{\alpha_2}}, \tag{4.34}$$

we have

$$\begin{aligned}\mathcal{L}g_{\boldsymbol{\alpha}}(x) &= \frac{1}{2}\alpha g_{\boldsymbol{\alpha}+\boldsymbol{e}_1}(x) + \frac{1-\tau}{2\tau}\alpha_1 g_{\boldsymbol{\alpha}-\boldsymbol{e}_1+\boldsymbol{e}_2}(x) - \frac{1}{2}(\alpha_2 + \tau^{-1}\alpha_1)g_{\boldsymbol{\alpha}}(x) \\ &= \frac{1}{2}\alpha[g_{\boldsymbol{\alpha}+\boldsymbol{e}_1}(x) - g_{\boldsymbol{\alpha}}(x)] + \frac{1-\tau}{2\tau}\alpha_1[g_{\boldsymbol{\alpha}-\boldsymbol{e}_1+\boldsymbol{e}_2}(x) - g_{\boldsymbol{\alpha}}(x)].\end{aligned}$$

By using the method of duality in Definition 1.8, and considering $g_{\cdot}(\cdot)$ as functions of the type configuration $\boldsymbol{\alpha}$ of ancestral lines taken from the stationary population, rather than as functions of the frequency x, these terms before $g_{\boldsymbol{\beta}}(x) - g_{\boldsymbol{\alpha}}(x)$ in the above equation can be interpreted as the transition rates of a Markov process. The transition rates of the ancestral dual process are

$$\begin{aligned}q(\boldsymbol{\alpha}, \boldsymbol{\alpha}+\boldsymbol{e}_1) &= \frac{1}{2}\alpha, \\ q(\boldsymbol{\alpha}, \boldsymbol{\alpha}-\boldsymbol{e}_1+\boldsymbol{e}_2) &= \frac{1-\tau}{2\tau}\alpha_1.\end{aligned}$$

The ancestral lines can be decomposed into real and virtual ones by an argument similar to that in the proof of Theorem 2.1. Hence, the transition rates of the ancestral dual process decomposed into real and virtual lines, are

$$\begin{aligned}q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) &= \frac{1}{2}(r+v), \\ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1+\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1-\tau}{2\tau}r_1, \\ q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1+\boldsymbol{e}_2)) &= \frac{1-\tau}{2\tau}v_1.\end{aligned}$$

□

Recall that one unit of time on the timescale of the deterministic process equals θ_1 units of time on the diffusion timescale (see the derivation of equation (4.2)). Thus the transition rates of the ancestral dual process of Theorem 4.16 on the diffusion

timescale are

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) &= \frac{1}{2}(r+v)\theta_1,\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1+\boldsymbol{e}_2,\boldsymbol{v})) &= \frac{1-\tau}{2\tau}r_1\theta_1,\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_1+\boldsymbol{e}_2)) &= \frac{1-\tau}{2\tau}v_1\theta_1.
\end{aligned}
\tag{4.35}
$$

Remark 4.17 (Connections to the forwards-in-time process)**.** The transition rates (4.35) of the ancestral process that is dual to a forwards-in-time deterministic process are exactly the higher-order terms in the transition rates (4.17) obtained from asymptotic expansions in θ_1.

Chapter 5

Selection-mutation balance in an additive selection model

In this chapter, we consider an additive selection model, which is a special case of assortative mating models that have been proposed by numerous authors to study speciation (see Crow and Felsenstein, 1968; Kondrashov and Shpak, 1998; Devaux and Lande, 2008, and references therein). We consider a population of fixed size, where each individual is composed of L diallelic loci. Let the two alleles on each locus be labelled 0 and 1. (The types are labelled 0 and 1 in this chapter, instead of 1 and 2 as in the previous chapters, because the type of an individual in our additive selection model is then the sum of the type of loci.) We assume that the reproductive ability of an individual is determined by the number of type 1 loci. There are $L + 1$ different types in the population. The frequencies of individuals in a large population can be approximated by a diffusion process, provided that time is rescaled appropriately. Etheridge and Lemaire (2011) present a theoretical treatment of the diffusion approximation for multi-locus models with assortative mating, recombination and mutation.

However, the particular problem we are concerned with in this chapter is the

limiting behaviour of the multi-dimensional ancestral dual process with the additive selection scheme under selection-mutation balance. We first identify the ancestral process with typed lines that is dual to the forwards-in-time diffusion. We then adopt the methods in the previous chapters to analyse the limit of the genealogies embedded in the ancestral dual process, under the assumption that the mutation coefficient $\theta \to \infty$ and the selection coefficient $\sigma \to \infty$ such that

$$\lim_{\theta\to\infty} \theta/\sigma = \tau, \tag{5.1}$$

where $0 < \tau < \infty$ is fixed. The ancestral dual process and its limit in such an additive selection model have never been investigated before.

The remaining part of this chapter is organised as follows. In Section 5.1, we introduce a multi-dimensional variant of the Moran model with additive selection and mutation. The diffusion approximation of the Moran model is derived in Proposition 5.1. However, the explicit form of the unique stationary distribution of the L-locus diffusion is unknown. In Section 5.2, we deduce that the diffusion process can be approximated as $\theta \to \infty$ under selection-mutation balance by a process satisfying a system of differential equations (5.6). In Section 5.3, we are then able to asymptotically expand the transition rates of the dual process in terms of the stable equilibrium point of the deterministic system. The asymptotic expansions are in Lemma 5.10. We then concentrate on the limiting behaviour of the ancestral dual process with typed lines under selection-mutation balance. In Section 5.4.1, we consider the one-locus case ($L = 1$) for simplicity. The evolution of the virtual lines on a faster timescale is shown in Theorem 5.12, which is supported by simulations in Figure 5.3; while the real genealogy follows a time-changed Kingman's coalescent as proved in Theorem 5.14. In Section 5.4.2, we analyse the ancestral dual process in the L-locus case. The main results are in Theorems 5.17 and 5.19. In the final section, Section 5.5, we extend the studies to the case where $L \to \infty$. The number of type 0

loci in an individual drawn from the stationary population is proved to be finite in the limit as $L \to \infty$ in Lemma 5.20. The corresponding stationary frequency of type j individuals, $j \in \mathbb{N}_0 := \{0, 1, 2 \dots\}$, is Poisson with parameters j and $\tau\mu_0$, where μ_0 is the probability that given a mutation occurs on a locus, the mutation changes the locus to type 0. Backwards in time, types in a sample are lost one by one on a faster timescale, and the sample is instantaneously converted to the one in which individuals are not composed of any type 0 locus. The real genealogy then follows Kingman's coalescent with coalescence rate $\binom{r}{2}\exp(-\tau\mu_0)$. These results provide new insight into the studies of background selection model. The genealogies in the background selection model have been studied by various authors; see, e.g, Charlesworth et al. (1993, 1995), Hudson and Kaplan (1994, 1996) and Durrett (2008). Those authors investigate various properties of the genealogy under selection-mutation balance, e.g., the nucleotide diversity and the average time to lose a sampled type. We provide a full description of the genealogies under the constant frequency assumption; see Remarks 5.21 and 5.23 for detailed discussions.

5.1 Moran model and diffusion approximation

This section is devoted to the presentation of a diallelic multi-locus Moran model with additive selection and mutation. The diffusion approximation of the Moran model in a large population is deduced in Proposition 5.1.

In the Moran model, we consider a haploid and panmictic population of fixed size N. The genotype of each individual is composed of a finite number, $L > 0$, of loci, with two alleles per locus, denoted 0 and 1. Hence, an individual can be represented by a L-tuple $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_L) \in \{0, 1\}^L$. The type of an individual is determined by the number of type 1 loci, i.e., the type of an individual $\boldsymbol{\xi}$ is $|\boldsymbol{\xi}|$. Let $E = \{0, 1, \dots, L\}$ be the finite set of possible types.

In each reproduction event, an individual of type j, $j \in E$ gives birth at rate $\lambda_j \geq 0$; and at the same time, an individual is chosen at random to die. Each locus is assumed to have an identical influence on the reproductive ability of an individual. We suppose that each type 0 locus *decreases* the reproductive ability by s, i.e.,

$$\lambda_j = 1 - (L - j)s, \quad j \in E, \tag{5.2}$$

where $0 \leq s \leq 1/L$.

Mutations are assumed to occur continuously in time at rate $u \geq 0$ on each individual, independently of the reproduction process. At each locus, the allele has a probability μ_0 of mutating (independently of its current type) to type 0, and a probability μ_1 to type 1; $\mu_0 + \mu_1 = 1$. Given a mutation occurs on an individual of type $i \in E$, the probability that the mutation changes this individual to type $j \in E$, is

$$p_{ij} = \begin{cases} \frac{L-i}{L}\mu_1, & \text{if } j = i+1, \\ \frac{i}{L}\mu_0, & \text{if } j = i-1, \\ \frac{L-i}{L}\mu_0 + \frac{i}{L}\mu_1, & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

It is clear that $(p_{ij})_{i,j\in E}$ is a transition matrix, since each $p_{ij} \geq 0$, and for any fixed i, $\sum_{j\in E} p_{ij} = 1$.

The population configuration at time $t \geq 0$ can be described by a vector $\boldsymbol{Z}(t) = (Z_j(t))_{j\in E}$, where $Z_j(t)$ is the number of type j individuals at time t. We have $|\boldsymbol{Z}(t)| = N$ for all $t \geq 0$. The transition rate of the Markov process $\{\boldsymbol{Z}(t), t \geq 0\}$ from $\boldsymbol{z}$ to $\boldsymbol{z} - \boldsymbol{e}_i + \boldsymbol{e}_j$, $i, j \in E$, $i \neq j$, is

$$q_N(\boldsymbol{z}, \boldsymbol{z} - \boldsymbol{e}_i + \boldsymbol{e}_j) = z_j\lambda_j\frac{z_i}{N} + z_i u p_{ij}.$$

To obtain a diffusion approximation for a large population, we speed up time by a factor of $N/2$, and correspondingly scale the coefficients in the model by N, i.e., we set $Ns = \sigma$ and $Nu = \theta$. Let $X_i^{(N)}(t) = Z_i(t)/N$ be the proportion of type i individuals

at time $t \geq 0$. The generator of the L-locus diffusion in the additive selection model in the following equation (5.4) is a special case of the general multidimensional diffusion that describes mutation and selection on a set of a finite number of alleles (see, e.g., Etheridge and Lemaire, 2011).

Proposition 5.1 (L-locus diffusion in the additive selection model)**.** *The process* $\{\boldsymbol{X}_N(Nt/2), t \geq 0\}$ *converges in distribution to a diffusion process* $\{\boldsymbol{X}(t), t \geq 0\}$*, as* $N \to \infty$*. The state space of the diffusion is the* L*-dimensional simplex*

$$\mathcal{S}_{L+1} \coloneqq \{\boldsymbol{x} \colon \boldsymbol{x} \in [0,1]^{L+1}, |\boldsymbol{x}| = 1\}.$$

The generator of the diffusion, evaluated on a twice continuously differentiable function $f \colon \mathcal{S}_{L+1} \to \mathbb{R}$*, is*

$$\mathcal{L}f(\boldsymbol{x}) = \frac{1}{2} \sum_{i,j \in E} a_{ij}(\boldsymbol{x}) \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} + \sum_{j \in E} b_j(\boldsymbol{x}) \frac{\partial f(\boldsymbol{x})}{\partial x_j}, \tag{5.4}$$

where

$$a_{ij}(\boldsymbol{x}) = \delta_{ij} x_i - x_i x_j$$

and

$$\begin{aligned} b_j(\boldsymbol{x}) = \frac{\theta}{2L} \big[& x_{j-1}(L-j+1)\mu_1 + x_{j+1}(j+1)\mu_0 - x_j((L-j)\mu_1 + j\mu_0) \big] \\ & + \frac{\sigma}{2} x_j \left(j - \sum_{k \in E} k x_k \right). \end{aligned}$$

We set $x_{-1} = x_{L+1} = 0$.

Proof. Let $\delta \boldsymbol{X}_N(t) = \boldsymbol{X}_N(t + \delta t) - \boldsymbol{X}_N(t)$ and $\boldsymbol{x} = \boldsymbol{z}/N$. The mean of $\delta X_j^{(N)}(t)$ satisfies

$$\begin{aligned} & \mathbb{E}\left[\delta X_j^{(N)}(t) \mid \boldsymbol{X}_N(t) = \boldsymbol{x}\right] \\ & \quad = \frac{1}{N} \sum_{k \in E} [q_N(\boldsymbol{z}, \boldsymbol{z} - \boldsymbol{e}_k + \boldsymbol{e}_j) - q_N(\boldsymbol{z}, \boldsymbol{z} - \boldsymbol{e}_j + \boldsymbol{e}_k)] \frac{N}{2} \delta t + o(\delta t) \\ & \quad = \frac{\theta}{2L} [x_{j-1}(L-j+1)\mu_1 + x_{j+1}(j+1)\mu_0 - x_j(L-j)\mu_1 - x_j j \mu_0] \delta t \end{aligned}$$

$$
\begin{aligned}
&+\sum_{k\in E}\left[x_jx_k\frac{N}{2}(1-(L-j)s)-x_jx_k\frac{N}{2}(1-(L-k)s)\right]\delta t+o(\delta t)\\
&=\frac{\theta}{2L}\big[x_{j-1}(L-j+1)\mu_1+x_{j+1}(j+1)\mu_0-x_j((L-j)\mu_1+j\mu_0)\big]\delta t\\
&\qquad+\frac{\sigma}{2}x_j\left(j-\sum_{k\in E}kx_k\right)\delta t+o(\delta t),
\end{aligned}
$$

where the probability of seeing more than one transition in a time interval of length δt, is $o(\delta t)$. The variance of $\delta X_j^{(N)}(t)$ satisfies

$$
\begin{aligned}
&\mathrm{Var}\left[\delta X_j^{(N)}(t)\mid \boldsymbol{X}_N(t)=\boldsymbol{x}\right]\\
&\quad=\frac{1}{N^2}\sum_{\substack{k\in E\\k\neq j}}[q_N(\boldsymbol{z},\boldsymbol{z}-\boldsymbol{e}_k+\boldsymbol{e}_j)+q_N(\boldsymbol{z},\boldsymbol{z}-\boldsymbol{e}_j+\boldsymbol{e}_k)]\frac{N}{2}\delta t+o(\delta t)\\
&\quad=\frac{\theta}{2NL}[x_{j-1}(L-j+1)\mu_1+x_{j+1}(j+1)\mu_0-x_j(L-j)\mu_1-x_jj\mu_0]\delta t\\
&\qquad+\frac{1}{2}\sum_{\substack{k\in E\\k\neq j}}[x_jx_k(1-(L-j)s)+x_jx_k(1-(L-k)s)]\delta t+o(\delta t)\\
&\quad=\left[x_j\sum_{\substack{k\in E\\k\neq j}}x_k+O(N^{-1})\right]\delta t+o(\delta t)\\
&\quad=[x_j(1-x_j)+O(N^{-1})]\delta t+o(\delta t).
\end{aligned}
$$

For the covariance, we have

$$
\begin{aligned}
&\mathrm{Cov}\left[\delta X_i^{(N)}(t),\delta X_{i+1}^{(N)}(t)\mid \boldsymbol{X}_N(t)=\boldsymbol{x}\right]\\
&\quad=-\frac{1}{N^2}[q_N(\boldsymbol{z},\boldsymbol{z}-\boldsymbol{e}_{i+1}+\boldsymbol{e}_i)+q_N(\boldsymbol{z},\boldsymbol{z}-\boldsymbol{e}_i+\boldsymbol{e}_{i+1})]\frac{N}{2}\delta t+o(\delta t)\\
&\quad=-\frac{1}{2}\Bigg[x_i(1-(L-i)s)x_{i+1}+x_{i+1}\frac{\theta}{N}(i+1)\mu_0\\
&\qquad\qquad+x_{i+1}(1-(L-i-1)s)x_i+x_i\frac{\theta}{N}(L-i)\mu_1\Bigg]\delta t+o(\delta t)\\
&\quad=-[x_ix_{i+1}+O(N^{-1})]\delta t+o(\delta t);
\end{aligned}
$$

and for $j\notin\{i-1,i,i+1\}$, we have

$$
\mathrm{Cov}\left[\delta X_i^{(N)}(t),\delta X_j^{(N)}(t)\mid \boldsymbol{X}^{(N)}(t)=\boldsymbol{x}\right]
$$

$$
\begin{aligned}
&= -\frac{1}{N^2}[q_N(\boldsymbol{z}, \boldsymbol{z} - \boldsymbol{e}_j + \boldsymbol{e}_i) + q_N(\boldsymbol{z}, \boldsymbol{z} - \boldsymbol{e}_i + \boldsymbol{e}_j)]\frac{N}{2}\delta t + o(\delta t) \\
&= -\frac{1}{2}[x_i(1 - (L-i)s)x_j + x_j(1 - (L-j)s)x_i]\delta t + o(\delta t) \\
&= -[x_i x_j + O(N^{-1})]\delta t + o(\delta t).
\end{aligned}
$$

Higher-order moments of $\delta \boldsymbol{X}(t)$ are $o(\delta t)$. Thus, as $N \to \infty$, we obtain the weak convergence to the diffusion with generator (5.4), according to Ethier and Kurtz (2005, Theorem 1.1 in Chapter 10). □

Example 5.2 (Diffusion in the one-locus case)**.** Consider the case of one locus with two alleles, i.e., $L = 1$. Let $x_0(t)$ be the frequency of type 0 individuals at time $t \geq 0$. For a twice continuously differentiable function $f\colon [0,1] \to \mathbb{R}$, it follows from Proposition 5.1 that the generator of the diffusion $\{x_0(t), t \geq 0\}$ on $[0,1]$, is

$$
\begin{aligned}
\mathcal{L}f(x_0) &= \frac{1}{2}x_0(1-x_0)\frac{\mathrm{d}^2 f(x_0)}{\mathrm{d}x_0^2} + \frac{1}{2}[\theta(\mu_0 x_1 - \mu_1 x_0) - \sigma x_0 x_1]\frac{\mathrm{d}f(x_0)}{\mathrm{d}x_0} \\
&= \frac{1}{2}x_0(1-x_0)\frac{\mathrm{d}^2 f(x_0)}{\mathrm{d}x_0^2} + \frac{1}{2}[-\sigma x_0(1-x_0) + \theta\mu_0 - \theta x_0]\frac{\mathrm{d}f(x_0)}{\mathrm{d}x_0}\,. \qquad (5.5)
\end{aligned}
$$

We recover the generator (1.10) of the diffusion in the two-type case with selection and mutation.

Remark 5.3 (Stationary distribution of the L-locus diffusion)**.** The L-locus diffusion specified by generator (5.4) with additive selection and recurrent mutation at each locus ($\theta\mu_0 > 0$ and $\theta\mu_1 > 0$), is ergodic (i.e., irreducible and aperiodic). The L-locus diffusion falls into a general class of diffusions with selection and mutation, of which the existence of a unique stationary distribution is proved by Ethier and Kurtz (1998). The unique stationary distribution $\pi(\boldsymbol{x})$ of the L-locus diffusion satisfies the Kolmogorov forward equation

$$
\frac{1}{2}\sum_{i,j\in E}\frac{\partial^2}{\partial x_i \partial x_j}[a_{ij}(\boldsymbol{x})\pi(\boldsymbol{x})] - \sum_{j\in E}\frac{\partial}{\partial x_j}[b_j(\boldsymbol{x})\pi(\boldsymbol{x})] = 0,
$$

where $a_{ij}(\boldsymbol{x})$ and $b_j(\boldsymbol{x})$ are the covariance and drift of the diffusion, respectively.

However, the explicit form of the stationary distribution $\pi(\boldsymbol{x})$ is unknown.

Recall that in the one-locus case in Chapter 4, we see in Remark 4.9 that the transition rates of the ancestral dual process with typed lines depend on the sampling probabilities $\mathscr{H}(\cdot,\cdot)$, which in turn depend on the stationary density $\pi(x)$ of the forwards-in-time diffusion process. Let X be a random variable with the density $\pi(x)$. We have shown that, for a configuration $(\boldsymbol{r},\boldsymbol{v})$ of ancestral lines,

$$\mathscr{H}(\boldsymbol{r},\boldsymbol{v}) \coloneqq \mathbb{E}[X^{r_1+v_1}(1-X)^{r_2+v_2}] \sim \tau^{r_1+v_1}(1-\tau)^{r_2+v_2},$$

as $\theta_1 \to \infty$ under selection-mutation balance, where τ is the stable equilibrium point of the deterministic process that approximates the diffusion. In other words, by letting $t \to \infty$, we obtain the stationary distribution of the diffusion process. Then, by letting $\theta_1 \to \infty$, the sampling probability $\mathscr{H}(\boldsymbol{r},\boldsymbol{v})$ can be approximated by $\tau^{r_1+v_1}(1-\tau)^{r_2+v_2}$. We can take another route. By letting $\theta_1 \to \infty$, the diffusion process can be approximated by a deterministic process. Then, by letting $t \to \infty$, the stable equilibrium point τ of the deterministic process can be used to approximate the sampling probability.

