Ancestries of a Recombining Diploid Population

Raazesh Sainudiin [*]
School of Mathematics and Statistics,
University of Canterbury, Christchurch, New Zealand,
r.sainudiin@math.canterbury.ac.nz

Bhalchandra D. Thatte
Departamento de Matemática
Universidade Federal de Minas Gerais,
Brasil, bhalchandra@mat.ufmg.br

Amandine Véber
Centre de Mathématiques Appliquées
Ecole Polytechique, Palaiseau, France,
amandine.veber@cmap.polytechnique.fr

UCDMS Research Report 2014/3, November 3, 2014

Abstract

We derive the exact one-step transition probabilities of the number of lineages that are ancestral to a random sample from the current generation of a bi-parental population that is evolving under the discrete Wright-Fisher model with n diploid individuals. Our model allows for a per-generation recombination probability of r. When $r = 1$, our model is equivalent to Chang's model [4] for the karyotic pedigree. When $r = 0$, our model is equivalent to Kingman's discrete coalescent model [16] for the cytoplasmic tree or sub-karyotic tree containing a DNA locus that is free of intra-locus recombination. When $0 < r < 1$ our model can be thought to track a sub-karyotic ancestral graph containing a DNA sequence from an autosomal chromosome that has an intra-locus recombination probability r. Thus, our family of models indexed by $r \in [0, 1]$ connects Kingman's discrete coalescent to Chang's pedigree in a continuous way as r goes from 0 to 1. For large populations, we also study three properties of the r-specific ancestral process: the time $\mathcal{T}_n$ to a most recent common ancestor (MRCA) of the population, the time $\mathcal{U}_n$ at which all individuals are either common ancestors to all present day individuals or ancestral to none of them, and the fraction of individuals that are common ancestors at time $\mathcal{U}_n$. These results generalize the three main results in [4]. When we appropriately rescale time and recombination probability by the population size, our model leads to the continuous time Markov chain called the ancestral recombination graph of Hudson [12] and Griffiths [9].

Mathematics Subject Classifications: Primary: 60C05, 60J85; Secondary: 92D15, 60J05;

[*] Corresponding Author & Current Address: 411 Malott Hall, Department of Mathematics, Cornell University, Ithaca, NY 14850, USA

1 Introduction

Suppose we take a sample of present-day humans and track back in time through their parents and their parents' parents, and so on. This is the pedigree of the sample where each individual will bifurcate into its father and mother in the previous generation. Recording only the number of individuals who are ancestors of the present-day sample (instead of the full genealogical relationships) as we trace back through time in the pedigree gives us a simpler ancestral process, of fundamental interest in population genetics as it is the most elementary description of the pedigree that relates the sampled individuals. It is this process on which we shall concentrate here.

Chang [4] studied this ancestral process when the sample consists of all present-day individuals in the population that is reproducing under a simplified mathematical model called *the two-parent Wright-Fisher model* [7, 24]. In the Wright-Fisher model, the population is made of a constant number n of diploid individuals, generations are non-overlapping and discrete, and the pedigree is formed by each individual in each generation choosing two parents uniformly and independently at random from the previous generation. His findings about the process were in stark contrast with the well-known results for the analogous ancestral process of Kingman's coalescent [16], in which each individual chooses only one parent (still independently and uniformly at random) in the previous generation, corresponding to reconstructing, say, the discrete coalescent tree describing the ancestry at a non-recombining autosomal locus within the pedigree. Firstly, the number of generations $\mathcal{T}_n$ before we find a *most recent common ancestor* (MRCA) of the whole population in Chang's pedigree process is about $\log_2(n)$, while it is of the order of n for Kingman's discrete coalescent. Secondly, if we continue further back through time past $\mathcal{T}_n$ then we reach a generation $\mathcal{U}_n$ (about $1.77\log_2(n)$ generations ago) at and beyond which every individual is either a common ancestor (CA) of the whole present population or is extinct. Finally, a randomly chosen individual in generation $\mathcal{U}_n$ is a CA with probability about 0.8. Once again this is in contrast with the coalescent where the MRCA is the only CA at $\mathcal{T}_n$ and only one individual in each generation beyond $\mathcal{T}_n$ is the CA and the other individuals are extinct.

Several conjectures generalized Chang's results to the case where individuals or genes have on average less than 2 parents in the previous generation. This happens for example when we are interested in a stretch of autosomal DNA that can recombine with probability r, or not with probability $1 - r$. In [5], Wiuf and Hein proposed that $\mathcal{T}_n$ should be approximately equal to $\log_{1+r}(n)$, on the basis that the mean number of parents for a given lineage is $2 \times r + 1 \times (1 - r) = 1 + r$. Indeed, following this idea the number of ancestors of a given individual should grow like $(1 + r)^t$ as we go backwards in time, and so should be of the order of n for $t \approx \log_{1+r}(n)$. At this time, the sets of ancestors of all present-day individuals overlap a lot and we are thus likely to find a common ancestor to all these individuals. However, this reasoning fails to take into account the much slower decay of the family of 'non-descendants' of a given potential CA, which is now linear instead of quadratic as in Chang's model. As we shall see in the proof of Theorem 2, this last stage adds another $-\log_{1-r}(n)$ to the time needed to find a most recent common ancestor to the whole population. The same question appears in [17], where the author argues that population growth and inbreeding, specified by $\overline{r}$ and $\overline{w}$, can cause the number of ancestors of an individual to grow at a rate $(\overline{r}\,\overline{w})^t$ instead of 2^t (at least during the first generations). There the inbreeding mechanism imposes some correlations between the choices of ancestors made by two distinct lineages, and it is not clear that $\log_{\overline{r}\,\overline{w}}(n)$ is a good approximation to the time of the MRCA in large populations.

In [22], the authors highlight the fact that in a diploid biparental population, the genealogy of a sample of individuals at a given locus is constrained by the organismal or karyotic population pedigree giving the ancestral history of the whole population. This means that the different transmission processes of various karyotic genetic material (autosomal fragments of homologous DNA sequences) are all embedded into the same parental pedigree. The relation between genetic and genealogical ancestries has been the object of several interesting papers [2, 8, 18, 22]. Most of them try to quantify the genetic contribution of each genealogical ancestor to the current population, given the pedigree. They show in particular that although 80% of individuals deep in the past are ancestors to all present-day individuals, a fraction (close to 1) of them will leave no genetic material to any of their descendants. Our work does not aim at quantifying these contributions, but instead formalizes the embedding of the material transmission into the parental pedigree. It also enables us to embed the transmission of a nested family of subsets of a given material into ancestral sub-graphs of the karyotic pedigree corresponding to this material.

We formally describe the reproduction in such a recombining Wright-Fisher model in Sects. 1.1 and 1.2. Before going into the mathematical description, we start with biologically concrete definitions to articulate various ancestral relations of a sample through time. Consider a biparental eukaryotic population of diploid individuals with discrete non-overlapping generations. Let $n \in \mathbb{N} := \{1, 2, 3, \ldots\}$ be the size of the diploid population. The n individuals in the population are labeled by $\{1, 2, \ldots, n\}$. Such a population with $n = 5$ is shown in Fig. 1. An individual i from generation t is identified with its zygocyte or zygote $\mathbb{I}_{i,t}$, a labeled 3-ball, i.e. the protoplasmic contents of an individual zygotic sphere in three dimensions that is bounded by its plasma membrane. Within the protoplasm of this zygote lies its nucleus or karyon $\mathbb{K}_{i,t}$, another labeled 3-ball that is bounded by its nuclear membrane. Finally, the cytoplasm of zygote i is given by $\mathbb{J}_{i,t} = \mathbb{I}_{i,t} \setminus \mathbb{K}_{i,t}$, i.e., the contents of the zygote excluding its nucleus. In order to emphasize the longer evolutionary time-scale in units of non-overlapping zygote generations we ignore the ontogenesis of the $\mathbb{I}_{i,t}$-rooted mitotic branching process to produce the mature adult individual $\mathbb{M}_{i,t}$. The diploid adult individuals from generation t meiotically produce gametes. These gametes in turn randomly form pairs to produce the diploid zygotes of generation $t + 1$ given by $\mathbb{I}_{i,t+1}$, for $i = 1, 2, \ldots, n$.

In Fig. 1 we illustrate the *zygotic ancestral graph* which contains the *cytoplasmic* as well as the *karyotic ancestral graphs* of the individuals from the present generation at the bottom. The cytoplasm and karyon of each individual zygote $i \in [5] := \{1, 2, 3, 4, 5\}$ from the present generation is labelled by $\{i\}^c$ and $\{i\}^k$, respectively. The respective ancestries in the past generations are obtained by taking unions of the offspring sets. For example, the second individual zygote one generation ago is the cytoplasmic ancestor of 1 and 2

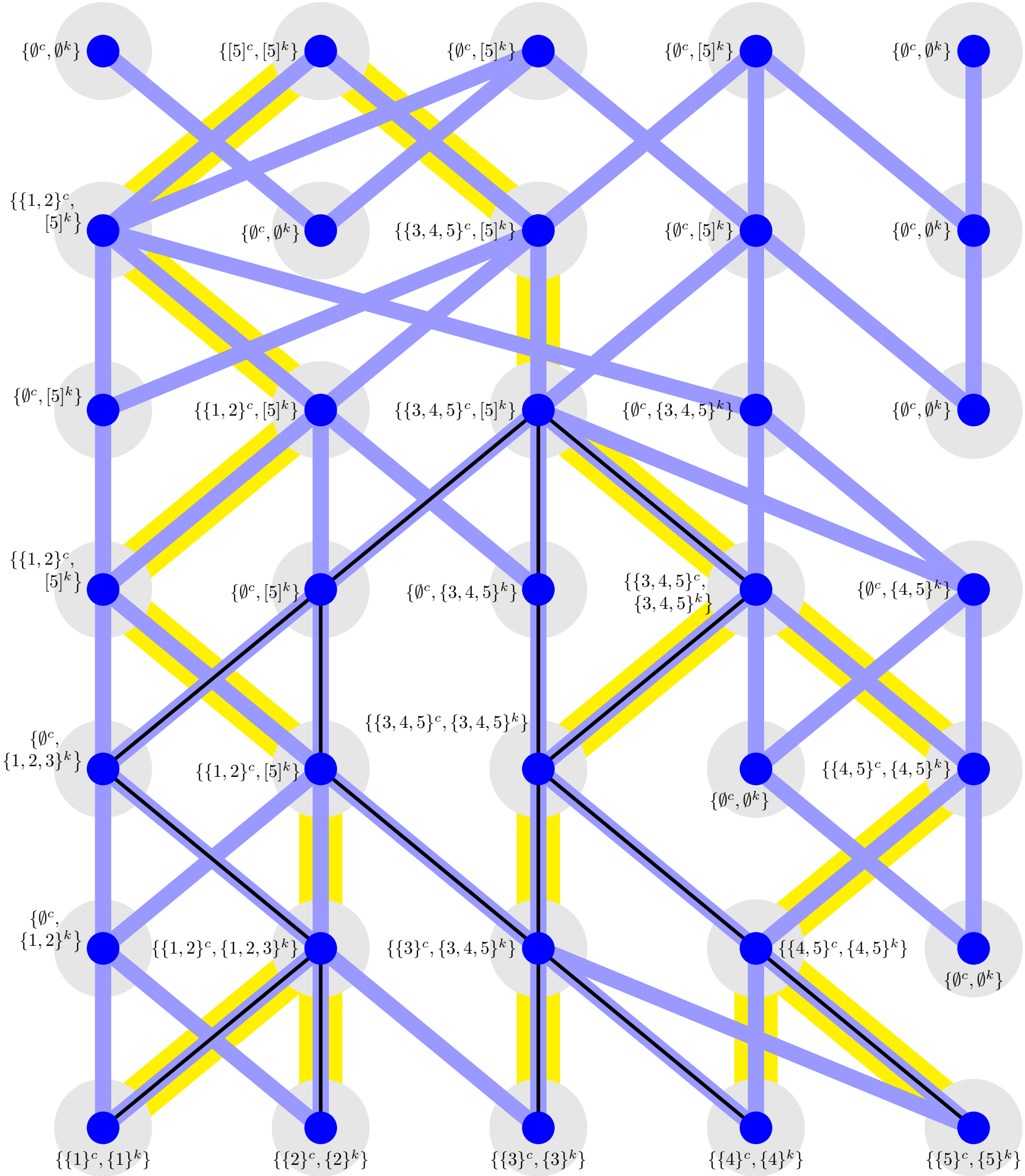
$\{\emptyset^c, \emptyset^k\}$
$\{[5]^c, [5]^k\}$
$\{\emptyset^c, [5]^k\}$
$\{\emptyset^c, [5]^k\}$
$\{\emptyset^c, \emptyset^k\}$
$\{\{1,2\}^c, [5]^k\}$
$\{\emptyset^c, \emptyset^k\}$
$\{\{3,4,5\}^c, [5]^k\}$
$\{\emptyset^c, [5]^k\}$
$\{\emptyset^c, \emptyset^k\}$
$\{\emptyset^c, [5]^k\}$
$\{\{1,2\}^c, [5]^k\}$
$\{\{3,4,5\}^c, [5]^k\}$
$\{\emptyset^c, \{3,4,5\}^k\}$
$\{\emptyset^c, \emptyset^k\}$
$\{\{1,2\}^c, [5]^k\}$
$\{\emptyset^c, [5]^k\}$
$\{\emptyset^c, \{3,4,5\}^k\}$
$\{\{3,4,5\}^c, \{3,4,5\}^k\}$
$\{\emptyset^c, \{4,5\}^k\}$
$\{\emptyset^c, \{1,2,3\}^k\}$
$\{\{1,2\}^c, [5]^k\}$
$\{\{3,4,5\}^c, \{3,4,5\}^k\}$
$\{\emptyset^c, \emptyset^k\}$
$\{\{4,5\}^c, \{4,5\}^k\}$
$\{\emptyset^c, \{1,2\}^k\}$
$\{\{1,2\}^c, \{1,2,3\}^k\}$
$\{\{3\}^c, \{3,4,5\}^k\}$
$\{\{4,5\}^c, \{4,5\}^k\}$
$\{\emptyset^c, \emptyset^k\}$
$\{\{1\}^c, \{1\}^k\}$
$\{\{2\}^c, \{2\}^k\}$
$\{\{3\}^c, \{3\}^k\}$
$\{\{4\}^c, \{4\}^k\}$
$\{\{5\}^c, \{5\}^k\}$

Figure 1: Zygotic, karyotic, cytoplasmic and sub-karyotic ancestral graphs of a Wright-Fisher population with five eukaryotic diploid individuals.

as well as the karyotic ancestor of 1, 2 and 3 from the present. This ancestral status is denoted by $\{1,2\}^c$ and $\{1,2,3\}^k$. The second individual two generations ago with $[5]^k$ is the karyotic MRCA of all five individuals in the present and thus $\mathcal{T}_5 = 2$ for the karyotic ancestral process in Fig. 1. One can also see that $\mathcal{U}_5 = 5$ since this is the first time from the present when each one of the 5 individual karyons is either a karyotic CA with $[5]^k$ or extinct with $\emptyset^k$. The cytoplasmic ancestry of all present day individuals in the bottom of the figure is shown in yellow, their karyotic ancestry, i.e. a pedigree, is shown by lineages tracing blue 3-balls, and an embedded sub-graph giving the *karyotic ancestral graph of a recombining autosomal locus* with per-generation recombination probability $r \in (0,1)$ is shown by black lines (this is only depicted until the MRCA is found for simplicity). Note how $\mathcal{T}_5 = \mathcal{U}_5 = 6$ for the cytoplasmic ancestry.

For example, consider a recombining fragment of autosomal DNA sequence. Such a fragment resides in the nucleus (karyon) of the sampled individual and has a probability r of recombining in one generation, and a probability $1 - r$ of not recombining, where r is allowed to take a fixed value in $[0,1]$. Hence, understanding the evolution of the genetic diversity of the population/sample at this locus requires to follow sometimes one parental karyotic line only, and sometimes both lines. The subsets of the material may correspond to different r's, the probability of having risen from two parental karyons in the previous generation decreasing with the length of the contiguous DNA sequence with constant per-site recombination probability per generation. In the end, the process with $r = 1$ traces back the number of karyotic ancestors of the sample back through time, but decreasing r enables us to concentrate on more precise transmission phenomena that are naturally embedded into the karyotic pedigree (see Fig. 1). The extreme case $r = 0$ then corresponds to focusing on the number of ancestors at a non-recombining locus.

As another example consider the cytoplasmic ancestry of the current population of zygotic mitochondria (the mitochondria present in each zygote today and their cytoplasmic ancestors from which they could have descended). Although mitochondria are predominantly maternally inherited in many animals, a state of mitochondrial heteroplasmy is also reached if paternal mitochondria enter the egg cytoplasm at fertilization. This is referred to as paternal leakage resulting in the coexistence of mitochondria from two unique ancestral lineages corresponding to both parents. Such biparental mitochondrial inheritance has been documented in mammals, birds, reptiles, fish, molluscs, nematodes, and arthropods, and is the norm in some bivalves (see [23] and the references therein). We can model the cytoplasmic ancestry of the mitochondria within the n individual zygotes at the present generation with a per generation paternal leakage probability r using our model (the yellow-edged cytoplasmic ancestral tree with $r = 0$ is shown in Fig. 1).

