Looking down in the ancestral selection graph:
A probabilistic approach to the common ancestor type distribution

Ute Lenz\textsuperscript{a,1}, Sandra Kluth\textsuperscript{b,2}, Ellen Baake\textsuperscript{b,\ast}, Anton Wakolbinger\textsuperscript{a,3}

\textsuperscript{a}Institut für Mathematik, Goethe-Universität Frankfurt, Box 111932, 60054 Frankfurt am Main, Germany
\textsuperscript{b}Faculty of Technology, Bielefeld University, Box 100131, 33501 Bielefeld, Germany

Abstract
In a (two-type) Wright-Fisher diffusion with directional selection and two-way mutation, let \(x\) denote today’s frequency of the beneficial type, and given \(x\), let \(h(x)\) be the probability that, among all individuals of today’s population, the individual whose progeny will eventually take over in the population is of the beneficial type. Earnhead [Earnhead, P., 2002. The common ancestor at a nonneutral locus. J. Appl. Probab. 39, 38-54] and Taylor [Taylor, J. E., 2007. The common ancestor process for a Wright-Fisher diffusion. Electron. J. Probab. 12, 808-847] obtained a series representation for \(h(x)\). We develop a construction that contains elements of both the ancestral selection graph and the lookdown construction and includes pruning of certain lines upon mutation. Besides interest in its own right, this construction allows a transparent derivation of the series coefficients of \(h(x)\) and gives them a probabilistic meaning.

Keywords: common ancestor type distribution, ancestral selection graph, lookdown graph, pruning, Wright-Fisher diffusion with selection and mutation

1. Introduction
The understanding of ancestral processes under selection and mutation is among the fundamental challenges in population genetics. Two central concepts are the ancestral selection graph (ASG) and the lookdown (LD) construction. The ancestral selection graph [Krone and Neuhauser, 1997; Neuhauser and Krone, 1997; see also Shiga and Uchiyama, 1986 for an analogous construction in a diffusion model with spatial structure] describes the set of lines that are potential ancestors of a sample of individuals taken from a present population. In contrast, the lookdown construction [Donnelly and Kurtz, 1999] is an integrated representation that makes all individual lines in a population explicit, together with the genealogies of arbitrary samples. See Etheridge (2011, Chapter 5) for an excellent overview of the area.

Both ASG and LD are important theoretical concepts as well as valuable tools in applications. Interest is usually directed towards the genealogy of a sample, backwards in time until the most recent common ancestor (MRCA). However, the ancestral line that continues beyond the MRCA into the distant past is of considerable interest on its own, not least because it

\textsuperscript{\ast}Corresponding author. Phone: +49–521–106–4896. E-mail: ebaake@techfak.uni-bielefeld.de (E. Baake).
\textsuperscript{1}E-mail: lenz@math.uni-frankfurt.de.
\textsuperscript{2}Present address: Department of Internal Medicine I and Center for Integrated Oncology, University Hospital Cologne, Kerpener Strasse 62, 50937 Koeln. E-mail: sandra.kluth@uni-bielefeld.de.
\textsuperscript{3}E-mail: wakolbin@math.uni-frankfurt.de.

Preprint submitted to Theoretical Population Biology September 3, 2014
links the genealogy (of a sample from a population) to the longer time scale of phylogenetic trees. The extended time horizon then shifts attention to the asymptotic properties. The stationary type distribution on the ancestral line may differ substantially from the stationary type distribution in the population. This mirrors the fact that the ancestral line consists of those individuals that are successful in the long run; thus, its type distribution is expected to be biased towards the favourable types.

When looking at the evolution of the system in (forward) time $[0, \infty)$, one may ask for properties of the so-called immortal line, which is the line of descent of those individuals whose offspring eventually takes over the entire population. In other words, the immortal line restricted to any time interval $[0, t]$ is the common ancestral line of the population back from the far future. It then makes sense to consider the type of the immortal line at time 0. To be specific, let us consider a Wright-Fisher diffusion with two types of which one is more and one is less fit. The common ancestor type (CAT) distribution at time 0, conditional on the type frequencies $(x, 1 - x)$, then has weights $(h(x), 1 - h(x))$, where $h(x)$ is the probability that the population ultimately consists of offspring of an individual of the beneficial type, when starting with a frequency $x$ of beneficial individuals at time 0. 

In the case without mutations (in which $h(x)$ coincides with the classical fixation probability of the beneficial type starting from frequency $x$), 

Mano (2009) and Pokalyuk and Pfaffelhuber (2013) have represented $h(x)$ in terms of the equilibrium ASG, making use of a time reversal argument (see Section 2.2). However, the generalisation to the case with mutation is anything but obvious. One purpose of this article is to solve this problem. A key ingredient will be a combination of the ASG with elements of the lookdown construction, which also seems of interest in its own right.

The paper is organised as follows. In Section 2, we start by briefly recapitulating the ASG (starting from the Moran model for definiteness). We then recall the Fearnhead-Taylor representation of $h(x)$ and give its explanation in terms of the equilibrium ASG in the case without mutations, inspired by Pokalyuk and Pfaffelhuber (2013). In Section 3, we prepare the scene by ordering the lines of the ASG in a specific way; in Section 4, we then represent the ordered ASG in terms of a fixed arrangement of levels, akin to a lookdown construction. In Section 5, a pruning procedure is described that reduces the number of lines upon mutation. The stationary number of lines in the resulting pruned LD-ASG will provide the desired connection to the (conditional) common ancestor type distribution. Namely, the tail probabilities of the number of lines turn out as the coefficients in the series representation. In Section 6, the graphical approach will directly reveal various monotonicity properties of the tail probabilities as functions of the model parameters, which translate into monotonicity properties of the common ancestor type distribution. Section 7 is an add-on, which makes the connection to Taylor’s boundary value problem for $h(x)$ explicit; Section 8 is devoted to some concluding remarks.
2. Concepts and models

2.1. The Moran model and its diffusion limit

Let us consider a haploid population of fixed size \( N \in \mathbb{N} \) in which each individual is characterised by a type \( i \in S := \{0, 1\} \). An individual of type \( i \) may, at any instant in continuous time, do either of two things: It may reproduce, which happens at rate 1 if \( i = 1 \) and at rate \( 1 + s_N \), \( s_N \geq 0 \), if \( i = 0 \); or it may mutate to type \( j \) at rate \( u_N \nu_j \), \( u_N \geq 0 \), \( 0 \leq \nu_j \leq 1 \), \( \nu_0 + \nu_1 = 1 \). If an individual reproduces, its single offspring inherits the parent’s type and replaces a randomly chosen individual, maybe its own parent. Concerning mutations, \( u_N \) is the total mutation rate and \( \nu_j \) the probability of a mutation to type \( j \). Note that the possibility of silent mutations from type \( j \) to type \( j \) is included.