In the following section, the analogy of this approximation is applied to the additive selection model, since the explicit form of the stationary distribution of the L-locus diffusion is unknown. The L-locus diffusion can be approximated by a deterministic process $\{\boldsymbol{x}(t), t \geq 0\}$ as $\theta \to \infty$ under selection-mutation balance. Then, by letting $t \to \infty$, we obtain the stable equilibrium point $\boldsymbol{x}^* = \lim_{t\to\infty} \boldsymbol{x}(t)$.

5.2 Deterministic trajectories

As $\theta \to \infty$ under selection-mutation balance (5.1), the path of the L-locus diffusion can be approximated by the trajectory of a system of differential equations as in the following equation (5.6).

Consider the diffusion process $\{\boldsymbol{X}(t), t \geq 0\}$ with the generator $\mathcal{L}_\theta$ given in equation (5.4) and initial value $\boldsymbol{X}(0) = \boldsymbol{\zeta} \in (0,1)^{L+1}$. For a twice continuously differentiable function $f\colon \mathcal{S}_{L+1} \to \mathbb{R}$, the generator of the time-changed process $\{\boldsymbol{X}(t/\theta), t \geq 0\}$ satisfies

$$\begin{aligned}\lim_{\theta\to\infty} \frac{1}{\theta}\mathcal{L}_\theta f(\boldsymbol{x}) = \lim_{\theta\to\infty} \Bigg\{ \sum_{j\in E} \Bigg(& \frac{1}{2L}\big[x_{j-1}(L-j+1)\mu_1 + x_{j+1}(j+1)\mu_0 - x_j((L-j)\mu_1 + j\mu_0)\big] \\ & + \frac{\sigma}{2\theta} x_j \Bigg(j - \sum_{k\in E} k x_k \Bigg) \Bigg) \frac{\partial f(\boldsymbol{x})}{\partial x_j} \\ & + \frac{1}{2\theta} \sum_{i,j\in E} (\delta_{ij} x_i - x_i x_j) \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j} \Bigg\} \\ = \sum_{j\in E} \Bigg\{ & \frac{1}{2L}\big[x_{j-1}(L-j+1)\mu_1 + x_{j+1}(j+1)\mu_0 - x_j((L-j)\mu_1 + j\mu_0)\big] \\ & + \frac{1}{2\tau} x_j \Bigg(j - \sum_{k\in E} k x_k \Bigg) \Bigg\} \frac{\partial f(\boldsymbol{x})}{\partial x_j}\,.\end{aligned}$$

Therefore, the time-changed diffusion process can be approximated by $\{\boldsymbol{x}(t), t \geq 0\}$ where $\boldsymbol{x}(t)$ is the solution to the following system of differential equations:

$$\begin{aligned}\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = & \frac{1}{2L}\big[x_{j-1}(t)(L-j+1)\mu_1 + x_{j+1}(t)(j+1)\mu_0 - x_j(t)((L-j)\mu_1 + j\mu_0)\big] \\ & + \frac{1}{2\tau} x_j(t) \Bigg(j - \sum_{k\in E} k x_k(t) \Bigg), \quad j \in E, \end{aligned} \tag{5.6}$$

with initial value $\boldsymbol{x}(0) = \boldsymbol{\zeta}$.

5.2.1 One-locus case

In the one-locus case, the corresponding differential equation is

$$\frac{\mathrm{d}x_0(t)}{\mathrm{d}t} = \frac{1}{2}\left[\mu_0 - x_0(t) - \frac{1}{\tau} x_0(t)(1 - x_0(t))\right] \tag{5.7}$$

with some initial value $x_0(0) = \zeta \in (0,1)$.

Note that equation (5.7) is different from the differential equation (4.2) deduced in Chapter 4. This is because in Chapter 4, we assume that the selection coefficient,

and the mutation coefficient for mutating to type 1, are large and balanced (see equation (4.1)). But in the present chapter, we assume that the selection coefficient, and the mutation coefficient for mutating to *both* types, are large and comparable (see equation (5.1)). The following stable equilibrium point is also different from that in equation (4.4) in Chapter 4. However, by setting $\mu_0 = 1$, we recover the results under the selection-mutation balance discussed in Chapter 4.

Lemma 5.4 (Stable equilibrium in the one-locus case)**.** *In the one-locus case, let* $x_0(t)$ *be the solution to the differential equation* (5.7) *with initial value* $x_0(0) = \zeta \in (0, 1)$. *Let* $\gamma = \sqrt{(1+\tau)^2 - 4\tau\mu_0}$. *We have*

$$x_0(t) = \frac{1}{2}\left[1 + \tau - \gamma + \frac{2\gamma(1 + \tau - \gamma - 2\zeta)}{1 + \tau - \gamma - 2\zeta - (1 + \tau + \gamma - 2\zeta)\exp(\gamma t/(2\tau))}\right]. \tag{5.8}$$

Let $x_0^* \coloneqq \lim_{t\to\infty} x_0(t)$. *Then,*

$$x_0^* = \frac{1}{2}(1 + \tau - \gamma) = \frac{1}{2}\Big(1 + \tau - \sqrt{(1+\tau)^2 - 4\tau\mu_0}\Big). \tag{5.9}$$

Furthermore, $0 < x_0^* < \min\{1, \tau\}$, *for* $\tau > 0$.

Proof. Let $\gamma = \sqrt{(1+\tau)^2 - 4\tau\mu_0}$. The differential equation (5.7) can be rewritten as

$$\frac{\mathrm{d}x_0(t)}{\mathrm{d}t} = a(x_0(t) - b)(x_0(t) - c),$$

where $a = 1/(2\tau)$, $b = (1 + \tau + \gamma)/2$ and $c = (1 + \tau - \gamma)/2$. The solution is then

$$\begin{aligned} x_0(t) &= c + \frac{(b-c)(c-\zeta)}{c - \zeta - (b-\zeta)e^{a(b-c)t}} \\ &= \frac{1}{2}\left[1 + \tau - \gamma + \frac{2\gamma(1 + \tau - \gamma - 2\zeta)}{1 + \tau - \gamma - 2\zeta - (1 + \tau + \gamma - 2\zeta)\exp(\gamma t/(2\tau))}\right]. \end{aligned}$$

The stable equilibrium point in equation (5.9) follows immediately.

Since $0 < \mu_0 < 1$ and $\tau > 0$, we have

$$|1 - \tau| < \gamma < 1 + \tau.$$

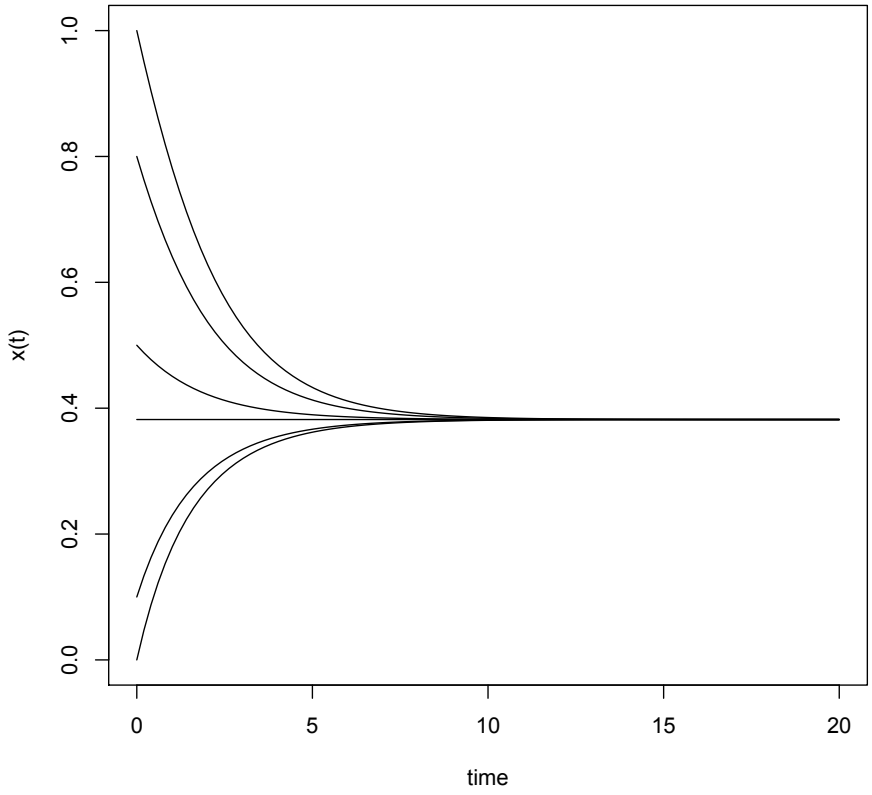


Figure 5.1: The solution to the differential equation (5.7) in the one-locus case. The coefficient values are $\tau = 2$ and $\mu_0 = 0.5$, and the initial value ζ is in $\{0, 0.1, 0.38, 0.5, 0.8, 1\}$. The stable equilibrium point is $x_0^* \approx 0.38$, which is calculated from equation (5.9).

It follows that

$$\frac{1}{2}(1 + \tau - |1 - \tau|) = \begin{cases} \tau, & \text{if } 0 < \tau < 1, \\ 1, & \text{if } \tau \geq 1. \end{cases}$$

Thus, we have $0 < x_0^* < \min\{1, \tau\}$, for $\tau > 0$. □

Figure 5.1 plots the solution $x_0(t)$ to the differential equation (5.7), and shows that as $t \to \infty$, $x_0(t)$ converges to the stable equilibrium point x_0^* given in equation (5.9).

Remark 5.5. Consider $x_0(t)$ which satisfies equation (5.7) and $x_1(t) = 1 - x_0(t)$ for all t. Let $\boldsymbol{x}(t) = (x_0(t), x_1(t))^T$, where we write $\boldsymbol{A}^T$ for the transpose of a vector (or

matrix) $\boldsymbol{A}$. We have

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{U}\boldsymbol{x}(t) + \big(\boldsymbol{v}^T\boldsymbol{x}(t)\big)\boldsymbol{x}(t), \tag{5.10}$$

where $\boldsymbol{v} = (0, -1/(2\tau))^T$ and

$$\boldsymbol{U} = \frac{1}{2}\begin{pmatrix} -\mu_1 & \mu_0 \\ \mu_1 & -\mu_0 \end{pmatrix} + \frac{1}{2\tau}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.11}$$

with $\mu_1 = 1 - \mu_0$. The matrix $\boldsymbol{U}$ is the sum of a matrix for mutation and a matrix representing selection differences. The solution and stable equilibrium of such a system of differential equations in population genetics have been known in the literature (see, e.g., O'Brien, 1985; Baake and Wagner, 2001). The idea is that we can use Thompson's trick (Thompson and McBride, 1974) by letting

$$\boldsymbol{y}(t) = \boldsymbol{x}(t)e^{-\int_0^t \boldsymbol{v}^T\boldsymbol{x}(u)\,\mathrm{d}u}$$

to turn equation (5.10) into the equation

$$\frac{\mathrm{d}\boldsymbol{y}(t)}{\mathrm{d}t} = \boldsymbol{U}\boldsymbol{y}(t) \tag{5.12}$$

with $\boldsymbol{y}(0) = \boldsymbol{x}(0)$. The general solution of the above equation is $\boldsymbol{y}(t) = \exp(\boldsymbol{U}t)\boldsymbol{y}(0)$. The solution for $\boldsymbol{x}(t)$ in equation (5.10) is provided by O'Brien (1985, Equation 21). Since the matrix $\boldsymbol{U}$ is primitive, the system will converge towards an equilibrium that is given by the Perron-Frobenius eigenvector, i.e., the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{U}$.

We define $\gamma = \sqrt{(1+\tau)^2 - 4\mu_0\tau}$. The eigenvalues of $\boldsymbol{U}$ are $(1 - \tau \pm \gamma)/(4\tau)$. The eigenvector corresponding to the largest eigenvalue $(1 - \tau + \gamma)/(4\tau)$ is

$$\boldsymbol{F}^* = \left(-\frac{1 - \mu_0\tau + \mu_1\tau - \gamma}{2\mu_1\tau}, 1 \right)^T.$$

Let $\|\boldsymbol{F}^*\|_1$ be the sum of all elements of $\boldsymbol{F}^*$. We then have

$$\lim_{t\to\infty} \boldsymbol{x}(t) = \frac{\boldsymbol{F}^*}{\|\boldsymbol{F}^*\|_1} = \frac{1}{2}\big(1 + \tau - \gamma, 1 - \tau + \gamma\big)^T,$$

which is the same as our result in Lemma 5.4.

5.2.2 L-locus case

In the L-locus case, the deterministic system (5.6) is in the form

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{F}(t, \boldsymbol{x}),$$

with initial value $\boldsymbol{x}(0) \in (0,1)^{L+1}$. Since $\boldsymbol{F}(t, \boldsymbol{x})$ is continuously differentiable with respect to $\boldsymbol{x}$ (all of the partial derivatives $\partial \boldsymbol{F}/\partial x_j$ exist and are continuous), there exists a unique solution to the initial value problem (5.6), according to Hale (1980, Theorem 3.1 in Chapter 1). The stable equilibrium point of the deterministic system (5.6) is deduced in the following theorem.

Theorem 5.6 (Stable equilibrium in the L-locus case)**.** *Let $\boldsymbol{x}(t)$ be the unique solution to the system* (5.6) *of differential equations with initial value $\boldsymbol{x}(0) \in (0,1)^{L+1}$. Define $\boldsymbol{x}^* = (x_j^*)_{j\in E}$ such that*

$$\boldsymbol{x}^* = \lim_{t\to\infty} \boldsymbol{x}(t).$$

We have

$$x_j^* = \binom{L}{j} \psi_{L,\tau,\mu_0}^j (1-\psi_{L,\tau,\mu_0})^{L-j}, \quad j \in E, \tag{5.13}$$

where

$$\psi_{L,\tau,\mu_0} = \frac{1}{2L}\Big(L - \tau + \sqrt{(L+\tau)^2 - 4L\tau\mu_0}\Big), \quad \tau > 0, \tag{5.14}$$

with $\max\{0, 1-\tau/L\} < \psi_{L,\tau,\mu_0} < 1$.

Remark 5.7. In our L-locus model, an individual is an unordered sample of L loci from the population. The constant ψ_{L,τ,μ_0} in equation (5.13) can be understood as the stationary frequency of type 1 loci.

Proof of Theorem 5.6. Since there exists a unique solution to the deterministic system

(5.6) with initial value $\boldsymbol{x}(0) \in (0,1)^{L+1}$, we propose a solution $(x_j(t))_{j\in E}$, where

$$x_j(t) = \binom{L}{j}\psi(t)^j(1-\psi(t))^{L-j}, \quad j \in E. \tag{5.15}$$

The aim is to find a differential equation for $\psi(t)$. For the brevity of notation, we write

$$x_{L,j}(t) \coloneqq \binom{L}{j}\psi(t)^j(1-\psi(t))^{L-j}.$$

By substituting equation (5.15) into the differential equation (5.6), the left-hand side of equation (5.6) becomes

$$\begin{aligned} LHS &= \left[j\binom{L}{j}\psi(t)^{j-1}(1-\psi(t))^{L-j} - (L-j)\binom{L}{j}\psi(t)^j(1-\psi(t))^{L-j-1}\right]\frac{\mathrm{d}\psi(t)}{\mathrm{d}t} \\ &= L[x_{L-1,j-1}(t) - x_{L-1,j}(t)]\,\frac{\mathrm{d}\psi(t)}{\mathrm{d}t}\,; \end{aligned}$$

and the right-hand side of equation (5.6) becomes

$$\begin{aligned} RHS = \frac{1}{2}\Bigg[&\frac{L}{\tau}\psi(t)(x_{L-1,j-1}(t) - x_{L,j}(t)) + (1-\psi(t))\mu_1 x_{L-1,j-1}(t) \\ &+ \psi(t)\mu_0 x_{L-1,j}(t) - (1-\psi(t))\mu_1 x_{L-1,j}(t) - \psi(t)\mu_0 x_{L-1,j-1}(t)\Bigg] \\ = \frac{1}{2}\Bigg[&\frac{L}{\tau}\psi(t)(1-\psi(t))(x_{L-1,j-1}(t) - x_{L-1,j}(t)) + (1-\psi(t))\mu_1 x_{L-1,j-1}(t) \\ &+ \psi(t)\mu_0 x_{L-1,j}(t) - (1-\psi(t))\mu_1 x_{L-1,j}(t) - \psi(t)\mu_0 x_{L-1,j-1}(t)\Bigg], \end{aligned}$$

where the second equality follows from the identity

$$x_{L,j}(t) = x_{L-1,j-1}(t)\psi(t) + x_{L-1,j}(t)(1-\psi(t)).$$

By dividing both sides of equation (5.6) by $[x_{L-1,j-1}(t) - x_{L-1,j}(t)]$, it follows that

$$\frac{\mathrm{d}\psi(t)}{\mathrm{d}t} = \frac{1}{2L}\left[\mu_1 - \psi(t) + \frac{L}{\tau}\psi(t)(1-\psi(t))\right], \tag{5.16}$$

with some initial value $\psi(0) = \psi_0$ satisfying

$$x_j(0) = \binom{L}{j}\psi_0^j(1-\psi_0)^{L-j}, \quad j \in E.$$

Let $\gamma = \sqrt{(L+\tau)^2 - 4L\tau\mu_0}$. The differential equation (5.16) can be rewritten as

$$\frac{\mathrm{d}\psi(t)}{\mathrm{d}t} = a(\psi(t) - b)(\psi(t) - c),$$

where $a = -1/(2\tau)$, $b = (L - \tau - \gamma)/(2L)$ and $c = (L - \tau + \gamma)/(2L)$. The solution is then

$$\psi(t) = c + \frac{(b-c)(c-\psi_0)}{c - \psi_0 - (b - \psi_0)e^{a(b-c)t}}.$$

Thus, we have

$$\psi^* \coloneqq \lim_{t\to\infty} \psi(t) = c = \frac{1}{2L}(L - \tau + \gamma).$$

Since $0 < \mu_0 < 1$, $\tau > 0$ and $L > 0$, we have

$$|L - \tau| < \gamma < L + \tau.$$

It follows that

$$\frac{1}{2L}(L - \tau + |L - \tau|) < \psi^* < 1.$$

Since

$$\frac{1}{2L}(L - \tau + |L - \tau|) = \begin{cases} 1 - \frac{\tau}{L}, & \text{if } 0 < \tau < L, \\ 0, & \text{if } \tau \geq L, \end{cases}$$

we have $\max\{0, 1 - \tau/L\} < \psi^* < 1$.

□

Example 5.8 (Deterministic approximation in the two-locus case)**.** We illustrate the deterministic approximation and Theorem 5.6 in the two-locus case. As $\theta \to \infty$ under selection-mutation balance (5.1), the path of the forwards-in-time diffusion can be approximated by the trajectory of the following system of differential equations:

$$\begin{aligned}
\frac{\mathrm{d}x_0(t)}{\mathrm{d}t} &= \frac{1}{2}\left[\frac{1}{\tau}x_0(t)(-x_1(t) - 2x_2(t)) - \mu_1 x_0(t) + \frac{1}{2}\mu_0 x_1(t)\right], \\
\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} &= \frac{1}{2}\left[\frac{1}{\tau}x_1(t)(1 - x_1(t) - 2x_2(t)) + \mu_1 x_0(t) - \frac{1}{2}x_1(t) + \mu_0 x_2(t)\right], \\
\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} &= \frac{1}{2}\left[\frac{1}{\tau}x_2(t)(2 - x_1(t) - 2x_2(t)) + \frac{1}{2}\mu_1 x_1(t) - \mu_0 x_2(t)\right].
\end{aligned} \tag{5.17}$$

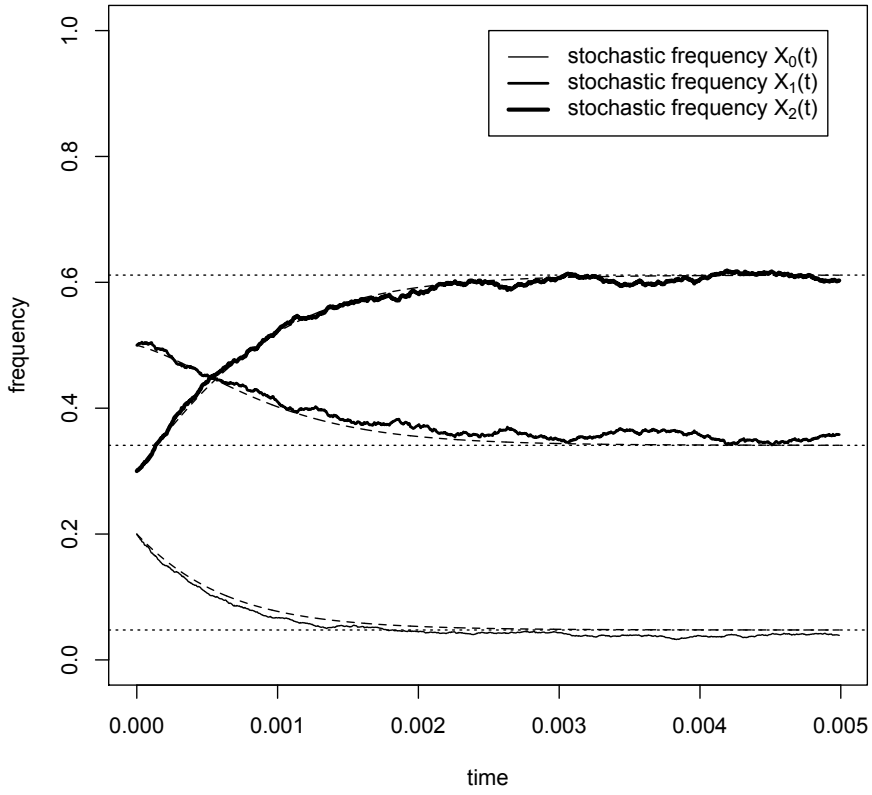


Figure 5.2: A path of the diffusion with generator (5.4) in the two-locus case, superimposed on the trajectory of the deterministic system (5.17). Time is measured on the diffusion timescale. In the simulation the diffusion is approximated by the multi-locus Moran model in Section 5.1 with population size $N = 15000$. The other coefficient values are $\theta = 2000$, $\tau = 0.5$ and $\mu_0 = 0.9$.