Other papers have considered biparental pedigrees and the corresponding numbers of ancestors or their sampling effects for a bisexual model comprised of a population of males and females [13, 14, 19, 22]. Each time they take always two parents (a mother and a father). Here we only consider monoecious individuals in accordance with Chang's model and want to be able to embed different transmission processes into the same parental pedigree by allowing r, the probability that an individual chooses two parents (instead of one), to take any fixed value in $[0,1]$. This is not considered carefully in former studies which focus on the fate of independent non-recombining loci [22, 18] or that of the genetic content of an individual within a pedigree [2, 8]. We defer more realistic bi-sexual generalizations of our model to the future.

Ancestral recombination graphs (ARGs) as we describe above in a discrete setting already exist in the literature, mostly in a continuous time setting [9, 12]. For a recent concise treatment of ARGs see [11]. These continuous time models arise as large-population-small-recombination-probability limits of models of the kind that we consider here, and are usually much easier to handle than their finite-population counterparts. However, a lot of efforts are now devoted to understanding the genetic consequences of population bottlenecks, or the evolution of endangered species which have typically reached a critically small census size, or the genealogical ancestry when sample sizes reach the population size, or the zygotic ancestry of a large recombining locus with a constant per-generational recombination probability that is independent of the population size. Hence, it seems rather important to formulate and study a model where population size is kept finite. The model we shall use here is a straightforward r-specific generalization of Chang's biparental Wright-Fisher model, whose transition probabilities (to our knowledge) have never been explicitly stated. The exact transition probability matrix for any given $r \in [0, 1]$ and $n \in \mathbb{N}$ of this ancestral size Markov chain will be the content of Theorem 1. We obtain this result by formulating the model forward in time and then developing its structure back in time using exact counts. Using rational arithmetic, we explicitly compute the probabilities back in time and obtain the exact stationary distribution of the number of ancestral lineages in small recombining populations. These computations are compared with direct simulations. We show that as r varies, qualitatively different stationary behaviour of the proportion of ancestral lineages of the sample occurs – there is a critical $x^*_{r,n}$ proportion of the population about which this walk concentrates as n gets large. We then study the large-n asymptotic properties of the ancestral size chain by using branching process approximations to generalize for any $r \in (0, 1)$ the results of Chang regarding $\mathcal{T}_n$ in Theorem 2, $\mathcal{U}_n$ in Theorem 3, and the probability of being a CA after $\mathcal{U}_n$ in Corollary 4. We finally show how a special limiting case of our ancestral size Markov chain as an ancestral birth-death process of the zygotic ancestry contains Hudson's and Griffiths' ARG in Theorem 5.

We emphasize that our model only traces the ancestry at the genealogical or physical level of karyotic and cytoplasmic enclosures of the genetic content in the sample. Our main reasons to study the process at such genealogical as opposed to genetic resolution are the following. First, this level of ancestral description provides the support graphs, i.e. the dominating counting measures on ancestral graphs, within which one can naturally embed the genetic ancestry of all DNA content of the current population via thinning constructions given in Remark 1. Second, such embeddings give the lumped Markov chains of the ancestral size process which can be delumped to the standard models of coalescent with recombination, mutation, selection, structure, demography, etc. Third, our resolution allows us to build a r-parametric homotopy between Kingman's discrete coalescent with $r = 0$ and Chang's pedigree with $r = 1$ and thereby unify classical models through explicit discrete time Markov transition probabilities as well as through continuous time Markov chains in limiting cases of large population and large-population-small-recombination. Finally, the karyotic and cytoplasmic enclosures of the genetic content provide the finest physical structure for genetic transmission. Thus, when individual zygotic labels are further mapped to physical and/or behavioral space, via parameterized territorial, dispersal and mate-choice operators for instance, one can consistently carry over the diploid biology of these ancestral graphs to more realistic models of population structure for the transmission of genes that are undergoing mutation, recombination and natural selection.

1.1 A Recombining Wright-Fisher Model

Recall that we consider a population of n diploid individuals. In the Wright-Fisher model [7, 24] of selectively neutral reproduction within a recombining, diploid, monoecious, panmictic population of constant size n, there are discrete and non-overlapping generations labeled by integers $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, 0, +1, +2, \ldots$ as we go forward in time. The current generation is labeled 0. Let r denote the probability of intra-locus recombination at an autosomal locus per meiotic generation. We model the lines of descent of the genetic material identified by this autosomal locus from one of the gametes of a diploid individual in generation $k+1$, that fertilized into a zygote in the following generation $k+2$. This is equivalent to modeling one of the two copies from a diploid individual in generation $k+2$ at this locus as a non-recombinant offspring of exactly one diploid individual in the previous generation k with probability $1-r$, or as a recombinant offspring of two distinct diploid individuals in generation k. Thus, we only trace the lines of descent for the autosomal locus up to its minimal topological enclosure by diploid karyons within individual zygotes. These parent-offspring choices occur independently among individuals of the same generation due to panmixia. Let us label the n diploid individuals (whose karyons are depicted by dark-blue circles in Fig. 1) in generation k using the label set $[n] := \{1, 2, \ldots, n\}$. For the lines of descent at our autosomal locus, let V_i denote the number of non-recombinant offspring of the diploid individual labeled i and $U_{i,j}$ denote the number of offspring that are recombinants of the diploid pair labeled by $\{i, j\}$ with $i < j$. Let also

$$V_\bullet := \sum_{i=1}^{n} V_i \quad \text{and} \quad U_{\bullet,\bullet} := \sum_{\{i,j \in [n] \colon i<j\}} U_{i,j}$$

denote the total number of non-recombinant and recombinant offspring, respectively. The lines of descent of the minimal set of diploid karyons that enclose the n homologous copies at the autosomal locus, with per-generation intra-locus recombination probability r, into the next generation follow from the multinomial random vector $(V, U) := (V_1, V_2, \ldots, V_n, U_{1,2}, U_{1,3}, \ldots, U_{n-1,n})$ of length $n + \binom{n}{2}$ such that $V_\bullet + U_{\bullet,\bullet} = n$ a.s. and for any realization v and u satisfying this constraint of constant n,

$$\begin{aligned} \mathbb{P}\big((V,U) = (v,u)\big) &= \mathbb{P}\left(V_1 = v_1, \ldots, V_n = v_n, U_{1,2} = u_{1,2}, \ldots, U_{n-1,n} = u_{n-1,n}\right) \\ &= \frac{n!}{v_1! \cdots v_n! u_{1,2}! \cdots u_{n-1,n}!} r^{u_{\bullet,\bullet}} \binom{n}{2}^{-u_{\bullet,\bullet}} (1-r)^{v_\bullet} \left(\frac{1}{n}\right)^{v_\bullet}. \end{aligned} \tag{1.1}$$

This reproduction scheme is independently and identically enforced in each generation to obtain the standard neutral Wright-Fisher model with recombination as we go forward in time.

In the absence of recombination, i.e. with $r = 0$, $V_\bullet = n$ and $U_{\bullet,\bullet} = 0$, the number of nonrecombinant offspring born to each of the n individuals of the previous generation is the symmetric multinomial random vector $V := (V_1, V_2, \dots, V_n)$ with,

$$\mathbb{P}\left[V_1 = v_1, V_2 = v_2, \dots, V_i = v_i, \dots, V_n = v_n\right] = \frac{n!}{v_1! \cdots v_n!} \left(\frac{1}{n}\right)^n . \tag{1.2}$$

This corresponds to the standard Wright-Fisher model with population size n.

When recombination is certain, i.e. with $r = 1$, $U_{\bullet,\bullet} = n$ and $V_\bullet = 0$, the number of recombinant offspring born to each pair of the n haploid individuals of the previous generation is given by the multinomial random vector $U := (U_{1,2}, U_{1,3}, \dots, U_{n-1,n})$, with

$$\mathbb{P}\left[U_{1,2} = u_{1,2}, \dots, U_{n-1,n} = u_{n-1,n}\right] = \frac{n!}{u_{1,2}! \cdots u_{n-1,n}!} \binom{n}{2}^{-n} . \tag{1.3}$$

This forward-in-time process is depicted for the case of $r = 1$ in Fig. 1 by the *karyotic population pedigree* with light-blue edges for lines of descent and dark-blue nodes for diploid individual karyons. This corresponds to Chang's pedigree model but without any possibility for self-fertilization.

1.2 Number of ancestral lineages of a sample

The Wright-Fisher model of Sect. 1.1 has a simple structure as we go back in time. We can choose to track (backwards in time) various aspects of the ancestry of a sample of individuals from current generation 0 that is embedded within the full karyotic population pedigree. Here, we simply track the number of lineages that are ancestral to our sample, perhaps the most foundational aspect of the ancestry.

The forward-time offspring distribution of Eq. (1.1) is equivalent to the backward-time scheme where each individual chooses a single parent uniformly at random from among the n individuals in the previous generation with probability $1 - r$, or chooses a parental pair uniformly at random from among the $n(n-1)/2$ pairs in the previous generation with probability r. Since the choices made at different generations are independent of each other, recording (backwards in time) the number of ancestors of a sample gives us a rather simple Markov chain $\{{}^{n,r}X(t)\}_{t \in \mathbb{Z}_-}$ over the state space of ancestral sizes $\mathbb{X} := \{1, 2, \dots, n\}$. Note that the time index set $\mathbb{Z}_- := \{0, -1, -2, \dots\}$ is negative to indicate the number of generations into the past. Note also that we only trace the ancestry of our autosomal locus up to its minimal topological enclosure by diploid individual karyons, and refer to such a locus-specific physical/karyotic genealogy as a *sub-karyotic ancestral graph*. This is shown in Fig. 1 as a sub-graph (with black lines as edges) that is embedded within the population's karyotic ancestral graph (with blue channels as edges and blue diploid karyons as nodes). This construction is under the assumption that we are not conditioning on a realization of the karyotic ancestral graph or population pedigree with $r = 1$. Otherwise, we need to embed the sub-karyotic ancestral graph of our locus within the karyotic ancestral graph by an r-specific thinning as described in Remark 1.

$\{^{n,r}X(t)\}_{t\in\mathbb{Z}_-}$ is the *lumped* Markov chain [15, Sec. 6.3, p. 123], that is at the most interesting lowest resolution, of finer Markov chains that describe the full genealogical relations of a random sample from the recombining Wright-Fisher model of Sect. 1.1 with further complications such as, coalescence within a karyon with probability 1/2, structure of the linked locus, mutation, etc. Some classical examples of such finer genealogical resolutions include: (i) the ARG of [12] that describes sampled loci as unit intervals of infinitely many sites and only tracks the recombining genealogy of the parts of the intervals that have genetic material in the sample – the genetic ancestral history, (ii) the two-locus ARG of [9] that allows recombination to occur between the loci but can track the complete genealogy of the two-locus samples including the genetic ancestral history, and (iii) the ARG of [10] that tracks the complete genealogy of the sampled unit intervals of infinitely many sites and generalizes the previous two models.

Hence, studying the number of ancestors $\{^{n,r}X(t)\}_{t\in\mathbb{Z}_-}$ is a necessary step towards understanding the rate at which recombination and coalescence events occur in the rescaled continuous-time approximations. It is also of huge importance in situations where very small population sizes would tend to keep genetic diversity very low in the absence of recombination. Indeed, the number of genetic ancestors of the whole population is a good indication of how efficiently recombination is able to foster or restore diversity. Note that the exact transition probability matrix $^{n,r}P$ of $\{^{n,r}X(t)\}_{t\in\mathbb{Z}_-}$ given in Theorem 1 is not available in the literature as most authors move directly to the diffusive limit and only define the continuous-time process in terms of exponential transition rates. This diffusive limit is thought to be valid, but never explicitly derived from $\{^{n,r}X(t)\}_{t\in\mathbb{Z}_-}$, as $n \to \infty$ and $r \to 0$ such that nr approaches a constant recombination rate. We establish this in Theorem 5 by showing the convergence of the transition probabilities in Theorem 1 to that of the standard continuous-time ARG model.

Remark 1. **(Thinning of ancestral graphs)** Let us remark that we can mathematically embed the sub-karyotic ancestral graph $\mathcal{A}$ of a given locus (with recombination probability r) into another sub-karyotic ancestral graph $\mathcal{A}'$ of some genetic material containing the locus of interest (with recombination probability $r' \geq r$) thanks to a simple rejection argument. Indeed, let us fix a realization of the latter. At each generation, for every lineage that chooses two parents in $\mathcal{A}'$ we decide that it chooses (the same) two parents in $\mathcal{A}$ with probability r/r', or it chooses only one of these parents (each with probability 1/2) with probability $1 - r/r'$. Coalescences are not modified. Averaging over the law of $\mathcal{A}'$, we obtain that each lineage recombines at a given generation with probability $r' \times (r/r') = r$ and so $\mathcal{A}$ does have the desired law. In such a nested construction, we see that the ancestor-counting process corresponding to $\mathcal{A}$ is always smaller than or equal to that corresponding to $\mathcal{A}'$, as we would expect from focusing on the ancestry of a smaller region of the genome. Moreover, we can embed multiple sub-karyotic ancestral graphs $\{\mathcal{A}_i\}_{i=1}^m$, corresponding to m independent unlinked loci each with intra-locus recombination probability $r_1, r_2, \ldots, r_m$, respectively, within the karyotic ancestral graph $\mathcal{K}$ with

$r = 1$, by independently thinning $\mathcal{K}$ through r_i to obtain $\mathcal{A}_i$ for each $i \in \{1, 2, \ldots, m\}$.

When studying large populations, it will be convenient to use the following slight modification of our model: when a lineage recombines, two parents are chosen independently and uniformly at random within the previous generation (instead of a pair of distinct individuals). The main difference is that one individual can then be chosen twice as the parent (allowing self-fertilization as in Chang's model), but since n is supposed to be large, this will happen with probability $\mathcal{O}(1/n)$, negligible in our analysis. For large n's, we use the following approach of Chang to study the ancestral process. Let the individuals in generation $t \geq 0$ (forward in time starting from some fixed generation $t = 0$) be denoted by $I_{t,i}, i = 1, 2, \ldots, n$. For any i, let the set of descendants in generation t of individual $I_{0,i}$ be denoted by $\mathcal{G}_t^i$, and the cardinality of $\mathcal{G}_t^i$ by G_t^i.

The probability that a given individual at generation $t + 1$ belongs to $\mathcal{G}_{t+1,i}$, or equivalently that it has at least one parent among $\mathcal{G}_t^i$ is

$$(1-r)\frac{G_t^i}{n} + r\left(1 - \left(1 - \frac{G_t^i}{n}\right)^2\right) = (1+r)\frac{G_t^i}{n} - r\frac{(G_t^i)^2}{n^2}.$$

Since the parental choices are made independently, the process $\{G_t^i\}_{t\in\mathbb{Z}_+}$ is thus a Markov chain with transition probabilities

$$(G_{t+1}^i \mid G_t^i) \sim \mathtt{Bin}\left(n, (1+r)\frac{G_t^i}{n} - r\frac{(G_t^i)^2}{n^2}\right), \tag{1.4}$$

where $\mathtt{Bin}(n, p)$ denotes the binomial distribution with parameters n, p. We drop the superscript i when there is no confusion. In particular, once we identify an individual, say I, who is unlikely to go extinct over generations, we follow the set of its descendants in each generation, and denote the sizes of these sets as G_t.

For the purposes of the proofs, we shall also consider another process: the individuals who are *not* descendants of i. In generation t, there are B_t^i such individuals. The same kind of calculations gives us that

$$(B_{t+1}^i \mid B_t^i) \sim \mathtt{Bin}\left(n, (1-r)\frac{B_t^i}{n} + r\frac{(B_t^i)^2}{n^2}\right). \tag{1.5}$$

Again, we drop the superscript when there is no confusion. In particular, we refer to the number of individuals in generation t that are not descendants of a chosen individual I in generation 0 by B_t. It will be convenient to study the process $(G_t)_{t\in\mathbb{Z}_+}$ until G_t is at least $n/2$, and then it will be more convenient to study $(B_t)_{t\in\mathbb{Z}_+}$.

2 Main Results

We fix the population size $n \in \mathbb{N}$ and the recombination probability $r \in (0, 1)$. We use the notation $x_+ := \max\{0, x\}$.