Figure 1: The Moran model with two-way mutation and selection. The types are indicated for the initial population (left) and the final one (right). Crosses represent mutations to type 1, circles mutations to type 0. Selective events are depicted as arrows with star-shaped heads.

The Moran model has a well-known graphical illustration as an interacting particle system (cf. Fig. 1). The individuals are represented by horizontal line pieces, with forward time running from left to right in the figure. Arrows indicate reproduction events with the parent at its tail and the offspring at its head. For later use, we decompose reproduction events into neutral and selective ones. Neutral arrows appear at rate \( 1/N \), selective arrows (those with a star-shaped arrowhead in Fig. 1) at rate \( s_N/N \) per ordered pair of lines, irrespective of their types. The rates specified above are obtained by the convention that neutral arrows may be used by all individuals, whereas selective arrows may only be used by type-0 individuals and are ignored otherwise. Mutations to type 0 are marked by circles, mutations to type 1 by crosses.

The usual diffusion rescaling in population genetics is applied, i.e. rates are rescaled such that \( \lim_{N \to \infty} N s_N = \sigma \) and \( \lim_{N \to \infty} Nu_N = \theta \), \( 0 \leq \sigma, \theta < \infty \), and time is sped up by a factor of \( N \). Let \( X_t \) be the frequency of type-0 individuals at time \( t \) in this diffusion limit. Then, the process \( (X_t)_{t \in \mathbb{R}} \) is a Wright-Fisher diffusion which is characterised by the drift coefficient \( a(x) = (1-x)\theta \nu_0 - x \theta \nu_1 + x(1-x)\sigma \) and the diffusion coefficient \( b(x) = 2x(1-x) \). The stationary density \( \pi \) is given by \( \pi(x) = C(1-x)^{\theta \nu_1 - 1} x^{\theta \nu_0 - 1} \exp(\sigma x) \), where \( C \) is a normalising constant (cf. Durrett (2008, Chapters 7, 8) or Ewens (2004, Chapters 4, 5)).

2.2. The ancestral selection graph

The ancestral selection graph was introduced by Krone and Neuhauser (1997) and Neuhauser and Krone (1997) to construct samples from a present population, together with their ancestrys, in the diffusion limit of the Moran model with mutation and selection. The basic idea...
is to understand selective arrows as unresolved reproduction events backwards in time: the descendant has two potential ancestors, the incoming branch (at the tail) and the continuing branch (at the tip), see also Fig. 2. The incoming branch is the ancestor if it is of type 0, otherwise the continuing one is ancestral. For a hands-on exposition, see Wakeley (2009, Chapter 7.1).

![Figure 2: Incoming branch (I), continuing branch (C), and descendant (D). The ancestor is marked bold.](image)

The ASG is constructed by starting from the (as yet untyped) sample and tracing back the lines of all potential ancestors. In the finite graphical representation, a neutral arrow that joins two potential ancestral lines appears at rate $2/N$ per currently extant pair of potential ancestral lines, then giving rise to a coalescence event, i.e. the two lines merge into a single one. In the same finite setting, a selective arrow that emanates from outside the current set of $n$ potential ancestral lines and hits this set appears at rate $n(N - n)s_N/N$. This gives rise to a branching event, i.e., viewed backwards in time, the line that is hit by the selective arrow splits into an incoming and continuing branch as described above. Thus in the diffusion limit, since $N - n \sim N$ as $N \to \infty$, the process $(Z_r)_{r \in \mathbb{R}}$, where $Z_r$ is the number of lines in the ASG at time $r = -t$, evolves backwards in time with rates

$$\hat{q}_Z(n, n-1) = n(n-1), \quad \hat{q}_Z(n, n+1) = n\sigma, \quad n \in \mathbb{N}. \quad (1)$$

(Throughout, backward time will be indicated by ‘$r$’, and corresponding rates marked by ‘$\hat{}$’.) At a coalescence event a randomly chosen pair of lines coalesces, at a branching event a randomly chosen line splits into two.

The (reversible) equilibrium distribution of the dynamics (1) turns out to be the Poisson($\sigma$)-distribution conditioned to $\{1, 2, \ldots\}$, i.e.

$$P(Z_r = n) = \frac{\sigma^n}{n!(\exp(\sigma) - 1)}, \quad n \in \mathbb{N}. \quad (2)$$

We may construct the equilibrium ASG as in Pokalyuk and Pfaffelhuber (2013) in a two-stage way: first take a random path $(Z_r)_{-\infty < r < \infty}$, and then fill in the branching and coalescence events, with a random choice of one of the $Z_r$ lines at each upward jump, and of one of the $(Z_r^2)$ pairs at each downward jump of $(Z_r)$. Mutation events (at rates $\theta\nu_0$ and $\theta\nu_1$) are superposed on the lines of the (equilibrium) ASG by Poisson processes with rates $\theta\nu_0$ and $\theta\nu_1$. Given

---

4Since our population size is $N$ (rather than $2N$), the selection coefficient $\sigma$ in our scenario corresponds to $\sigma/2$ and our $n(n-1)$ to $n(n-1)/2$ in Krone and Neuhauser (1997).
the frequency $x$ of the beneficial type at time 0, one then assigns types to the lines of the ASG in the (forward) time interval $[0, \infty)$ by first drawing the types of the lines at time 0 independently and identically distributed (i.i.d.) from the weights $(x, 1-x)$, and propagates the types forward in time, respecting the mutation events. In this way, the (backward in time) branching events may now be resolved into the true parent and a fictitious parent.

Note that there are various ways to illustrate the same realisation of the ASG graphically. See, for instance, Fig. 3 with backward time $r$ running from right to left. The left and right panels of Fig. 3 represent the same realisation of the ASG, but differ in the ordering of the lines.

2.3. The common ancestor

In the population, at any time $t$, there almost surely exists a unique individual that is, at some time $s > t$, ancestral to the whole population; cf. Fig. 4. The descendants of this individual become fixed, and we call it the common ancestor at time $t$. The lineage of these distinguished individuals over time defines the so-called ancestral (or immortal) line.