The stable equilibrium point calculated from equation (5.13) is (x_0^*, x_1^*, x_2^*), where

$$\begin{aligned} x_0^* &= (1-\psi_{2,\tau,\mu_0})^2 = \frac{1}{16}\Big(2+\tau-\sqrt{(2+\tau)^2-8\tau\mu_0}\Big)^2, \\ x_1^* &= 2\psi_{2,\tau,\mu_0}(1-\psi_{2,\tau,\mu_0}) = \frac{1}{4}\Big(4\tau\mu_0-\tau^2-2\tau+\tau\sqrt{(2+\tau)^2-8\tau\mu_0}\Big), \\ x_2^* &= \psi_{2,\tau,\mu_0}^2 = \frac{1}{16}\Big(2-\tau+\sqrt{(2+\tau)^2-8\tau\mu_0}\Big)^2. \end{aligned} \tag{5.18}$$

Figure 5.2 shows that the deterministic system (5.17) is a good approximation of the diffusion process, and the frequencies stabilise to the point (5.18) as $t \to \infty$.

Recall that in Remark 5.5, we discussed the calculations of stable equilibrium in

the one-locus case using the techniques of O'Brien (1985) and Baake and Wagner (2001). The techniques also work for the general L-locus case. However, in order to fit our additive selection model into those authors' formulation, we need to consider an ordered version where the type of an individual is represented by a sequence of zeros and ones instead of the sum of the number of ones.

We first consider the two-locus case ($L = 2$) to illustrate the ideas. The sequences of loci are ordered such that the possible types are $(0,0), (0,1), (1,0)$ and $(1,1)$. Recall equation (5.12). Let $\boldsymbol{y}(t)$ be the (scaled) frequencies of those types in the population. We still have

$$\frac{\mathrm{d}\boldsymbol{y}(t)}{\mathrm{d}t} = \boldsymbol{\mathcal{H}}\boldsymbol{y}(t).$$

But for ordered sequences,

$$\boldsymbol{\mathcal{H}} = \frac{1}{4}\begin{pmatrix} -2\mu_1 & \mu_0 & \mu_0 & 0 \\ \mu_1 & -1 & 0 & \mu_0 \\ \mu_1 & 0 & -1 & \mu_0 \\ 0 & \mu_1 & \mu_1 & -2\mu_0 \end{pmatrix} + \frac{1}{2\tau}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Let

$$\boldsymbol{U} = \frac{1}{4}\begin{pmatrix} -\mu_1 & \mu_0 \\ \mu_1 & -\mu_0 \end{pmatrix} + \frac{1}{2\tau}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.19}$$

be the one-site Hamiltonian. We then have

$$\boldsymbol{\mathcal{H}} = \sum_{i=1}^{2} \boldsymbol{H}_i,$$

where $\boldsymbol{H}_1 = \boldsymbol{U} \otimes \boldsymbol{I}$ and $\boldsymbol{H}_2 = \boldsymbol{I} \otimes \boldsymbol{U}$ with $\boldsymbol{I}$ being the identity matrix and $\otimes$ representing the tensor product. The eigenvalues of $\boldsymbol{U}$ are $\lambda_{1,2} = (2 - \tau \pm \gamma)/(8\tau)$ where $\gamma = \sqrt{(2+\tau)^2 - 8\mu_0\tau}$. We denote the corresponding eigenvectors by $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. The eigenvector $\boldsymbol{v}_1$ is

$$\boldsymbol{v}_1 = \left(-\frac{2 + (1 - 2\mu_0)\tau - \gamma}{2\mu_1\tau}, 1 \right)^T,$$

which implies

$$\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|_1} = \frac{1}{4}\Big(2+\tau-\gamma, 2-\tau+\gamma\Big)^T.$$

As argued by Baake and Wagner (2001, Section 6), the maximum eigenvalue for $\boldsymbol{\mathcal{H}}$ is then $2\lambda_1$ and the corresponding Perron-Frobenius eigenvector is

$$\begin{aligned}\boldsymbol{F}^* &= \boldsymbol{v}_1 \otimes \boldsymbol{v}_1 \\ &= \left(\frac{2+(1-2\mu_0)\tau-\gamma)^2}{4\mu_1^2\tau^2}, -\frac{2+(1-2\mu_0)\tau-\gamma}{2\mu_1\tau}, -\frac{2+(1-2\mu_0)\tau-\gamma}{2\mu_1\tau}, 1\right)^T.\end{aligned}$$

It then follows that the stable equilibrium point of the frequency of the ordered sequences is

$$\frac{\boldsymbol{F}^*}{\|\boldsymbol{F}^*\|_1} = \frac{1}{8}\left(\frac{1}{2}(2+\tau-\gamma)^2, 4\tau\mu_0-\tau^2-2\tau+\tau\gamma, 4\tau\mu_0-\tau^2-2\tau+\tau\gamma, \frac{1}{2}(2-\tau+\gamma)^2\right)^T.$$

Since type 0 of the unordered sequences corresponds to type $(0,0)$ of the ordered sequences, type 1 corresponds to types $(0,1)$ and $(1,0)$, and type 2 corresponds to type $(1,1)$, we recover the values in equations (5.18).

For a general L, the one-site Hamiltonian becomes

$$\boldsymbol{U} = \frac{1}{2L}\begin{pmatrix}-\mu_1 & \mu_0 \\ \mu_1 & -\mu_0\end{pmatrix} + \frac{1}{2\tau}\begin{pmatrix}0 & 0 \\ 0 & 1\end{pmatrix}. \tag{5.20}$$

The mutation-selection operator $\boldsymbol{\mathcal{H}}$ becomes $\boldsymbol{\mathcal{H}} = \sum_{i=1}^{L} \boldsymbol{H}_i$ where

$$\boldsymbol{H}_i = \boldsymbol{I} \otimes \cdots \otimes \boldsymbol{I} \otimes \boldsymbol{U} \otimes \boldsymbol{I} \cdots \otimes \boldsymbol{I}$$

with $\boldsymbol{U}$ in the ith place. The eigenvalues of $\boldsymbol{U}$ are $\lambda_{1,2} = (L-\tau \pm \gamma_L)/(4L\tau)$ where $\gamma_L = \sqrt{(L+\tau)^2 - 4L\tau\mu_0}$; and the eigenvector corresponding to λ_1 is

$$\boldsymbol{v}_1 = \left(-\frac{L+(1-2\mu_0)\tau-\gamma}{2\mu_1\tau}, 1\right)^T,$$

which implies

$$\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|_1} = \frac{1}{2L}\Big(L+\tau-\gamma_L, L-\tau+\gamma_L\Big)^T.$$

The first and second elements in the above vector are indeed our ψ_{L,τ,μ_0} and $1-\psi_{L,\tau,\mu_0}$ in equation (5.14). Following a similar argument, the Perron-Frobenius eigenvector of $\boldsymbol{\mathcal{H}}$ is $\boldsymbol{F}^* = \otimes^L \boldsymbol{v}_1$. It is then not difficult to see that the stable equilibrium of the ordered sequences must satisfy the binomial formula derived in equation (5.13).

5.3 Ancestral dual process with typed lines

We now focus on the ancestral dual process with typed lines in the additive selection model under selection-mutation balance. The transition rates of the multi-dimensional ancestral dual process will be given in the following equations (5.21). The asymptotic form of these transition rates will be computed in Lemma 5.10.

The L-locus diffusion specified by generator (5.4) fits the formulation of the multi-dimensional diffusion processes with selection and mutation of Sato (1978). Etheridge and Griffiths (2009, Equation 24) derive that the transition rates of the corresponding ancestral dual process with typed lines can be written in the form of the unconditional rate of an event multiplied by the sampling probability of ancestral lines given the event, then divided by the sampling probability before the event.

Consider the ancestral process $\{(\boldsymbol{r}(t), \boldsymbol{v}(t)), t \geq 0\}$ that is dual to the L-locus diffusion with generator (5.4), where $\boldsymbol{r}(t)$ count the numbers of real lines and $\boldsymbol{v}(t)$ count the numbers of virtual lines at time t in the past. By a decomposition of the ancestral lines into real and virtual lines (cf. Theorem 2.1), the transition rates of the ancestral dual process $\{(\boldsymbol{r}(t), \boldsymbol{v}(t)), t \geq 0\}$ are

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_j)) &= \frac{1}{2}(r+v)(L-j)\sigma\frac{\mathscr{H}(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_j)}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_j,\boldsymbol{v})) &= \frac{1}{2}r_j(r_j-1)\frac{\mathscr{H}(\boldsymbol{r}-\boldsymbol{e}_j,\boldsymbol{v})}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_j)) &= \frac{1}{2}v_j(2r_j+v_j-1)\frac{\mathscr{H}(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_j)}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{v})) &= \frac{1}{2L}r_j(j+1)\theta\mu_0\frac{\mathscr{H}(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{v})}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},
\end{aligned}
$$

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{v})) &= \frac{1}{2L} r_j (L-j+1)\theta\mu_1 \frac{\mathscr{H}(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{v})}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j)) &= \frac{1}{2L} v_j (j+1)\theta\mu_0 \frac{\mathscr{H}(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j)}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j)) &= \frac{1}{2L} v_j (L-j+1)\theta\mu_1 \frac{\mathscr{H}(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j)}{\mathscr{H}(\boldsymbol{r},\boldsymbol{v})},
\end{aligned}
\tag{5.21}
$$

where $j \in E$, $r = |\boldsymbol{r}|$, $v = |\boldsymbol{v}|$ and $\mathscr{H}(\cdot,\cdot)$ are the sampling probabilities defined in the following equation (5.22). Note that we abuse the notation by letting the indices of the vectors $\boldsymbol{r}$ and $\boldsymbol{v}$ start from 0. So, for example, $\boldsymbol{r}-\boldsymbol{e}_0 = (r_0, r_1, \ldots, r_L) - (1, 0, \ldots, 0) = (r_0 - 1, r_1, \ldots, r_L)$.

Let X_j be the stationary frequency of type j individuals in the population governed by the L-locus diffusion, and $\pi(\boldsymbol{x})$ be the density of $(X_j)_{j\in E}$. By analogy to the sampling probability (2.5) in the one-locus case, the probability of obtaining an ordered type configuration associated with the decomposition $(\boldsymbol{r},\boldsymbol{v})$ in the ancestral dual process $\{(\boldsymbol{r}(t),\boldsymbol{v}(t)), t \geq 0\}$ at any fixed time, is

$$
\begin{aligned}
\mathscr{H}(\boldsymbol{r},\boldsymbol{v}) &= \mathbb{E}\left[\prod_{j\in E} X_j^{r_j+v_j}\right]\\
&= \int_{\boldsymbol{x}\in\mathcal{S}_{L+1}} x_0^{r_0+v_0} x_1^{r_1+v_1}\cdots x_L^{r_L+v_L}\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.
\end{aligned}
\tag{5.22}
$$

The sampling probability is stationary; hence it is independent of time.

The stationary density $\pi(\boldsymbol{x})$ of the L-locus diffusion is required in order to compute the sampling probabilities in the transition rates (5.21). However, the explicit form of $\pi(\boldsymbol{x})$ is unknown, as mentioned in Remark 5.3. Nevertheless, as the path of the L-locus diffusion can be approximated by the trajectory of the deterministic system (5.6), the sampling probabilities can be approximated in terms of the stable equilibrium point (5.13) of the deterministic system. We have the following lemma for the sampling probabilities in the L-locus ancestral dual process.

Lemma 5.9 (Sampling probability under selection-mutation balance)**.** *The sampling*

probability defined in equation (5.22) *satisfies*

$$\mathscr{H}(\boldsymbol{r},\boldsymbol{v}) \sim \prod_{j\in E} \left(x_j^*\right)^{r_j+v_j}, \tag{5.23}$$

as $\theta \to \infty$ *under selection-mutation balance* (5.1)*, where* $(x_j^*)_{j\in E}$ *is the stable equilibrium point in equation* (5.13) *of the deterministic system* (5.6)*.*

Given Lemma 5.9, we can write down the asymptotic expansions of the transition rates (5.21) of the L-locus ancestral dual process under selection-mutation balance in the following lemma.

Lemma 5.10. *Consider the transition rates* (5.21) *of the* L*-locus ancestral dual process. As* $\theta \to \infty$ *under selection-mutation balance* (5.1)*, these transition rates satisfy*

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_j)) \sim \frac{x_j^*}{2\tau}(r+v)(L-j)\theta, \tag{5.24}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_j,\boldsymbol{v})) \sim \frac{1}{2x_j^*}r_j(r_j-1), \tag{5.25}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_j)) \sim \frac{1}{2x_j^*}v_j(2r_j+v_j-1), \tag{5.26}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{v})) \sim \frac{x_{j+1}^*}{2x_j^*}r_j\frac{j+1}{L}\theta\mu_0, \tag{5.27}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{v})) \sim \frac{x_{j-1}^*}{2x_j^*}r_j\frac{L-j+1}{L}\theta\mu_1, \tag{5.28}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j)) \sim \frac{x_{j+1}^*}{2x_j^*}v_j\frac{j+1}{L}\theta\mu_0, \tag{5.29}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j)) \sim \frac{x_{j-1}^*}{2x_j^*}v_j\frac{L-j+1}{L}\theta\mu_1, \tag{5.30}$$

where $j \in E$*,* $r = |\boldsymbol{r}|$*,* $v = |\boldsymbol{v}|$ *and* $(x_j^*)_{j\in E}$ *is the stable equilibrium point in equation* (5.13)*.*

Proof. This lemma follows immediately from substituting equation (5.23) into the transition rates (5.21) and some minor algebra. □

5.4 Genealogies under selection-mutation balance

In this section, we study the genealogies embedded in the ancestral dual process with transition rates (5.24) to (5.30) in the limit as $\theta \to \infty$ under selection-mutation balance. To illustrate the idea, the one-locus case where $L = 1$ is considered in the following section. The general L-locus case is presented in Section 5.4.2.

5.4.1 One-locus case

In the one-locus case, the differential equation that approximates the diffusion process with generator (5.5) is given in equation (5.7). The corresponding stable equilibrium point is derived in equation (5.9). According to Lemmas 5.9 and 5.10, we consider the ancestral dual process with the following transition rates:

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_0)) = \frac{x_0^*}{2\tau}(r + v)\theta, \tag{5.31}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_0, \boldsymbol{v})) = \frac{1}{2x_0^*} r_0(r_0 - 1), \tag{5.32}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_1, \boldsymbol{v})) = \frac{1}{2x_1^*} r_1(r_1 - 1), \tag{5.33}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_0)) = \frac{1}{2x_0^*} v_0(2r_0 + v_0 - 1), \tag{5.34}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_1)) = \frac{1}{2x_1^*} v_1(2r_1 + v_1 - 1), \tag{5.35}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_1 - \boldsymbol{e}_0, \boldsymbol{v})) = \frac{x_1^*}{2x_0^*} r_0 \theta \mu_0, \tag{5.36}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_0 - \boldsymbol{e}_1, \boldsymbol{v})) = \frac{x_0^*}{2x_1^*} r_1 \theta \mu_1, \tag{5.37}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1 - \boldsymbol{e}_0)) = \frac{x_1^*}{2x_0^*} v_0 \theta \mu_0, \tag{5.38}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_0 - \boldsymbol{e}_1)) = \frac{x_0^*}{2x_1^*} v_1 \theta \mu_1, \tag{5.39}$$

where $r = r_0 + r_1$, $v = v_0 + v_1$, x_0^* is the stable equilibrium point given in equation (5.9) and $x_1^* = 1 - x_0^*$.

The transition rates (5.31) to (5.39) are composed of terms of different orders;

thus, we can apply the idea of separation of timescales used in previous chapters to study the limiting behaviour of the ancestral dual process.

Remark 5.11. The transitions with rates (5.32), (5.33), (5.36) and (5.37) that change the configuration of real lines do *not* depend on the configuration of virtual lines. Thus, the real genealogy is independent of the evolution of the virtual lines. However, the evolution of the virtual lines *does* depend on the numbers of real lines.

The limiting behaviour of the real genealogy will be deduced in Theorem 5.14. Before proceeding to that, we will give a description of the behaviour of the virtual lines. We will see in Theorem 5.12 that, if time is rescaled, the evolution of the virtual lines is also independent of the real lines, and can be approximated by a deterministic process.

In the ancestral dual process for a finite sample $\boldsymbol{n}$, the numbers of real lines $\boldsymbol{r}(t)$ are always $O(1)$, in fact, $\mathbf{0} \leq \boldsymbol{r}(t) \leq \boldsymbol{n}$ for all t. When the numbers of virtual lines $\boldsymbol{v}(t)$ are also $O(1)$, the transition rates (5.31) to (5.39) are composed of terms of $O(\theta)$ and $O(1)$. The transitions with rates of $O(\theta)$ occur before those with rates of $O(1)$. When time is rescaled such that one unit of time on the faster timescale corresponds to θ units of time on the original timescale, the possible transitions and rates are

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_0)) = \frac{x_0^*}{2\tau}(r+v), \tag{5.40}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_0,\boldsymbol{v})) = \frac{x_1^*}{2x_0^*}r_0\mu_0, \tag{5.41}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_0-\boldsymbol{e}_1,\boldsymbol{v})) = \frac{x_0^*}{2x_1^*}r_1\mu_1, \tag{5.42}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_0)) = \frac{x_1^*}{2x_0^*}v_0\mu_0, \tag{5.43}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_0-\boldsymbol{e}_1)) = \frac{x_0^*}{2x_1^*}v_1\mu_1. \tag{5.44}$$

The mutations with rates (5.41) to (5.44) do not change the total number of real lines nor the total number of virtual lines. Thus, the branching rate (5.40) is not affected by these mutations. The total number of virtual lines follows a birth

process with birth rate (5.40). The mutations among the virtual lines, with transitions $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_1 - \boldsymbol{e}_0)$ and $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_0 - \boldsymbol{e}_1)$, occur at rates that are of the same order as those of the branching events. Thus, the virtual lines of both types will keep growing in such a process, and will eventually reach the level of $O(\theta)$. The coalescence rates (5.34) and (5.35) for virtual lines then become very large ($O(\theta^2)$), and these coalescence events take effect on the faster timescale. We will see in the following Theorem 5.12, the numbers of virtual lines will stabilise on the level of $O(\theta)$.

Consider the ancestral dual process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{v}_\theta(t)), t \geq 0\}$ with transition rates (5.31) to (5.39). Define $\boldsymbol{\kappa}_\theta(t) = \boldsymbol{v}_\theta(t)/\theta$. The process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{\kappa}_\theta(t)), t \geq 0\}$ has the state space $S = \{0, 1, \ldots, n_1\} \times \{0, 1, \ldots, n_2\} \times \mathbb{R}_+ \times \mathbb{R}_+$.