Theorem 1. *The exact transition probabilities of the ancestral process* $\{{}^{n,r}X(t)\}_{t\in\mathbb{Z}_-}$ *are*

$$
\begin{aligned}
{}^{n,r}P_{i,j} &= \binom{n}{j} \sum_{k=(j-i)_+}^{i} \frac{\binom{i}{k} r^k (1-r)^{i-k}}{2^k n^{i-k} \binom{n}{2}^k} \sum_{m=0}^{j} (-1)^{j-m} \binom{j}{m} m^i (m-1)^k \qquad \text{if } 1 < j \le 2i, \\
&= \frac{(1-r)^i}{n^{i-1}} \qquad \text{if } j = 1, \\
&= 0 \qquad \text{otherwise.}
\end{aligned}
$$

Theorem 2. *Let $\mathcal{T}_n$ denote the number of generations, counting back in time from the present, to an MRCA to all present-day individuals. Then for every $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\big[(1-\epsilon)C(r)\ln n \le \mathcal{T}_n \le (1+\epsilon)C(r)\ln n\big] = 1,$$

where

$$C(r) := \frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}.$$

Theorem 3. *Let $\mathcal{U}_n$ denote the number of generations, counting back in time from the present, to a generation in which each individual is either a CA to all present-day individuals or an ancestor of no present-day individual. Let $\varrho = \varrho(r)$ be the unique solution in $(0,1)$ to the equation $x = e^{-(1+r)(1-x)}$, and recall the definition of $C(r)$ given in the statement of Theorem 2. Then for every $\epsilon > 0$,*

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\Bigg[(1-\epsilon)\bigg(C(r) - \frac{1}{\ln((1+r)\varrho)}\bigg) \ln n \le \mathcal{U}_n \\
\le (1+\epsilon)\bigg(C(r) - \frac{1}{\ln((1+r)\varrho)} - \frac{1}{\ln(1-r)}\bigg) \ln n\Bigg] = 1.
\end{aligned}
$$

Note that $\varrho(r)$ defined as in Theorem 3 is the extinction probability of a Galton-Watson process with offspring distribution $\texttt{Poisson}(1+r)$ (see Appendix A and Chapter 1 of [1] for a definition and first properties of Galton-Watson processes – in Appendix A, we also argue that $(1+r)\varrho < 1$, hence the minus sign in the expression involving the log of this quantity). Indeed, the proofs of Theorems 2 and 3 show that a given individual will be common ancestor to the whole population in the future if in the $\mathcal{O}(\ln n)$ generations following him, his family survives and grows exponentially. We shall show that at the early stages of this family, its size can be approximated by a Galton-Watson process with offspring distribution $\texttt{Poisson}(1+r)$. This gives us the following result.

Corollary 4. *An individual chosen uniformly at random from $\mathcal{U}_n$ generations ago is a CA of the current population with probability tending to $1-\varrho$ as n tends to infinity, where $\varrho = \varrho(r)$ is the unique solution in $(0,1)$ to the equation $x = e^{-(1+r)(1-x)}$.*

Our final result relates the recombining Wright-Fisher model defined in Sect. 1.1 to the number of lineages in the classical ancestral recombination graph (ARG) with recombination rate $\rho > 0$ appearing in [9, 12]. This process $Z = \{Z(t),\ t \ge 0\}$ is the continuous-time jump process with values in $\mathbb{N}$, that jumps from z to $z+1$ at rate ρz and from z to $z-1$ at rate $\binom{z}{2}$ (this rate being 0 when $z = 1$).

Theorem 5. *Suppose that the recombination probability r decreases with increasing population size n in such a way that $\rho := nr$ remains a constant. Then as n tends to infinity, the ancestral process $\{{}^{n,\rho/n}X(\lfloor nt \rfloor),\ t \geq 0\}$ converges in distribution towards Z. The convergence is in the sense of weak convergence in the space $D_{\mathbb{N}}[0,\infty)$ of all càdlàg paths with values in $\mathbb{N}$.*

In other words, the standard continuous-time ARG model can be recovered from the recombining Wright-Fisher model in the regime of parameters where recombination is *weak* and population size is large. In this case, we need to consider the population ancestry over time intervals of length $\mathcal{O}(n)$ (as in the coalescent approximation of the non-recombining Wright-Fisher model).

Remark 2. In our models when two lineages share a common ancestral karyon we consider them to have genealogically coalesced. If one is interested in the genetic coalescence of samples of homologous autosomal DNA sequences then we need to introduce an additional factor of 1/2 for the probability that the two lineages found in the same diploid karyon actually come from one of the two copies of DNA. Thus, our results can be interpreted with or without this factor of 1/2.

3 Simulations

We performed exact rational arithmetic [21] to compute the stationary distribution

$$\lim_{t\to\infty} \left({}^{n,r}\bar{P}\right)^t = {}^{n,r}\pi$$

of the rescaled random walk

$$\{{}^{n,r}X(t)/n\}_{t\in\mathbb{Z}_-}$$

on $\{1/n, 2/n, \ldots, 1\}$ for small values of $n \in \{10, 50, 100\}$. From these exact computations of the stationary distributions depicted in Fig. 2 the following qualitative observations stand out. First, for a given n and r there is a focal or critical ratio

$${}^{n,r}x^* = \operatorname*{arg\,max}_{x\in[0,1]} {}^{n,r}\pi(x)$$

of the number of CAs for a given n and r about which the probability mass of the stationary distribution ${}^{n,r}\pi$ of the rescaled random walk $\{{}^{n,r}X(t)\}_{t\in\mathbb{Z}_-}/n$ on $\{1/n, 2/n, \ldots, 1\}$ is concentrated and therefore the random walk hovers about this focal ratio ${}^{n,r}x^*$. Second, as n approaches infinity, the stationary distribution ${}^{n,r}\pi$ seems to converge weakly towards a Dirac mass at some $x_r^* \in (0,1)$. Corollary 4 suggests that $x_r^* = 1 - \varrho(r)$ and this corresponds to what we observe from computations with the stationary distribution ${}^{n,r}\pi$ for n as large as 100 and also from simulations for n as large as 10^5. Third, the boundary behaviour is as expected: when r approaches 0 we recover the non-recombining coalescent with a fraction of CAs very close to 0 (recall that eventually there is only one CA in the Kingman coalescent) and when r approaches 1 we recover the Chang's karyotic pedigree model with a fraction of CAs close to $1 - \varrho(2) \approx 0.8$.

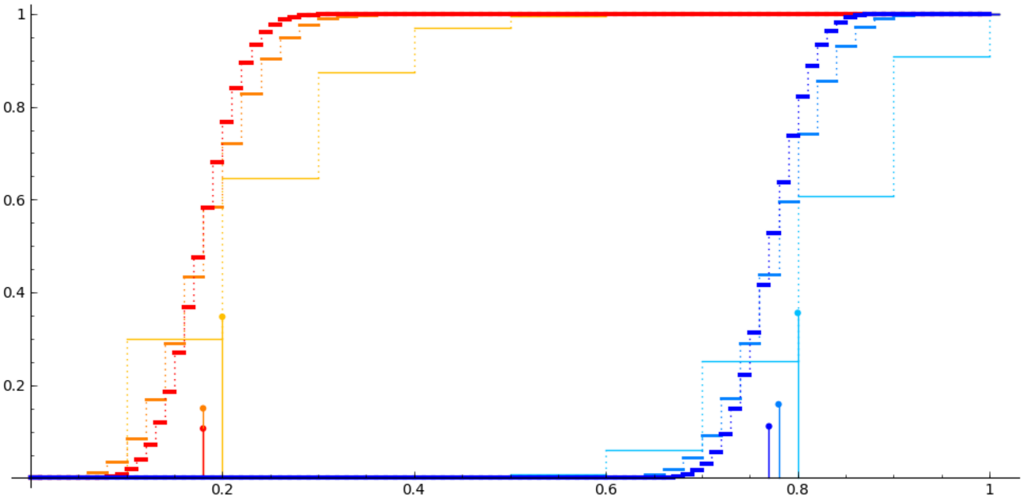

Figure 2: CDF of the stationary distribution of the number of ancestors for $n = 10, 50, 100$ with $r = 1/10$ (red distributions closer to 0) and with $r = 9/10$ (blue distributions closer to 1). The thickness of the CDF increases with n. The mass at the most likely state is shown by a stem plot for each distribution.

r	population size n in simulation			${}^{100,r}x^*$	$1 - \varrho(r)$
	10^2	10^4	10^5		
0.001	0.01(.01, .02)	0.0019(.0014, .0025)	0.002(.0018, .0022)	0.01	0.002
0.01	0.02(.01, .04)	0.019(.016, .022)	0.02(.019, .02)	0.02	0.0197
0.05	0.08(.04, .15)	0.093(.089, .097)	0.093(.091, .095)	0.09	0.0937
0.1	0.18(.07, .23)	0.17(.17, .18)	0.18(.17, .18)	0.18	0.176
0.25	0.34(.23, .45)	0.37(.36, .38)	0.37(.37, .37)	0.37	0.371
0.5	0.56(.44, .65)	0.58(.57, .6)	0.58(.58, .58)	0.59	0.583
0.75	0.71(.64, .77)	0.71(.7, .72)	0.71(.71, .72)	0.72	0.713
0.95	0.77(.71, .85)	0.78(.78, .79)	0.78(.78, .79)	0.79	0.783
0.99	0.8(.72, .87)	0.79(.79, .8)	0.79(.79, .8)	0.8	0.794
0.999	0.8(.69, .91)	0.8(.79, .8)	0.8(.79, .8)	0.8	0.797

Table 1: The median (minimum, maximum), based on 25 simulations, of the fraction of CAs at time $\mathcal{U}_n$, ${}^{100,r}x^*$ and $1 - \varrho(r)$ for different values of r and n.

Table 1 gives a tabular summary of 25 simulations of the fraction of CAs at $\mathcal{U}_n$ for a range of r and n values. It is compared with ${}^{100,r}x^*$, the value at which the stationary distribution of the ancestral Markov chain attains the maximum probability for $n = 100$. We see a nice pattern here that shows the maximum of the stationary distribution ${}^{n,r}\pi(x)$ and the concentrating fraction of CAs at $\mathcal{U}_n$ from simulations for various values of r as n ranges in $\{10^2, 10^4, 10^5\}$. According to Corollary 4, for a large population with $n = 10^5$ the fraction of CAs at $\mathcal{U}_n$ based on 25 simulations is highly concentrated about $x^*_r = 1 - \varrho(r)$.

In Table 2 the approximation for $\mathcal{T}_n$ is working well uniformly over r as n increases, albeit slower for smaller r. In Table 3 the approximation for $\mathcal{U}_n$ is also improving as n increases. One needs much larger n due to the $\log\log$ terms that have been dropped in the limits in Theorems 2 and 3.

r	population size n in simulation		
	10^3	10^4	10^5
0.001	$1245(795, 2741)$	$4564(3146, 6665)$	$8186(7369, 9711)$
	$13815; 0.0901$	$18420; 0.248$	$23025; 0.356$
0.01	$469(296, 683)$	$842(758, 1135)$	$1229(1143, 1416)$
	$1381; 0.34$	$1842; 0.457$	$2302; 0.534$
0.05	$147(119, 215)$	$224(202, 253)$	$310(297, 347)$
	$276; 0.533$	$368; 0.609$	$460; 0.674$
0.1	$82(65, 100)$	$123(113, 145)$	$165(159, 173)$
	$138; 0.594$	$184; 0.668$	$230; 0.717$
0.25	$38(34, 44)$	$54(51, 62)$	$71(70, 75)$
	$54; 0.704$	$73; 0.74$	$91; 0.78$
0.5	$21(20, 23)$	$29(27, 31)$	$37(37, 38)$
	$27; 0.778$	$36; 0.806$	$45; 0.822$
0.75	$14(14, 15)$	$20(19, 21)$	$25(25, 25)$
	$17; 0.824$	$23; 0.87$	$28; 0.893$
0.95	$11(11, 12)$	$15(15, 15)$	$19(19, 19)$
	$12; 0.917$	$16; 0.938$	$21; 0.905$
0.99	$11(10, 11)$	$14(14, 15)$	$18(17, 18)$
	$11; 1$	$15; 0.933$	$19; 0.947$
0.999	$11(10, 11)$	$14(14, 14)$	$18(17, 18)$
	$10; 1.1$	$14; 1$	$18; 1$

Table 2: The median (minimum, maximum) based on 25 simulations of $\mathcal{T}_n$ for different values of r and n. In the second row for each r we compare with the limit and the limiting ratio, given in Theorem 2 by $\lfloor C(r)\ln n\rfloor; \text{median}/\lfloor C(r)\ln n\rfloor$, where $C(r) = \left(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\right)$.

r	population size n in simulation		
	10^3	10^4	10^5
0.001	$2344(1153, 4494)$	$9273(6249, 13376)$	$15543(13543, 19634)$
	$[20724, 27628]$; 0.097	$[27632, 36838]$; 0.288	$[34540, 46047]$; 0.386
0.01	$931(624, 1429)$	$1626(1449, 1995)$	$2333(2089, 2786)$
	$[2073, 2760]$; 0.385	$[2764, 3681]$; 0.505	$[3455, 4601]$; 0.579
0.05	$297(238, 380)$	$420(368, 502)$	$559(498, 645)$
	$[415, 550]$; 0.616	$[554, 733]$; 0.653	$[692, 917]$; 0.695
0.1	$154(140, 195)$	$227(208, 268)$	$301(285, 328)$
	$[208, 273]$; 0.64	$[277, 365]$; 0.707	$[347, 456]$; 0.75
0.25	$74(63, 83)$	$104(93, 121)$	$128(122, 150)$
	$[83, 107]$; 0.779	$[111, 143]$; 0.819	$[139, 179]$; 0.805
0.5	$41(35, 57)$	$54(50, 61)$	$67(65, 75)$
	$[41, 51]$; 0.891	$[55, 68]$; 0.878	$[69, 86]$; 0.865
0.75	$28(26, 31)$	$37(34, 41)$	$46(44, 51)$
	$[27, 32]$; 0.949	$[36, 43]$; 0.937	$[45, 53]$; 0.939
0.95	$22(19, 27)$	$29(27, 33)$	$36(33, 39)$
	$[20, 23]$; 1.02	$[27, 30]$; 1.02	$[34, 38]$; 1
0.99	$21(19, 24)$	$27(25, 36)$	$34(32, 37)$
	$[19, 20]$; 1.08	$[25, 27]$; 1.04	$[32, 34]$; 1.03
0.999	$21(19, 24)$	$27(25, 30)$	$33(32, 37)$
	$[18, 19]$; 1.14	$[24, 26]$; 1.08	$[31, 32]$; 1.05

Table 3: The median (minimum, maximum) based on 25 simulations of $\mathcal{U}_n$ for different values of r and n. In the second row for each r we compare with the limit and the limiting ratio, given in Theorem 3 by, $[\lfloor \underline{D}(r)\ln n\rfloor, \lfloor \overline{D}(r)\ln n\rfloor]$; median$/((\overline{D}(r)\ln n+\underline{D}(r)\ln n)/2)$, where $\underline{D}(r) = C(r) - \frac{1}{\ln((1+r)\varrho)}$ and $\overline{D}(r) = C(r) - \frac{1}{\ln((1+r)\varrho)} - \frac{1}{\ln(1-r)}$.

4 Proof of Theorem 1

Notice first that each individual of the sample chooses at most 2 parents, and so ${}^{n,r}P_{i,j} = 0$ whenever $j > 2i$. Likewise, if $j = 1$ none of the current lineages can have recombined, there are n potential 'single' parents and n^i potential allocations of parents for i non-recombining individuals, so that

$$^{n,r}P_{i,1} = (1-r)^i \frac{n}{n^i}.$$

Let us thus consider the case $1 < j \leq 2i$. Let I be a fixed set of haploid lineages in the current generation. Let $|I| = i$. We are interested in the probability $\mathbb{P}(j|I)$ that these

i lineages descend from exactly j ancestral haploid lineages in the previous generation (see Fig. 3).

First, we have

$$\mathbb{P}(j|I) = \sum_{J:|J|=j} \mathbb{P}(J|I) = \binom{n}{j}\mathbb{P}(J_0|I), \tag{4.1}$$

where $\mathbb{P}(J|I)$ denotes the probability that the set of lineages ancestral to I in the previous generation is J and $J_0 := \{1, \dots, j\}$ is taken as a typical block (here the labels of the individuals have no influence on their fates and the above probability is the same for any set J of j labels).

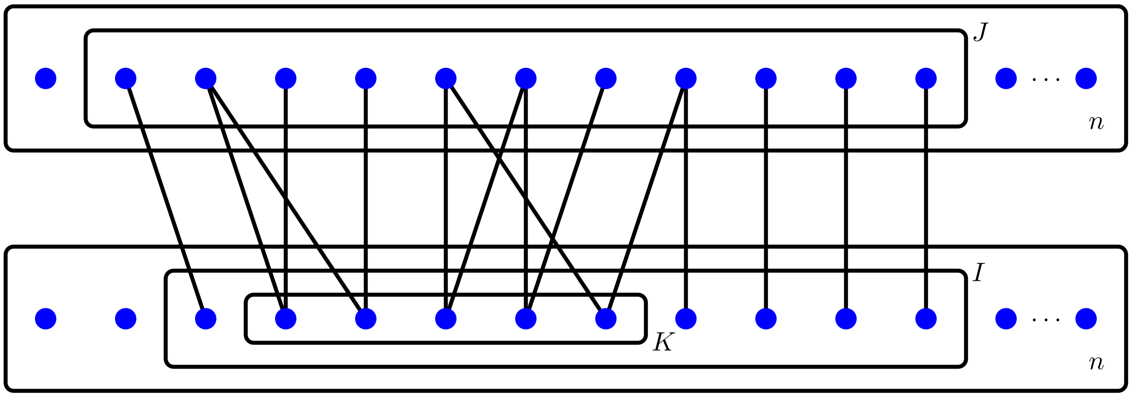

Figure 3: Combinatorial structure diagram for a panmictic monoecious sample genealogy within a recombining Wright-Fisher population of constant diploid size n in one generation from a set I to a set J with a subset K of I being recombinants.

Let $\mathbb{P}(J_0|I, K)$ be the probability that the set of lineages ancestral to I in the previous generation is J_0 given that lineages in a fixed subset K of I are recombinants and the lineages in $I \setminus K$ are non-recombinants. Since each individual is a recombinant offspring with probability r, we obtain

$$\mathbb{P}(J_0|I) = \sum_{K \subseteq I} r^{|K|}(1-r)^{|I|-|K|}\mathbb{P}(J_0|I, K) = \sum_{k=0}^{i} r^k (1-r)^{i-k}\binom{i}{k}\mathbb{P}(J_0|I, K_k), \tag{4.2}$$

where for any k, K_k denotes a given subset of I of size k.