Looking at the population at time $t$, say $t = 0$, we are interested in $h(x)$, the probability that the common ancestor is of type 0, given $X_0 = x$. Equivalently, one may understand $h(x)$ as the fixation probability of the offspring of all type-0 individuals regardless of the offspring’s types, if $X_0 = x$. Working in a stationary picture, $h(x)$ does depend on the type-0 frequency $x$ at that time but not on the time itself. According to previous results by Fearnhead (2002) and Taylor (2007), it reads

$$h(x) = x + \sum_{n \geq 1} a_n x (1-x)^n,$$  \hspace{1cm} (3)
where the coefficients $a_n$, $n \geq 1$, are characterised by the recursion

$$(n + 1 + \theta \nu_1)a_{n+1} - (n + 1 + \sigma + \theta)a_n + \sigma a_{n-1} = 0 \quad (4)$$

under the constraints

$$1 = a_0 \geq a_1 \geq \cdots, \quad \lim_{n \to \infty} \frac{a_{n+1}}{a_n} = 0. \quad (5)$$

These results were reviewed by Baake and Bialowons (2008), and re-obtained with the help of a descendant process (forward in time) by Kluth et al. (2013).

In order to get a handle on the representation (3) and the recursion (4) in terms of the equilibrium ASG, one observes that the type of the common ancestor at time $t = 0$ may be recovered in the following way. In the equilibrium ASG marked with the mutation events (as described in Section 2.3), assign i.i.d types to the lines at time $0$ and propagate them forward in time, respecting the mutation events. The immortal line is then encoded in the realisation of the marked ASG.

The event of fixation of the beneficial type is easily described in the case without mutations. First, recall that, as stated in Section 2.3, the number $Z_0$ of lines in the equilibrium ASG at time $0$ is Poisson distributed with parameter $\sigma$, conditioned to be positive. Next, note that the equilibrium ASG with probability 1 has bottlenecks, i.e. times at which it consists of a single line. Let $t_0$ be the smallest among all the non-negative times at which there is a bottleneck (see Fig. 5). This way, the unique individual is identified that is the true ancestor of the single individual at forward time $t_0$ and, at the same time, of the entire equilibrium ASG at any later time (and ultimately of the entire population).

![Figure 5: A realisation of the equilibrium ASG, with its first bottleneck (after time 0) at time $t_0$.](image)

As observed by Mano (2009) and Pokalyuk and Pfaffelhuber (2013), type 0 becomes fixed if and only if the single line at time $t_0$ carries type 0, and this, in turn, happens if and only if at least one ancestral line at time $t = 0$ is of type 0. The latter probability is $1 - (1 - x)Z_0$, given that the frequency of type-0 individuals is $x$ at this time. Therefore, with the help of (2), the fixation probability can be obtained as

$$h(x) = \mathbb{E}(1 - (1 - x)Z_0) = \frac{1}{\exp(\sigma) - 1} \sum_{n \geq 1} \left(1 - (1 - x)^n\right) \frac{\sigma^n}{n!} = \frac{1 - \exp(-\sigma x)}{1 - \exp(-\sigma)}, \quad (6)$$

which coincides with the classical result of Kimura (1962). Putting $\gamma_n := \mathbb{P}(Z_0 \geq n + 1)$,
$n \geq 0$, the left-hand side of (6) may also be expressed as

$$h(x) = \sum_{n \geq 1} (1 - (1 - x)^n) \left[ \Pr(Z_0 \geq n) - \Pr(Z_0 \geq n + 1) \right] = x + \sum_{n \geq 1} \gamma_n x (1 - x)^n,$$

which is the representation (3). (Indeed, one checks readily that the tail probabilities $\gamma_n$ satisfy the recursion (4) in the case $\theta = 0$.) The elegance of this approach lies in the fact that one does not need to know the full representation of the ASG, in particular one does not need to distinguish between incoming and continuing branches. As soon as mutations are included, however, keeping track of the hierarchy of the branches becomes a challenge. We thus aim at an alternative representation of the ASG that allows for an orderly bookkeeping leading to a generalisation of the idea above, and yields a graphical interpretation of (3)-(5). This will be achieved in the next three sections.

3. The ordered ASG

In the previous section, we have reminded ourselves that one may represent the same realisation of one ASG in different ways. In the following, we propose a construction, denoted by ordered ASG, which is obtained backwards in time from a given realisation of the ASG as follows (compare Fig. 6).

- Coalescence: Each coalescence event is represented by a (neutral) arrow pointing from the lower participating line to the upper one. The (single) parental line continues the lower one of the two lines.

- Branching: A selective arrow with star-shaped head is pointed towards the splitting line at a branching event. The incoming branch is always placed directly beneath the continuing branch at the tail of the arrow; in particular, there are no lines between incoming and continuing branch at the time of the branching event.

- Mutation: Mutations, symbolised here by circles and crosses, occur along the lines as in the original ASG.

The ordered ASG corresponding to both representations in Fig. 3 is shown in Fig. 6.

![Figure 6: The ordered ASG corresponding to Fig. 3 or directly deduced from Fig. 1](image-url)
4. The lookdown ASG

To each point in the ordered ASG, let us introduce two coordinates: its time and its level, the latter being an element of \( \{0, 1, 2, \ldots\} \). Since this is in close analogy to ideas known from the lookdown processes by Donnelly and Kurtz (1999a,b), we call this construction the lookdown ASG (LD-ASG). It can be obtained backwards in time from a given realisation of the ordered ASG, or it may as well be constructed in distribution via Poissonian elements representing coalescence, branching, and mutation. The two possibilities are described in Sections 4.1 and 4.2 respectively.

4.1. Construction from a given realisation of the ordered ASG

Backwards in time, the realisation of the LD-ASG corresponding to a given realisation of the ordered ASG is obtained in the following way. Start with all \( n \) individuals respectively lines that are present in the (ordered) ASG and place them at levels 0 to \( n-1 \) by adopting their vertical order from the ordered ASG. Then let the following events happen.

- **Coalescence**: Coalescence events between levels \( i \) and \( j > i \) are treated the same way as in the ordered ASG: The remaining branch continues at level \( i \). In addition, all lines at levels \( n > j \) are shifted one level downwards to \( n-1 \) (cf. Fig. 7, left).

- **Branching**: A selective arrow with star-shaped head in the ordered ASG is translated into a star at the level \( i \) of the branching line. The incoming branch emanates out of the star at the same level and all lines at levels \( n \geq i \) are pushed one level upwards to \( n+1 \). In particular, the continuing branch is shifted to level \( i+1 \) (cf. Fig. 7, right).