Theorem 5.12 (Virtual lines under selection-mutation balance in the one-locus case)**.** *Assume $\boldsymbol{\kappa}_\theta(0) \neq \mathbf{0}$. As $\theta \to \infty$, the process $\{(\boldsymbol{r}_\theta(t/\theta), \boldsymbol{\kappa}_\theta(t/\theta)), t \geq 0\}$ converges in finite-dimensional distribution to a process $\{(\boldsymbol{r}(t), \boldsymbol{\kappa}(t)), t \geq 0\}$ with state space S running on the faster timescale where one unit of time corresponds to θ units of time on the original timescale.*

Suppose $h(\boldsymbol{r}, \boldsymbol{\kappa})\colon S \to \mathbb{R}$ is continuously differentiable with respect to $\boldsymbol{\kappa}$. The process $\{(\boldsymbol{r}(t), \boldsymbol{\kappa}(t)), t \geq 0\}$ is specified by the generator

$$\begin{aligned}
\mathcal{L}h(\boldsymbol{r}, \boldsymbol{\kappa}) = {} & [h(\boldsymbol{r} - \boldsymbol{e}_0 + \boldsymbol{e}_1, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{x_1^*}{2x_0^*} r_0 \mu_0 \\
& + [h(\boldsymbol{r} + \boldsymbol{e}_0 - \boldsymbol{e}_1, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{x_0^*}{2x_1^*} r_1 \mu_1 \\
& + \frac{x_0^*}{2\tau}(\kappa_0 + \kappa_1) \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_0} - \frac{1}{2x_0^*}\kappa_0^2 \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_0} - \frac{1}{2x_1^*}\kappa_1^2 \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_1} \\
& + \frac{x_1^*}{2x_0^*}\kappa_0 \mu_0 \left[\frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_1} - \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_0} \right] \\
& + \frac{x_0^*}{2x_1^*}\kappa_1 \mu_1 \left[\frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_0} - \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_1} \right],
\end{aligned} \tag{5.45}$$

where x_0^* is the stable equilibrium point in equation (5.9)*, and $x_1^* = 1 - x_0^*$.*

Proof. The proof follows that same idea as that in the proof of Theorem 4.11. We

consider the process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{\kappa}_\theta(t)), t \geq 0\}$ that has the transition rates (5.31) to (5.39) and infinitesimal generator $\mathcal{L}_\theta$. The generator of the time-changed process $\{(\boldsymbol{r}_\theta(t/\theta), \boldsymbol{\kappa}_\theta(t/\theta)), t \geq 0\}$ is then $\theta^{-1}\mathcal{L}_\theta$. We apply the generator to a test function $h(\boldsymbol{r}, \boldsymbol{\kappa})$, and then take the limit of $\theta^{-1}\mathcal{L}_\theta h(\boldsymbol{r}, \boldsymbol{\kappa})$ as $\theta \to \infty$ to obtain equation (5.45).

□

Remark 5.13. Given the generator (5.45), on the faster timescale, the Markov jump process $\{\boldsymbol{r}(t), t \geq 0\}$ for the numbers of real lines, and the continuous process $\{\boldsymbol{\kappa}(t), t \geq 0\}$ for the scaled numbers of virtual lines, are independent.

In the dual process, the scaled numbers of virtual lines evolve backwards in time on the faster timescale, according to the following system of differential equations:

$$\begin{aligned}
\frac{\mathrm{d}\kappa_0(t)}{\mathrm{d}t} &= \frac{1}{2}\left(\frac{x_0^*}{\tau}(\kappa_0(t) + \kappa_1(t)) - \kappa_0^2(t)\frac{1}{x_0^*} - \kappa_0(t)\mu_0\frac{x_1^*}{x_0^*} + \kappa_1(t)\mu_1\frac{x_0^*}{x_1^*}\right), \\
\frac{\mathrm{d}\kappa_1(t)}{\mathrm{d}t} &= \frac{1}{2}\left(\kappa_0(t)\mu_0\frac{x_1^*}{x_0^*} - \kappa_1^2(t)\frac{1}{x_1^*} - \kappa_1(t)\mu_1\frac{x_0^*}{x_1^*}\right).
\end{aligned} \tag{5.46}$$

The possible equilibrium point $\boldsymbol{\kappa}^* = (\kappa_0^*, \kappa_1^*)$ of the system (5.46) satisfies the following system of equations:

$$\begin{aligned}
0 &= \frac{x_0^*}{\tau}(\kappa_0^* + \kappa_1^*) - (\kappa_0^*)^2\frac{1}{x_0^*} - \kappa_0^*\mu_0\frac{x_1^*}{x_0^*} + \kappa_1^*\mu_1\frac{x_0^*}{x_1^*} \\
0 &= \kappa_0^*\mu_0\frac{x_1^*}{x_0^*} - (\kappa_1^*)^2\frac{1}{x_1^*} - \kappa_1^*\mu_1\frac{x_0^*}{x_1^*},
\end{aligned}$$

which has two solutions, namely $(0, 0)$ and

$$\boldsymbol{\kappa}^* = \frac{1}{\tau}\big((x_0^*)^2, x_0^* x_1^*\big). \tag{5.47}$$

Given $0 < \mu_0 < 1$ and $\tau > 0$, we have $\max\{0, 1 - \tau\} < x_1^* < 1$. Thus $\boldsymbol{\kappa}^* > \mathbf{0}$. The Jacobian matrix of the system (5.46) is

$$J(\kappa_0(t), \kappa_1(t)) = \begin{pmatrix} \frac{x_0^*}{\tau} - \frac{2\kappa_0(t) + \mu_0 x_1^*}{x_0^*} & \frac{x_0^*}{\tau} + \mu_1\frac{x_0^*}{x_1^*} \\ \mu_0\frac{x_1^*}{x_0^*} & -\frac{2\kappa_1(t) + \mu_1 x_0^*}{x_1^*} \end{pmatrix}.$$

The eigenvalues of $J(0, 0)$ are $x_0^*/\tau > 0$ and $-(\mu_1 x_0^*/x_1^* + \mu_0 x_1^*/x_0^*) < 0$; and the

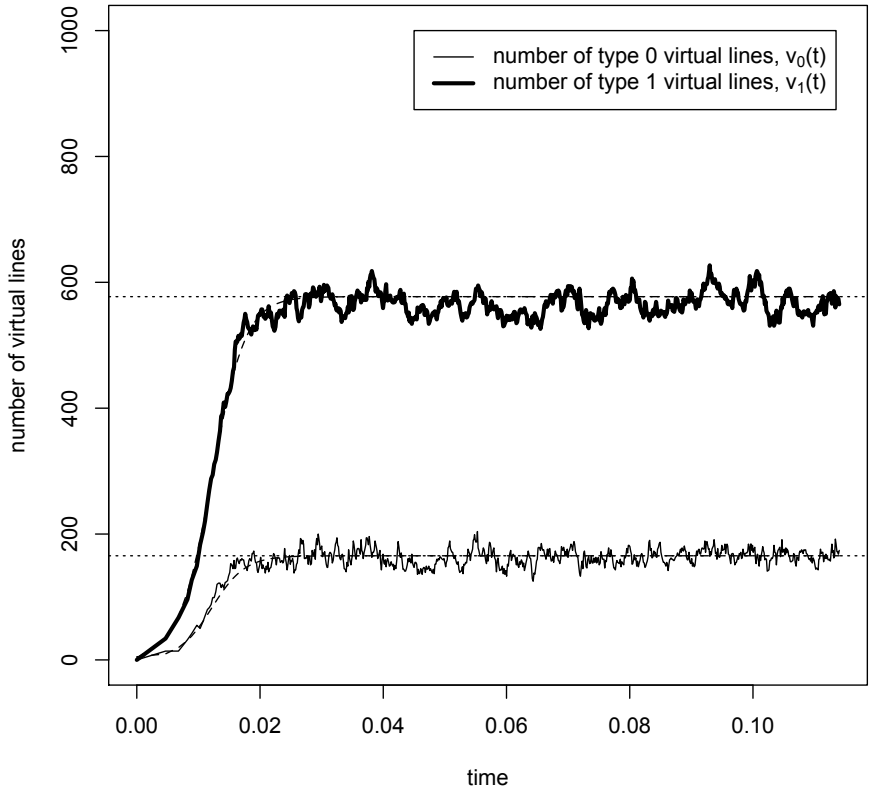


Figure 5.3: The numbers of virtual lines in the ancestral dual process with full transition rates in the one-locus case, superimposed on the trajectory of the deterministic system (5.46). Time is measured on the original timescale. The sample is $(5, 3)$, and the coefficient values are $\theta = 1000$, $\tau = 0.3$ and $\mu_0 = 0.8$. The horizontal dotted lines represent the stable equilibrium point $(\theta\kappa_0^*, \theta\kappa_1^*)$, where κ_0^* and κ_1^* are calculated from equation (5.47).

eigenvalues of $J(\kappa_0^*, \kappa_1^*)$ are $-x_0^*/\tau < 0$ and $-2x_0^*/\tau - \mu_1 x_0^*/x_1^* - \mu_0 x_1^*/x_0^* < 0$. It follows Hale (1980, Theorem 4.2 in Chapter 3) that the positive equilibrium point $\boldsymbol{\kappa}^*$ in equation (5.47) is locally asymptotically stable, while $(0, 0)$ is not even locally asymptotically stable.

Figure 5.3 plots a path of the numbers of virtual lines in the ancestral dual process with full transition rates in the one-locus case, superimposed on the trajectory of the deterministic system (5.46). The simulation provides additional support to our argument.

The total number of virtual lines will eventually become absorbed into one via the events occurring on the original timescale. (This follows because if we do not distinguish the types and the real and virtual lines and set the mutation coefficients to zero, the ancestral dual process is just a birth-death process with quadratic death rate and linear birth rate, see Etheridge and Griffiths, 2009, Equation 37). These events for virtual lines on the original timescale have been ignored in our approximation of the transition rates, as they are complicated to handle and our interests focus primarily on the real genealogy.

We have described the behaviour of the virtual lines in the limit. According to Remark 5.11, for the real genealogy, it is sufficient to consider the marginal process of real ancestral lines in which the transition rates are

$$q_\theta(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{e}_0) = \frac{1}{2x_0^*} r_0(r_0 - 1), \tag{5.48}$$

$$q_\theta(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{e}_1) = \frac{1}{2x_1^*} r_1(r_1 - 1), \tag{5.49}$$

$$q_\theta(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_1 - \boldsymbol{e}_0) = \frac{x_1^*}{2x_0^*} r_0 \theta \mu_0, \tag{5.50}$$

$$q_\theta(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_0 - \boldsymbol{e}_1) = \frac{x_0^*}{2x_1^*} r_1 \theta \mu_1, \tag{5.51}$$

where x_0^* is the stable equilibrium point in equation (5.9), and $x_1^* = 1 - x_0^*$.

The ancestral process with transition rates (5.48) to (5.51) in the limit as $\theta \to \infty$, is the same as the strong migration limit (when migration is caused by mutation) in the structured coalescent (see Nagylaki, 1980; Notohara, 1993; Nordborg and Krone, 2002). Nordborg and Krone (2002) use the idea of separation of timescales to study the strong migration limit of the ancestral dual process. In the process with transition rates (5.48) to (5.51), as $\theta \to \infty$, the mutation events with transition rates (5.50) and (5.51) of $O(\theta)$ occur much more quickly than the coalescence events with transition rates (5.48) and (5.49) of $O(1)$. The number of mutation events between the coalescence events tends to infinity. Thus, in the limit, the mutation events approach an equilibrium

on the faster timescale between the coalescence events, and the total number of real lines reduces by one when a coalescence event occurs on the original timescale. In the terminology of Nordborg and Krone (2002), the structure associated with types *collapses* to a one-type Kingman's coalescent where the coalescence rate is an average over the stationary distribution of the mutation process.

The following theorem shows that in the limit as $\theta \to \infty$ under selection-mutation balance, the real genealogy in the ancestral dual process follows Kingman's coalescent with the coalescence rate given in equation (5.52). The theorem is obtained by applying the theorem of Nordborg and Krone (2002) to the context of our strong mutation limit.

Let $\{(r_0^{(\theta)}(t), r_1^{(\theta)}(t)), t \geq 0\}$ be the ancestral process for a sample $(n_0, n_1) \in \mathbb{Z}_+^2$ of size $n = n_0 + n_1 \geq 1$, which has the transition rates (5.48) to (5.51). Define $r_\theta(t) = r_0^{(\theta)}(t) + r_1^{(\theta)}(t)$. The process $\{r_\theta(t), t \geq 0\}$ has state space $S = \{1, 2, \ldots, n\}$ and initial state n.

Theorem 5.14. *The process $\{r_\theta(t), t \geq 0\}$ converges in distribution, as $\theta \to \infty$, to a process $\{r(t), t \geq 0\}$. $\{r(t), t \geq 0\}$ is a death process with state space S, initial value n and death rate*

$$q(r, r-1) = \binom{r}{2}\left(\frac{(1-\phi)^2}{x_0^*} + \frac{\phi^2}{x_1^*}\right) \equiv \binom{r}{2}\frac{\mu_0^2\psi^3 + \mu_1^2(1-\psi)^3}{(\mu_0\psi^2 + \mu_1(1-\psi)^2)^2}, \quad r \geq 2, \quad (5.52)$$

where

$$\phi = \frac{\mu_0\psi^2}{\mu_0\psi^2 + \mu_1(1-\psi)^2},$$

$x_0^* = 1 - \psi$, $x_1^* = \psi$ *and*

$$\psi = \frac{1}{2}\Big(1 - \tau + \sqrt{(1+\tau)^2 - 4\tau\mu_0}\Big),$$

as given in equation (5.9).

Proof. Consider the ancestral process $\{(r_0^{(\theta)}(t), r_1^{(\theta)}(t)), t \geq 0\}$ with transition rates

(5.48) to (5.51). According to Nordborg and Krone (2002, Theorem 1), we first consider the events with transition rates (5.50) and (5.51) that occur on the faster timescale where one unit of time corresponds to θ units of time on the original timescale. For fixed r, $2 \leq r \leq n$, there exists a Markov process $\{(r_0^{(r)}(t), r_1^{(r)}(t)), t \geq 0\}$ (running on the faster timescale) that has the transition rates

$$q_r((r_0, r_1), (r_0 - 1, r_1 + 1)) = \frac{x_1^*}{2x_0^*} r_0 \mu_0,$$
$$q_r((r_0, r_1), (r_0 + 1, r_1 - 1)) = \frac{x_0^*}{2x_1^*} r_1 \mu_1.$$

The state space of the process $\{(r_0^{(r)}(t), r_1^{(r)}(t)), t \geq 0\}$ is $S_r = \{(r_0, r_1) \colon (r_0, r_1) \in \mathbb{Z}_+^2, r_0 + r_1 = r\}$. We call the process $\{(r_0^{(r)}(t), r_1^{(r)}(t)), t \geq 0\}$ the level-r mutation process.

The level-r mutation process $\{(r_0^{(r)}(t), r_1^{(r)}(t)), t \geq 0\}$, in which only mutation events occur, is ergodic so it has a unique stationary distribution. The stationary distribution is, for $(r_0, r_1) \in S_r$,

$$\pi_r(r_0, r_1) = \binom{r}{r_0} (1 - \phi)^{r_0} \phi^{r_1}, \tag{5.53}$$

where

$$\phi = \mu_0 \frac{x_1^*}{x_0^*} \left(\mu_0 \frac{x_1^*}{x_0^*} + \mu_1 \frac{x_0^*}{x_1^*} \right)^{-1}.$$

Then, by equation (5.9) for x_0^*, we can write

$$\phi = \frac{\mu_0 \psi^2}{\mu_0 \psi^2 + \mu_1 (1 - \psi)^2}, \tag{5.54}$$

where

$$\psi = \frac{1}{2} \Big(1 - \tau + \sqrt{(1 + \tau)^2 - 4\tau\mu_0} \Big).$$

We can verify that the stationary distribution (5.53) satisfies the detailed balance equations as follows. For $r_0 = 1, 2, \ldots, r$, we have

$$\pi_r(r_0, r_1) q_r((r_0, r_1), (r_0 - 1, r_1 + 1))$$

$$
\begin{aligned}
&= \binom{r}{r_0}(1-\phi)^{r_0}\phi^{r-r_0} \times \frac{x_1^*}{2x_0^*} r_0\theta\mu_0 \\
&= \binom{r}{r_0-1}(1-\phi)^{r_0-1}\phi^{r-r_0+1} \times \frac{1}{2}(r-r_0+1)\theta\mu_1\frac{x_0^*}{x_1^*} \\
&= \pi_r(r_0-1, r_1+1)q_r((r_0-1, r_1+1),(r_0, r_1));
\end{aligned}
$$

and for $r_0 = 0, 1, \ldots, r-1$, we have

$$
\begin{aligned}
&\pi_r(r_0, r_1)q_r((r_0, r_1),(r_0+1, r_1-1)) \\
&\quad= \binom{r}{r_0}(1-\phi)^{r_0}\phi^{r-r_0} \times \frac{x_0^*}{2x_1^*} r_1\theta\mu_1 \\
&\quad= \binom{r}{r_0+1}(1-\phi)^{r_0+1}\phi^{r-r_0-1} \times \frac{1}{2}(r_0+1)\theta\mu_0\frac{x_1^*}{x_0^*} \\
&\quad= \pi_r(r_0+1, r_1-1)q_r((r_0+1, r_1-1),(r_0, r_1)).
\end{aligned}
$$

Equation (5.53) is then proved.

In the ancestral process $\{(r_0^{(\theta)}(t), r_1^{(\theta)}(t)), t \geq 0\}$ with transition rates (5.48) to (5.51), the transitions with rates of $O(1)$ are

$$
q(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{e}_0) = \frac{1}{2x_0^*} r_0(r_0-1), \tag{5.55}
$$

$$
q(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{e}_1) = \frac{1}{2x_1^*} r_1(r_1-1). \tag{5.56}
$$

These transitions correspond to the coalescence events occurring on the original timescale.

Consider the process $\{r_\theta(t), t \geq 0\}$, where $r_\theta(t) = r_0^{(\theta)}(t) + r_1^{(\theta)}(t)$. According to Nordborg and Krone (2002, Theorem 1), as $\theta \to \infty$, we have the weak convergence of the process $\{r_\theta(t), t \geq 0\}$ to the death process of Kingman's coalescent with an effective coalescence rate $q(r, r-1)$. $q(r, r-1)$ can be computed as the coalescence rates (5.55) and (5.56) averaged over the stationary distribution (5.53) of the level-r mutation process. Therefore,

$$
q(r, r-1) = \sum_{r_0=0}^{r} \frac{1}{2x_0^*} r_0(r_0-1)\pi_r(r_0, r-r_0) + \sum_{r_1=0}^{r} \frac{1}{2x_1^*} r_1(r_1-1)\pi_r(r-r_1, r_1)
$$

$$
\begin{aligned}
&= \frac{(1-\phi)^2}{2x_0^*} r(r-1) + \frac{\phi^2}{2x_1^*} r(r-1) \\
&= \binom{r}{2} \left(\frac{(1-\phi)^2}{x_0^*} + \frac{\phi^2}{x_1^*} \right) \\
&= \binom{r}{2} \frac{\mu_0^2 \psi^3 + \mu_1^2 (1-\psi)^3}{(\mu_0 \psi^2 + \mu_1 (1-\psi)^2)^2}, \qquad (5.57)
\end{aligned}
$$

for $r \geq 2$, where the last equality follows from equations (5.9) and (5.54). The proof is then completed. □

Remark 5.15.

1. By equation (5.53), when there are a total of two lines in the level-2 mutation process, the probability that these two lines are both type 0 is $(1-\phi)^2$, and the probability that these two lines are both type 1 is ϕ^2. The terms ϕ^2 and $(1-\phi)^2$ appear in the formula for $q(r, r-1)$ in equation (5.52), because two ancestral lines must be of the same type (through mutations) in order for them to coalesce. The squared terms that have a similar interpretation also appear in the relevant formulae in the strong-migration models (see Nagylaki, 2000; Bahlo and Griffiths, 2001).

2. The selection-mutation balance we assume in this chapter is different from that in Chapter 4. However, by setting $\mu_0 = 1$, we recover the selection-mutation balance discussed in Chapter 4. When $\mu_0 = 1$, according to equation (5.9), the forwards-in-time frequencies will approach a stable equilibrium point $(x_0^*, x_1^*) = (\tau, 1-\tau)$, $0 < \tau < 1$. Moreover, as we have $\phi = 1$ when $\mu_1 = 1 - \mu_0 = 0$, by Theorem 5.14, the real genealogy under such a selection-mutation balance follows Kingman's coalescent with coalescence rate

$$
\binom{r}{2} \frac{1}{1-\tau}, \quad r \geq 2.
$$

We then recover the result derived in Theorem 4.13 in Chapter 4.

5.4.2 L-locus case

In the previous section, we have discussed the limiting behaviour of the ancestral dual process for a sample under selection-mutation balance (5.1) in the one-locus case. In this section, we generalise the investigation to the L-locus case. We will give a heuristic description first, and the formal results will be deduced in Theorems 5.17 and 5.19.