Let us thus describe how to calculate $\mathbb{P}(J_0|I, K_k)$ for a given $k \in \{0, \dots, i\}$. Let $B(J|I, K_k)$ be the set of bipartite graphs with vertex set $I \cup J$, with bipartition $J|I$, such that the vertices in K_k are of degree 2, the vertices in $I \setminus K_k$ are of degree 1, and no vertices in J is isolated. This set is empty if $j > 2k + (i-k) = k + i$, and so we may consider only the case where $k \geq (j-i)_+$. Since the parents are chosen uniformly at random, we have

$$\mathbb{P}(J_0|I, K_k) = \frac{|B(J_0|I, K_k)|}{n^{i-k}\binom{n}{2}^k}. \tag{4.3}$$

Indeed, there are exactly $n^{i-k}\binom{n}{2}^k$ ways in which the lineages in I choose their ancestral lineages from the previous generation so that lineages in K_k are recombinants and the lineages in $I \setminus K_k$ are non-recombinants.

Therefore, combining Eqs. (4.1), Eq. (4.2) and Eq. (4.3) (and since only the cardinalities of $|I| = i, |J_0| = j$ and $|K_k| = k$ matter) we have

$$\mathbb{P}(j|i) = \binom{n}{j} \sum_{k=(j-i)_+}^{i} \binom{i}{k} r^k (1-r)^{i-k} \frac{|B(j|i,k)|}{n^{i-k}\binom{n}{2}^k}, \tag{4.4}$$

where we have extended the notation $B(J|I,K)$ into $B(j|i,k)$ in a natural way. We can count $|B(j|i,k)|$ by the inclusion-exclusion formula , given by the next lemma.

Lemma 6. *For arbitrary values of i, j, k, we can count $|B(j|i,k)|$ by the following formula:*

$$|B(j|i,k)| = \sum_{m=0}^{j} (-1)^{j-m} \binom{j}{m} \binom{m}{2}^k m^{i-k} \tag{4.5}$$

Proof. Suppose that the sets of vertices I, J, K, with cardinalities i, j, k, respectively, are fixed. Let A be the set of parents of vertices in I. For $L \subseteq J, |L| = l$, let $P(L)$ denote the number of bipartite graphs such that $A \subseteq L$ and let $Q(L)$ denote the number of bipartite graphs such that $A = L$. We have

$$P(L) = \sum_{M \subseteq L} Q(M).$$

Also,

$$P(L) = \binom{l}{2}^k l^{i-k},$$

since, for a fixed L, there are $\binom{l}{2}$ ways to choose 2 distinct parents of each vertex in K and l ways to choose 1 parent of each vertex in $I \backslash K$.

By Möbius inversion [3, Ch. 5], we have

$$Q(L) = \sum_{M \subseteq L} (-1)^{l-m} P(M) = \sum_{M \subseteq L} (-1)^{l-m} \binom{m}{2}^k m^{i-k}. \tag{4.6}$$

When $L = J$, Equation ((4.6)) implies that

$$\begin{aligned} |B(j|i,k)| = Q(J) &= \sum_{M \subseteq J} (-1)^{j-m} \binom{m}{2}^k m^{i-k} \\ &= \sum_{m} (-1)^{j-m} \binom{j}{m} \binom{m}{2}^k m^{i-k} \end{aligned}$$

□

Finally, Eqs. (4.4) and Eq. (4.5) for $(j-i)_+ \leq k \leq i$, give the (n,r)-specific exact one-step transition probability matrix

$$^{n,r}P = (\ ^{n,r}P_{i,j}\ )_{i,j\in\{1,2,\ldots,n\}}\ ,$$

as

$$^{n,r}P_{i,j} = \binom{n}{j} \sum_{k=(j-i)_+}^{i} \left(\binom{i}{k} r^k (1-r)^{i-k} \left(\frac{\sum_{m=0}^{j} (-1)^{j-m} \binom{j}{m} \binom{m}{2}^k m^{i-k}}{n^{i-k} \binom{n}{2}^k} \right) \right) \quad , \tag{4.7}$$

for the ancestral size Markov chain $\{^{n,r}X(t)\}_{t\in\mathbb{Z}_-}$.

5 Proof of Theorem 2

The proof of Theorem 2 is done in a number of steps as follows. In the next two sections, we fix $\epsilon \in (0,1/2)$.

1. **Stage G1:** For some individual $I := I_{0,i}$ in generation 0, G_t^i reaches at least $(\ln n)^2$ after $\mathcal{T}_n^{(G1)}$ generations, where $\mathcal{T}_n^{(G1)}$ is about $2\ln\ln n/\ln(1+r)$ with high probability. Moreover, the probability that the family of I eventually goes extinct is negligible.

2. **Stage G2:** The number of descendants of I increases from $(\ln n)^2$ to $g_2 n$ in $\mathcal{T}_n^{(G2)}$ generations, where $g_2 \in (0,1/2)$ is a well-chosen constant depending on the ϵ-precision that we want, and $\mathcal{T}_n^{(G2)}$ is about $\ln n/\ln(1+r)$. More precisely, the probabilities

$$\mathbb{P}\left[\mathcal{T}_n^{(G2)} > \left(1+\frac{\epsilon}{2}\right)\frac{\ln n}{\ln(1+r)}\right] \qquad \text{and} \qquad \mathbb{P}\left[\mathcal{T}_n^{(G2)} < \left(1-\frac{\epsilon}{2}\right)\frac{\ln n}{\ln(1+r)}\right]$$

   are both $o(1/n)$.

3. **Stage G3:** The number of descendants of I increases from $g_2 n$ to $n/2$ in $\mathcal{T}_n^{(G3)}$ generations, where $\mathcal{T}_n^{(G3)} \leq \ln\ln n$ with probability $1-o(1/n)$.

4. **Stage B1:** The number of non-descendants of I decreases from at most $n/2$ to at most $b_1 n$ in $\mathcal{T}_n^{(B1)}$ generations, where $b_1 \in (0,1/2)$ is another well-chosen constant and $\mathcal{T}_n^{(B1)} \leq \ln\ln n$ with probability $1-o(1/n)$.

5. **Stage B2:** The number of non-descendants of I decreases from at most $b_1 n$ to $(\ln n)^2$ in $\mathcal{T}_n^{(B2)}$ generations, where $\mathcal{T}_n^{(B2)}$ is about $-\ln n/\ln(1-r)$ generations. More precisely, the probabilities

$$\mathbb{P}\left[\mathcal{T}_n^{(B2)} > \left(1+\frac{\epsilon}{2}\right)\frac{\ln n}{-\ln(1-r)}\right] \qquad \text{and} \qquad \mathbb{P}\left[\mathcal{T}_n^{(B2)} < \left(1-\frac{\epsilon}{2}\right)\frac{\ln n}{-\ln(1-r)}\right]$$

   are both $o(1/n)$.

6. **Stage B3:** The non-descendants of I go extinct: the number of non-descendants of I decreases from at most $(\ln n)^2$ to 0 in $\mathcal{T}_n^{(B3)}$ generations, where $\mathcal{T}_n^{(B3)} \approx -2\ln\ln n/\ln(1-r)$ with high probability.

All these results combined show that, with probability tending to 1, the first time at which an individual becomes CA to the whole population is bounded from above by

$$\mathcal{O}(\ln\ln n) + \left(1+\frac{\epsilon}{2}\right)\frac{\ln n}{\ln(1+r)} + \mathcal{O}(1) - \left(1+\frac{\epsilon}{2}\right)\frac{\ln n}{\ln(1-r)},$$

which entails the upper bound in Theorem 2. This is detailed in Sect. 5.7. The lower bound is why we need the probabilities corresponding to the $\mathcal{O}(\ln n)$-long phases to be $o(1/n)$. Indeed, this will guarantee that with probability tending to 1, *none* of the n families born from our initial individuals can reach size $g_2 n$ in less than $(1-\epsilon/2)\ln n/\ln(1+r)$ generations, and neither can any of the (at most n) families reaching size $n/2$ increase to $n-(\log n)^2$ in less than $-(1-\epsilon/2)\ln n/\ln(1-r)$. Thus, with probability tending to 1 any individual needs at least $(1-\epsilon/2)\ln n\big(1/\ln(1+r) - 1/\ln(1-r)\big)$ generations to become a CA to the whole population.

Here we follow rather closely the proof of Theorem 1 in [4]. The first 4 stages are nearly identical to Stages 1 to 4 in [4], and so we only give a sketch of their proofs to recall the philosophy behind the maths. The main difference comes in the last two stages. Indeed, in [4] the rate of decrease of the number of non-descendants of I is quadratic. Here, because a fraction $1-r$ of the population picks only one parent in the previous generation, the rate of decrease is linear and extinction takes of the order of $\ln n$ generations to occur (hence the additional $-\ln n/\ln(1-r)$ term compared to Chang's result for $r=1$). We shall give more details about the proofs concerning Stages B_2 and B_3, although most of the arguments are similar to those used in the other stages.

Before considering the different stages one after the other, let us recall the classical Bernstein's inequality. We shall use it in the regime where G or B is large, to show that the behaviour of these processes is very close to their expectations.

Lemma 7. (Bernstein's inequality) *If* $X \sim \texttt{Bin}(n,p)$ *and* $x>0$*, then*

$$\mathbb{P}[X \geq np + x] \leq \exp\left\{\frac{-x^2}{2np(1-p)+(2/3)x}\right\},$$

and

$$\mathbb{P}[X \leq np - x] \leq \exp\left\{\frac{-x^2}{2np(1-p)+(2/3)x}\right\}.$$

5.1 Stage G1: Finding an individual I that has at least $(\ln n)^2$ descendants after $o(\ln n)$ generations

The proof is identical to that of Stage 1 in [4] and is based on the following argument (that we do not detail much).

By Lemma 18 applied with $b_n = (\ln n)^2$ and $m_n = \frac{3}{\ln(1+r)} \ln\ln n$, the probability that the process G starting at 1 reaches $(\ln n)^2$ in less than m_n generations is asymptotically equivalent to the same probability for a Galton-Watson process with offspring distribution $\texttt{Poisson}(1+r)$. By a standard martingale argument, the limsup of the latter is bounded from below by $1-\varrho > 0$, where we recall that $\varrho = \varrho(r)$ is the extinction probability of the Galton-Watson process. Therefore, if every m_n generations we test whether the individual labelled by 1 at that time has at least $(\ln n)^2$ descendants another m_n generations later, we obtain a geometric trial that ends with a success in a finite number of steps with probability 1. In particular, this means that the probability that *no* individuals at the origin have at least $(\ln n)^2$ descendants after $(\ln\ln n)m_n = o(\ln n)$ generations is bounded from above (for any $\delta \in (0, \varrho)$) by $(1-\varrho+\delta)^{\ln\ln n} \to 0$ as $n \to \infty$. In formula:

$$\lim_{n\to\infty} \mathbb{P}\left[\mathcal{T}_n^{(G1)} > \frac{3}{\ln(1+r)} (\ln\ln n)^2\right] = 0. \tag{5.1}$$

From now on, we call I such a thriving individual and $(G_t)_{t\in\mathbb{Z}_+}$ the process of its family size.

5.2 Stage G2: From $G_t \geq (\ln n)^2$ to $G_t \geq g_2 n$

Here, we use the fact that G is already large enough to behave roughly like its expectation. Indeed, suppose $G_0 \geq (\ln n)^2$ and let $\eta \in (0,1)$ and $g_2 > 0$ be such that $\eta > rg_2$ (we shall fix these quantities more precisely later). Since $G_{t+1}|G_t \sim \texttt{Bin}(n, (1+r)G_t/n - r(G_t/n)^2)$, Lemma 7 (with $x = \eta G_t - rG_t^2/n > 0$ if $G_t \leq g_2 n$) tells us that for any $t \in \mathbb{Z}_+$,

$$\begin{aligned}
&\mathbb{P}[G_{t+1} < (1+r-\eta)G_t\,,\, (\ln n)^2 \leq G_t \leq g_2 n] \\
&\leq \mathbb{E}\left[\exp\left\{\frac{-(\eta G_t - rG_t^2/n)^2}{2n\big[(1+r)\frac{G_t}{n} - r\frac{G_t^2}{n^2}\big]\big[1-(1+r)\frac{G_t}{n} + r\frac{G_t^2}{n^2}\big] + \frac{2}{3}(\eta G_t - rG_t^2/n)}\right\}\mathbf{1}_{\{(\ln n)^2 \leq G_t \leq g_2 n\}}\right] \\
&\leq \mathbb{E}\left[\exp\left\{-\frac{(\eta - rg_2)^2 G_t^2}{2(1+r)G_t + 2\eta G_t/3}\right\}\mathbf{1}_{\{(\ln n)^2 \leq G_t \leq g_2 n\}}\right] \leq \exp\left\{-\frac{(\eta - rg_2)^2(\ln n)^2}{2(1+r) + 2\eta/3}\right\}.
\end{aligned}$$

Let $m_n := \left\lceil\frac{\ln n - 2\ln\ln n + \ln g_2}{\ln(1+r-\eta)}\right\rceil$. If $G_{t+1} \geq (1+r-\eta)G_t$ for every $t \leq m_n$, necessarily we have

$$G_{m_n} \geq (1+r-\eta)^{m_n} G_0 \geq g_2 n.$$

Hence, as in the proof of Proposition 9 in [4] we can write

$$\begin{aligned}
\mathbb{P}[G_t < g_2 n, \forall t \leq m_n \,|\, G_0 \geq (\ln n)^2] &\leq \sum_{i=0}^{m_n-1} \mathbb{P}[G_{t+1} < (1+r-\eta)G_t\,,\, (\ln n)^2 \leq G_t \leq g_2 n] \\
&\leq m_n e^{-C(r,\eta,g_2)(\ln n)^2} = o(1/n).
\end{aligned} \tag{5.2}$$

Finally, if we choose $\eta > 0$ and g_2 small enough so that

$$\ln(1+r-\eta) > \frac{\ln(1+r)}{1+\epsilon/2} \qquad \text{and} \qquad g_2 < \frac{\eta}{r}, \tag{5.3}$$

we can conclude from Eq. (5.2) that:

Lemma 8. *Suppose* $G_0 \geq (\ln n)^2$ *and let* $\mathcal{T}_n^{(G2)} := \inf\{t : G_t \geq g_2 n\}$. *As* n *tends to infinity, we have*

$$\mathbb{P}\left[\mathcal{T}_n^{(G2)} > \left(1+\frac{\epsilon}{2}\right)\frac{\ln n}{\ln(1+r)}\right] = o\left(\frac{1}{n}\right).$$

The same reasoning using the second Bernstein inequality gives us (up to taking g_2 even smaller):

Lemma 9. *Suppose that* $(\ln n)^2 \leq G_0 \leq (\ln n)^3$. *As* n *tends to infinity, we have*

$$\mathbb{P}\left[\mathcal{T}_n^{(G2)} < \left(1-\frac{\epsilon}{2}\right)\frac{\ln n}{\ln(1+r)}\right] = o\left(\frac{1}{n}\right).$$

5.3 Stage G3: From $G_t \geq g_2 n$ to $G_t \geq n/2$

The very same reasoning as in Stage G2 gives us now

Lemma 10. *Suppose* $G_0 \geq g_2 n$ *and let* $\mathcal{T}_n^{(G3)} := \inf\{t : G_t \geq n/2\}$. *As* n *tends to infinity, we have*

$$\mathbb{P}\big[\mathcal{T}_n^{(G3)} > \ln\ln n\big] = o\left(\frac{1}{n}\right).$$

Remark 3. In fact $\ln\ln n$ is a very crude upper bound here, and could be replaced by $\lceil((\ln(1/2) - \ln g_2)/\ln(1+r-\eta)\rceil$ as in the previous stage. The advantage of $\ln\ln n$ is that it does not depend on ϵ, contrary to η and g_2.