- **Mutation**: Mutations (symbolised again as circles and crosses) are taken from the ordered ASG.

![Figure 7: Coalescence (left) and branching event (right) in the LD-ASG.](image)

Fig. 8 gives a realisation that corresponds to the realisation of the ordered ASG in Fig. 6. Note that we obviously have a bijection between realisations of the ordered ASG and the LD-ASG and that the LD-ASG is just a neat arrangement of the ordered ASG.

4.2. Construction from elements of Poisson point processes

The LD-ASG may, in distribution, as well be constructed backwards in time via the elements ‘arrows’, ‘stars’, ‘circles’ and ‘crosses’ arising as representations of independent Poisson point processes:
• **Coalescence:** For each ordered pair of levels \((i, j)\), where \(i < j\) and level \(j\) is occupied by a line, arrows from level \(i\) to \(j\) emerge independently according to a Poisson point process \(\Gamma_{ij}^\uparrow\) at rate 2. An arrow from \(i\) to \(j\) is understood as a coalescence of the lines at levels \(i\) and \(j\) to a single line on level \(i\). In addition, all lines at levels \(n > j\) are shifted one level downwards to \(n - 1\) (cf. Fig. 7, left).

• **Branching:** On each occupied level \(i\) stars appear according to independent Poisson point processes \(\Gamma_i^\ast\) at rate \(\sigma\). A star at level \(i\) indicates a branching event, where a new line, namely the incoming branch, is inserted at level \(i\) and all lines at levels \(n \geq i\) are pushed one level upwards to \(n + 1\). In particular, the continuing branch is shifted to level \(i + 1\) (cf. Fig. 7, right).

• **Mutation:** Mutations to type 0 and type 1, i.e. circles and crosses, occur via independent Poisson point processes \(\Gamma_i^\circ\) at rate \(\theta \nu_0\) and \(\Gamma_1^\times\) at rate \(\theta \nu_1\), respectively, on each occupied level \(i\).

The independent superposition of these Poisson point processes and their effects on the lines characterises the LD-ASG. Let \(K_r\) be the highest occupied level at time \(r\). Since \(K_r = Z_r - 1\) (recall that \((Z_r)_{r \in \mathbb{R}}\) is the line counting process of the (ordered) ASG), the transition rates of the process \((K_r)_{r \in \mathbb{R}}\), which also evolves backwards in time, follow directly from (1) and are given by

\[
\hat{q}_{K}(n, n - 1) = \hat{q}_{Z}(n + 1, n) = n(n + 1),
\]

\[
\hat{q}_{K}(n, n + 1) = \hat{q}_{Z}(n + 1, n + 2) = (n + 1)\sigma,
\]  

when in state \(n, n \in \mathbb{N}_0\).

Note that, although we will ultimately rely on the ASG in equilibrium only, neither the ordering of the ASG nor the LD-ASG construction are restricted to the equilibrium situation. The equilibrium comes back in when we search for the immortal line, which will be done next.

**4.3. The immortal line in the LD-ASG in the case without mutations**

We consider a realisation \(\mathcal{L}\) of the equilibrium LD-ASG, write \(K_r\) for its highest occupied level at (backward) time \(r\), and again write

\[
t_0 = t_0(\mathcal{L}) := -\sup\{r \leq 0 : K_r = 0\}
\]

for the smallest (forward) time at which \(\mathcal{L}\) has a ‘bottleneck’, see Fig. 9.
The level of the immortal line at time 0 does not only depend on \( L \), but also on the types \( I^0, \ldots, I^{K_0} \in \{0, 1\} \) that are assigned to the levels \( 0, 1, \ldots, K_0 \) at time 0. Irrespective of this type assignment, we will consider a distinguished line which is constructed as follows: Back from each bottleneck of \( L \), this line goes up one level at each branching event that happens at a level smaller or equal to its current level, and follows the coalescence events in a lookdown manner, see the bold line in the right panel of Fig. 11. In particular, this line follows the continuing branch whenever it is hit by a branching event at its current level. For reasons that will become clear in the following section, we will call this distinguished line the immune line.

The rule for the choice of the ‘true ancestral line’ depending on the types at a (backward) branching event that is depicted in Fig. 2 translates into the lookdown framework as illustrated in Fig. 10.

Now let us distinguish a few cases:

If \( I^0 = \cdots = I^{K_0} = 1 \), then the immortal line forward from time 0 equals the immune line, since at any (backwards in time) branching event it must constitute the continuing branch (this is the case illustrated in the right panel of Fig. 11).

If \( I^0 = 0 \), then, since the line at level 0 is incoming to all branching events at that level, the immortal line is the level 0 line.

If \( K^0 \geq 1, I^0 = 1, I^1 = 0 \), the line starting at level 1 is parental to the branching event with the line at level 0. Since it is incoming to all other branching events by which it is affected, it is parental to them as well and thus immortal (compare Fig. 11 left panel). Continuing this chain of thought, for all \( n \in \{0, \ldots, K_r\} \) the event \( \{I^0 = \cdots = I^{n-1} = 1, I^n = 0\} \) results in
Figure 11: LD-ASG with types. The level of the immortal line (bold) starting out from time 0 depends on the type assignment at time 0.

the immortal line being at level \( n \) at time 0. We thus have proved the following

**Proposition 1** *In the absence of mutations, the level of the immortal line in the equilibrium LD-ASG with types assigned at time 0 is either the lowest type-0 level or, if all lines at time 0 are of type 1, it is the level of the immune line, i.e. the line that constitutes the continuing branch in all branching events by which it is affected (cf. Fig. 11).*

Proposition 1 specifies the immortal line in the case of selection only. The aim in the next section is to establish an analogous statement in the case of selection and mutation.

5. The pruned equilibrium LD-ASG and the CAT distribution

We now consider the equilibrium LD-ASG marked with the mutation events. Working backwards from the bottleneck time \( t_0 \) (see Eq. (8) and Figs. 8 and 9), we see that the mutation events that occur along the lines may eliminate some of them as candidates for being the immortal line. Cutting away certain branches that carry no information has been used, explicitly or implicitly, in various investigations of the ASG (e.g., by Slade (2000), Fearnhead (2002), Arthreya and Sward (2005), and Etheridge and Griffiths (2009)), but our construction requires a specific pruning procedure which we now describe.