We consider the ancestral dual process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{v}_\theta(t)), t \geq 0\}$, where $r_j^{(\theta)}$, $j \in E = \{0, 1, \ldots, L\}$ denotes the number of real ancestral lines of type j, and $v_j^{(\theta)}$ denotes the number of virtual ancestral lines of type j. The process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{v}_\theta(t)), t \geq 0\}$ has the following transition rates, as derived in Lemma 5.10:

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_j)) = \frac{x_j^*}{2\tau}(r + v)(L - j)\theta, \tag{5.58}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_j, \boldsymbol{v})) = \frac{1}{2x_j^*} r_j(r_j - 1), \tag{5.59}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_j)) = \frac{1}{2x_j^*} v_j(2r_j + v_j - 1), \tag{5.60}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j, \boldsymbol{v})) = \frac{x_{j+1}^*}{2x_j^*} r_j \frac{j+1}{L}\theta\mu_0, \tag{5.61}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j, \boldsymbol{v})) = \frac{x_{j-1}^*}{2x_j^*} r_j \frac{L-j+1}{L}\theta\mu_1, \tag{5.62}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j)) = \frac{x_{j+1}^*}{2x_j^*} v_j \frac{j+1}{L}\theta\mu_0, \tag{5.63}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j)) = \frac{x_{j-1}^*}{2x_j^*} v_j \frac{L-j+1}{L}\theta\mu_1, \tag{5.64}$$

where $j \in E$, $r = |\boldsymbol{r}|$, $v = |\boldsymbol{v}|$ and $(x_j^*)_{j \in E}$ is the stable equilibrium point given in equation (5.13).

Remark 5.16. The transitions with rates (5.59), (5.61) and (5.62) that change the configuration of real lines do *not* depend on the configuration of virtual lines. The real genealogy is thus independent of the evolution of the virtual lines. But the evolution of the virtual lines *does* depend on the numbers of real lines. However, we will see in

Theorem 5.17 that, if we rescale the time, the virtual lines can be treated independently of the real lines.

The transition rates (5.58) to (5.64) are composed of terms of $O(1)$ and terms of a higher order. We separate the timescales according to the order of the terms in the transition rates. The transitions with rates of a higher order occur before the transitions with rates of $O(1)$. The numbers of real lines are always $O(1)$. At time that is sufficiently close to 0, the numbers of virtual lines are also $O(1)$. The transition rates (5.58) to (5.64) then consist of terms of $O(\theta)$ and terms of $O(1)$. On the faster timescale where one unit of time corresponds to θ units of time on the original timescale, the possible transitions and rates are

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_j)) = \frac{x_j^*}{2\tau}(r+v)(L-j), \tag{5.65}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{v})) = \frac{x_{j+1}^*}{2x_j^*}r_j\frac{j+1}{L}\mu_0, \tag{5.66}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{v})) = \frac{x_{j-1}^*}{2x_j^*}r_j\frac{L-j+1}{L}\mu_1, \tag{5.67}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j)) = \frac{x_{j+1}^*}{2x_j^*}v_j\frac{j+1}{L}\mu_0, \tag{5.68}$$

$$q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j)) = \frac{x_{j-1}^*}{2x_j^*}v_j\frac{L-j+1}{L}\mu_1. \tag{5.69}$$

The branching rates (5.65) are not affected by the mutation events with transition rates (5.66) to (5.69), because these mutation events do not change the total number of real lines nor the total number of virtual lines. Thus, the total number of virtual lines follows a birth process. Since the rates (5.68) and (5.69) at which the virtual lines mutate are of the same order as the branching rates, all virtual lines will keep growing in such a process. When the numbers of virtual lines reach the level of $O(\theta)$, the rates (5.60) of the coalescence events that eliminate the virtual lines become $O(\theta^2)$. These coalescence events take effect on the faster timescale; the numbers of virtual lines then stabilise on the level of $O(\theta)$.

Consider the ancestral dual process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{v}_\theta(t)), t \geq 0\}$ with transition rates (5.58) to (5.64) for a sample $\boldsymbol{n} \in \mathbb{Z}_+^{L+1}$ of size $|\boldsymbol{n}| \geq 1$. Define $\kappa_j^{(\theta)}(t) = v_j^{(\theta)}(t)/\theta$, $j \in E$. We have the following theorem regarding the rescaled and time-changed process $\{(\boldsymbol{r}_\theta(t/\theta), \boldsymbol{\kappa}_\theta(t/\theta)), t \geq 0\}$ with state space $S = \{0, 1, \ldots, n_0\} \times \{0, 1, \ldots, n_1\} \times \cdots \times \{0, 1, \ldots, n_L\} \times \mathbb{R}_+^{L+1}$.

Theorem 5.17 (Virtual lines under selection-mutation balance)**.** *Assume $\boldsymbol{\kappa}_\theta(0) \neq \mathbf{0}$. As $\theta \to \infty$ under selection-mutation balance* (5.1)*, the process $\{(\boldsymbol{r}_\theta(t/\theta), \boldsymbol{\kappa}_\theta(t/\theta)), t \geq 0\}$ converges in finite-dimensional distribution to a process $\{(\boldsymbol{r}(t), \boldsymbol{\kappa}(t)), t \geq 0\}$ with state space S running on the faster timescale where one unit of time corresponds to θ units of time on the original timescale.*

Suppose $h(\boldsymbol{r}, \boldsymbol{\kappa})\colon S \to \mathbb{R}$ is continuously differentiable with respect to $\boldsymbol{\kappa}$. The process $\{(\boldsymbol{r}(t), \boldsymbol{\kappa}(t)), t \geq 0\}$ is specified by the following generator:

$$\begin{aligned}
\mathcal{L}h(\boldsymbol{r}, \boldsymbol{\kappa}) = \sum_{j \in E} \Bigg\{ & [h(\boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{x_{j+1}^*}{2x_j^*} r_j \frac{j+1}{L} \mu_0 \\
& + [h(\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{x_{j-1}^*}{2x_j^*} r_j \frac{L-j+1}{L} \mu_1 \Bigg\} \\
+ \sum_{j \in E} \Bigg\{ & \frac{x_j^*}{2\tau} |\boldsymbol{\kappa}| (L-j) \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_j} - \frac{1}{2x_j^*} \kappa_j^2 \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_j} \\
& + \frac{x_{j+1}^*}{2x_j^*} \kappa_j \frac{j+1}{L} \mu_0 \left(\frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_{j+1}} - \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_j} \right) \\
& + \frac{x_{j-1}^*}{2x_j^*} \kappa_j \frac{L-j+1}{L} \mu_1 \left(\frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_{j-1}} - \frac{\partial h(\boldsymbol{r}, \boldsymbol{\kappa})}{\partial \kappa_j} \right) \Bigg\},
\end{aligned}$$

where $(x_j^*)_{j \in E}$ is the stable equilibrium point given in equation (5.13)*.*

Proof. Consider the process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{\kappa}_\theta(t)), t \geq 0\}$ that has the transition rates (5.58) to (5.64) and generator $\mathcal{L}_\theta$. Suppose $h(\boldsymbol{r}, \boldsymbol{\kappa})\colon S \to \mathbb{R}$ is twice continuously differentiable with respect to $\boldsymbol{\kappa}$. For $\kappa_j = v_j/\theta$, $j \in E$, we have

$$\mathcal{L}_\theta h(\boldsymbol{r}, \boldsymbol{\kappa}) = \sum_{j \in E} [h(\boldsymbol{r} - \boldsymbol{e}_j, \boldsymbol{\kappa}) - h(\boldsymbol{r}, \boldsymbol{\kappa})] \times \frac{1}{2x_j^*} r_j (r_j - 1)$$

$$
\begin{aligned}
&+\sum_{j\in E}[h(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j+1}}{2x^*_j}r_j\frac{j+1}{L}\mu_0\theta\\
&+\sum_{j\in E}[h(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j-1}}{2x^*_j}r_j\frac{L-j+1}{L}\mu_1\theta\\
&+\sum_{j\in E}[h(\boldsymbol{r},\boldsymbol{\kappa}+\boldsymbol{e}_j/\theta)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_j}{2\tau}\left(\frac{r}{\theta}+|\boldsymbol{\kappa}|\right)(L-j)\theta^2\\
&+\sum_{j\in E}[h(\boldsymbol{r},\boldsymbol{\kappa}-\boldsymbol{e}_j/\theta)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2x^*_j}\kappa_j\left(\frac{2r_j}{\theta}+\kappa_j-\frac{1}{\theta}\right)\theta^2\\
&+\sum_{j\in E}[h(\boldsymbol{r},\boldsymbol{\kappa}+\boldsymbol{e}_{j+1}/\theta-\boldsymbol{e}_j/\theta)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j+1}}{2x^*_j}\kappa_j\frac{j+1}{L}\mu_0\theta^2\\
&+\sum_{j\in E}[h(\boldsymbol{r},\boldsymbol{\kappa}+\boldsymbol{e}_{j-1}/\theta-\boldsymbol{e}_j/\theta)-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j-1}}{2x^*_j}\kappa_j\frac{L-j+1}{L}\mu_1\theta^2\\
=&\sum_{j\in E}[h(\boldsymbol{r}-\boldsymbol{e}_j,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{1}{2x^*_j}r_j(r_j-1)\\
&+\sum_{j\in E}[h(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j+1}}{2x^*_j}r_j\frac{j+1}{L}\mu_0\theta\\
&+\sum_{j\in E}[h(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j-1}}{2x^*_j}r_j\frac{L-j+1}{L}\mu_1\theta\\
&+\sum_{j\in E}\left[\frac{1}{\theta}\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}+\frac{1}{2\theta^2}\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j^2}+O(\theta^{-3})\right]\times\frac{x^*_j}{2\tau}\left(\frac{r}{\theta}+|\boldsymbol{\kappa}|\right)(L-j)\theta^2\\
&+\sum_{j\in E}\left[-\frac{1}{\theta}\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}+\frac{1}{2\theta^2}\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j^2}+O(\theta^{-3})\right]\times\frac{1}{2x^*_j}\kappa_j\left(\frac{2r_j}{\theta}+\kappa_j-\frac{1}{\theta}\right)\theta^2\\
&+\sum_{j\in E}\left[\frac{1}{\theta}\left(\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_{j+1}}-\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}\right)+\frac{1}{2\theta^2}\left(\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j^2}-2\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j\partial\kappa_{j+1}}+\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_{j+1}^2}\right)\right.\\
&\qquad\left.+O(\theta^{-3})\right]\times\frac{x^*_{j+1}}{2x^*_j}\kappa_j\frac{j+1}{L}\mu_0\theta^2\\
&+\sum_{j\in E}\left[\frac{1}{\theta}\left(\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_{j-1}}-\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}\right)+\frac{1}{2\theta^2}\left(\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j^2}-2\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j\partial\kappa_{j-1}}+\frac{\partial^2h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_{j-1}^2}\right)\right.\\
&\qquad\left.+O(\theta^{-3})\right]\times\frac{x^*_{j-1}}{2x^*_j}\kappa_j\frac{L-j+1}{L}\mu_1\theta^2.
\end{aligned}
$$

The generator of the time-changed process $\{(\boldsymbol{r}_\theta(t/\theta),\boldsymbol{\kappa}_\theta(t/\theta)),t\geq 0\}$ then satisfies

$$
\lim_{\theta\to\infty}\frac{1}{\theta}\mathcal{L}_\theta h(\boldsymbol{r},\boldsymbol{\kappa})=\sum_{j\in E}\left\{[h(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j+1}}{2x^*_j}r_j\frac{j+1}{L}\mu_0\right.
$$

$$
\begin{aligned}
&\quad + [h(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{\kappa})-h(\boldsymbol{r},\boldsymbol{\kappa})]\times\frac{x^*_{j-1}}{2x^*_j}r_j\frac{L-j+1}{L}\mu_1\Bigg\}\\
&+\sum_{j\in E}\Bigg\{\frac{x^*_j}{2\tau}|\boldsymbol{\kappa}|(L-j)\,\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}-\frac{1}{2x^*_j}\kappa_j^2\,\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}\\
&\qquad+\frac{x^*_{j+1}}{2x^*_j}\kappa_j\frac{j+1}{L}\mu_0\bigg(\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_{j+1}}-\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}\bigg)\\
&\qquad+\frac{x^*_{j-1}}{2x^*_j}\kappa_j\frac{L-j+1}{L}\mu_1\bigg(\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_{j-1}}-\frac{\partial h(\boldsymbol{r},\boldsymbol{\kappa})}{\partial\kappa_j}\bigg)\Bigg\}\\
&=\mathcal{L}h(\boldsymbol{r},\boldsymbol{\kappa}).
\end{aligned}
$$

The proof is thus completed. □

It follows from Theorem 5.17 that on the faster timescale, we can treat the process for the virtual lines separately from the one for the real lines. We consider the virtual lines first; the real genealogy will be described in Theorem 5.19. The scaled numbers of virtual lines evolve on the faster timescale according to the following system of differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}\kappa_j(t)}{\mathrm{d}t}=\frac{1}{2}\Bigg[&(L-j)\frac{x^*_j}{\tau}\sum_{l\in E}\kappa_l(t)-\kappa_j^2(t)\frac{1}{x^*_j}+\kappa_{j-1}(t)\frac{j}{L}\mu_0\frac{x^*_j}{x^*_{j-1}}-\kappa_j(t)\frac{j+1}{L}\mu_0\frac{x^*_{j+1}}{x^*_j}\\
&-\kappa_j(t)\frac{L-j+1}{L}\mu_1\frac{x^*_{j-1}}{x^*_j}+\kappa_{j+1}(t)\frac{L-j}{L}\mu_1\frac{x^*_j}{x^*_{j+1}}\Bigg],
\end{aligned}
\tag{5.70}
$$

where $j\in E$. The equilibrium point $\boldsymbol{\kappa}^*\coloneqq\lim_{t\to\infty}\boldsymbol{\kappa}(t)$ of the above deterministic system (5.70) satisfies the following system of equations:

$$
\begin{aligned}
0=(L-j)\frac{x^*_j}{\tau}\sum_{l\in E}\kappa^*_l-(\kappa^*_j)^2\frac{1}{x^*_j}+\kappa^*_{j-1}\frac{j}{L}\mu_0\frac{x^*_j}{x^*_{j-1}}-\kappa^*_j\frac{j+1}{L}\mu_0\frac{x^*_{j+1}}{x^*_j}\\
-\kappa^*_j\frac{L-j+1}{L}\mu_1\frac{x^*_{j-1}}{x^*_j}+\kappa^*_{j+1}\frac{L-j}{L}\mu_1\frac{x^*_j}{x^*_{j+1}},
\end{aligned}
$$

where $j\in E$. The solutions to the above algebraic system are $\mathbf{0}$ and

$$
\kappa^*_j=\frac{L(1-\psi)}{\tau}x^*_j,\quad j\in E,
\tag{5.71}
$$

where x^*_j is the forwards-in-time stable equilibrium point given in equation (5.13) and

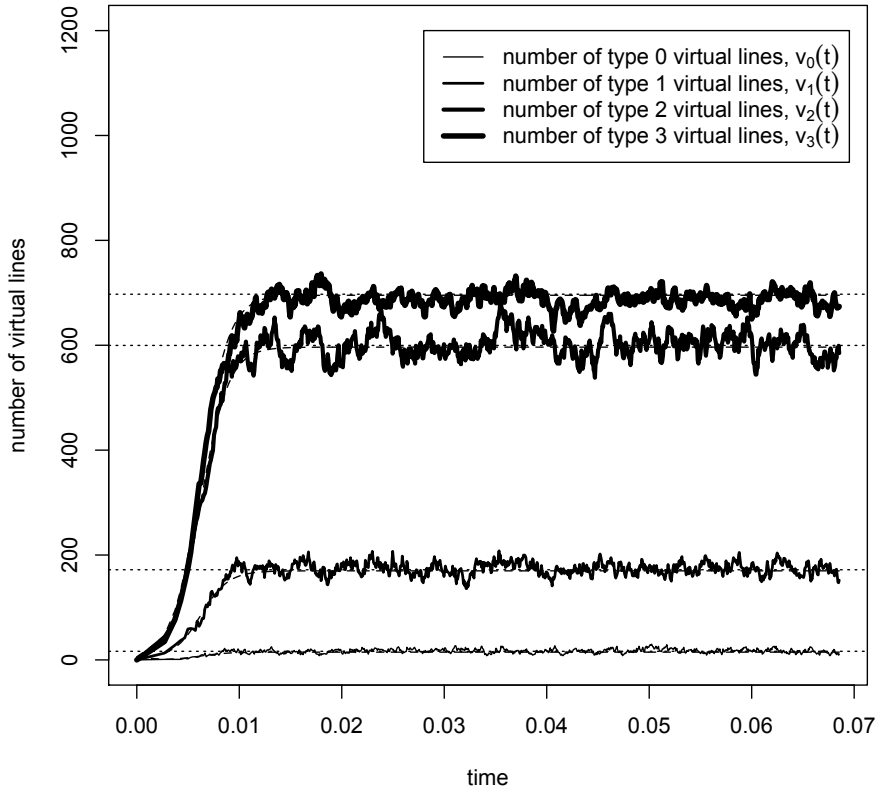


Figure 5.4: The numbers of virtual lines in the ancestral dual process with transition rates (5.58) to (5.64) in the three-locus case, superimposed on the trajectory of the deterministic system (5.70). Time is measured on the original timescale. The sample is $(5, 3, 6, 1)$, and the coefficient values are $\theta = 2000$, $\tau = 0.9$ and $\mu_0 = 0.8$. The horizontal dotted lines represent the stable equilibrium point $\theta\boldsymbol{\kappa}^*$, where $\boldsymbol{\kappa}^*$ is calculated from equation (5.71).

ψ is given in equation (5.14).

Figure 5.4 plots a path of the numbers of virtual lines in the ancestral dual process with transition rates (5.58) to (5.64) in the three-locus case. The path is superimposed on the trajectory of the deterministic system (5.70). We observe that the numbers of virtual lines increase rapidly from time 0, and then stabilise to $\theta\boldsymbol{\kappa}^*$, where $\boldsymbol{\kappa}^*$ is given in equation (5.71), when the coalescence events with rates (5.26) for virtual lines become effective on the faster timescale. Thus, our argument is supported by simulations.

We now concentrate on the real genealogy. According to Remark 5.16, it is sufficient to consider the marginal process of the real ancestral lines, $\{\boldsymbol{r}_\theta(t), t \geq 0\}$, which has the transitions rates

$$q_\theta(\boldsymbol{r}, \boldsymbol{r} - \boldsymbol{e}_j) = \frac{1}{2x_j^*} r_j(r_j - 1), \tag{5.72}$$

$$q_\theta(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j) = \frac{x_{j+1}^*}{2x_j^*} r_j \frac{j+1}{L} \theta\mu_0, \tag{5.73}$$

$$q_\theta(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j) = \frac{x_{j-1}^*}{2x_j^*} r_j \frac{L-j+1}{L} \theta\mu_1, \tag{5.74}$$

where $j \in E$, and $(x_j^*)_{j \in E}$ is the stable equilibrium point given in equation (5.13).

We use the idea of separation of timescales to study the limit of the process $\{\boldsymbol{r}_\theta(t), t \geq 0\}$ as $\theta \to \infty$. Nordborg and Krone (2002) use the same idea to deduce the strong migration limit in the structured coalescent process, which can be adopted to our process when migration is caused by mutation. The transition rates (5.73) and (5.74) are $O(\theta)$. These mutation events occur much more rapidly than the coalescence events with rates (5.72) of $O(1)$. In the limit as $\theta \to \infty$, the mutation events approach an equilibrium on the faster timescale between the coalescence events, and the total number of real lines reduces by one when a coalescence event occurs on the original timescale. According to Nordborg and Krone (2002), the real genealogy follows Kingman's coalescent, where the coalescence rate is an average over the stationary distribution of the mutation process.

We first consider the mutation process on the faster timescale. For fixed $r, 2 \leq r \leq n$ where n is the sample size, there exists a Markov process $\{\boldsymbol{r}_r(t), t \geq 0\}$ with state space $S_r = \{\boldsymbol{r} \colon \boldsymbol{r} \in \mathbb{Z}_+^{L+1}, |\boldsymbol{r}| = r\}$ and transition rates

$$q_r(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j) = \frac{x_{j+1}^*}{2x_j^*} r_j \frac{j+1}{L} \mu_0, \tag{5.75}$$

$$q_r(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j) = \frac{x_{j-1}^*}{2x_j^*} r_j \frac{L-j+1}{L} \mu_1, \tag{5.76}$$

where $j \in E$. We call the process $\{\boldsymbol{r}_r(t), t \geq 0\}$ the level-r mutation process. The

mutation process is ergodic; hence it has a unique stationary distribution, which is specified in the following lemma.

Lemma 5.18. *Define*

$$\gamma_j = \binom{L}{j}\phi^j(1-\phi)^{L-j}, \qquad j \in E, \tag{5.77}$$

where

$$\phi = \frac{\mu_0\psi^2}{\mu_0\psi^2 + \mu_1(1-\psi)^2},$$

with ψ given in equation (5.14).