5.4 Stage B1: From $B_t \leq n/2$ to $B_t \leq b_1 n$

Recall that in Chang's notation, B_t is the number of individuals in generation t that are not descendants of I. Hence, $(B_{t+1} \mid B_t)$ is binomially distributed as

$$(B_{t+1} \mid B_t) \sim \texttt{Bin}\left(n, (1-r)\frac{B_t}{n} + r\frac{B_t^2}{n^2}\right).$$

In this stage, B_t decreases almost deterministically (being of the order of n) at a rate which is at least $1 - r + r/2 = 1 - r/2$. Hence, exactly as in Stage G3 we have:

Lemma 11. *Suppose* $B_0 \leq n/2$ *and let* $b_1 \in (0, 1/2)$. *Set* $\mathcal{T}_n^{(B1)} := \inf\{t : B_t \leq b_1 n\}$. *As* n *tends to infinity, we have*

$$\mathbb{P}\big[\mathcal{T}_n^{(B1)} > \ln\ln n\big] = o\left(\frac{1}{n}\right).$$

5.5 Stage B2: From $B_t \le b_1 n$ to $B_t \le (\ln n)^2$

Here we can proceed exactly as in Stage G2. Let $\eta \in (0, 1-r)$ and fix $b_1 > 0$ such that $\eta > rb_1$. Using again Lemma 7, we can write that

$$\mathbb{P}\big[B_{t+1} > (1-r+\eta)B_t\,,\ (\ln n)^2 \le B_t \le b_1 n\big] \le \exp\bigg\{-\frac{(\eta - rb_1)^2(\ln n)^2}{2+2\eta/3}\bigg\}.$$

Let $m_n := \lceil \frac{\ln n - 2\ln\ln n + \ln b_1}{-\ln(1-r+\eta)} \rceil$. If $B_{t+1} \le (1-r+\eta)B_t$ for every $t \le m_n$, we must have

$$G_{m_n} \le (1-r+\eta)^{m_n} B_0 \le (\ln n)^2.$$

Hence,

$$\mathbb{P}\big[B_t > (\ln n)^2,\ \forall t \le m_n \,|\, B_0 \le b_1 n\big] \le m_n e^{-C'(r,\eta,b_1)(\ln n)^2} = o(1/n),$$

and if we choose $\eta > 0$ and b_1 small enough so that

$$\ln(1-r+\eta) \le \frac{\ln(1-r)}{1+\epsilon/2} \qquad \text{and} \qquad b_1 < \frac{\eta}{r}, \tag{5.4}$$

we obtain that

Lemma 12. *Suppose $B_0 \le b_1 n$ and let $\mathcal{T}_n^{(B2)} := \inf\{t : B_t \le (\ln n)^2\}$. Then as $n \to \infty$,*

$$\mathbb{P}\bigg[\mathcal{T}_n^{(B2)} > \Big(1+\frac{\epsilon}{2}\Big)\frac{\ln n}{-\ln(1-r)}\bigg] = o\Big(\frac{1}{n}\Big).$$

Likewise, up to taking smaller η and b_1, we have

Lemma 13. *Suppose $n/\log n \le B_0 \le b_1 n$. Then as $n \to \infty$,*

$$\mathbb{P}\bigg[\mathcal{T}_n^{(B2)} < \Big(1-\frac{\epsilon}{2}\Big)\frac{\ln n}{-\ln(1-r)}\bigg] = o\Big(\frac{1}{n}\Big).$$

5.6 Stage B3: Extinction of $(B_t)_{t\in\mathbb{Z}_+}$

Suppose that $B_0 \le (\ln n)^2$ and let us show that the number $\mathcal{T}_n^{(B3)}$ of generations that $(B_t)_{t\in\mathbb{Z}_+}$ needs to reach 0 is of the order of $\ln\ln n$ at most. By Lemma 18(ii) (with $\alpha = 0$ and $\gamma = 1/4$, say) we have

$$\mathbb{P}_{B_0}\big[\mathcal{T}_n^{(B3)} > C\ln\ln n\big] = \mathbb{P}_{B_0}\big[\tau_0^{Y-} > C\ln\ln n\big](1+o(1)),$$

where τ_0^{Y-} is the first time at which a $\texttt{Poisson}(1-r)$ Galton-Watson process becomes extinct (starting from the same initial condition B_0). Now, by Lemma 16 and the branching property of Galton-Watson processes, we can write that

$$\begin{aligned}\mathbb{P}_{B_0}\big[\tau_0^{Y-} \le C\ln\ln n\big] &= \mathbb{P}_1\big[\tau_0^{Y-} \le C\ln\ln n\big]^{B_0}\\ &\ge \big(1-(1-r)^{C\ln\ln n}\big)^{(\ln n)^2} = e^{-(\ln n)^2[(1-r)^{C\ln\ln n} + o((1-r)^{C\ln\ln n})]},\end{aligned}$$

which will tend to 1 as $n \to \infty$ whenever $C > -2/\ln(1-r)$.

5.7 Proof of Theorem 2: putting the stages together

Proof of the upper bound in Theorem 2. Let us call $\mathcal{E}_i$ the event that everything goes well during stage i. That is,

$$\begin{aligned}
\mathcal{E}_1 &= \left\{ \text{ a successful individual } I \text{ exists and } \mathcal{T}_n^{(G1)} \le \frac{3(\ln\ln n)^2}{\ln(1+r)} \right\};\\
\mathcal{E}_2 &= \left\{ G \text{ does not reach } 0 \text{ and } \mathcal{T}_n^{(G2)} \le \frac{(1+\frac{\epsilon}{2})\ln n}{\ln(1+r)} \right\};\\
\mathcal{E}_3 &= \left\{ G \text{ does not reach } 0 \text{ and } \mathcal{T}_n^{(G3)} \le \ln\ln n \right\};\\
\mathcal{E}_4 &= \left\{ B \text{ does not reach } n \text{ and } \mathcal{T}_n^{(B1)} \le \ln\ln n \right\};\\
\mathcal{E}_5 &= \left\{ B \text{ does not reach } n \text{ and } \mathcal{T}_n^{(B2)} \le \frac{(1+\frac{\epsilon}{2})\ln n}{-\ln(1-r)} \right\};\\
\mathcal{E}_6 &= \left\{ B \text{ does not reach } n \text{ and } \mathcal{T}_n^{(B3)} \le C\ln\ln n \right\}.
\end{aligned}$$

Notice that the events $\{G \text{ does not reach } 0\}$ and $\{B \text{ does not reach } n\}$ both coincide with the survival of the family of I during the stage considered. Since if all these events hold we have

$$\begin{aligned}
\mathcal{T}_n &\le \left(1+\frac{\epsilon}{2}\right)\left(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\right)\ln n + \frac{3(\ln\ln n)^2}{\ln(1+r)} + (2+C)\ln\ln n\\
&\le (1+\epsilon)\left(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\right)\ln n,
\end{aligned}$$

when n is large enough, we can write that

$$\begin{aligned}
\mathbb{P}\left[\mathcal{T}_n > (1+\epsilon)\left(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\right)\ln n\right] &\le \sum_{i=1}^{5} \mathbb{P}\big[\mathcal{E}_1,\, \ldots,\, \mathcal{E}_{i-1} \text{ hold, } \mathcal{E}_i \text{ does not}\big]\\
&\le o(1) + o(1/n) \to 0
\end{aligned}$$

as $n \to \infty$, where we have used the results of the previous paragraphs in the last line. This gives us the desired upper bound on $\mathcal{T}_n$. □

Proof of the lower bound in Theorem 2. We want to show that for any individual j living at time 0 (with family size process $(G_t^j)_{t\in\mathbb{Z}_+}$),

$$\begin{aligned}
\mathbb{P}\Bigg[&G^j \text{ goes from above } (\ln n)^2 \text{ to above } n-(\ln n)^2 \text{ in less than}\\
&\qquad (1-\epsilon)\left(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\right)\ln n \text{ generations}\Bigg] = o\left(\frac{1}{n}\right).
\end{aligned} \tag{5.5}$$

More precisely, we want to show that if G^j takes off and grows above $(\ln n)^2$ at some time (we call σ_n^j the first such time), then starting at this new value $G_{\sigma_n^j}^j$ the time that G^j needs to reach $n - (\ln n)^2$ is at least $(1-\epsilon)\big(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\big)\ln n$ with probability $1 - o(1/n)$. Once this is shown, we can then write

$$\mathbb{P}\bigg[\mathcal{T}_n < (1-\epsilon)\bigg(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\bigg)\ln n\bigg] \\ \leq \sum_{j=1}^{n} \mathbb{P}_1\bigg[G^j \text{ reaches } n-(\ln n)^2 \text{ in less than } (1-\epsilon)\bigg(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\bigg)\ln n\bigg] = o(1)$$

as $n \to \infty$, and the lower bound on $\mathcal{T}_n$ is proved.

Let us thus show Eq. (5.5). The only difficulty here is to control the size of G^j at the first time σ_n^j at which it goes above $(\ln n)^2$, and at the first time θ_n^j at which it goes above $(1-b_1)n$. Indeed, then Lemmas 9 and 13 tell us that each of the long stages takes the appropriate minimal amount of time with probability $1-o(1/n)$. Now, writing $G := G^j_{\sigma_n^j - 1}$ for the value of G^j just before the first jump over $(\ln n)^2$ to simplify the notation (hence $G < (\ln n)^2$ by definition), by Lemma 7 applied with $x = (\ln n)^3 - (1+r)G + r^2G^2/n$ we have

$$\mathbb{P}_1\Big[G^j_{\sigma_n^j} > (\log n)^3\Big] = \mathbb{P}_1\bigg[\mathtt{Bin}\bigg(n, (1+r)G - r^2\frac{G^2}{n^2}\bigg) > (\ln n)^3\bigg] \\ \leq \mathbb{E}\bigg[\exp\bigg\{-\frac{\big[(\ln n)^3 - (1+r)G + r^2G^2/n\big]^2}{2n\big[(1+r)\frac{G}{n} - \frac{r^2G^2}{n^2}\big]\big[1-(1+r)\frac{G}{n} + \frac{r^2G^2}{n^2}\big] + \frac{2}{3}\big[(\ln n)^3 - (1+r)G + \frac{r^2G^2}{n}\big]}\bigg\}\bigg] \\ \leq e^{-C_r(\ln n)^3} = o\bigg(\frac{1}{n}\bigg),$$

where the last inequality uses the fact that $G < (\ln n)^2$. Likewise, writing B for the value $B^j_{\theta_n^j - 1}$ of B^j just before the jump that makes it smaller than $b_1 n$, we have

$$\mathbb{P}_1\Big[G^j_{\theta_n^j} > n(1-1/\ln n)\Big] \leq \mathbb{P}_1\bigg[\mathtt{Bin}\bigg(n, (1-r)B + r\frac{B^2}{n^2}\bigg) < \frac{n}{\ln n}\bigg] \leq e^{-C'_r n} = o\bigg(\frac{1}{n}\bigg).$$

Summing up the above and using Lemmas 9 and 13, we obtain that the quantity in the l.h.s. of Eq. (5.5) is bounded by

$$\mathbb{P}_1\Big[G^j_{\sigma_n^j} > (\log n)^3\Big] + \mathbb{P}_1\bigg[G^j_{\sigma_n^j} \leq (\log n)^3\,,\, \mathcal{T}_n^{(G2)} < (1-\epsilon)\frac{\ln n}{\ln(1+r)}\bigg] \\ + \mathbb{P}_1\Big[G^j_{\theta_n^j} > n(1-1/\ln n)\Big] + \mathbb{P}_1\bigg[G^j_{\theta_n^j} \leq n(1-1/\ln n)\,,\, \mathcal{T}_n^{(B2)} < (1-\epsilon)\frac{\ln n}{-\ln(1-r)}\bigg] \\ = o\bigg(\frac{1}{n}\bigg),$$

as desired. The proof of the lower bound is thus complete. □

6 Proof of Theorem 3

Here again, we follow Chang's proof rather closely and the main difference comes in the last part, where all the families G^i having taken off eventually reach n.

Let us give the outline of the proof. As before, the stages that are very close to Chang's ones will not be detailed much. Recall that $\varrho = \varrho(r)$ is the extinction probability of a Galton-Watson process with offspring distribution $\texttt{Poisson}(1+r)$, that is the unique solution in $(0,1)$ to the equation $x = e^{-(1+r)(1-x)}$.

1. In about $-\ln n/\ln((1+r)\varrho)$ generations, each individual in generation 0 has either at least $(\ln n)^2$ descendants or has no descendants.
2. Each successful individual has at least $n-(\ln n)^2$ descendants $\left(\frac{1}{\ln(1+r)} - \frac{1}{\ln(1-r)}\right)\ln n$ generations later.
3. After another $-\ln n/\ln(1-r)$ generations at most, all successful individuals have become CA's to the whole population.

The main distinction between Theorem 2 and Theorem 3 is that at the end of Stage G1 of Theorem 2, we only require that at least one individual should have $(\ln n)^2$ descendants. At the end of the first stage of Theorem 3, we require each individual in generation 0 to have become successful or extinct, which is why we expect this stage to take longer than Stage G1 of Theorem 2. Similarly, we expect Stage 3 of Theorem 3 to take longer than Stage B3 of Theorem 2 since all the families of successful individuals have to reach n and not just one. Stage 2 of Theorem 3 is already detailed in Stages $G2$ to $B2$ of Theorem 2 and so we do not analyze it here.

6.1 Stage 1: extinction or 'explosion' of the G^i's

As a start, we show that by time $-\frac{1+\epsilon/2}{\ln((1+r)\varrho)}\ln n$, all the families generated by an individual alive at time 0 are either extinct or have reached size $(\ln n)^2$. That is,

Lemma 14. *Define*

$$\tau^i_{0,b} := \inf\left\{t : G^i_t = 0 \text{ or } G^i_t \geq b\right\},$$

and let

$$A_n := \bigcup_{i=1}^n \left\{\tau^i_{0,(\ln n)^2} > -\frac{1+\epsilon/2}{\ln((1+r)\varrho)}\ln n\right\},$$

Then

$$\lim_{n\to\infty} \mathbb{P}[A_n] = 0.$$

The proof is identical to that of Lemma 17 in [4] and is based on the following ideas. First, by Lemma 18(i), the probability that a given family size has not reached 0 or $(\ln n)^2$ by the prescribed time is asymptotically the same as the corresponding probability for a

$\texttt{Poisson}(1+r)$ Galton-Watson process Y. Now, since $\{\tau^Y_{0,(\ln n)^2} > t\} \subset \{0 < Y_t < (\ln n)^2\}$, Lemma 17($iii$) guarantees that for any $\delta > 0$ and n sufficiently large,

$$\ln \mathbb{P}_1\Big[\tau^Y_{0,(\ln n)^2} > -\frac{1+\epsilon/2}{\ln((1+r)\varrho)}\ln n\Big] \leq \big[\ln((1+r)\varrho)+\delta\big]\,\frac{1+\epsilon/2}{-\ln((1+r)\varrho)}\,\ln n,$$

and by choosing δ small enough, we can conclude that for any $i \in \{1,\dots,n\}$,

$$\mathbb{P}_1\Big[\tau^i_{0,(\ln n)^2} > \Big(-\frac{1+\epsilon/2}{\ln((1+r)\varrho)}\Big)\ln n\Big] = o\Big(\frac{1}{n}\Big).$$

Summing over i yields the result of Lemma 14.

Let us now show that at time $-\frac{1-\epsilon/2}{\ln((1+r)\varrho)}\ln n$, there are still a lot of individuals whose family sizes lie between 1 and $(\ln n)^2 - 1$. This corresponds to the following lemma.

Lemma 15. *Let* $t_n := \big\lfloor -\frac{1-\epsilon/2}{\ln((1+r)\varrho)}\ln n\big\rfloor$ *and let*

$$N_n := \mathrm{Card}\big(\big\{i \in \{1,\dots,n\} : G^i_{t_n} \in \{1,\dots,(\ln n)^2-1\}\big\}\big)$$

There exists $\gamma \in (0,1)$ *such that*

$$\lim_{n\to\infty} \mathbb{P}[N_n > n^\gamma] = 1.$$

Proof. We may proceed as in the proof of Lemmas 19 and 20 in [4]. Instead, to shorten a bit the proof (although our arguments are in fact similar to Chang's ones), we observe that the proof of Lemma 18 works also when we compare the pair of processes (G^1, G^2) to the pair (Y^1, Y^2) of independent Galton-Watson processes with offspring distribution $\texttt{Poisson}(1+r)$. That is, the transition probabilities of both processes starting at $(1,1)$ (and with values in $\mathbb{Z}_+\times\mathbb{Z}_+$) are equivalent as long as we look at times sufficiently small for both G^i's to remain negligible compared to n (the reader not enthralled by the idea of checking the transition probabilities may instead call on an easy modification of Theorem 2.2 in [19]). This is the case with the timescale t_n and so we can write that (using also the exchangeability of the family sizes)

$$\mathbb{E}[N_n] = n\mathbb{P}_1\big[1 \leq G^1_{t_n} < (\ln n)^2\big] = n\mathbb{P}_1\big[1 \leq Y^1_{t_n} < (\ln n)^2\big](1+o(1)),$$

and likewise

$$\begin{aligned}\mathrm{Var}(N_n) \sim n(n-1)\mathbb{P}_{(1,1)}\big[1 \leq Y^1_{t_n}, Y^2_{t_n} < (\ln n)^2\big] + n\mathbb{P}_1\big[1 \leq Y^1_{t_n} < (\ln n)^2\big]\\ - n^2\mathbb{P}_1\big[1 \leq Y^1_{t_n} < (\ln n)^2\big]^2\\ = n\mathbb{P}_1\big[1 \leq Y^1_{t_n} < (\ln n)^2\big]\big(1-\mathbb{P}_1\big[1 \leq Y^1_{t_n} < (\ln n)^2\big]\big),\end{aligned}$$

since $\mathbb{P}_{(1,1)}[1 \leq Y^1_{t_n}, Y^2_{t_n} < (\ln n)^2] = \mathbb{P}_1[1 \leq Y^1_{t_n} < (\ln n)^2]^2$ by independence of Y^1 and Y^2.