5.1. Pruning the LD-ASG

Denote by \( t' \) the time at which the most recent event back from \( t_0 \) occurs. If this event was a branching or coalescing event (denoted by a star or an arrow, respectively), proceed according to the instructions given in Chapter 4 (see Fig. 7). Note that the immune line always serves as continuing branch at a branching event and therefore moves up one level at each branching event that happens at a level smaller or equal to its current level. When included in a coalescence event with a line at a lower level, it is shifted downwards. In this way, the immune line may drop to levels below the top.

If the event at time \( t' \) was a mutation, this mutation can be either deleterious or beneficial. First consider the case in which this mutation is deleterious (symbolised by a cross). Since there is no mutation between times \( t' \) and \( t_0 \), Proposition 1 applies (with time 0 replaced by time \( t' \)), showing that the line that carries the deleterious mutation at time \( t' \) cannot be the immortal one unless it is the immune line. In our search for the true ancestor at time 0 of the single line that starts from time \( t_0 \), we can therefore erase the line that carries the deleterious mutation at time \( t' \), unless this line is the immune one.

Next consider the case in which the mutation occurring at time \( t' \) is beneficial (symbolised by a circle) and happens at level \( \ell' \), say. Then, again appealing to Proposition 1 we see that
none of the lines that occupy levels \( \ell' > \ell \) at time \( t' \) can be parental to the single line starting from time \( t_0 \). We can therefore erase all these lines from the list of candidates. If all lines were assigned the deleterious type just left of \( t' \), then the line on level \( \ell' \) would become the only one with a beneficial type and therefore would be the immortal one. Thus, the immune line is relocated to level \( \ell' \) at time \( t' \).

Figure 12: Pruning procedure in the LD-ASG due to beneficial (left) and deleterious (right) mutations.

Proceeding to the next mutation event on the remaining lines back from time \( t' \) (which happens at time \( t'' \) at level \( \ell'' \), say), we can iterate this procedure: if the mutation is deleterious, the line is killed unless it is the immune one. If the mutation is beneficial, all the lines at higher levels are killed, with the line back from \( t'' \) at level \( \ell'' \) being declared the new immune line. Having worked back to \( t = 0 \), we arrive at the pruned LD-ASG between times \( 0 \) and \( t_0 \). Note that except for the immune line, due to the pruning procedure there are no mutations on any line of the pruned LD-ASG. In other words, each line present at time \( 0 \) is the immune line, or has no mutations on it between times \( 0 \) and \( \zeta \), where \( \zeta \) is the time when that line was incoming to a branching event with the immune line. Note also that beneficial mutations can only be present on the current top level of the pruned LD-ASG.

5.2. Finding the immortal line in the pruned LD-ASG

For a given assignment of types at time \( 0 \) one can now identify the immortal line in the same way as in the LD-ASG in the case of selection only (Proposition 1). We have thus proved

**Theorem 2** The level of the immortal line in the pruned equilibrium LD-ASG with types assigned at time \( 0 \) is either the lowest type-0 level (if there exists such a level) or otherwise (if all lines are of type 1) the level of the immune line.

As in Section 4.2 we can construct the pruned LD-ASG (together with the level of the immune line) backwards in time in a Markovian way, using the Poisson processes \( \Gamma_{ij} \) and \( \Gamma_i^* \) (for all
occupied levels \(i\) and \(j\), cf. Fig. 7, and \(\Gamma_i^x\) and \(\Gamma_i^y\), where the pruning procedure is applied as described above (cf. Fig. 12). Fig. 13 gives a realisation that corresponds to the realisation of the LD-ASG in Fig. 8.

The process \((L_r)_{r \in \mathbb{R}}\), where \(L_r\) is the level of the top line at time \(r\), evolves backwards in time with transition rates

\[
\begin{align*}
\hat{q}_L(n, n-1) &= n(n+1) + n\theta + \nu_0, \\
\hat{q}_L(n, n+1) &= (n+1)\sigma, \\
\hat{q}_L(n, n-\ell) &= \theta \nu_0, \quad 2 \leq \ell \leq n, \quad n \in \mathbb{N}_0.
\end{align*}
\]

(9)

**Remark 3** The process \(L\) is stochastically dominated by the the process \(K\) (the highest level of the unpruned LD-ASG), and thus by \(Z\) (the line counting process of the unpruned ASG).

In fact, using the above-described pruning procedure in a time-stationary picture between all the bottlenecks of the equilibrium ASG line counting process \(Z = (Z_r)_{r \in \mathbb{R}}\), we obtain that \(L_r \leq K_r = Z_r - 1\) for all \(r \in \mathbb{R}\).

The stochastic dynamics induced by (9) has a unique equilibrium distribution, which we denote by \(\rho\). In the following, let \(L = (L_r)_{r \in \mathbb{R}}\) be the time-stationary process with jump rates (9), hence

\[\rho_n = \mathbb{P}(L_0 = n), \quad n \in \mathbb{N}_0.\]

5.3. The CAT distribution via the pruned equilibrium LD-ASG

With the help of Lemma 2, it is now possible to provide an interpretation of the probability \(h(x)\) that the common ancestor is of type 0, given that the frequency of the beneficial type at time 0 is \(x\). Let \(I_k \in \{0, 1\}\) be the type that is assigned to the individual at level \(k \in \{0, \ldots, L_0\}\) in the pruned equilibrium LD-ASG at time 0. Conditional on the initial frequency of the beneficial type being \(x\), these types are assigned in an i.i.d. manner with \(\mathbb{P}(I_k = 0) = x\).

The quantity \(h(x)\) thus is the probability that at least one of a random number of i.i.d. trials is a success, where the random number of trials is \(L_0 + 1\) and the success probability is \(x\). A decomposition of \(h(x)\) according to the first level which is occupied by type 0 yields

\[
h(x) = \sum_{n \geq 0} \mathbb{P}(I^n = 0, I^k = 1 \ \forall k < n, \ L_0 \geq n) = \sum_{n \geq 0} x(1-x)^n \mathbb{P}(L_0 \geq n).
\]

(10)

We have thus proved

**Theorem 4** Given the frequency of the beneficial type at time 0 is \(x\), the probability that the common ancestor at time 0 is of beneficial type is

\[
h(x) = \sum_{n \geq 0} x(1-x)^n \alpha_n,
\]

(11)

where \((\alpha_n)\) are the tail probabilities of the distribution of the number of lines in the pruned equilibrium LD-ASG, i.e.

\[
\alpha_n := \mathbb{P}(L_0 \geq n) = \sum_{i \geq n} \rho_i, \quad n \in \mathbb{N}_0,
\]

(12)
and $\rho = (\rho_n)$ is the probability vector obeying

$$\rho Q = 0,$$  \hspace{1cm} (13)

with $Q$ being the generator matrix determined by the jump rates $\mathbf{q}$.