The unique stationary distribution of the level-r mutation process $\{\boldsymbol{r}_r(t), t \geq 0\}$ (running on the faster timescale), is multinomial with parameters r and $(\gamma_j)_{j\in E}$; i.e., for $\boldsymbol{r} \in S_r$, the stationary distribution is

$$\pi_r(\boldsymbol{r}) = \frac{r!}{r_0!r_1!\cdots r_L!}\prod_{j\in E}\gamma_j^{r_j}. \tag{5.78}$$

Proof. As given in equation (5.13), the forwards-in-time stable equilibrium point $(x_j^*)_{j\in E}$ satisfies

$$x_j^* = \binom{L}{j}\psi^j(1-\psi)^{L-j}, \quad j \in E,$$

where ψ is given in equation (5.14) as

$$\psi = \frac{1}{2L}\Big(L - \tau + \sqrt{(L+\tau)^2 - 4L\tau\mu_0}\Big).$$

Thus, γ_j defined in equation (5.77) can be rewritten as

$$\gamma_j = \left(\frac{x_j^*}{x_0^*}\right)^2 \frac{j!}{L_{[j]}}\left(\frac{\mu_0}{\mu_1}\right)^j \gamma_0, \tag{5.79}$$

for $j = 1, 2, \ldots, L$, where

$$\gamma_0 = \Big[\sum_{j\in E}\left(\frac{x_j^*}{x_0^*}\right)^2 \frac{j!}{L_{[j]}}\left(\frac{\mu_0}{\mu_1}\right)^j\Big]^{-1}.$$

We then verify that the probability mass function (5.78) with γ_j in the form of

equation (5.79) satisfies the following detailed balance equations:

$$\pi_r(\boldsymbol{r})q_r(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j) = \pi_r(\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j)q_r(\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j, \boldsymbol{r}), \tag{5.80}$$

for $j = 1, 2, \ldots, L$, and

$$\pi_r(\boldsymbol{r})q_r(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j) = \pi_r(\boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j)q_r(\boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j, \boldsymbol{r}), \tag{5.81}$$

for $j = 0, 1, \ldots, L-1$, where the transition rates $q_r(\cdot, \cdot)$ are those given in equations (5.75) and (5.76).

By substituting the probability mass function (5.78) and the transition rates into the left-hand side of equation (5.80), we have, for $j = 1, 2, \ldots, L$,

$$\begin{aligned}
LHS &= \frac{r!}{r_0!r_1!\cdots r_L!}\gamma_0^{r_0}\gamma_1^{r_1}\cdots\gamma_L^{r_L} \times \frac{x_{j-1}^*}{2x_j^*} r_j \frac{L-j+1}{L}\mu_1 \\
&= \frac{r!}{r_0!r_1!\cdots(r_{j-1}+1)!(r_j-1)!\cdots r_L!}\gamma_0^{r_0}\gamma_1^{r_1}\cdots\gamma_{j-1}^{r_{j-1}+1}\gamma_j^{r_j-1}\cdots\gamma_L^{r_L} \\
&\quad \times \frac{x_j^*}{2x_{j-1}^*}(r_{j-1}+1)\frac{j}{L}\mu_0 \times \frac{\gamma_j}{\gamma_{j-1}}\left(\frac{x_{j-1}^*}{x_j^*}\right)^2 \frac{L-j+1}{j}\frac{\mu_1}{\mu_0} \\
&= \pi_r(\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j) \times \frac{x_j^*}{2x_{j-1}^*}(r_{j-1}+1)\frac{j}{L}\mu_0 \\
&= RHS.
\end{aligned}$$

Similarly, by substituting the probability mass function (5.78) and the transition rates into the left-hand side of equation (5.81), we have, for $j = 0, 1, \ldots, L-1$,

$$\begin{aligned}
LHS &= \frac{r!}{r_0!r_1!\cdots r_L!}\gamma_0^{r_0}\gamma_1^{r_1}\cdots\gamma_L^{r_L} \times \frac{x_{j+1}^*}{2x_j^*} r_j \frac{j+1}{L}\mu_0 \\
&= \frac{r!}{r_0!r_1!\cdots(r_j-1)!(r_{j+1}+1)!\cdots r_L!}\gamma_0^{r_0}\gamma_1^{r_1}\cdots\gamma_j^{r_j-1}\gamma_{j+1}^{r_{j+1}+1}\cdots\gamma_L^{r_L} \\
&\quad \times \frac{x_j^*}{2x_{j+1}^*}(r_{j+1}+1)\frac{L-j}{L}\mu_1 \times \frac{\gamma_j}{\gamma_{j+1}}\left(\frac{x_{j+1}^*}{x_j^*}\right)^2 \frac{j+1}{L-j}\frac{\mu_0}{\mu_1} \\
&= \pi_r(\boldsymbol{r} + \boldsymbol{e}_{j+1} - \boldsymbol{e}_j) \times \frac{x_j^*}{2x_{j-1}^*}(r_{j-1}+1)\frac{j}{L}\mu_0 \\
&= RHS.
\end{aligned}$$

Finally, it is easy to see from equation (5.79) that

$$\sum_{j\in E}\gamma_j = 1.$$

The proof is thus completed. □

Given the stationary distribution of the level-r mutation process on the faster timescale, we can now present the following Theorem 5.19 for the real genealogy in the additive selection model in the limit as $\theta \to \infty$ under selection-mutation balance. Consider the process $\{\boldsymbol{r}_\theta(t), t \geq 0\}$ with transition rates (5.72) to (5.74). Nordborg and Krone (2002) show that in the limit as $\theta \to \infty$, the structure associated with types in the process $\{\boldsymbol{r}_\theta(t), t \geq 0\}$ collapses to Kingman's coalescent in which all ancestral lines are exchangeable. Define $r_\theta(t) = |\boldsymbol{r}_\theta(t)|$. The process $\{r_\theta(t), t \geq 0\}$ for a sample $(n_1, n_2) \in \mathbb{Z}_+^2$ of size $n = n_1 + n_2 \geq 1$, has state space $S = \{1, 2, \ldots .n\}$ and initial state n.

Theorem 5.19 (Real lines in the additive selection model under selection-mutation balance)**.** *As $\theta \to \infty$, the process $\{r_\theta(t), t \geq 0\}$ converges in distribution to $\{r(t), t \geq 0\}$. $\{r(t), t \geq 0\}$ is a death process with state space S, initial state n and death rate*

$$q(r, r-1) = \binom{r}{2}\sum_{j\in E}\frac{\gamma_j^2}{x_j^*} \equiv \binom{r}{2}\left(\frac{\mu_0^2\psi^3 + \mu_1^2(1-\psi)^3}{(\mu_0\psi^2 + \mu_1(1-\psi)^2)^2}\right)^L, \quad r \geq 2,$$

where $(\gamma_j)_{j\in E}$ is given in equation (5.77)*, $(x_j^*)_{j=0}^L$ is the stable equilibrium point given in equation* (5.13) *and ψ is given in equation* (5.14)*.*

Proof. According to Nordborg and Krone (2002, Theorem 1), as $\theta \to \infty$, we have the weak convergence of the process $\{r_\theta(t), t \geq 0\}$ to the death process of Kingman's coalescent with an effective coalescence rate $q(r, r-1)$. $q(r, r-1)$ can be computed as the coalescence rates (5.72) (on the original timescale) averaged over the stationary distribution of the level-r mutation process (on the faster timescale) derived in Lemma 5.18. Let $(R_j)_{j\in E}$ be a random vector distributed with the probability mass

function (5.78). We have

$$\begin{aligned} q(r, r-1) &= \sum_{j \in E} \mathbb{E}\left[\frac{1}{2x_j^*} R_j(R_j - 1)\right] \\ &= \binom{r}{2} \sum_{j \in E} \frac{\gamma_j^2}{x_j^*} \\ &= \binom{r}{2} \sum_{j \in E} \binom{L}{j} \left(\frac{\mu_0^2 \psi^3}{(\mu_0 \psi^2 + \mu_1 (1-\psi)^2)^2}\right)^j \left(\frac{\mu_1^2 (1-\psi)^3}{(\mu_0 \psi^2 + \mu_1 (1-\psi)^2)^2}\right)^{L-j}, \\ &= \binom{r}{2} \left(\frac{\mu_0^2 \psi^3 + \mu_1^2 (1-\psi)^3}{(\mu_0 \psi^2 + \mu_1 (1-\psi)^2)^2}\right)^L, \end{aligned}$$

where the third equality follows from equations (5.13) and (5.77), and the last equality follows from the binomial theorem. The proof is then completed. □

5.5 Limit as $L \to \infty$

In this section, we consider the additive selection model in which each individual is composed of an infinite number of diallelic loci, i.e., we let $L \to \infty$. We derive in Lemma 5.20 that as $L \to \infty$, there is a *finite* number of type 0 loci in an individual drawn randomly from the stationary population. Thus, we need to assign the type to an individual according to the number of type 0 loci, instead of type 1 loci as in the previous sections. We also deduce that, backwards in time, types in a sample are lost one by one on a faster timescale, and the sample is instantaneously converted to the one in which individuals are not composed of any type 0 locus. The real genealogy then follows Kingman's coalescent with coalescence rate $\binom{r}{2} \exp(-\tau \mu_0)$. These results provide new insight into the studies of background selection model (see Remarks 5.21 and 5.23).

For now, let L still be finite to see how the way we assign types to individuals changes the settings of the Moran model, the diffusion and the ancestral dual process. Recall that $E = \{0, 1, \ldots, L\}$ is the set of possible types.

In the Moran model, since each type 0 locus decreases the reproductive ability of an individual by s, we assume each individual of type j gives birth at rate

$$\lambda_j = 1 - js, \quad j \in E,$$

where $0 \leq s \leq 1/L$; cf. equation (5.2). At the same time as reproduction, an individual is picked uniformly at random from the population to die.

Mutations still occur continuously in time at rate $u \geq 0$ on each individual, independently of reproduction. At each locus, the allele has a probability μ_0 of mutating to type 0, and a probability μ_1 to type 1, independently of its current type; $\mu_0 + \mu_1 = 1$. Given a mutation occurs on an individual of type $i \in E$, the probability that the mutation changes this individual to type $j \in E$, is

$$p_{ij} = \begin{cases} \frac{L-i}{L}\mu_0, & \text{if } j = i+1, \\ \frac{i}{L}\mu_1, & \text{if } j = i-1, \\ \frac{L-i}{L}\mu_1 + \frac{i}{L}\mu_0, & \text{if } j = i, \\ 0, & \text{otherwise;} \end{cases}$$

cf. equation (5.3).

Let $\tilde{Z}_j(t)$ be the number of type j individuals in the population at time t. Define $Y_j^{(N)}(t) = \tilde{Z}_j(t)/N$. Suppose $Ns = \sigma$, and $Nu = \theta$. By slightly altering the calculations in Proposition 5.1, we can show that the process $\{\boldsymbol{Y}_N(Nt/2), t \geq 0\}$ converges in distribution to a diffusion process $\{\boldsymbol{Y}(t), t \geq 0\}$ on the L-dimensional simplex $\mathcal{S}_{L+1}$, as $N \to \infty$. For a twice continuously differentiable function $f \colon \mathcal{S}_{L+1} \to \mathbb{R}$, the generator of the diffusion is

$$\mathcal{L}f(\boldsymbol{y}) = \frac{1}{2} \sum_{i,j \in E} \tilde{a}_{ij}(\boldsymbol{y}) \frac{\partial^2 f(\boldsymbol{y})}{\partial y_i \partial y_j} + \sum_{j \in E} \tilde{b}_j(\boldsymbol{y}) \frac{\partial f(\boldsymbol{y})}{\partial y_j}, \tag{5.82}$$

where

$$\tilde{a}_{ij}(\boldsymbol{y}) = \delta_{ij} y_i - y_i y_j,$$

and

$$\tilde{b}_j(\boldsymbol{y}) = \frac{\theta}{2L}\big[y_{j-1}(L-j+1)\mu_0 + y_{j+1}(j+1)\mu_1 - y_j((L-j)\mu_0 + j\mu_1)\big] - \frac{\sigma}{2}y_j\bigg(j - \sum_{k=0}^{L} k y_k\bigg).$$

We set $y_{-1} = y_{L+1} = 0$.

Then, similar to the deterministic approximation of the L-locus diffusion with generator equation (5.4) by equation (5.6), as $\theta \to \infty$ under selection-mutation balance (5.1), the time-changed process $\{\boldsymbol{Y}(t/\theta), t \geq 0\}$ with initial value $\boldsymbol{Y}(0) = \boldsymbol{\zeta} \in (0,1)^{L+1}$, can be approximated by the solution $\boldsymbol{y}(t)$ to the following system of differential equations:

$$\frac{\mathrm{d}y_j(t)}{\mathrm{d}t} = \frac{1}{2L}\big[y_{j-1}(t)(L-j+1)\mu_0 + y_{j+1}(t)(j+1)\mu_1 - y_j(t)((L-j)\mu_0 + j\mu_1)\big] - \frac{1}{2\tau}y_j(t)\bigg(j - \sum_{k=0}^{L} k y_k(t)\bigg),$$

where $j \in E$, with initial value $\boldsymbol{y}(0) = \boldsymbol{\zeta}$.

By an argument similar to that in Theorem 5.6, there exists $\boldsymbol{y}^* = \lim_{t\to\infty} \boldsymbol{y}(t)$, where

$$y_j^* = \binom{L}{j}\psi_L^j(1-\psi_L)^{L-j}, \qquad j \in E, \tag{5.83}$$

with

$$\psi_L = \frac{1}{2L}\Big(L + \tau - \sqrt{(L+\tau)^2 - 4L\tau\mu_0}\Big). \tag{5.84}$$

Then, according to Lemma 5.10, the transition rates of the ancestral process that is dual to the diffusion with generator (5.82) satisfy

$$\begin{aligned} q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_j)) &\sim \frac{y_j^*}{2\tau}(r+v)j\theta, \\ q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_j,\boldsymbol{v})) &\sim \frac{1}{2y_j^*}r_j(r_j-1), \\ q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}-\boldsymbol{e}_j)) &\sim \frac{1}{2y_j^*}v_j(2r_j+v_j-1), \end{aligned}$$

$$
\begin{aligned}
q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j,\boldsymbol{v})) &\sim \frac{y^*_{j+1}}{2y^*_j} r_j \frac{j+1}{L}\theta\mu_1,\\
q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j,\boldsymbol{v})) &\sim \frac{y^*_{j-1}}{2y^*_j} r_j \frac{L-j+1}{L}\theta\mu_0,\\
q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j+1}-\boldsymbol{e}_j)) &\sim \frac{y^*_{j+1}}{2y^*_j} v_j \frac{j+1}{L}\theta\mu_1,\\
q_\theta((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_{j-1}-\boldsymbol{e}_j)) &\sim \frac{y^*_{j-1}}{2y^*_j} v_j \frac{L-j+1}{L}\theta\mu_0,
\end{aligned}
\tag{5.85}
$$

where $j \in E$.

In the following lemma, we prove that as $L \to \infty$, the stationary frequency of individuals with j type 0 loci is

$$
\frac{(\tau\mu_0)^j}{j!}e^{-\tau\mu_0}, \quad j \in \mathbb{N}_0.
$$

The above quantity can also be interpreted as the distribution of the number of type 0 loci in an individual of infinite length chosen randomly from the stationary distribution.

Lemma 5.20. *Consider ψ_L given in equation* (5.84)*. For $0 < \tau < \infty$ and $0 < \mu_0 < 1$, we have*

$$
\lim_{L\to\infty} L\psi_L = \tau\mu_0. \tag{5.86}
$$

Furthermore, the stationary frequencies y^*_j in equation (5.83) *satisfy*

$$
\lim_{L\to\infty} y^*_j = \frac{(\tau\mu_0)^j}{j!}e^{-\tau\mu_0}, \quad j \in \mathbb{N}_0. \tag{5.87}
$$

Proof. Consider

$$
\psi_L = \frac{1}{2L}\Big(L+\tau-\sqrt{(L+\tau)^2-4L\tau\mu_0}\Big),
$$

as in equation (5.84), with $0 < \tau < \infty$ and $0 < \mu_0 < 1$. By letting $k = 1/L$, we have

$$
\psi_L = \frac{1}{2}\Big(1+k\tau-\sqrt{(1+k\tau)^2-4k\tau\mu_0}\Big) \coloneqq g(k).
$$

By applying the Taylor expansion to the function $g(k)$ at the point $k = 0$, it follows

that

$$g(k) = \tau\mu_0 k - \mu_0\mu_1\tau^2 k^2 + O(k^3),$$

which implies

$$\psi_L = \frac{\tau\mu_0}{L} - \frac{\mu_0\mu_1\tau^2}{L^2} + O(L^{-3}).$$

The limit in equation (5.86) then follows.

There is a well-known fact that the binomial distribution with parameters n and p converges to the Poisson distribution with parameter np, as $n \to \infty$ and $p \to 0$, but np remains fixed. Hence, we have

$$y_j^* = \binom{L}{j} \psi_L^j (1-\psi_L)^{L-j} \to \frac{(\tau\mu_0)^j}{j!} e^{-\tau\mu_0},$$

as $L \to \infty$. □

The result in Lemma 5.20 can be related to the studies of Muller's ratchet and the background selection model. Consider an asexual population in which chromosomes are passed down as indivisible blocks. When only deleterious mutations can occur, once an individual has a certain number of mutations, it will never have fewer mutations (forwards in time). This effect is known as Muller's ratchet, as when all individuals in the current "best" class have accumulated additional mutations, the ratchet clicks. Muller (1964), Haigh (1978) and Etheridge et al. (2009) investigate the properties of this mechanism (e.g., the average time between clicks of the ratchet). The studies of Muller's ratchet are often based on the background selection model as follows. The population is of a large size, in which the type of an individual is determined by the number of deleterious mutations accumulated. In a reproduction event, each individual chooses a parent from the previous generation as in the classic Wright-Fisher model. The relative fitness of a type j individual is assumed to be $(1-s)^j$, where $0 < s < 1$ and $j \in \mathbb{N}_0$, so selection is against deleterious alleles maintained by mutation. The number of extra mutations that occur between two successive generations is assumed

to follow a Poisson distribution with rate λ. The aforementioned authors find that when selection and the deleterious mutation are balanced, the frequency of type j individuals in the stationary population is Poisson with parameters j and λ/s. We can compare this result with ours, as in the following remark.

Remark 5.21 (Connections to previous work)**.** The Poisson distribution shows up both in the classic derivation of Muller's ratchet and in our additive selection model. In the background selection model, the Poisson stationary distribution of the frequencies in the population is shown by Haigh (1978, Equation 6). In our model, as the stable equilibrium frequencies satisfy a binomial formula in equation (5.13), we arrive in a Poisson distribution when $L \to \infty$. More specifically, in our Moran model, type 0 loci are deleterious. As $L \to \infty$, there are an infinite number of type 1 loci but only a finite number of type 0 loci. Hence, the mutation rate for a type 1 locus to type 0 is approximately $u\mu_0$, while the mutation rate for a type 0 locus to type 1 is approximately 0. Therefore, $u\mu_0$ in our Moran model has an identical role as the coefficient λ in the background selection model. The reproductive ability of an individual composed of j type 0 loci is $1 - js$ in our Moran model. We have $1 - js \approx (1 - s)^j$ for small s, which is the relative fitness assumed in the background selection model. Lemma 5.20 shows that the stationary frequency of an individual (of infinite length) having j type 0 loci is Poisson with parameters j and $\tau\mu_0$, which is approximately $u\mu_0/s$ since we assume $s = \sigma/N$, $u = \theta/N$ and $\lim_{\theta\to\infty} \sigma/\theta = \tau$.

We now concentrate on the ancestral dual process. We will see in Remark 5.23 that our results for the limiting genealogy are consistent with the arguments and simulations in the literature for the genealogy in the corresponding background selection model. The novelty of our work is that we obtain a full description of the genealogy using the ancestral dual process with typed lines and the time-scaling theorems developed in the previous sections.

Given the transition rates (5.85) and the limiting values of y_j^* in equation (5.87),

as $L \to \infty$, we consider an ancestral dual process $\{(\boldsymbol{r}_\theta(t), \boldsymbol{v}_\theta(t)), t \geq 0\}$ that has the following transition rates:

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_j)) = \frac{y_j^*}{2\tau}(r + v)j\theta, \tag{5.88}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} - \boldsymbol{e}_j, \boldsymbol{v})) = \frac{1}{2y_j^*}r_j(r_j - 1), \tag{5.89}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} - \boldsymbol{e}_j)) = \frac{1}{2y_j^*}v_j(2r_j + v_j - 1), \tag{5.90}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j, \boldsymbol{v})) = \frac{1}{2\tau}jr_j\theta, \tag{5.91}$$

$$q_\theta((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j)) = \frac{1}{2\tau}jv_j\theta, \tag{5.92}$$

where $j \in \mathbb{N}_0$, and

$$y_j^* = \frac{(\tau\mu_0)^j}{j!}e^{-\tau\mu_0},$$

as given in equation (5.87). Note that by assuming the stationary frequencies in the population are constants, we implicitly neglect the effect of Muller's ratchet.