Now, by Lemma 17(iii), for any $\delta > 0$ and n large enough we have

$$\mathbb{P}_1\big[1 \leq Y^1_{t_n} < (\ln n)^2\big] \geq e^{t_n(\ln((1+r)\varrho)-\delta)} = n^{-(1-\epsilon/2)(1-\delta/\ln((1+r)\varrho))} > n^{-1+\frac{5\epsilon}{12}}$$

where the last inequality holds if we choose δ small enough. Likewise,

$$\mathbb{P}_1\big[1 \leq Y^1_{t_n} < (\ln n)^2\big] < n^{-1+\frac{7\epsilon}{12}}.$$

Hence,

$$\mathbb{E}[N_n] \geq n^{\frac{5\epsilon}{12}}, \qquad \mathrm{Var}(N_n) \leq n^{\frac{7\epsilon}{12}},$$

and by the Markov inequality

$$\mathbb{P}\Big[|N_n - \mathbb{E}[N_n]| > n^{\frac{\epsilon}{3}}\Big] \leq n^{-\frac{2\epsilon}{3}}\mathrm{Var}(N_n) \leq n^{-\frac{\epsilon}{12}} \to 0.$$

Taking $\gamma < \frac{5\epsilon}{12}$ yields the desired result. □

Since N_n tends to infinity with probability 1 and since the survival probability of a Galton-Watson process with offspring distribution $\texttt{Poisson}(1+r)$ is strictly greater than zero, with probability tending to 1 at least one (and in fact a lot) of the N_n families will reach size $(\ln n)^2$ and grow nearly deterministically to n. This fact will be used in the proof of the lower bound on $\mathcal{U}_n$, since it shows that these families have to be taken into account in the time needed to reach a state where all individuals at time 0 are either CA's or extinct.

6.2 Stage 3: extinction of the families of 'non-descendants'

Here, we bound the remaining amount of time needed to see the number B^i of 'non-descendants' of a given individual i go from at most $(\ln n)^2$ down to 0.

To this end, let us use Lemma 18(ii), the branching property of Galton-Watson processes and then Lemma 16 to write that for any $0 < x \leq (\ln n)^2$

$$\begin{aligned}
&\mathbb{P}_x\left[B \text{ becomes extinct before time } \frac{-(1+\epsilon/2)\ln n}{\ln(1-r)}\right] \\
&\qquad \approx \mathbb{P}_x\left[Y^- \text{ becomes extinct before time } \frac{-(1+\epsilon/2)\ln n}{\ln(1-r)}\right] \\
&\qquad = \mathbb{P}_1\left[Y^- \text{ becomes extinct before time } \frac{-(1+\epsilon/2)\ln n}{\ln(1-r)}\right]^x \\
&\qquad \geq \Big(1 - C(1-r)^{-(1+\frac{\epsilon}{2})\ln n/(\ln(1-r))}\Big)^x \geq 1 - \frac{C'(\ln n)^2}{n^{1+\epsilon/2}},
\end{aligned}$$

so that

$$\mathbb{P}_x\left[B \text{ does not become extinct before time } \frac{-(1+\epsilon/2)\ln n}{\ln(1-r)}\right] = o\left(\frac{1}{n}\right) \tag{6.1}$$

uniformly in $0 < x \leq (\ln n)^2$. Here, as in Lemma 18, Y^- denotes a Galton-Watson process with offspring distribution $\texttt{Poisson}(1-r)$. As in the previous section, this rapid decay of the probability of slow extinction will guarantee that with probability tending to one, all B^i's starting below $(\ln n)^2$ will become extinct in less than $-(1+\epsilon/2)\ln n/\ln(1-r)$ steps.

6.3 Proof of Theorem 3

Let us start by the lower bound. From Lemma 15, we know that with probability tending to one, there are $N_n \geq n^\gamma$ individuals at time 0 whose family sizes at time $t_n := \left\lfloor \frac{1-\epsilon/2}{-\ln((1+r)\varrho)} \ln n \right\rfloor$ belong to $\{1, \dots, (\ln n)^2 - 1\}$ (we call these families *slow*). Furthermore, the probability that such a family will eventually have more than $(\ln n)^2$ descendants is bounded from below by the probability that a $\texttt{Poisson}(1+r)$ Galton-Watson process survives (starting from $G^i_{t_n} \geq 1$), which itself is bounded from below by $1 - \varrho$. We shall see below that this suffices to guarantee that the number $\bar{N}_n$ of those families that eventually reach size $(\ln n)^2$ is positive with probability tending to 1. A successful family then grows nearly deterministically to $n - (\ln n)^2$ in about $C(r) \ln n$ generations and reaches n in at least $\mathcal{O}(\ln \ln n)$ generations with probability tending to 1 (by comparison with a $\texttt{Poisson}(1-r)$ Galton-Watson process). This yields

$$\begin{aligned}\mathbb{P}\big[\mathcal{U}_n < t_n + C(r)(1-\epsilon/2)\ln n\big] &\leq \mathbb{P}[N_n < n^\gamma] + \mathbb{P}\big[\bar{N}_n = 0,\, N_n \geq n^\gamma\big] \\ &\quad + \mathbb{P}\big[\text{all the } \bar{N}_n \text{ successful slow families reach size } n - (\ln n)^2 \text{ in less than} \\ &\qquad\qquad C(r)(1-\epsilon/2)\ln n \text{ generations}, \mathcal{E}_1, \mathcal{E}_2\big] \\ &\quad + \mathbb{P}[\text{all the successful slow families become extinct before reaching } n, \mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3],\end{aligned} \tag{6.2}$$

where

$$\begin{aligned}&\mathcal{E}_1 := \{N_n \geq n^\gamma\}, \qquad \mathcal{E}_2 := \{\bar{N}_n \geq 1\}, \\ &\mathcal{E}_3 := \{\text{some of the } \bar{N}_n \text{ successful slow families reach size } n - (\ln n)^2 \text{ in more than} \\ &\qquad\quad C(r)(1-\epsilon/2)\ln n \text{ generations}\}.\end{aligned}$$

Now, the first term in the r.h.s. of Eq. (6.2) tends to 0 by Lemma 15.

For the second term, an easy adaptation of the proof of Theorem 2.3 in [19] shows that for any $k \geq 1$,

$$\limsup_{n\to\infty} \mathbb{P}\big[\bar{N}_n = 0,\, N_n \geq n^\gamma\big] \leq \varrho^k.$$

Indeed, if we choose one individual among each of the $N_n \geq n^\gamma$ sets of individuals at time t_n constituting the slow families, the subfamilies they subsequently produce converge to independent $\texttt{Poisson}(1+r)$ Galton-Watson processes and the probability that all of them become extinct (which bounds our probability of interest) is precisely ϱ^k. Since this is valid for any $k \geq 1$, the second term in the r.h.s. of Eq. (6.2) tends to zero.

The third term is bounded by

$$n\,\mathbb{P}_{(\ln n)^2}[G \text{ grows to } n - (\ln n)^2 \text{ in less than } C(r)(1-\epsilon/2)\ln n \text{ generations}] = o(1)$$

by the analysis of Stages $G2$ to $B2$ of Theorem 2. Finally, the last term is bounded by the probability that a given family starting at size $n-(\ln n)^2$ becomes extinct before reaching n, which is itself bounded by the probability that the associated process $(B_t)_{t\in\mathbb{Z}_+}$, starting at $(\ln n)^2$, grows to $(\ln n)^5$ before reaching 0. By Lemma 18, the latter is equivalent to the probability of the same event for a $\texttt{Poisson}(1-r)$ Galton-Watson process (that we denote by Y below). Let us show that this event cannot occur with probability tending to 1 as $n \to \infty$. First, by Lemma 16 we know that

$$\mathbb{P}_1[Y \text{survives until time } C \ln\ln n] \leq (1-r)^{C\ln\ln n} = (\ln n)^{-C|\ln(1-r)|},$$

which combined with the independence of the subfamilies generated by individuals of the same generation tells us that

$$\mathbb{P}_{(\ln n)^2}[Y \text{survives until time } C \ln\ln n] \leq (\ln n)^{2-C|\ln(1-r)|} \to 0$$

whenever $C|\ln(1-r)| > 2$. Second, by Lemma 17(ii) $(Y_t/(1-r)^t)_{t\in\mathbb{Z}_+}$ is a nonnegative martingale. By Doob's maximal inequality, we thus have that

$$\mathbb{P}_1\left[\sup_{t\leq C\ln\ln n} \frac{Y_t}{(1-r)^t} > (\ln n)^3\right] \leq \frac{\mathbb{E}_1[Y_0] + 2\mathbb{E}_1[Y_{C\ln\ln n}/(1-r)^{C\ln\ln n}]}{(\ln n)^3} = \frac{2}{(\ln n)^3}.$$

Since $(1-r)^t \leq 1$ for any t, this entails

$$\mathbb{P}_1\left[\sup_{t\leq C\ln\ln n} Y_t > (\ln n)^3\right] \leq \frac{2}{(\ln n)^3},$$

and so again we obtain that (here 'subfamilies' refer to the families generated by each of the $(\ln n)^2$ initial individuals)

$$\mathbb{P}_{(\ln n)^2}\left[\text{at least one subfamily grows to } (\ln n)^3 \text{ before time } C\ln\ln n\right] \to 0$$

as $n \to \infty$. Summing up what we have proved, and using the fact that one of the $(\ln n)^2$ subfamilies has to reach size $(\ln n)^3$ for Y to reach $(\ln n)^5$, we can conclude that

$$\mathbb{P}_{(\ln n)^2}\left[Y \text{ ever reaches } (\ln n)^5\right] \to 0$$

and the last term in Eq. (6.2) tends to 0 as well. This completes the proof of the lower bound.

Let us now turn to the upper bound. We shall use the facts that by time $\frac{1+\epsilon/2}{-\ln((1+r)\varrho)} \ln n$, all family sizes have reached either 0 or $(\ln n)^2$ (cf. Lemma 14), and that the successful families will need at most another $(C(r) - 1/\ln(1-r))(1+\epsilon/2)\ln n$ generations to contain the whole current population (cf. Stages $G2$ to $B2$ of Theorem 2 and Stage 3 above).

That is, we have (in the notation of Lemma 14):

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{U}_n > (1+\epsilon)\left(\frac{-1}{\ln((1+r)\varrho)} + C(r) - \frac{1}{\ln(1-r)}\right)\ln n\right] \\
&\quad \le \mathbb{P}[A_n] + \mathbb{P}[\text{at least one successful family does not reach size } n-(\ln n)^2 \text{ in} \\
&\qquad\qquad\qquad C(r)(1+\epsilon/2)\ln n \text{ generations}] \\
&\qquad + \mathbb{P}\bigg[\text{at least one successful family having reached size } n-(\ln n)^2 \text{ needs more than} \\
&\qquad\qquad\qquad -\frac{1+\epsilon/2}{\ln(1-r)}\ln n \text{ generations to reach } n\bigg] \\
&\quad \le \mathbb{P}[A_n] + n\,\mathbb{P}_{(\ln n)^2}[G \text{ does not reach } n-(\ln n)^2 \text{ by time } C(r)(1+\epsilon/2)\ln n] \\
&\qquad + n\,\mathbb{P}_{(\ln n)^2}\left[B \text{ does not become extinct before time } \frac{-(1+\epsilon/2)}{\ln(1-r)}\ln n\right] \to 0
\end{aligned}$$

as $n\to\infty$ by Lemmas 14, 8, 10, 11, 12 and Eq. (6.1). The upper bound is proved.

7 Proof of Corollary 4

All we need to show is that the individual $I_{0,1}$ labeled by 1 in generation 0 becomes a CA with probability tending to $1-\varrho$. Indeed, since in our model individuals are exchangeable, the probability that individual i at time 0 becomes a CA is the same for all $i \in \{1,\dots,n\}$ and is thus equal to the probability that an individual chosen at random is a CA at time $\mathcal{U}_n$.

Now, recalling the different results obtained in the proofs of Theorems 2 and 3, we can write

$$\begin{aligned}
\mathbb{P}[I_{0,1} \text{ becomes a } CA] = \mathbb{P}[G^1 \text{ reaches } (\ln n)^2] \times \mathbb{P}[G^1 \text{ reaches } n-(\ln n)^2 \,|\, G^1\text{reaches } (\ln n)^2] \\
\times\, \mathbb{P}[G^1 \text{ reaches } n \,|\, G^1 \text{ reaches } n-(\ln n)^2]. \qquad (7.1)
\end{aligned}$$

By Lemma 18(i), the first term in the r.h.s. of Eq. (7.1) is equivalent to the same probability for a Poisson$(1+r)$ Galton-Watson process Y starting at 1. But for any $k\in\mathbb{Z}_+$ and any n sufficiently large we have

$$\mathbb{P}_1[Y \text{ survives}] \le \mathbb{P}_1[Y \text{ reaches } (\ln n)^2] \le \mathbb{P}_1[Y \text{ reaches } k],$$

and so

$$\mathbb{P}_1[Y \text{ survives}] \le \liminf_{n\to\infty} \mathbb{P}_1[Y \text{ reaches } (\ln n)^2] \le \limsup_{n\to\infty} \mathbb{P}_1[Y \text{ reaches } (\ln n)^2] \le \mathbb{P}_1[Y \text{ reaches } k].$$

Since these inequalities hold for any k, taking the limit $k\to\infty$ and observing that the events $\{Y \text{ reaches } k\}$ form a decreasing family whose intersection over $k\in\mathbb{Z}_+$ equals $\{Y \text{ survives}\}$, we obtain that

$$\lim_{n\to\infty} \mathbb{P}_1[Y \text{ reaches } (\ln n)^2] = \mathbb{P}_1[Y \text{ survives}] = 1-\varrho.$$

Next, by the analysis carried out in Stages $G2$ to $B2$ of Theorem 2, the second term in the r.h.s. of Eq. (7.1) tends to 1 as n tends to infinity.

Finally, using the same reasoning as in the study of the third term in the r.h.s. of Eq. (6.2), we can write

$$\begin{aligned}
&\mathbb{P}[G^1 \text{ does not reach } n \,|\, G^1 \text{ reaches } n-(\ln n)^2] \\
&\qquad \leq \sup_{x\leq(\ln n)^2} \mathbb{P}_x[B^1 \text{ survives during at least } \ln\ln n \text{ generations}] \\
&\qquad\qquad + \sup_{x\leq(\ln n)^2} \mathbb{P}_x[B^1 \text{ reaches } (\ln n)^5 \text{ in less than } \ln\ln n \text{ generations}] \to 0
\end{aligned}$$

as $n \to \infty$ (here we have used again Lemma 18(ii) to compare the process B^1 to a $\texttt{Poisson}(1-r)$ Galton-Watson process).

Combining these 3 steps, we obtain that

$$\lim_{n\to\infty} \mathbb{P}[I_{0,1} \text{ becomes a } CA] = 1-\varrho,$$

which ends the proof of Corollary 4.

8 Proof of Theorem 5

8.1 Approximate probabilities for large n and any r for small samples

The n-specific probability of i ancestral sample lineages in the current generation becoming j ancestral lineages in the previous generation can be obtained from Eq. (4.4), and is given by

$$ {}^{n,r}P_{i,j} = \begin{cases}
(1-r)^i\binom{i}{2}\frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right) & \text{if } j=i-1, \\[2ex]
\binom{i}{s}r^s(1-r)^{i-s}\left(1-\left\{\binom{i+s}{2}-s\right\}\frac{1}{n}\right) & \\
\qquad +\binom{i}{s+1}r^{s+1}(1-r)^{i-s-1}\frac{(i-s-1)(i+3s+2)+4s(s+1)}{2(n-1)} & \\
\qquad +\mathcal{O}\left(\frac{1}{n^2}\right) & \text{if } j=i+s, 0\leq s\leq i, \\[2ex]
\mathcal{O}\left(\frac{1}{n^2}\right) & \text{if } j\leq i-2, \\[2ex]
0 & \text{otherwise.}
\end{cases} \tag{8.1}$$

Eq. (8.1) is proved in Appendix B. Observe that the probability that at least 3 lineages have a common ancestor in the previous generation is of the order of $\mathcal{O}(1/n^2)$, and so is the probability that at least two pairs of lineages have a common ancestor, these ancestors being different. Hence, the probability that the number of lineages decreases by more than 1 due to some coalescence (independently of how many new lineages are created by recombination at the same time) is of order $\mathcal{O}(n^{-2})$. This explains the cases $j \le i-2$, but also tells us that the other formulae correspond to at most one coalescence (and potentially many recombinations).

8.2 Hudson-Griffiths' ARG as a special limiting case

Let us prove Theorem 5. Recall that we assume that $r = \rho/n$ for some $\rho > 0$. Plugging in this relation into Eq. (8.1), we obtain that:

1. When $j = i-1$,
$$^{n,\frac{\rho}{n}}P_{i,i-1} = \frac{1}{n}\left(1-\frac{\rho}{n}\right)^{i}\binom{i}{2} + \mathcal{O}\left(\frac{1}{n^2}\right) = \frac{1}{n}\binom{i}{2} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

2. When $j = i$,
$$^{n,\frac{\rho}{n}}P_{i,i} = \left(1-\frac{\rho}{n}\right)^{i}\left(1-\frac{1}{n}\binom{i}{2}\right) + \mathcal{O}\left(\frac{1}{n^2}\right) = 1 - \frac{1}{n}\left\{\binom{i}{2} + i\rho\right\} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

3. When $j = i+1$,
$$^{n,\frac{\rho}{n}}P_{i,i+1} = i\,\frac{\rho}{n}\left(1-\frac{\rho}{n}\right)^{i-1}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) + \mathcal{O}\left(\frac{1}{n^2}\right) = \frac{i\rho}{n} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

4. When $j \ge i+2$,
$$^{n,\frac{\rho}{n}}P_{i,j} = \mathcal{O}\left(\frac{1}{n^2}\right).$$

By Eq. (8.1), the other transition probabilities are either 0 or $\mathcal{O}(n^{-2})$. As a consequence, seeing $^{n,\frac{\rho}{n}}P$ as an infinite matrix, we can write
$$^{n,\frac{\rho}{n}}P = \mathrm{Id} + \frac{1}{n}A + \frac{1}{n^2}B^n, \tag{8.2}$$
where Id denotes the identity matrix, A is defined by
$$A_{i,i} = -\binom{i}{2} - i\rho, \qquad A_{i,i-1} = \binom{i}{2}, \qquad A_{i,i+1} = i\rho,$$
and $A_{i,j} = 0$ otherwise, and for any $i_* \in \mathbb{Z}_+$, all the coefficients of the submatrix $(B^n)_{1\le i,j\le i_*}$ are bounded by some $C(i_*) > 0$ uniformly in n.