Comparing (3) and (11), it is clear that the tail probabilities $(\alpha_n)$ must obey the Fearnhead recursion. The proof of the following proposition gives a direct argument for this.

**Proposition 5** The tail probabilities $(\alpha_n)$ obey the recursion (4).

**Proof:** For $n \in \mathbb{N}$, the $n^{th}$ entry of the vector $\rho Q$ is

$$(\rho Q)_n = \rho_{n-1} q_L(n-1, n) + \rho_n q_L(n-1, n) + \sum_{j \geq n+2} \rho_j q_L(j, n) - \rho_n \left[ q_L(n, n-1) + q_L(n, n+1) + \sum_{\ell=0}^{n-2} q_L(n, \ell) \right].$$

Plugging in the jump rates (9), this can be written as

$$0 = \rho_{n-1} n \sigma + \rho_n \left[ (n+1)(n+2) + (n+1) \theta \nu_1 \right] + \sum_{j \geq n+1} \rho_j \theta \nu_0 - \rho_n \left[ n(n+1) + n \theta + (n+1) \sigma \right].$$

Rearranging the summands and writing this in terms of the tail probabilities (12), we obtain

$$0 = n \{ -\alpha_n [n+1+\theta+\sigma] + \alpha_{n+1} [n+1+\theta \nu_1] + \alpha_{n-1} \sigma \} + (n+1) \{ \alpha_{n+1} [n+2+\theta+\sigma] - \alpha_{n+2} [n+2+\theta \nu_1] - \alpha_n \sigma \},$$

which we abbreviate by

$$n(\alpha F)_n = (n+1)(\alpha F)_{n+1}$$

with the (tridiagonal) matrix $F$ that appears in the recursion (4). In view of these equalities, the proposition is proved if we can show that $\lim_{n \to \infty} n(\alpha F)_n = 0$, or, even better, that

$$\lim_{n \to \infty} n^2 \alpha_n = 0.$$  \hspace{1cm} (14)

To see (14), recall that as stated in Remark 3, $L_0$ is stochastically dominated by the number $Z_0$ of lines in the equilibrium ASG, which has distribution (2). In particular, $L_0$ has a finite third moment. From this, (14) is immediate since for any non-negative integer-valued random variable $\xi$ one has $E[\xi(\xi-1)(\xi-2)] = 3 \sum_{n=0}^{\infty} n(n-1) E(\xi > n)$. Thus, Proposition (5) is proved.

Let us note in passing that the tail probabilities also satisfy the constraint (5), i.e. $\lim_{n \to \infty} \alpha_{n+1}/\alpha_n = 0$. To see this, define $\lambda_n := \alpha_n/\alpha_{n-1}$, $n \in \mathbb{N}$, $\lambda_0 = 1$, and rearrange recursion (4),

$$\lambda_{n-1} = \frac{\sigma}{(n+\theta+\sigma) - \lambda_n (n+\theta \nu_1)}, \quad n \in \mathbb{N}. $$  \hspace{1cm} (15)

Then, for $n \geq 2\sigma - \theta$, an easy calculation shows

$$0 < \frac{\sigma^2}{n+\sigma} - \frac{\lambda_n}{\lambda_{n-1}} + \frac{\sigma}{n+\sigma},$$
which is equivalent to $\lambda_n < \lambda_{n-1}$. Thus, $\lambda_n$ is strictly decreasing for $n$ large enough. In addition, as $(\alpha_n)$ are tail probabilities, $\lambda_n \in [0, 1]$ for all $n$ and one can conclude that the limit $\lambda := \lim_{n \to \infty} \lambda_n$ exists and is in $[0, 1)$. Rewriting (15) we get

$$
\lambda = \lim_{n \to \infty} \frac{\sigma}{n(1 - \lambda) + \theta \nu_1 (1 - \lambda) + \sigma + \theta \nu_0}.
$$

(16)

The only solution to (16) is $\lambda = 0$, so we have shown that $\lim_{n \to \infty} \alpha_{n+1}/\alpha_n = 0$.

6. Monotonicities in the model parameters

The (conditional) probability $h(x)$ that the immortal line at time 0 carries the beneficial type does not only depend on the frequency $x$ of this type but also on three parameters: selection coefficient $\sigma$, mutation rate $\theta$, and mutation probability $\nu_1$ to the deleterious type. As shown in Fig. 14, some monotonicity properties apply. Since $h(x) = \sum_{n \geq 0} \alpha_n x(1 - x)$ depends on the tail probabilities $(\alpha_n)$ monotonically, an increase of $\alpha_n$ for all $n \in \mathbb{N}$ yields an increase of $h(x)$ as well. Let us now explain how the dependency of the tail probabilities on the three parameters can be understood in terms of the pruned equilibrium LD-ASG.

To this end, we consider the tail probabilities as functions of the parameters, i.e., $\alpha_n = \alpha_n(\sigma, \theta, \nu_1)$.

- If $\sigma_1 > \sigma_2$, then $\alpha_n(\sigma_1, \theta, \nu_1) > \alpha_n(\sigma_2, \theta, \nu_1)$. This is due to the fact that higher selection coefficients result in higher intensities of the Poisson point process $\Gamma^*$ of stars (compare Section 4.2). Since each star indicates the birth of a line in the pruned LD-ASG, in distribution more lines are born, which increases the tail probabilities of the top level $L$.

- For $\theta_1 > \theta_2$, one observes $\alpha_n(\sigma, \theta_1, \nu_1) < \alpha_n(\sigma, \theta_2, \nu_1)$. This is because each mutation results in deleting lines from the pruned LD-ASG, and a higher mutation rate results in more lines being cut away (in distribution). This decreases the tail probabilities for $L$.

- For $\nu_{1,1} > \nu_{1,2}$, one has $\alpha_n(\sigma, \theta, \nu_{1,1}) > \alpha_n(\sigma, \theta, \nu_{1,2})$. The reason is that increasing $\nu_1$ (at constant $\theta$) means replacing each circle in a realisation of the Poisson point processes $\Gamma^\circ$ by a cross (with a given probability), which thus adds to $\Gamma^x$. Since the pruning procedure can cut away more than one line at each circle but at most one line at each cross, we get, in distribution, more lines at higher $\nu_1$, which explains the increased tail probabilities.