Remark 5.22. Lemma 5.20 shows that, as $L \to \infty$, there are only a finite number of type 0 loci in an individual. Thus, forwards in time, given a mutation occurs on such an individual, the probability that the mutation acts on a type 0 locus and changes it to type 1 tends to zero. Therefore, in the ancestral dual process, there are only mutations with transitions $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j, \boldsymbol{v})$ and $(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j)$ that decrease the number of type 0 loci.

We note that the transitions with rates (5.89) and (5.91) that change the configuration of real lines do not depend on the configuration of virtual lines. Thus, the real genealogy is independent of the evolution of the virtual lines. On the faster timescale, the real lines evolve according to the transition rates

$$q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_{j-1} - \boldsymbol{e}_j, \boldsymbol{v})) = \frac{1}{2\tau}jr_j, \quad j \in \mathbb{N}_0.$$

It is not difficult to see that the types in the sample are lost one by one because of

the mutations on the faster timescale. In the equilibrium of the mutation process on the faster timescale, there are only individuals of type 0, i.e., having no type 0 loci.

Then, from time 0+ on the original timescale, the real genealogy follows a time-changed Kingman's coalescent with coalescence rate

$$\frac{1}{2y_0^*} r_0(r_0 - 1) \equiv \frac{1}{2} r_0(r_0 - 1) e^{-\tau \mu_0}. \tag{5.93}$$

Remark 5.23 (Connections to previous work)**.**

1. The nucleotide site diversity, i.e., the probability that a pair of randomly chosen chromosomes differ at a given nucleotide site, indicates the shape of the underlying genealogy (see Tavaré, 2004, Chapter 5). Charlesworth et al. (1993) study the nucleotide site diversity in the background selection model (see Remark 5.21). The authors show that, in a large population of size N, under selection-mutation balance, the nucleotide site diversity is the same as that observed in a neutral population with an effective population size Nf_0, where f_0 is the equilibrium frequency of individuals not having any mutation. In the context of our dual process as $L \to \infty$ under selection-mutation balance, their result means the shape of the real genealogy is determined by Kingman's coalescent with the coalescence rate given in equation (5.93).

2. Durrett (2008, p. 213) argues that in the background selection model under selection-mutation balance, the average time it takes for a sampled type j individual to become type 0 backwards in time, is smaller than or equal to $\log(j+1)/s \ll N$. This corresponds to the time spent by our limiting ancestral dual process on the faster timescale, which is negligible on the original timescale where the coalescence events occur.

3. Walczak et al. (2012), Desai et al. (2012) and Nicolaisen and Desai (2012) study the effect of purifying selection in the background selection model, in which

these authors assume there is a selection-mutation balance so the stationary frequency of type j individuals (individuals that have accumulated j mutations) is a constant determined by a Poisson formula similar to that in equation (5.87). These authors show that for a certain range of parameters corresponding to very strong selection, our approximation of the limiting genealogy with the coalescence rate (5.93) does hold well. However, these authors also prove that for moderately strong selection, the genealogy is neutral but with a time-dependent effective population size.

Appendix A

Proofs

A.1 Proof of Lemma 2.7

The confluent hypergeometric function of the first kind, $M(\theta_1, \theta, -\sigma)$, is defined as

$$M(\theta_1, \theta, -\sigma) = \frac{1}{B(\theta_1, \theta_2)} \int_0^1 x^{\theta_1 - 1}(1-x)^{\theta_2 - 1} e^{-\sigma x} \, \mathrm{d}x,$$

where $\theta_1 > 0$, $\theta_2 > 0$, $\theta = \theta_1 + \theta_2$ and $\sigma > 0$. After the change of variable by setting $y = \sigma x$, the above equation can be rewritten as

$$M(\theta_1, \theta, -\sigma) = \frac{1}{B(\theta_1, \theta_2)\sigma^{\theta_1}} \int_0^\sigma y^{\theta_1 - 1} \left(1 - \frac{y}{\sigma}\right)^{\theta_2 - 1} e^{-y} \, \mathrm{d}y.$$

By applying the Taylor expansion to the function $g(y) = (1 - y/\sigma)^{\theta_2 - 1}$ about the point $y = 0$, we have

$$\left(1 - \frac{y}{\sigma}\right)^{\theta_2 - 1} = \sum_{k=0}^{\infty} \frac{y^k}{k!} g^{(k)}(0) = \sum_{k=0}^{\infty} \frac{(1-\theta_2)_{(k)}}{k!} \left(\frac{y}{\sigma}\right)^k,$$

where $g^{(k)}(0)$ denotes the kth derivative of g evaluated at the point 0.

Therefore, we have

$$M(\theta_1, \theta, -\sigma) = \frac{1}{B(\theta_1, \theta_2)\sigma^{\theta_1}} \int_0^\sigma e^{-y} y^{\theta_1 - 1} \sum_{k=0}^{\infty} \frac{(1-\theta_2)_{(k)}}{k!} \left(\frac{y}{\sigma}\right)^k \mathrm{d}y$$

$$
\begin{aligned}
&= \frac{1}{B(\theta_1,\theta_2)\sigma^{\theta_1}} \sum_{k=0}^{\infty} \frac{(1-\theta_2)_{(k)}}{k!\sigma^k} \int_0^{\sigma} e^{-y} y^{\theta_1+k-1}\,\mathrm{d}y \\
&= \frac{1}{B(\theta_1,\theta_2)\sigma^{\theta_1}} \sum_{k=0}^{\infty} \frac{(1-\theta_2)_{(k)}}{k!\sigma^k} \left(\Gamma(\theta_1+k) - \int_{\sigma}^{\infty} e^{-y} y^{\theta_1+k-1}\,\mathrm{d}y \right) \\
&= \frac{\Gamma(\theta)}{\Gamma(\theta_2)\sigma^{\theta_1}} \sum_{k=0}^{\infty} \frac{\theta_{1(k)}(1-\theta_2)_{(k)}}{k!\sigma^k} \\
&\quad - e^{-\sigma} \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\sigma} \sum_{k=0}^{\infty} \frac{(1-\theta_2)_{(k)}}{k!} \left(\sum_{j=0}^{2} \frac{(\theta_1+k)_{[j]}}{\sigma^j} + O(\sigma^{-3}) \right) \\
&= \frac{\Gamma(\theta)}{\Gamma(\theta_2)\sigma^{\theta_1}} \left[1 + \frac{\theta_1(1-\theta_2)}{\sigma} + \frac{\theta_1(\theta_1+1)(1-\theta_2)(2-\theta_2)}{2\sigma^2} + O(\sigma^{-3}) \right],
\end{aligned}
$$

where the fourth equality follows Olver et al. (2010, Equation 8.11.2).

A.2 Proof of Lemma 2.8

We show how the transition rates (2.6) to (2.10) of the ancestral dual process with typed lines can be asymptotically expanded for large σ. The calculations are done by substituting equation (2.22) for the confluent hypergeometric function into the transition rates (2.6) to (2.10). It follows that

$$
\begin{aligned}
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r},\boldsymbol{v}+\boldsymbol{e}_1)) &= \frac{1}{2}\sigma(r+v)\frac{(r_1+v_1+\theta_1)}{r+v+\theta}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_1+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \\
&= \frac{1}{2}\sigma(r+v)\frac{(r_1+v_1+\theta_1)}{r+v+\theta} \times \frac{\Gamma(r+v+1+\theta)}{\Gamma(r_2+v_2+\theta_2)\sigma^{r_1+v_1+1+\theta_1}} \\
&\quad \times [1+\sigma^{-1}(r_1+v_1+1+\theta_1)(1-r_2-v_2-\theta_2)+O(\sigma^{-2})] \\
&\quad \times \frac{\Gamma(r_2+v_2+\theta_2)\sigma^{r_1+v_1+\theta_1}}{\Gamma(r+v+\theta)} \\
&\quad \times [1+\sigma^{-1}(r_1+v_1+\theta_1)(1-r_2-v_2-\theta_2)+O(\sigma^{-2})]^{-1} \\
&= \frac{1}{2}(r+v)(r_1+v_1+\theta_1)[1+\sigma^{-1}(1-r_2-v_2-\theta_2)]+O(\sigma^{-2}); \\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_1,\boldsymbol{v})) &= \frac{1}{2}r_1(r_1-1)\frac{r+v-1+\theta}{r_1+v_1-1+\theta_1}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_1+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \\
&= \frac{1}{2}r_1(r_1-1)\frac{r+v-1+\theta}{r_1+v_1-1+\theta_1} \times \frac{\Gamma(r+v-1+\theta)}{\Gamma(r_2+v_2+\theta_2)\sigma^{r_1+v_1-1+\theta_1}}
\end{aligned}
$$

$$
\begin{aligned}
&\times \Big[1+\sigma^{-1}(r_1+v_1-1+\theta_1)(1-r_2-v_2-\theta_2)\\
&\quad+\frac{1}{2}\sigma^{-2}(r_1+v_1-1+\theta_1)(r_1+v_1+\theta_1)\\
&\qquad\times(1-r_2-v_2-\theta_2)(2-r_2-v_2-\theta_2)\\
&\quad+O(\sigma^{-3})\Big]\times\frac{\Gamma(r_2+v_2+\theta_2)\sigma^{r_1+v_1+\theta_1}}{\Gamma(r+v+\theta)}\\
&\times\Big[1+\sigma^{-1}(r_1+v_1+\theta_1)(1-r_2-v_2-\theta_2)\\
&\quad+\frac{1}{2}\sigma^{-2}(r_1+v_1+\theta_1)(r_1+v_1+1+\theta_1)\\
&\qquad\times(1-r_2-v_2-\theta_2)(2-r_2-v_2-\theta_2)\\
&\quad+O(\sigma^{-3})\Big]^{-1}\\
&=\frac{1}{2}r_1(r_1-1)\frac{\sigma}{r_1+v_1-1+\theta_1}\\
&\quad\times[1-\sigma^{-1}(1-r_2-v_2-\theta_2)\\
&\qquad+\sigma^{-2}(r_1+v_1+\theta_1)(r_2+v_2-1+\theta_2)+O(\sigma^{-3})]\\
&=\frac{r_1(r_1-1)}{2(r_1+v_1-1+\theta_1)}\\
&\quad\times[\sigma+r_2+v_2-1+\theta_2+\sigma^{-1}(r_1+v_1+\theta_1)(r_2+v_2-1+\theta_2)]\\
&\quad+O(\sigma^{-2});\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}-\boldsymbol{e}_2,\boldsymbol{v}))&=\frac{1}{2}r_2(r_2-1)\frac{r+v-1+\theta}{r_2+v_2-1+\theta_2}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_2+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}\\
&=\frac{1}{2}r_2(r_2-1)\frac{r+v-1+\theta}{r_2+v_2-1+\theta_2}\times\frac{\Gamma(r+v-1+\theta)}{\Gamma(r_2+v_2-1+\theta_2)\sigma^{r_1+v_1+\theta_1}}\\
&\quad\times[1+\sigma^{-1}(r_1+v_1+\theta_1)(2-r_2-v_2-\theta_2)+O(\sigma^{-2})]\\
&\quad\times\frac{\Gamma(r_2+v_2+\theta_2)\sigma^{r_1+v_1+\theta_1}}{\Gamma(r+v+\theta)}\\
&\quad\times[1+\sigma^{-1}(r_1+v_1+\theta_1)(1-r_2-v_2-\theta_2)+O(\sigma^{-2})]^{-1}\\
&=\frac{1}{2}r_2(r_2-1)[1+\sigma^{-1}(r_1+v_1+\theta_1)]+O(\sigma^{-2});\\
q((\boldsymbol{r},\boldsymbol{v}),(\boldsymbol{r}+\boldsymbol{e}_1-\boldsymbol{e}_2,\boldsymbol{v}))&=\frac{1}{2}r_2\theta_2\frac{r_1+v_1+\theta_1}{r_2+v_2-1+\theta_2}\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_1-\boldsymbol{e}_2+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2} r_2 \theta_2 \frac{r_1 + v_1 + \theta_1}{r_2 + v_2 - 1 + \theta_2} \times \frac{\Gamma(r + v + \theta)}{\Gamma(r_2 + v_2 - 1 + \theta_2)\sigma^{r_1 + v_1 + 1 + \theta_1}} \\
&\quad \times [1 + \sigma^{-1}(r_1 + v_1 + 1 + \theta_1)(2 - r_2 - v_2 - \theta_2) + O(\sigma^{-2})] \\
&\quad \times \frac{\Gamma(r_2 + v_2 + \theta_2)\sigma^{r_1 + v_1 + \theta_1}}{\Gamma(r + v + \theta)} \\
&\quad \times [1 + \sigma^{-1}(r_1 + v_1 + \theta_1)(1 - r_2 - v_2 - \theta_2) + O(\sigma^{-2})]^{-1} \\
&= \frac{1}{2} \sigma^{-1} r_2 \theta_2 (r_1 + v_1 + \theta_1) \\
&\quad \times [1 + \sigma^{-1}(r_1 + v_1 + \theta_1 + 2 - r_2 - v_2 - \theta_2) + O(\sigma^{-2})] \\
&= \frac{1}{2} \sigma^{-1} r_2 \theta_2 (r_1 + v_1 + \theta_1) + O(\sigma^{-2});
\end{aligned}
$$

and

$$
\begin{aligned}
q((\boldsymbol{r}, \boldsymbol{v}), (\boldsymbol{r} + \boldsymbol{e}_2 - \boldsymbol{e}_1, \boldsymbol{v})) &= \frac{1}{2} r_1 \theta_1 \frac{r_2 + v_2 + \theta_2}{r_1 + v_1 - 1 + \theta_1} \frac{\nu(\boldsymbol{r} + \boldsymbol{v} + \boldsymbol{e}_2 - \boldsymbol{e}_1 + \boldsymbol{\theta})}{\nu(\boldsymbol{r} + \boldsymbol{v} + \boldsymbol{\theta})} \\
&= \frac{1}{2} r_1 \theta_1 \frac{r_2 + v_2 + \theta_2}{r_1 + v_1 - 1 + \theta_1} \times \frac{\Gamma(r + v + \theta)}{\Gamma(r_2 + v_2 + 1 + \theta_2)\sigma^{r_1 + v_1 - 1 + \theta_1}} \\
&\quad \times [1 - \sigma^{-1}(r_1 + v_1 - 1 + \theta_1)(r_2 + v_2 + \theta_2) + O(\sigma^{-2})] \\
&\quad \times \frac{\Gamma(r_2 + v_2 + \theta_2)\sigma^{r_1 + v_1 + \theta_1}}{\Gamma(r + v + \theta)} \\
&\quad \times [1 + \sigma^{-1}(r_1 + v_1 + \theta_1)(1 - r_2 - v_2 - \theta_2) + O(\sigma^{-2})]^{-1} \\
&= \frac{1}{2} r_1 \theta_1 \frac{\sigma}{r_1 + v_1 - 1 + \theta_1} \times [1 + \sigma^{-1}(r_2 + v_2 + \theta_2 - r_1 - v_1 - \theta_1) \\
&\qquad + \sigma^{-2}(r_1 + v_1 + \theta_1)(r_2 + v_2 - 1 + \theta_2) + O(\sigma^{-3})] \\
&= \frac{1}{2} \frac{r_1 \theta_1}{r_1 + v_1 - 1 + \theta_1} \times [\sigma + r_2 + v_2 + \theta_2 - r_1 - v_1 - \theta_1 \\
&\qquad + \sigma^{-1}(r_1 + v_1 + \theta_1)(r_2 + v_2 - 1 + \theta_2)] + O(\sigma^{-2}).
\end{aligned}
$$

The other equations follow analogously. The proof is then completed.

A.3 Proof of Lemma 3.13

Let $\boldsymbol{\theta} > 0$ be a three-dimensional vector representing the mutation coefficients. We consider the function

$$\nu(\boldsymbol{\theta}) = \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \int_0^1 \int_0^{1-x_1} e^{-\sigma_1 x_1 - \sigma_2 x_2} x_1^{\theta_1 - 1} x_2^{\theta_2 - 1} (1 - x_1 - x_2)^{\theta_3 - 1} \, \mathrm{d}x_2 \, \mathrm{d}x_1,$$

as defined in equation (3.29).

After the changes of variables by setting $x_1 = y_1/|\sigma_1|$ and $x_2 = y_2/\sigma_2$, we have

$$\begin{aligned}\nu(\boldsymbol{\theta}) = {} & \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)\sigma_1^{\theta_1}\sigma_2^{\theta_2}} \\ & \times \int_0^{\sigma_1} \int_0^{(1-y_1/\sigma_1)\sigma_2} e^{-y_1 - y_2} y_1^{\theta_1 - 1} y_2^{\theta_2 - 1} \left(1 - \frac{y_1}{\sigma} - \frac{y_2}{\sigma}\right)^{\theta_3 - 1} \mathrm{d}y_2 \, \mathrm{d}y_1.\end{aligned}$$

By applying the Taylor expansion to the function

$$\left(1 - \frac{y_1}{\sigma_1} - \frac{y_2}{\sigma_2}\right)^{\theta_3 - 1}$$

about the point $y_1 = 0$ and $y_2 = 0$, we have

$$\left(1 - \frac{y_1}{\sigma_1} - \frac{y_2}{\sigma_2}\right)^{\theta_3 - 1} = \sum_{i=0}^{\infty} \sum_{j=0}^{i} \binom{i}{j} \frac{(1 - \theta_3)_{(i)}}{i!} \left(\frac{y_1}{\sigma_1}\right)^{i-j} \left(\frac{y_2}{\sigma_2}\right)^{j}.$$

It follows that

$$\begin{aligned}\nu(\boldsymbol{\theta}) = {} & \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)\sigma_1^{\theta_1}\sigma_2^{\theta_2}} \\ & \times \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{1}{\sigma_1^{i-j}\sigma_2^{j}} \binom{i}{j} \frac{(1 - \theta_3)_{(i)}}{i!} \int_0^{\sigma_1} \int_0^{(1-y_1/\sigma_1)\sigma_2} e^{-y_1 - y_2} y_1^{\theta_1 + i - j - 1} y_2^{\theta_2 + j - 1} \, \mathrm{d}y_2 \, \mathrm{d}y_1 \\ = {} & \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)\sigma_1^{\theta_1}\sigma_2^{\theta_2}} \\ & \times \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{1}{\sigma_1^{i-j}\sigma_2^{j}} \binom{i}{j} \frac{(1 - \theta_3)_{(i)}}{i!} \int_0^{\sigma_1} y_1^{\theta_1 + i - j - 1} e^{-y_1} \left[\int_0^{(1-y_1/\sigma_1)\sigma_2} e^{-y_2} y_2^{\theta_2 + j - 1} \, \mathrm{d}y_2 \right] \mathrm{d}y_1 \\ = {} & \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)\sigma_1^{\theta_1}\sigma_2^{\theta_2}}\end{aligned}$$

$$
\begin{aligned}
&\times \sum_{i=0}^{\infty}\sum_{j=0}^{i}\frac{1}{\sigma_1^{i-j}\sigma_2^{j}}\binom{i}{j}\frac{(1-\theta_3)_{(i)}}{i!}\Gamma(\theta_1+i-j)\Gamma(\theta_2+j)+o(\sigma_1^{-2})\\
&=\frac{\Gamma(\theta)}{\Gamma(\theta_3)\sigma_1^{\theta_1}\sigma_2^{\theta_2}}\sum_{i=0}^{\infty}\sum_{j=0}^{i}\binom{i}{j}\frac{(1-\theta_3)_{(i)}\theta_{1(i-j)}\theta_{2(j)}}{i!\sigma_1^{i-j}\sigma_2^{j}}+o(\sigma_1^{-2})\\
&=\frac{\Gamma(\theta)}{\Gamma(\theta_3)\sigma_1^{\theta_1}\sigma_2^{\theta_2}}\Bigg[1+\frac{(1-\theta_3)\theta_1}{\sigma_1}+\frac{(1-\theta_3)\theta_2}{\sigma_2}\\
&\qquad+\frac{(1-\theta_3)(2-\theta_3)}{2}\left(\frac{\theta_1(1+\theta_1)}{\sigma_1^2}+\frac{2\theta_1\theta_2}{\sigma_1\sigma_2}+\frac{\theta_2(1+\theta_2)}{\sigma_2^2}\right)+o(\sigma_1^{-2})\Bigg],
\end{aligned}
$$

where the third equality holds according to Olver et al. (2010, Equation 8.11.2).