To prove Theorem 5, we shall use an auxiliary process $Z^{n,k}$ living in the finite state space $\{1,\dots,k\}$ (contrary to $^{n,\frac{\rho}{n}}X$ which can take arbitrarily large values). We shall first prove that, as n tends to infinity, the process $\{Z^{n,k}(\lfloor nt\rfloor),\, t\ge 0\}$ converges in distribution towards the jump process $\{Z^{\infty,k}(t),\, t\geq 0\}$ which jumps from z to $z-1$ at rate $\binom{z}{2}$ and from z to $z+1$ at rate $z\rho$, until it goes above $k+1$ and falls into a cemetery state Δ in which it remains stuck forever. We shall then show that for any initial condition $x\in\mathbb{N}$ of ${}^{n,\frac{\rho}{n}}X$, any $T>0$ and any $\epsilon>0$, there exists $k_*=k_*(x,T,\epsilon)$ such that

$$\liminf_{n\to\infty}\,\mathbb{P}_x\left[\{{}^{n,\frac{\rho}{n}}X(\lfloor nt\rfloor),\, 0\leq t\leq T\}=\{Z^{n,k_*}(\lfloor nt\rfloor),\, 0\leq t\leq T\}\right]\geq 1-\epsilon. \tag{8.3}$$

In particular, Z^{n,k_*} does not reach the cemetery state Δ before time nT since ${}^{n,\frac{\rho}{n}}X$ does not. But on the event that Δ is not reached, Z^{k_*} and Hudson-Griffiths' ARG have the same law, and so Theorem 5 will be proved.

Let us proceed with the first part of our plan. Let us fix $k\in\mathbb{N}$ and for any $n\in\mathbb{N}$, let us define the process $Z^{n,k}$ as jumping like ${}^{n,\frac{\rho}{n}}X$ as long as it remains below k, and then jumping into some cemetery state Δ (where it remains stuck forever) whenever it is supposed to go over $k+1$. More precisely, $Z^{n,k}$ lives in $\{1,\ldots,k,\Delta\}$ and if $\Pi^{n,k}$ denotes its $(k+1)\times(k+1)$ transition matrix, we have

$$\Pi^{n,k}_{i,j}={}^{n,\frac{\rho}{n}}P_{i,j}\ \text{ if } 1\leq i,j\leq k,\quad \Pi^{n,k}_{i,\Delta}=\sum_{j=k+1}^{\infty}{}^{n,\frac{\rho}{n}}P_{i,j}\ \text{ if } 1\leq i\leq k,\quad \Pi^{n,k}_{\Delta,\Delta}=1.$$

As in Eq. (8.2), we have

$$\Pi^{n,k}=\mathrm{Id}+\frac{1}{n}\,A^{[k]}+\frac{1}{n^2}\,B^{n,k},$$

where $A^{[k]}_{i,j}=A_{i,j}$ for $1\leq i,j\leq k$, $A^{[k]}_{i,\Delta}=\sum_{j=k+1}^{\infty}A_{i,j}$, $A^{[k]}_{\Delta,\Delta}=0$, and all the coefficients of $B^{n,k}$ are bounded by some $C(k)>0$ uniformly in n. Hence, Möhle's Lemma [20] applies to $\Pi^{n,k}$ and enables us to conclude that for any $t\geq 0$,

$$\lim_{N\to\infty}\left(\Pi^{n,k}\right)^{\lfloor nt\rfloor}=e^{tA^{[k]}}.$$

In words, the semigroup associated to the process $\{Z^{n,k}(\lfloor nt\rfloor),\, t\geq 0\}$ converges towards that of the process $\{Z^{\infty,k}(t),\, t\geq 0\}$ with infinitesimal generator $A^{[k]}$. That is:

- If $z\in\{2,\ldots,k-1\}$, $Z^{\infty,k}$ jumps from z to $z+1$ at rate ρz and from z to $z-1$ at rate $\binom{z}{2}$;
- $Z^{\infty,k}$ jumps from 1 to 2 at rate ρ and cannot jump to 0;
- $Z^{\infty,k}$ jumps from k to Δ at rate ρk and from k to $k-1$ at rate $\binom{k}{2}$;
- $Z^{\infty,k}$ is absorbed at Δ.

Furthermore, since $\{1,\ldots,k,\Delta\}$ is a finite space, this convergence is uniform (i.e., $\left(\Pi^{n,k}\right)^{\lfloor nt\rfloor}f$ converges uniformly towards $e^{tA^{[k]}}f$ for any continuous function f on $\{1,\ldots,k,\Delta\}$), and so Theorem 4.2.12 of [6] tells us that $\{Z^{n,k}(\lfloor nt\rfloor),\, t\geq 0\}$ converges in distribution towards $Z^{\infty,k}$ in the space $D_{\{1,\ldots,k,\Delta\}}[0,\infty)$ of all càdlàg processes with values in $\{1,\ldots,k,\Delta\}$.

As concerns the second step, let us fix an initial value $x \in \mathbb{N}$, a time horizon $T > 0$ and $\epsilon \in (0,1)$. Recall that $Z = \{Z(t),\, t \geq 0\}$ denotes Hudson-Griffiths' ARG (with values in $\mathbb{N}$). Since Z grows at a linear rate and decreases at a quadratic rate, there exists $k_* = k_*(x, T, \epsilon)$ such that

$$\mathbb{P}_x\left[\sup_{t\in[0,T]} Z(t) \geq k_*\right] \leq \frac{\epsilon}{2}. \tag{8.4}$$

Observing that we can construct the processes Z^{∞,k_*} and Z in such a way that they coincide until the first time at which both leave $\{1,\dots,k_*\}$, we can write that Eq. (8.4) holds for Z^{∞,k_*} too. Now, $\sup_{[0,T]}$ is a continuous function on $D_{\{1,\dots,k_*,\Delta\}}[0,\infty)$ and so the convergence in distribution of $Z^{n,k_*}(\lfloor n\cdot\rfloor)$ towards Z^{∞,k_*} gives us the existence of n_* such that for every $n \geq n_*$,

$$\mathbb{P}_x\left[\sup_{t\in[0,T]} Z^{n,k_*}(\lfloor nt\rfloor) \geq k_*\right] \leq \epsilon.$$

Since we can construct Z^{n,k_*} and ${}^{n,\frac{\rho}{n}}X$ in such a way that they coincide until the first time at which both leave $\{1,\dots,k_*\}$, the above inequality yields Eq. (8.3) and the proof of Theorem 5 is complete.

Acknowledgements

B.T. was partly supported by post-doctoral fellowships at Department of Statistics, University of Oxford, UK (EPSRC grant EP/E05885X/1) and at Instituto de Matemática e Estatística, Universidade de São Paulo, Brasil (CNPq Processo 151782/2010-5 & MaCLinC). B.T. would like to thank Professor Yoshiharu Kohayakawa for many useful discussions related to Theorems 2, 3 and Corollary 4. R.S. is partly supported by a Sabbatical Grant from College of Engineering, University of Canterbury and a Visiting Scholarship at Department of Mathematics, Cornell University, Ithaca, NY, USA. R.S. thanks Robert C. Griffiths for an introduction to his ARGs and Alison Etheridge for discussions on stationary behaviour of the ancestral size chain. A.V. was supported by the ANR project MANEGE (ANR-09-BLAN-0215) and R.S. and A.V. were supported in part by the chaire Modélisation Mathématique et Biodiversité of Veolia Environnement-École Polytechnique-Museum National d'Histoire Naturelle-Fondation X.

References

[1] K.B. Athreya and P.E. Ney. *Branching processes.* Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040].

[2] N.H. Barton and A.M. Etheridge. The relation between reproductive value and genetic contribution. *Genetics*, 188:953–973, 2011.

[3] P.J. Cameron. *Combinatorics: topics, techniques, algorithms.* Cambridge University Press, Cambridge, 1994.

[4] J.T. Chang. Recent common ancestors of all present-day individuals. *Adv. in Appl. Probab.*, 31(4):1002–1038, 1999. With discussion and reply by the author.

[5] P. Donnelly, C. Wiuf, J. Hein, M. Slatkin, W.J. Ewens, and J.F.C. Kingman. Discussion: Recent common ancestors of all present-day individuals. *Adv. Appl. Probab.*, 31:1027–1035, 1999.

[6] S.N. Ethier and T.G. Kurtz. *Markov processes: Characterization and Convergence.* Wiley, 1986.

[7] R. Fisher. *The Genetical Theory of Natural Selection.* Clarenson, Oxford, 1930.

[8] S. Gravel and M. Steel. The existence and abundance of 'ghost' ancestors in biparental populations. *Preprint*, 2014.

[9] R.C. Griffiths. The two-locus ancestral graph. In IV Basawa and RL Taylor, editors, *Selected Proceedings of the Sheffield Symposium on Applied Probability: Held at the University of Sheffield, Sheffield, August 16–19, 1989*, IMS Lecture Notes – Monograph Series, Volume 18, pages 100–117. Institute of Mathematical Statistics, 1991.

[10] R.C. Griffiths and P. Marjoram. An ancestral recombination graph. In P Donnelly and S Tavaré, editors, *Progress in Population Genetics and Human Evolution*, IMA Volumes in Mathematics and its Applications, Volume 87, pages 257–270. Springer Verlag, 1997.

[11] D. Gusfield. *ReCombinatorics: The Algorithmics of Ancestral Recombination Graphs and Explicit Phylogenetic Networks.* MIT Press, 2014.

[12] R.R. Hudson. Properties of a neutral allele model with lntragenic recombination. *Theoretical Population Biology*, 23:183–201, 1983.

[13] K. Kämmerle. Looking forwards and backwards in a bisexual Moran model. *Journal of Applied Probability*, 26(4):pp. 880–885, 1989.

[14] K. Kämmerle. The extinction probability of descendants in bisexual models of fixed population size. *Journal of Applied Probability*, 28(3):pp. 489–502, 1991.

[15] J.G. Kemeny and J.L. Snell. *Finite Markov chains.* D. van Nostrand Company, Inc., Princeton, 1960.

[16] J.F.C. Kingman. The Coalescent. *Stochastic Processes and their Applications*, 13:235–248, 1982.

[17] J. Lachance. Inbreeding, pedigree size, and the most recent common ancestor of humanity. *J. Theor. Biol.*, 261:238–247, 2009.

[18] F.A. Matsen and S.N. Evans. To what extant does genealogical ancestry imply genetic ancestry? *Theor. Pop. Biol.*, 74:182–190, 2008.

[19] M. Möhle. Forward and backward processes in bisexual models with fixed population sizes. *J. Appl. Probab.*, 31(2):309–332, 1994.

[20] M. Möhle. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Probab.*, 30:493–512, 1998.

[21] W. A. Stein et al. *Sage Mathematics Software (Version 4.2.1).* The Sage Development Team, 2009. `http://www.sagemath.org`.

[22] J. Wakeley, L. King, B.S. Low, and S. Ramachandran. Gene genealogies within a fixed pedigree, and the robustness of Kingman's coalescent. *Genetics*, 190:1433–1445, 2012.

[23] D.J. White, J.N. Wolff, M. Pierson, and N.J. Gemmell. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, 17:4925–4942, 2008.

[24] S. Wright. Evolution in Mendelian populations. *Genetics*, 16:97–159, 1931.

A Galton-Watson processes with Poissonian offspring distribution

In this section, we collect a few facts about Galton-Watson processes. These facts are used in the proofs of Theorems 2 and 3. Recall that a Galton-Watson process with offspring distribution μ (a probability distribution on $\mathbb{Z}_+$) counts the number of individuals alive in each generation in a population evolving as follows: each individual in generation $k \geq 0$ gives birth to a random number of descendants with law μ, independently of each other; generation $k+1$ is then made of all these offspring. If the population becomes extinct at some time, its remains extinct for all later generations (and by extension we say that the Galton-Watson process *becomes extinct* - stuck at 0 - at that time).

The following lemmas summarize Lemmas 4 and 16 in [4] (with minor modifications, since Chang's results are for a Galton-Watson process with offspring distribution $\texttt{Poisson}(2)$) and some well-known general results from Chapter 1 of [1]. Therefore, we do not give all of their proofs here. We write m for the expectation of μ, σ^2 for its variance, and $\mathbb{P}_i$ (and $\mathbb{E}_i$) for the law under which the population starts with i individuals.

Lemma 16. *Let $(Y_t)_{t\in\mathbb{Z}_+}$ be a Galton-Watson process with $m<1$. Let $\tau_0 := \inf\{t : Y_t = 0\}$ be the extinction time of $(Y_t)_{t\in\mathbb{Z}_+}$. Then for any $k \in \mathbb{Z}_+$,*

$$\mathbb{P}_1[\tau_0 > k] < m^k.$$

Furthermore, if $\sigma^2 < \infty$ we have for large k

$$\mathbb{P}_1[\tau_0 > k] \geq \frac{1-m}{\sigma^2}\, m^{k+1}.$$

When μ is the law $\texttt{Poisson}(\lambda)$ for some $\lambda > 0$, we have $m = \lambda = \sigma^2 < \infty$.

Lemma 17. *Let $(Y_t)_{t\in\mathbb{Z}_+}$ be a Galton-Watson process with offspring distribution $\texttt{Poisson}(\lambda)$ for a given $\lambda > 0$. Let ψ be the probability generating function of X, where X is a random variable with law $\texttt{Poisson}(\lambda)$. That is, for any $z \in [0,1]$, $\psi(z) = \mathbb{E}[z^X] = e^{-\lambda(1-z)}$. Let also ϱ be the smallest solution in $[0,1]$ to $\psi(x) = x$. Then*

(i) The probability $\mathbb{P}_1[\tau_0 < \infty]$ that $(Y_t)_{t\in\mathbb{Z}_+}$ becomes extinct in finite time, starting with 1 individual, is equal to $\varrho > 0$.

(ii) The Markov chain $(Y_t/\lambda^t)_{t\in\mathbb{Z}_+}$ is a martingale. As t tends to infinity, it converges a.s. to a random variable M satisfying $\{M > 0\} = \{Y \text{ survives for ever}\}$.

(iii) Let $(b_t)_{t\geq 0}$ be a sequence of positive integers such that $\ln b_t = o(t)$ as $t \to \infty$. Then

$$\lim_{t\to\infty} \frac{\ln \mathbb{P}_1[1 \leq Y_t \leq b_t]}{t} = \ln(\lambda\varrho).$$

The first two points in Lemma 17 hold in fact for any Galton-Watson process (with λ replaced by its mean offspring distribution in (ii)). The last point says in essence that, up to polynomial prefactors, the probability that Y_t is still positive but grows less than exponentially in t decays like $e^{-|\ln(\lambda\varrho)|t}$ as $t \to \infty$. Observe that the product $\lambda\varrho$ is always less than one (except in the critical case $\lambda = 1$ in which we shall not be interested), since when $\lambda < 1$ we have $\varrho = 1$, and when $\lambda > 1$, $\lambda\varrho = \psi'(\varrho) < 1$ (ψ is strictly convex, $\psi(0) > 0$ and $\varrho < 1$ is the smallest positive value at which $\psi(x) = x$, the largest being $x = 1$).

Finally, let us give a comparison result between a single family size $(G_t)_{t\in\mathbb{Z}_+}$ (resp., the size of the family of non-descendants $(B_t)_{t\in\mathbb{Z}_+}$) and a Galton-Watson process with offspring distribution $\texttt{Poisson}(1+r)$ (resp., $\texttt{Poisson}(1-r)$).

Lemma 18. *Let $(Y_t^+)_{t\in\mathbb{Z}_+}$ (resp., $(Y_t^-)_{t\in\mathbb{Z}_+}$) be a Galton-Watson process with offspring distribution $\texttt{Poisson}(1+r)$ (resp., $\texttt{Poisson}(1-r)$). For any $b > 0$, let $\tau_b^Y := \inf\{t : Y_t \geq b\}$, $\tau_{0,b}^Y := \inf\{t : Y_t \geq b \text{ or } Y_t = 0\}$ and $\tau_0^Y := \inf\{t : Y_t = 0\}$ (where $Y = Y^+$ or Y^-). Define the same quantities for the processes G and B. Here again, $\mathbb{P}_i$ denotes the probability measure under which the process under consideration starts at i.*

(i) If k and b grow with n in such a way that $kb^2 = o(n)$, then as $n \to \infty$

$$\mathbb{P}_1\big[\tau_b^G > k\big] = \mathbb{P}_1\big[\tau_b^{Y^+} > k\big](1 + o(1)) \quad \textit{and} \quad \mathbb{P}_1\big[\tau_{0,b}^G > k\big] = \mathbb{P}_1\big[\tau_{0,b}^{Y^+} > k\big](1 + o(1)).$$

(ii) If for some $\alpha \in (0, 1/4)$ and $\gamma \in (2\alpha, 1/2)$ we have $i = \mathcal{O}(n^\alpha)$ and $k = o(n^{1-2\gamma})$, then as $n \to \infty$

$$\mathbb{P}_i\big[\tau_0^B > k\big] = \mathbb{P}_i\big[\tau_0^{Y^-} > k\big](1 + o(1)).$$

Remark 4. Note that $i = \mathcal{O}(n^\alpha)$ means that i is bounded by a constant times n^α, which allows to take i constant in n or growing more slowly than n^α.