An illustration of the probability weights $(\rho_n)$ of $L$ (i.e., $\rho_n = \alpha_n - \alpha_{n+1}, n \geq 0$) for various parameter combinations is also included in Fig. 14 (bottom).

7. Taylor’s representation of the CAT distribution via a boundary value problem

Taylor (2007) shows by analytic methods (see also Kluth et al. (2013)) that the (conditional) common ancestor type probabilities $h(x)$ arise as the solution of the boundary value problem

$$
\tilde{A}h(x) = 0, \quad 0 < x < 1
$$

$$
\lim_{x \to 0} h(x) = 0, \quad \lim_{x \to 1} h(x) = 1,
$$

(17)

(18)
Figure 14: Probability $h(x)$ that the immortal line at time 0 carries the beneficial type, given the frequency of this type is $x$ (top), tail probabilities $\alpha_n = P(L \geq n)$ of the stationary distribution of the highest occupied level in the pruned LD-ASG (middle), and probability weights $\rho_n = P(L = n)$ (bottom), $n \geq 0$. Results are shown for different combinations of selection coefficient $\sigma$, mutation rate $\theta$, and mutation probability $\nu_1$ to type 1.
where, for $\phi \in C^2([0,1], \mathbb{R})$,
\[
\tilde{A}\phi(x) = A\phi(x) + \theta \nu_0 \frac{1-x}{x} (\phi(0) - \phi(x)) + \theta \nu_1 \frac{x}{1-x} (\phi(1) - \phi(x)),
\]
and $A$ is the generator of the Wright-Fisher diffusion
\[
A\phi(x) := (x(1-x)) \frac{d^2}{dx^2} + (\theta \nu_0 (1-x) - \theta \nu_1 x + \sigma x (1-x)) \frac{d}{dx} \phi,
\]
see Taylor (2007, Proposition 2.4). Together with his Proposition 2.5, Taylor then suggests the following interpretation of (17): Given the frequency of the beneficial type at time 0 is $x$, sample two lineages at time 0, one of the beneficial type and one of the unfavourable type, and trace them back into the past. He writes (compare Fig. 15, left): “By comparing this generator with that of the structured coalescent for a sample of size 2 ... [with two different alleles], it is evident that the type of the common ancestor has the same distribution as the type of the sampled lineage which is of the more ancient mutant origin.” While Taylor here proposes to take the past type frequency path $(X_t)_{t \geq 0} = (X_r)_{r \geq 0}$ as a background for the structured coalescent, his idea leads to a direct derivation and interpretation of (17) when applied to the future frequency path $X := (X_t)_{t \geq 0} = (X_{-r})_{r \geq 0}$.

![Figure 15: Left: Taylor’s interpretation of (17). The ‘past’ type frequency path $(X_r)_{r \geq 0}$ of the beneficial type is shown as a background process. One lineage of the beneficial type (type-0 lineage at the bottom), and one of the unfavourable type (type-1 lineage at the top) are sampled. The cross symbolises a mutation of the type-1 line, the circle a mutation of the type-0 line. Since here the first mutation backwards in time is a cross, the type-0 line is of the more ancient mutant origin. Right: In the background of $X = (X_t)_{t \geq 0} = (X_r)_{r \leq 0}$, the CAT process $(C_r)_{-\infty < r \leq 0}$ (bold line) jumps to 1 at the points of $\Pi^{0,X}$ (dots at the top), and to 0 at the points of $\Pi^{1,X}$ (dots at the bottom). Here, $C_0 = 0$, because the last jump before time 0 was at time $T$ to type 0.

We take the chance to briefly explain this here, as an add-on to Taylor (2007) and to the approach developed in the previous sections. We abbreviate $b(x) := (1-x)/x$. Then, conditioned on the path $X = (X_t)_{t \geq 0}$ of the frequency of the beneficial type, the CAT process $(C_r)_{-\infty < r \leq 0}$ is a $\{0,1\}$-valued jump process with jump rates $\lambda_{0,X}^r := \theta \nu_0 b(X_t)$ from 0 to 1 and $\lambda_{1,X}^r := \theta \nu_1 b(X_t)^{-1}$ from 1 to 0; this dynamics arises when restricting the structured coalescent investigated by Barton et al. (2004) and Taylor (2007) to a single ancestral lineage. Thus, conditioned on $X$, the process $C$ can be obtained from two Poisson point processes $\Pi^{0,X}$ and $\Pi^{1,X}$ on $\mathbb{R}_+$ with the time-inhomogeneous intensities $\lambda_{0,X}$ and $\lambda_{1,X}$; backwards in time, the process $C$ jumps to 1 at the points of $\Pi^{0,X}$, and to 0 at the points of $\Pi^{1,X}$ (compare
\[
\tilde{X}_0 = X_0 = x \\
0 \quad T \quad t
\]

Figure 16: The process \( \tilde{X} \) follows \( X \) till the minimal point \( T \in \Pi^{0,X} \cup \Pi^{1,X} \). It then jumps to the boundary. Here, the minimal point is in \( \Pi^{1,X} \) and therefore we have a jump to 1.

Let us assume that \( X_0 \in (0, 1) \). Then, due to the boundedness of the rates \( \lambda^{0,X} \) and \( \lambda^{1,X} \) as long as \( X \) is strictly off the boundary \( \{0, 1\} \), there is a minimal point in \( \Pi^{0,X} \cup \Pi^{1,X} \), which we denote by \( T \). Let us define

\[
\tilde{X}_t := X_t \text{ for } 0 \leq t < T, \quad \tilde{X}_T := \begin{cases} 1 & \text{if } T \in \Pi^{1,X}, \\ 0 & \text{if } T \in \Pi^{0,X}, \end{cases} \quad \tilde{X}_t := \tilde{X}_T \text{ for } t > T.
\]

We then have

\[
\{C_0 = 0\} = \{T \in \Pi^{1,X}\} = \{\tilde{X}_T = 1\},
\]

and consequently, for \( x \in (0, 1) \),

\[
h(x) = P(C_0 = 0 \mid X_0 = x) = P(\tilde{X}_T = 1 \mid \tilde{X}_0 = x).
\]

In order to relate this probability to (17), we consider the process \( \tilde{X} \) with generator \( \tilde{A} \) given by (19), and put

\[
\tilde{T} := \inf\{t > 0 : \tilde{X} \text{ jumps to the boundary } \{0, 1\}\}.
\]

Since, for the same initial distribution at time 0, the process \( (\tilde{X}_t)_{0 \leq t \leq \tilde{T}} \) has the same distribution as \( (\tilde{X}_t)_{0 \leq t \leq \tilde{T}} \), we obtain from (20) for \( x \in (0, 1) \)

\[
h(x) = P(C_0 = 0 \mid X_0 = x) = P(\tilde{X}_T = 1 \mid \tilde{X}_0 = x) = P(\tilde{X}_{\tilde{T}} = 1 \mid \tilde{X}_0 = x).
\]

This shows that \( h \) satisfies (17). The boundary conditions (18) are explained by the fact that the jump rates \( \theta \nu_0 b(x) \) and \( \theta \nu_1 b(x)^{-1} \) converge to \( \infty \) as \( x \) converges to 0 and 1, respectively.