A.4 Proof of Lemma 3.14

Consider the ratios $\nu(\cdot)/\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})$ in the transition rates in equations (3.28). The ratios in the branching rates satisfy, for $i,j=1,2$ and $i\neq j$,

$$
\begin{aligned}
&\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}\\
&=\frac{\Gamma(r+v+1+\theta)\Gamma(r_3+v_3+\theta_3)\sigma_i^{r_i+v_i+\theta_i}\sigma_j^{r_j+v_j+\theta_j}}{\Gamma(r+v+\theta)\Gamma(r_3+v_3+\theta_3)\sigma_i^{r_i+v_i+1+\theta_i}\sigma_j^{r_j+v_j+\theta_j}}\\
&\quad\times\left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i+1+\theta_i)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right)\\
&\quad\times\left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i+\theta_i)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right)^{-1}\\
&=\frac{r+v+\theta}{\sigma_i}\left(1+\frac{1-r_3-v_3-\theta_3}{\sigma_i}+o(\sigma_i^{-1})\right).
\end{aligned}
$$

For the ratios appearing in the coalescence rates for individuals of type 1 and type 2, we have, for $i,j=1,2$ and $i\neq j$,

$$
\begin{aligned}
&\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}\\
&=\frac{\Gamma(r+v-1+\theta)\Gamma(\theta_3+\alpha_3)\sigma_i^{r_i+v_i+\theta_i}\sigma_j^{r_j+v_j+\theta_j}}{\Gamma(r+v+\theta)\Gamma(\theta_3+\alpha_3)\sigma_i^{r_i+v_i-1+\theta_i}\sigma_j^{r_j+v_j+\theta_j}}\\
&\quad\times\left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i-1+\theta_i)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right)
\end{aligned}
$$

$$\times \left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i+\theta_i)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right)^{-1}$$
$$=\frac{\sigma_i}{r+v-1+\theta}\left(1+\frac{r_3+v_3-1+\theta_3}{\sigma_i}+o(\sigma_i^{-1})\right).$$

The ratio in the coalescence rate for individuals of the favoured type 3 is slightly different from the previous two ratios. This ratio satisfies

$$\frac{\nu(\boldsymbol{r}+\boldsymbol{v}-\boldsymbol{e}_3+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}$$
$$=\frac{\Gamma(r+v-1+\theta)\Gamma(r_3+v_3+\theta_3)\sigma_1^{r_1+v_1+\theta_1}\sigma_2^{r_2+v_2+\theta_2}}{\Gamma(r+v+\theta)\Gamma(r_3+v_3-1+\theta_3)\sigma_1^{r_1+v_1+\theta_1}\sigma_2^{r_2+v_2+\theta_2}}$$
$$\times \left(1+\frac{(2-r_3-v_3-\theta_3)(r_1+v_1+\theta_1)}{\sigma_1}+\frac{(2-r_3-v_3-\theta_3)(r_2+v_2+\theta_2)}{\sigma_2}\right)$$
$$\times \left(1+\frac{(1-r_3-v_3-\theta_3)(r_1+v_1+\theta_1)}{\sigma_1}+\frac{(1-r_3-v_3-\theta_3)(r_2+v_2+\theta_2)}{\sigma_2}\right)^{-1}$$
$$=\frac{r_3+v_3-1+\theta_3}{r+v-1+\theta}\left(1+\frac{r_1+v_1+\theta_1}{\sigma_1}+\frac{r_2+v_2+\theta_2}{\sigma_2}+o(\sigma_1^{-1})\right).$$

The ratios in the mutation rates between individuals of type 1 and type 2 satisfy, for $i,j=1,2$ and $i\neq j$,

$$\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_j+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}$$
$$=\frac{\Gamma(r+v+\theta)\Gamma(\theta_3+\alpha_3)\sigma_i^{r_i+v_i+\theta_i}\sigma_j^{r_j+v_j+\theta_j}}{\Gamma(r+v+\theta)\Gamma(\theta_3+\alpha_3)\sigma_i^{r_i+v_i+\theta_i+1}\sigma_j^{r_j+v_j+\theta_j-1}}$$
$$\times \left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i+\theta_i+1)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j-1)}{\sigma_j}\right)$$
$$\times \left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i+\theta_i)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right)^{-1}$$
$$=\frac{\sigma_j}{\sigma_i}\left(1+\frac{1-r_3-v_3-\theta_3}{\sigma_i}+\frac{r_3+v_3-1+\theta_3}{\sigma_j}+o(\sigma_1^{-1})\right).$$

The ratios in the mutation rates of the two disfavoured types to the favoured type 3 satisfy, for $i,j=1,2$ and $i\neq j$,

$$\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_i-\boldsymbol{e}_3+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})}$$

$$
\begin{aligned}
&= \frac{\Gamma(r+v+\theta)\Gamma(r_3+v_3+\theta_3)\sigma_i^{r_i+v_i+\theta_i}\sigma_j^{r_j+v_j+\theta_j}}{\Gamma(r+v+\theta)\Gamma(r_3+v_3+\theta_3-1)\sigma_i^{r_i+v_i+\theta_i+1}\sigma_j^{r_j+v_j+\theta_j}} \\
&\quad \times \left(1+\frac{(2-r_3-v_3-\theta_3)(r_i+v_i+\theta_i+1)}{\sigma_i}+\frac{(2-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right) \\
&\quad \times \left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i+\theta_i)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right)^{-1} \\
&= \frac{r_3+v_3+\theta_3-1}{\sigma_i}\left(1+\frac{r_i+v_i+\theta_i-\theta_3-r_3-v_3+2}{\sigma_i}+\frac{r_j+v_j+\theta_j}{\sigma_j}+o(\sigma_1^{-1})\right).
\end{aligned}
$$

We also have, for $i,j=1,2$ and $i\neq j$,

$$
\begin{aligned}
&\frac{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{e}_3-\boldsymbol{e}_i+\boldsymbol{\theta})}{\nu(\boldsymbol{r}+\boldsymbol{v}+\boldsymbol{\theta})} \\
&= \frac{\Gamma(r+v+\theta)\Gamma(r_3+v_3+\theta_3)\sigma_i^{r_i+v_i+\theta_i}\sigma_j^{r_j+v_j+\theta_j}}{\Gamma(r+v+\theta)\Gamma(r_3+v_3+\theta_3+1)\sigma_i^{r_i+v_i+\theta_i-1}\sigma_j^{r_j+v_j+\theta_j}} \\
&\quad \times \left(1-\frac{(r_3+v_3+\theta_3)(r_i+v_i+\theta_i-1)}{\sigma_i}-\frac{(r_3+v_3+\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right) \\
&\quad \times \left(1+\frac{(1-r_3-v_3-\theta_3)(r_i+v_i+\theta_i)}{\sigma_i}+\frac{(1-r_3-v_3-\theta_3)(r_j+v_j+\theta_j)}{\sigma_j}\right)^{-1} \\
&= \frac{\sigma_i}{r_3+v_3+\theta_3}\left(1+\frac{r_3+v_3+\theta_3-r_i-v_i-\theta_i}{\sigma_i}-\frac{r_j+v_j+\theta_j}{\sigma_j}+o(\sigma_1^{-1})\right).
\end{aligned}
$$

The lemma then follows.

Bibliography

Abramowitz, M. and I. A. Stegun (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Dover Books on Mathematics. New York: Dover Publications.

Baake, E. and R. Bialowons (2008). Ancestral processes with selection: Branching and Moran models. In R. Bürger, C. Maes, and J. Miekisz (Eds.), *Banach Center Publications*, Volume 80, pp. 33–52. Warsaw: Institute of Mathematics, Polish Academy of Sciences.

Baake, E. and H. Wagner (2001). Mutation-selection models solved exactly with methods of statistical mechanics. *Genetical Research 78*, 93–117.

Bahlo, M. and R. C. Griffiths (2001). Coalescence time for two genes from a subdivided population. *Journal of Mathematical Biology 43*, 397–410.

Barton, N. H. (1998). The effect of hitch-hiking on neutral genealogies. *Genetical Research 72*, 123–133.

Barton, N. H., D. E. Briggs, J. A. Eisen, D. B. Goldstein, and N. H. Patel (2007). *Evolution*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Barton, N. H. and A. M. Etheridge (2004). The effect of selection on genealogies. *Genetics 166*, 1115–1131.

Barton, N. H., A. M. Etheridge, and A. K. Sturm (2004). Coalescence in a random background. *Annals of Applied Probability 14*, 754–785.

Bellman, R. (1997). *Introduction to Matrix Analysis* (2nd ed.), Volume 19 of *Classics in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics.

Charlesworth, B., M. T. Morgan, and D. Charlesworth (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics 134*, 1289–1303.

Charlesworth, D., B. Charlesworth, and M. Morgan (1995). The pattern of neutral molecular variation under the background selection model. *Genetics 141*, 1619–1632.

Coop, G. and R. C. Griffiths (2004). Ancestral inference on gene trees under selection. *Theoretical Population Biology 66*, 219–232.

Crow, J. F. and J. Felsenstein (1968). The effect of assortative mating on the genetic composition of a population. *Eugenics Quarterly 15*, 85–97.

Darden, T., N. L. Kaplan, and R. R. Hudson (1989). A numerical method for calculating moments of coalescent times in finite populations with selection. *Journal of Mathematical Biology 27*, 355–368.

Desai, M. M., L. E. Nicolaisen, A. M. Walczak, and J. B. Plotkin (2012). The structure of allelic diversity in the presence of purifying selection. *Theoretical Population Biology 81*, 144–157.

Devaux, C. and R. Lande (2008). Incipient allochronic speciation due to non-selective assortative mating by flowering time, mutation and genetic drift. *Proceedings of The Royal Society B 275*, 2723–2732.

Donnelly, P. (1984). The transient behaviour of the Moran model in population genetics. *Mathematical Proceedings of the Cambridge Philosophical Society 95*, 349–358.

Donnelly, P. and T. G. Kurtz (1999). Genealogical processes for Fleming-Viot models with selection and recombination. *Annals of Applied Probability 9*, 1091–1148.

Donnelly, P., M. Nordborg, and P. Joyce (2001). Likelihoods and simulation methods for a class of nonneutral population genetics models. *Genetics 159*, 853–867.

Donnelly, P. and S. Tavaré (1995). Coalescents and genealogical structure under neutrality. *Annual Review of Genetics 29*, 401–421.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow (1998). Rates of spontaneous mutation. *Genetics 148*, 1667–1686.

Durrett, R. (2008). *Probability Models for DNA Sequence Evolution* (2nd ed.). Proability and Its Applications. New York: Springer.

Durrett, R. and J. Schweinsberg (2004). Approximating selective sweeps. *Theoretical Population Biology 66*, 129–138.

Etheridge, A. M. (2011). *Some Mathematical Models from Population Genetics: Ecole d'été de probabilités de Saint-Flour XXXIX - 2009*, Volume 2012 of *Lecture Notes in Mathematics*. Berlin: Springer.

Etheridge, A. M. and R. C. Griffiths (2009). A coalescent dual process in a Moran model with genic selection. *Theoretical Population Biology 75*, 320–330.

Etheridge, A. M., R. C. Griffiths, and J. E. Taylor (2010). A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theoretical Population Biology 78*, 77–92.

Etheridge, A. M. and S. Lemaire (2011). Diffusion approximation of a multilocus model with assortative mating. *Electronic Journal of Probability 16*, 2122–2181.

Etheridge, A. M., P. Pfaffelhuber, and A. Wakolbinger (2009). How often does the ratchet click? Facts, heuristics, asymptotics. In J. Blath, P. Mörters, and M. Scheutzow (Eds.), *Trends in Stochastic Analysis*, Volume 353 of *London Mathematical Society Lecture Note Series*, pp. 365–390. Cambridge: Cambridge University Press.

Ethier, S. N. and R. C. Griffiths (1990). On the two-locus sampling distribution. *Journal of Mathematical Biology 29*, 131–159.

Ethier, S. N. and T. G. Kurtz (1998). Coupling and ergodic theorems for Fleming-Viot processes. *Annals of Probability 26*, 533–561.

Ethier, S. N. and T. G. Kurtz (2005). *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: Wiley-Interscience.

Ewens, W. (2004). *Mathematical Population Genetics: Theoretical Introduction* (2nd ed.), Volume 27 of *Interdisciplinary Applied Mathematics*. New York: Springer-Verlag.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology 3*, 87–112.

Fearnhead, P. (2002). The common ancestor at a nonneutral locus. *Journal of Applied Probability 39*, 38–54.

Fisher, R. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Fu, Y. X. and W. H. Li (1993). Statistical tests of neutrality of mutations. *Genetics 133*, 693–709.

Gillespie, J. H. (1986). Variability of evolutionary rates of DNA. *Genetics 113*, 1077–1091.

Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford Series in Ecology and Evolution. New York: Oxford University Press.

Golding, G. B. (1997). The effect of purifying selection on genealogies. In P. Donnelly and S. Tavaré (Eds.), *Progress in Population Genetics and Human Evolution*, Volume 87 of *IMA Volumes in Mathematics and its Applications*, pp. 271–285. New York: Springer.

Griffiths, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theoretical Population Biology 17*, 37–50.

Griffiths, R. C. (1989). Genealogical-tree probabilities in the infinitely-many-sites model. *Journal of Mathematical Biology 27*, 667–680.

Griffiths, R. C. and P. Marjoram (1996). Ancestral inference from samples of DNA sequences with recombination. *Journal of Computational Biology 3*, 479–502.

Griffiths, R. C. and P. Marjoram (1997). An ancestral recombination graph. In P. Donnelly and S. Tavaré (Eds.), *Progress in Population Genetics and Human Evolution*, Volume 87 of *IMA Volumes in Mathematics and its Applications*, pp. 257–270. New York: Springer.

Griffiths, R. C. and S. Tavaré (1994). Simulating probability distributions in the coalescent. *Theoretical Population Biology 46*, 131–159.

Haigh, J. (1978). The accumulation of deleterious genes in a population - Muller's ratchet. *Theoretical Population Biology 14*, 251–267.

Hale, J. K. (1980). *Ordinary Differential Equations* (2nd ed.). Huntington, New York: R. E. Krieger Publishing Company.

Hartl, D. L. and R. B. Campbell (1982). Allele multiplicity in simple Mendelian disorders. *American Journal of Human Genetics 34*, 866–873.

Hudson, R. R. (1991). Gene genealogies and the coalescent process. In D. Futuyama and J. Antonovics (Eds.), *Oxford Surveys in Evolutionary Biology*, Volume 7, pp. 1–44. New York: Oxford University Press.

Hudson, R. R. and N. L. Kaplan (1994). Gene trees with background selection. In G. B. Golding (Ed.), *Nonneutral Evolution: Theories and Molecular Data*, pp. 140–152. New York: Springer.

Hudson, R. R. and N. L. Kaplan (1996). The coalescent process and background selection. In P. H. Harvey, A. J. L. Brown, J. M. Smith, and S. Nee (Eds.), *New Uses for New Phylogenies*, pp. 57–65. Oxford: Oxford University Press.

Kaplan, N. L., T. Darden, and R. R. Hudson (1988). The coalescent process in models with selection. *Genetics 120*, 819–829.

Kimura, M. and J. F. Crow (1964). The number of alleles that can be maintained in a finite population. *Genetics 49*, 725–738.

Kingman, J. F. C. (1982a). The coalescent. *Stochastic Processes and their Applications 13*, 235–248.

Kingman, J. F. C. (1982b). On the genealogy of large populations. *Journal of Applied Probability 19*, 27–43.

Kondrashov, A. S. and M. Shpak (1998). On the origin of species by means of assortative mating. *Proceedings of the Royal Society B 265*, 2273–2278.

Krone, S. M. and C. Neuhauser (1997). Ancestral processes with selection. *Theoretical Population Biology 51*, 210–237.

Lercher, M. J. and L. D. Hurst (2002). Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in Genetics 18*, 337–340.

Lessard, S. and J. Wakeley (2004). The two-locus ancestral graph in a subdivided population: convergence as the number of demes grows in the island model. *Journal of Mathematical Biology 48*, 275–292.

Liggett, T. M. (2004). *Interacting Particle Systems*, Volume 276 of *Grundlehren der mathematischen Wissenschaften*. New York: Springer.

López, J. L. and P. J. Pagola (2010). The confluent hypergeometric functions $M(a, b; z)$ and $U(a, b; z)$ for large b and z. *Journal of Computational and Applied Mathematics 233*, 1570–1576.

Möhle, M. (1998). A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Advances in Applied Probability 30*, 493–512.

Möhle, M. (1998). Robustness results for the coalescent. *Journal of Applied Probability 35*, 438–447.

Moran, P. A. P. (1958). Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society 54*, 60–71.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 1*, 2–9.

Nagylaki, T. (1980). The strong-migration limit in geographically structured populations. *Journal of Mathematical Biology 9*, 101–114.

Nagylaki, T. (2000). Geographical invariance and the strong-migration limit in subdivided populations. *Journal of Mathematical Biology 41*, 123–142.

Neuhauser, C. and S. M. Krone (1997). The genealogy of samples in models with selection. *Genetics 145*, 519–534.

Nicolaisen, L. E. and M. M. Desai (2012). Distortions in genealogies due to purifying selection. *Molecular Biology and Evolution 29*, 3589–3600.

Nordborg, M. (2007). Coalescent theory. In D. J. Balding, M. Bishop, and C. Cannings (Eds.), *Handbook of Statistical Genetics* (3rd ed.), Volume 2, Chapter 25, pp. 843–877. Chichester: Wiley.

Nordborg, M. and P. Donnelly (1997). The coalescent process with selfing. *Genetics 146*, 1185–1195.

Nordborg, M. and S. M. Krone (2002). Separation of time scales and convergence to the coalescent in structured populations. In M. Slatkin and M. Veuille (Eds.), *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot*, Oxford Biology Readers, pp. 194–232. Oxford: Oxford University Press.

Norman, M. F. (1974). A central limit theorem for Markov processes that move by small steps. *Annals of Probability 2*, 1065–1074.

Notohara, M. (1990). The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology 29*, 59–75.

Notohara, M. (1993). The strong-migration limit for the genealogical process in geographically structured populations. *Journal of Mathematical Biology 31*, 115–122.

O'Brien, P. (1985). A genetic model with mutation and selection. *Mathematical Biosciences 73*, 239–251.

Olver, F. W. J., D. W. Lozier, R. F. Boisvert, and C. W. Clark (2010). *NIST Handbook of Mathematical Functions*. Cambridge: Cambridge University Press.

Przeworski, M., B. Charlesworth, and J. D. Wall (1999). Genealogies and weak purifying selection. *Molecular Biology and Evolution 16*, 246–252.

Ripley, B. (2006). *Stochastic Simulation*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: Wiley-Interscience.

Sato, K. (1978). Convergence to a diffusion of a multi-allelic model in population genetics. *Advances in Applied Probability 10*, 538–562.

Sawyer, S. (1983). A stability property of the Ewens sampling formula. *Journal of Applied Probability 20*, 449–459.

Slade, P. F. (2000). Simulation of selected genealogies. *Theoretical Population Biology 57*, 35–49.

Stephan, W., T. H. E. Wiehe, and M. W. Lenz (1992). The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. *Theoretical Population Biology 41*, 237–254.

Stephens, M. and P. Donnelly (2003). Ancestral inference in population genetics models with selection. *Australian and New Zealand Journal of Statistics 45*, 395–423.

Takahata, N. (1988). The coalescent in two partially ioslated diffusion populations. *Genetical Research 52*, 213–222.

Tavaré, S. (2004). Ancestral inference in population genetics. In J. Picard (Ed.), *Lectures on Probability Theory and Statistics: Ecole d'été de probabilités de Saint-Flour XXXI - 2001*, Volume 1837 of *Lecture Notes in Mathematics*, pp. 1–188. New York: Springer.

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly (1997). Inferring coalescence times from DNA sequence data. *Genetics 145*, 505–518.

Thompson, C. and J. McBride (1974). On Eigen's theory of the self-organization of

matter and the evolution of biological macromolecules. *Mathematical Biosciences 21*, 127–142.

Wakeley, J. (2008). Conditional gene genealogies under strong purifying selection. *Molecular Biology and Evolution 25*, 2615–2626.

Wakeley, J. (2009). *Coalescent Theory: An Introduction*. Greenwood Village, Colorado: Roberts & Co. Publishers.

Wakeley, J. and O. Sargsyan (2009). The conditional ancestral selection graph with strong balancing selection. *Theoretical Population Biology 75*, 355–364.

Walczak, A. M., L. E. Nicolaisen, J. B. Plotkin, and M. M. Desai (2012). The structure of genealogies in the presence of purifying selection: A fitness-class coalescent. *Genetics 190*, 753–779.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology 7*, 256–276.

Wilkinson-Herbots, H. M. (1998). Genealogy and subpopulation differentiation under various models of population structure. *Journal of Mathematical Biology 37*, 535–585.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics 16*, 97–159.

Yu, F., A. M. Etheridge, and C. Cuthbertson (2012). Fixation probability for competing selective sweep. *Electronic Journal of Probability 17*, 1–36.