In words, despite the dependency between the different family sizes in our original model, the early development of a single family $(G_t)_{t\in\mathbb{Z}_+}$ is very close to that of a $\texttt{Poisson}(1+r)$ Galton-Watson process. Likewise, as soon as there remains much less than n individuals in B, the extinction of this subpopulation occurs in the same way as in a $\texttt{Poisson}(1-r)$ Galton-Watson process.

Proof. The proof of (i) is similar to that of Lemma 3 in [4], and so we omit it here. The proof of (ii) follows the same lines but is a bit more complex. Our main aim is to show that as long as the processes do not grow too much and we do not look at too many generations, their transition probabilities are equivalent. Then, since Y^- starting below n^α will not grow beyond n^γ before going extinct with very high probability, neither will B and extinction will occur roughly at the same time (in distribution) for both.

Let us thus consider $x, y \leq n^\gamma$. Recall that conditionally on $B_t = x$, $B_{t+1} \sim \texttt{Bin}(n, (1-r)\frac{x}{n} + r\frac{x^2}{n^2})$. Since the sum of x independent $\texttt{Poisson}(1-r)$ random variables has the law $\texttt{Poisson}((1-r)x)$, we have for any $t \in \mathbb{Z}_+$

$$\frac{\mathbb{P}[B_{t+1} = y \,|\, B_t = x]}{\mathbb{P}[Y^-_{t+1} = y \,|\, Y^-_t = x]} = \frac{\binom{n}{y}\big((1-r)\frac{x}{n} + r\frac{x^2}{n^2}\big)^y \big(1-(1-r)\frac{x}{n} - r\frac{x^2}{n^2}\big)^{n-y}}{e^{-(1-r)x}(1-r)^y x^y/y!}$$
$$= \frac{n!}{(n-y)!n^y}\left(1 + \frac{r}{1-r}\frac{x}{n}\right)^y \exp\left\{(1-r)x + (n-y)\ln\left(1-(1-r)\frac{x}{n} - r\frac{x^2}{n^2}\right)\right\}$$

But $x, y \leq n^\gamma \ll n$, and so a first order Taylor expansion gives us that

$$\left(1 + \frac{r}{1-r}\frac{x}{n}\right)^y = e^{\frac{r}{1-r}\frac{xy}{n} + \mathcal{O}\left(\frac{yx^2}{n^2}\right)} \leq e^{\frac{2r}{1-r}n^{2\gamma-1}}$$

and

$$\exp\left\{(1-r)x + (n-y)\ln\left(1-(1-r)\frac{x}{n} - r\frac{x^2}{n^2}\right)\right\} \leq e^{C\left(\frac{x^2+xy}{n}\right) + \mathcal{O}\left(\frac{yx^2}{n^2}\right)} \leq e^{C'n^{2\gamma-1}}.$$

Together with the fact that $n!/((n-y)!\,n^y) \leq 1$, we obtain that

$$\frac{\mathbb{P}[B_{t+1} = y \,|\, B_t = x]}{\mathbb{P}[Y^-_{t+1} = y \,|\, Y^-_t = x]} \leq e^{C_r n^{2\gamma-1}} \tag{A.1}$$

for a constant $C_r > 0$ independent of x and y (recall that $\gamma < 1/2$).

The same analysis, separating the cases $y = 0, 1$ and $y \geq 2$ and using the fact that for $1 < y \leq n^\gamma$ we have

$$\frac{n!}{(n-y)!n^y} = \prod_{j=0}^{y-1}\left(1 - \frac{j}{n}\right) \geq e^{-\frac{y(y-1)}{n}} \geq e^{-1/n},$$

shows that

$$\frac{\mathbb{P}[B_{t+1} = y \,|\, B_t = x]}{\mathbb{P}[Y^-_{t+1} = y \,|\, Y^-_t = x]} \geq e^{C'_r n^{-1}} \tag{A.2}$$

for a constant C'_r independent of x and y. Putting together Eq. (A.1) and Eq. (A.2), we obtain that

$$\mathbb{P}[B_{t+1} = y \,|\, B_t = x] = \mathbb{P}[Y^-_{t+1} = y \,|\, Y^-_t = x]\big(1 + o(n^{2\gamma-1})\big),$$

where the remainder is uniform in $x \in \{1, \dots, n^\gamma\}$ and $y \in \{0, \dots, n^\gamma\}$. As a consequence, for any $x_0, \dots, x_{k-1} \in \{1, \dots, n^\gamma\}$ and $x_k \in \{0, \dots, n^\gamma\}$, we have

$$\begin{aligned}\mathbb{P}[B_0 = x_0, \dots, B_k = x_k] &= \mathbb{P}[Y^-_0 = x_0, \dots, Y^-_k = x_k]\big(1 + o(n^{2\gamma-1})\big)^k \\ &= \mathbb{P}[Y^-_0 = x_0, \dots, Y^-_k = x_k]\, e^{o(kn^{2\gamma-1})}.\end{aligned}$$

Summing over all paths corresponding to the event considered and using the fact that $kn^{2\gamma-1} = o(1)$ as $n \to \infty$, we can write that

$$\mathbb{P}_i\big[\tau^B_{0,n^\gamma} > k\big] = \mathbb{P}_i\big[\tau^{Y^-}_{0,n^\gamma} > k\big]\big(1 + o(1)\big)$$

and

$$\mathbb{P}_i\big[\tau^B_{0,n^\gamma} \le k\,;\, B_{\tau_{0,n^\gamma}} \ge n^\gamma\big] = \mathbb{P}_i\big[\tau^{Y^-}_{0,n^\gamma} \le k\,;\, Y^-_{\tau_{0,n^\gamma}} \ge n^\gamma\big]\big(1 + o(1)\big),$$

the latter being the probabilities that the process leaves $\{1, \dots, n^\gamma - 1\}$ before time k and by going over n^γ. Now,

$$\mathbb{P}_i\big[\tau^B_0 > k\big] = \mathbb{P}_i\big[\tau^B_{0,n^\gamma} > k\big] + \mathbb{P}_i\big[\tau^B_{0,n^\gamma} \le k\,;\, \tau^B_0 > k\big]. \tag{A.3}$$

From the above, the first term in the r.h.s. is equal to the corresponding term for Y^- up to a vanishing error term. As concerns the second quantity in the r.h.s., it is bounded by

$$\mathbb{P}_i\big[\tau^B_{0,n^\gamma} \le k\,;\, B_{\tau_{0,n^\gamma}} \ge n^\gamma\big] = \mathbb{P}_i\big[\tau^{Y^-}_{0,n^\gamma} \le k\,;\, Y^-_{\tau_{0,n^\gamma}} \ge n^\gamma\big]\big(1 + o(1)\big)$$

To finish the proof, let us show that the probability that Y^-, starting below n^α, reaches n^γ before becoming extinct tends to 0 as $n \to \infty$. Together with Eq. (A.3), this will give us the desired result since Eq. (A.3) holds also with B replaced by Y^-.

Since Y^- cannot grow beyond n^γ unless one of the $i \le Cn^\alpha$ families emanating from an initial individual reaches $n^{\gamma-\alpha}/C$, we have

$$\mathbb{P}_i\big[\tau^{Y^-}_{0,n^\gamma} \le k\,;\, Y^-_{\tau_{0,n^\gamma}} \ge n^\gamma\big] \le i\mathbb{P}_1\big[Y^- \text{ ever reaches } n^{\gamma-\alpha}/C\big] \le n^\alpha \sum_{j=1}^{\infty} \mathbb{P}_1\big[Y^-_j \ge n^{\gamma-\alpha}/C\big].$$

But $\mathbb{E}_1[Y^-_j] = (1-r)^j$, and so the Markov inequality applied to each term in the sum gives us

$$\mathbb{P}_i\big[\tau^{Y^-}_{0,n^\gamma} \le k\,;\, Y^-_{\tau_{0,n^\gamma}} \ge n^\gamma\big] \le n^\alpha \sum_{j=0}^{\infty} Cn^{\alpha-\gamma}(1-r)^j = \frac{C}{r}\, n^{2\alpha-\gamma}.$$

As $\gamma > 2\alpha$, this quantity goes to zero as n tends to infinity. □

B Proof of Eq. (8.1)

Here we derive the approximation of Eq. (8.1) in detail.

Note first that we have the following approximation:

$$\frac{n_{[j]}}{n^j} := \frac{n(n-1)\cdots(n-(j-1))}{n^j} = \prod_{k=1}^{j-1}\left(1-\frac{k}{n}\right) = 1 - \binom{j}{2}\frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

We will first consider some special cases. Fix I, J, K as before such that $|I| = i, |K| = k, |J| = j$. For $M \subseteq I$, let $A(M)$ be the set of parents of vertices in M.

Lemma 19.

$$|B(i+s|i,s)| = \frac{(i+s)!}{2^s}. \tag{B.1}$$

Proof. We have $|I| = i, |K| = s, |J| = i+s$. In this case, no two vertices in I have a common parent, hence

$$|B(i+s|i,s)| = \binom{i+s}{2s}\binom{2s}{\underbrace{2,2,\ldots,2}_{s \text{ times}}}(i-s)! = \frac{(i+s)!}{2^s}.$$

The first factor is the number of ways to select $A(K)$. The second factor is the number of ways to assign parents to vertices in K, each vertex being assigned 2 distinct parents. The last factor is the number of ways to assign parents to vertices in $I\backslash K$, each vertex being assigned a distinct parent. □

Lemma 20.

$$|B(i+s|i,s+1)| = \frac{(i+s)!((i-s-1)(i+3s+2)+4s(s+1))}{2^{s+2}} \tag{B.2}$$

Proof. We have 3 cases.

Case 1: $|A(K)| = 2s+2$, $|A(I\backslash K)| = i-s-2$, and $A(I\backslash K) = J\backslash A(K)$. The contribution to $|B(i+s|i,s+1)|$ in this case is

$$\binom{i+s}{2s+2}\binom{2s+2}{\underbrace{2,2,\ldots,2}_{(s+1) \text{ times}}}\binom{i-s-1}{2}(i-s-2)! = \frac{(i+s)!(i-s-1)(i-s-2)}{2^{s+2}} \tag{B.3}$$

The first factor is the number of ways to select $A(K)$. The second factor is the number of ways to assign parents to vertices in K, each vertex being assigned 2 distinct parents. The last two factors together give the number of ways to assign parents to vertices in $I\backslash K$; which is the number of onto maps from $A(I\backslash K)$ to $J\backslash A(K)$.

Case 2: $|A(K)| = 2s+2$, $|A(I\backslash K)| = i-s-1$, and $|A(K) \cap A(I\backslash K)| = 1$. The contribution to $|B(i+s|i, s+1)|$ in this case is

$$\binom{i+s}{2s+2}\binom{2s+2}{\underbrace{2,2,\ldots,2}_{(s+1)\text{ times}}}(2s+2)(i-s-1)! = \frac{(i+s)!(i-s-1)(2s+2)}{2^{s+1}} \tag{B.4}$$

The first two factors are as in Case 1. The third factor is the number of ways to select the (only) vertex in $A(K) \cap A(I\backslash K)$. The last factor is the number of ways to assign distinct parents to vertices in $I\backslash K$.

Case 3: $|A(K)| = 2s+1$ and $A(I\backslash K) = J\backslash A(K)$. We have $|A(I\backslash K)| = i-s-1$. Also, there are $s-1$ vertices in K with two distinct parents each, 2 vertices in K with one common parent, and $i-s-1$ vertices in $I\backslash K$ with distinct parents. The contribution to $|B(i+s|i, s+1)|$ in this case is

$$\binom{i+s}{2s-2, 3, i-s-1}\binom{s+1}{2}6\binom{2s-2}{\underbrace{2,2,\ldots,2,}_{s-1\text{ times}}}(i-s-1)! = \frac{(i+s)!s(s+1)}{2^{s}} \tag{B.5}$$

The first factor gives the number of partitions of J in 3 parts as described above. The second factor is the number of ways to select the two vertices in K that have a common parent; and the third factor is the number of ways to assign 2 parents to each of them, with one parent in common. The forth factor is the number of ways to assign parents to the remaining $s-1$ vertices in K, each vertex being assigned 2 distinct parents. The last factor is the number of ways to assign distinct parents to vertices in $I\backslash K$.

Now $|B(i+s, i, s+1)|$ is obtained by adding the contributions in Eqs. (B.3),(B.4) and (B.5). □

Let us now return to the Proof of Eq. (8.1) and consider the case $j \geq i$. We must have $k \geq j-i$, i.e., more recombinants than additional lineages. To find an approximation of ${}^{n,r}P_{i,j}$, we use the expression obtained in Theorem 1 (more precisely, we use Eq. (4.4)). We first evaluate the order of magnitude (in n) of each term appearing in the sum over $k \in \{j-i, \ldots, i\}$. We have

$$\binom{n}{j}\frac{1}{n^{i-k}\binom{n}{2}^k} = \frac{2^k n!}{j!(n-j)!n^i(n-1)^k} = \frac{2^k}{j!n^{i-j}(n-1)^k}\left(1-\binom{j}{2}\frac{1}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right). \tag{B.6}$$

Hence, the term corresponding to k will be of the order of $\mathcal{O}(n^{-2})$ whenever $k \geq j-i+2$. From now on, we thus consider the terms $k = j-i$ and $k = j-i+1$ only. Let us write $j = i+s$, with $0 \leq s \leq i$. Suppose first that $k = s$. Using again the notation $|B(j|i,k)|$ for the number of bipartite graphs defined in Sect. 4 (where we replaced the sets I, J, K

by their cardinalities since only these quantities matter), we can write

$$\begin{aligned}
&\binom{n}{i+s}\frac{1}{n^{i-s}\binom{n}{2}^s}\binom{i}{s}r^s(1-r)^{i-s}|B(i+s|i,s)| \\
&\qquad= \left(1-\binom{i+s}{2}\frac{1}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right)\frac{2^s}{(i+s)!(1-1/n)^s}\binom{i}{s}r^s(1-r)^{i-s}\frac{(i+s)!}{2^s} \\
&\qquad= \binom{i}{s}r^s(1-r)^{i-s}\left(1-\binom{i+s}{2}\frac{1}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right)\left(1+\frac{s}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right) \\
&\qquad= \binom{i}{s}r^s(1-r)^{i-s}\left(1-\left\{\binom{i+s}{2}-s\right\}\frac{1}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right),
\end{aligned}$$

where the first equality uses Eq. (B.6) and Eq. (B.1).

Let us now suppose that $k = s+1$. This time, we have

$$\begin{aligned}
&\binom{n}{i+s}\frac{1}{n^{i-s-1}\binom{n}{2}^{s+1}}\binom{i}{s+1}r^{s+1}(1-r)^{i-s-1}|B(i+s|i,s+1)| \\
&\quad= \frac{2^{s+1}}{(i+s)!(n-1)(1-1/n)^s}\left(1-\binom{i+s}{2}\frac{1}{n}+\mathcal{O}\left(\frac{1}{n^2}\right)\right)\binom{i}{s+1}r^{s+1}(1-r)^{i-s-1} \\
&\qquad\qquad\times\frac{(i+s)!\big((i-s-1)(i+3s+2)+4s(s+1)\big)}{2^{s+2}} \\
&\quad= \frac{1}{2(n-1)}\binom{i}{s+1}r^{s+1}(1-r)^{i-s-1}\big((i-s-1)(i+3s+2)+4s(s+1)\big)+\mathcal{O}\left(\frac{1}{n^2}\right),
\end{aligned}$$

where we have used Eq. (B.2). Combining the above, we obtain the desired approximation when $j \geq i$.

Next, we consider the case where $j = i - s$, with $s > 0$. Using again Eq. (B.6), we see that the terms appearing in the sum over k in the expression of ${}^{n,r}P_{i,j}$ will be $\mathcal{O}(1/n^2)$ whenever $s + k \geq 2$, which will be the case whenever $k \geq 1$ or $k = 0$ and $s \geq 2$. We thus concentrate on the case $k = 0$ and $s = 1$ only, corresponding to the scenario where a single pair of lineages coalesces and no recombinations occur. Since there are $\binom{i}{2}$ possible choices for the pair of lineages that coalesces and then $(i-1)!$ possible allocations of parents, we have $|B(i-1|i,0)| = (i-1)!\binom{i}{2}$ and thus

$$\begin{aligned}
\binom{n}{i-1}\frac{1}{n^i}(1-r)^i|B(i-1|i,0)| &= \frac{1}{(i-1)!\,n}\left(1-\mathcal{O}\left(\frac{1}{n}\right)\right)(1-r)^i(i-1)!\binom{i}{2} \\
&= \frac{1}{n}\binom{i}{2}(1-r)^i\left(1-\mathcal{O}\left(\frac{1}{n}\right)\right).
\end{aligned}$$

This entails the approximation given on the first line of Eq. (8.1).