8. Conclusion

The aim of this contribution was to find a transparent graphical method to identify the common ancestor in a model with selection and mutation, and in this way to obtain the type distribution on the immortal line at some initial time, given the type frequencies in the population at that time. Our construction relies on the following key ingredients. We start from the equilibrium ASG (without types), and from the insight that the immortal line is the one that is ancestral to the first bottleneck of this ASG. Identifying this ancestral line
had previously appeared to be difficult, since it requires to keep track of the hierarchy of (incoming and continuing) branches, which quickly may become confusing. We overcome this problem by ordering the lines, in this way introducing a lookdown version of the ASG and a neat arrangement of the lines according to their hierarchy. Next, we mark the lines of the equilibrium ASG by the mutation events and, working backwards in time, apply a pruning procedure, which cuts away those branches that cannot be ancestral. Finally, we assign types at time 0 to the lines of the resulting pruned LD-ASG by drawing the types of its lines from the initial frequency and thus determine the type of the immortal line at time 0.

This equilibrium lookdown ASG is the principal (and new) tool in our analysis: backward in time, the top level in the pruned ASG performs a Markov chain whose equilibrium distribution can be computed, and the tail probabilities of this equilibrium distribution are shown to obey the Fearnhead-Taylor recursion. This provides the link to the simulation algorithm described by Fearnhead (2002) for the common ancestor type distribution in the stationary case. More precisely, our Theorems 2 and 4 together connect the LD-ASG to the simulation algorithm and thus provide a probabilistic derivation for it. At the same time, they imply a generalisation to an arbitrary rather than a stationary initial type distribution. Furthermore, Theorem 4 sheds new light on the series representation for the conditional CAT distribution, whose coefficients now emerge as the tail probabilities of the number of lines in the pruned LD-ASG. As a nice by-product, the graphical approach directly reveals various monotonicity properties of the tail probabilities depending on the model parameters, which translate into monotonicity properties of the common ancestor type distribution.

We believe that the pruned equilibrium (lookdown) ASG has potential for the graphical analysis of type distributions and genealogies also beyond the applications considered in the present paper. Let us also emphasise that, unlike Fearnhead’s original approach which builds on the stationary type process (and unlike other pruning procedures that work in a stationary typed situation), we start out from the untyped lookdown ASG which then is marked and pruned, with the assignment of types at the fixed (initial) time being delayed until the very last step of the construction. This is essential to be able to assign the types i.i.d. with a given frequency, and in this way to arrive at the desired probabilistic derivation of the conditional common ancestor type distribution.

Acknowledgements

The authors thank Tom Kurtz and Peter Pfaffelhuber for stimulating and fruitful discussions. This project received financial support from Deutsche Forschungsgemeinschaft (Priority Programme SPP 1590 ‘Probabilistic Structures in Evolution’, grants no. BA 2469/5-1 and WA 967/4-1).

References

Arthreya, S. R., Swart, J. M., 2005. Branching-coalescing particle systems. Prob. Theory Relat. Fields 131, 376–414.

Baake, E., Bialowons, R., 2008. Ancestral processes with selection: Branching and Moran models. Banach Center Publications 80, 33–52.
Barton, N. H., Etheridge, A. M., Sturm, A. K., 2004. Coalescence in a random background. 
Ann. Appl. Prob. 14, 754–785.

Donnelly, P., Kurtz, T. G., 1999. Particle representations for measure-valued population models. Ann. Prob. 27, 166–205.

Donnelly, P., Kurtz, T. G., 1999. Genealogical processes for Fleming-Viot models with selection and recombination. Ann. Appl. Prob. 9, 1091–1148.

Durrett, R., 2008. Probability Models for DNA Sequence Evolution. Springer.

Etheridge, A., 2011. Some Mathematical Models from Population Genetics: École D’Été de Probabilités de Saint-Flour XXXIX-2009, volume 2012. Springer.

Etheridge, A., Griffiths, R., 2009. A coalescent dual process in a moran model with genic selection. Theor. Pop. Biol. 75, 320 – 330. Sam Karlin: Special Issue.

Ewens, W. J., 2004. Mathematical Population Genetics 1: I. Theoretical Introduction, volume 27. Springer, 2 edition.

Fearnhead, P., 2002. The common ancestor at a nonneutral locus. J. Appl. Prob. 39, 38–54.

Kimura, M., 1962. On the probability of fixation of mutant genes in a population. Genetics 47, 713.

Kluth, S., Hustedt, T., Baake, E., 2013. The common ancestor process revisited. Bull. of Math. Biol. 75, 2003–2027.

Krone, S. M., Neuhauser, C., 1997. Ancestral processes with selection. Theor. Pop. Biol. 51, 210 – 237.

Mano, S., 2009. Duality, ancestral and diffusion processes in models with selection. Theor. Pop. Biol. 75, 164 – 175.

Neuhauser, C., Krone, S. M., 1997. The genealogy of samples in models with selection. Genetics 145, 519–534.

Pokalyuk, C., Pfaffelhuber, P., 2013. The ancestral selection graph under strong directional selection. Theor. Pop. Biol. 87, 25 – 33.

Shiga, T., Uchiyama, K., 1986. Stationary states and their stability of the stepping stone model involving mutation and selection. Prob. Theory Rel. Fields 73, 87–117.

Slade, P., 2000. Simulation of selected genealogies. Theor. Pop. Biol. 57, 35 – 49.

Taylor, J. E., 2007. The common ancestor process for a Wright-Fisher diffusion. Electron. J. Probab. 12, 808–847.

Wakeley, J., 2009. Coalescent Theory: An Introduction, Greenwood Village, Colorado.