OXFORD GENETICS

Bursts of coalescence within population pedigrees whenever big families occur

Dimitrios Diamantidis,[1] Wai-Tong (Louis) Fan,[1,2] Matthias Birkner,[3] John Wakeley[2,*]

[1]Department of Mathematics, Indiana University, Bloomington, IN 47405, USA
[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[3]Institut für Mathematik, Johannes-Gutenberg-Universität, 55099 Mainz, Germany

*Corresponding author: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. Email: wakeley@fas.harvard.edu

We consider a simple diploid population-genetic model with potentially high variability of offspring numbers among individuals. Specifically, against a backdrop of Wright–Fisher reproduction and no selection, there is an additional probability that a big family occurs, meaning that a pair of individuals has a number of offspring on the order of the population size. We study how the pedigree of the population generated under this model affects the ancestral genetic process of a sample of size two at a single autosomal locus without recombination. Our population model is of the type for which multiple-merger coalescent processes have been described. We prove that the conditional distribution of the pairwise coalescence time given the random pedigree converges to a limit law as the population size tends to infinity. This limit law may or may not be the usual exponential distribution of the Kingman coalescent, depending on the frequency of big families. But because it includes the number and times of big families, it differs from the usual multiple-merger coalescent models. The usual multiple-merger coalescent models are seen as describing the ancestral process marginal to, or averaging over, the pedigree. In the limiting ancestral process conditional on the pedigree, the intervals between big families can be modeled using the Kingman coalescent but each big family causes a discrete jump in the probability of coalescence. Analogous results should hold for larger samples and other population models. We illustrate these results with simulations and additional analysis, highlighting their implications for inference and understanding of multilocus data.

Keywords: coalescent theory; population pedigree; genealogy; multiple mergers; ancestral inference

Introduction

Population-genetic background

Population geneticists routinely make inferences about the past by applying statistical models to DNA sequences or other genetic data. Because past events have already occurred, these models describe what might have happened. They are necessary because patterns of variation in DNA provide only indirect evidence about the past. But the decisions made in building these statistical models have important consequences for inference. A key question has received little attention: when and how should some parts of the past be treated as random variables, while others are viewed as fixed objects? Our particular concern here will be with the treatment of pedigrees, or the reproductive relationships among diploid individuals.

With limited exceptions the statistical models of population genetics have inherited the initial decisions which Fisher (1922, 1930) and Wright (1931) made in deriving allele frequency spectra and probability density functions of allele frequencies at stationarity. They modeled neutral alleles as well as those under selection in a large well-mixed population which in the simplest case was assumed to be of constant size over time. Accordingly it has been common in population genetics to think of population sizes as fixed, not random. Today's coalescent hidden Markov models, for example, infer a fixed trajectory of population sizes over time under the assumption of neutrality (Li and Durbin 2011; Sheehan *et al.* 2013; Wang *et al.* 2020; Schweiger and Durbin 2023).

Although coalescent models reflect later developments and were a significant shift in thinking for the field, fundamentally they depend on the same assumptions as the classical models of Fisher and Wright (Ewens 1990; Möhle 1999). This is clear even in the earliest treatments of ancestral genetic processes by Malécot (1941, 1946, 1948). What coalescent theory did was to broaden the scope of population genetics beyond forward-time models of changes in allele counts or frequencies to include gene genealogies constructed by series of common-ancestor events backward in time (Kingman 1982; Hudson 1983a, 1983b; Tajima 1983). Mathematically, the forward-time and backward-time models of population genetics are dual to each other (Möhle 1999).

Most importantly for our purposes here, Fisher (1922, 1930) and Wright (1931) obtained their predictions about genetic variation by averaging over an assumed random process of reproduction. The particular random process they used is now called the Wright–Fisher model (Ewens 2004). Because the outcome of the process of reproduction is a pedigree, their method is equivalent to averaging over the random pedigree of the population. That they did this without explanation in this context is somewhat curious given the attention to pedigrees in Fisher's infinitesimal model of quantitative genetics (Fisher 1918; Barton *et al.* 2017) and in Wright's method of path coefficients whose very purpose was to make predictions conditional on pedigrees (Wright 1921a, 1921b, 1921c, 1921d, 1921e, 1922).

The pedigree of the entire population is the set of reproductive relationships of all individuals for all time when reproduction is biparental. The corresponding graph is a genealogy in the usual sense. It has been referred to as an organismal pedigree (Ball et al. 1990) and the population pedigree (Wollenberg and Avise 1998; Wakeley et al. 2012; Ralph 2019). Here we simply call it the pedigree. Patterns of genetic variation depend on the pedigree because genetic inheritance happens within it. In particular, transmission of an autosomal genetic locus forward in time through the pedigree occurs by Mendel's law of independent segregation. Multilocus transmission follows Mendel's law of independent assortment or is mediated by recombination if the loci are linked. These processes, which may also be viewed backward in time, are conditional on the pedigree.

Within any pedigree, many possible uniparental paths can be traced backward in time from each individual. If there are two mating types, for example karyotypic females (F) and karyotypic males (M), then one such path might be depicted $F \rightarrow F \rightarrow M \rightarrow F \rightarrow M \rightarrow \cdots$ (Avise and Wollenberg 1997). For the ancestry of a single allele at an autosomal locus in a single individual, applying Mendel's law of independent segregation backward in time generates these uniparental paths with equal probabilities $1/2^g$ for any path extending g generations into the past. When two such paths meet in the same individual, then with equal probability, $1/2$, the alleles either coalesce in that individual or remain distinct. Thus coalescence is conditional on the pedigree, and many possible gene genealogies are embedded in any one pedigree. Some loci, such as the mitochondrial genome and the Y chromosome in humans, are strictly uniparentally inherited. They follow only paths $F \rightarrow F \rightarrow F \rightarrow \cdots$ and $M \rightarrow M \rightarrow M \rightarrow \cdots$, respectively, and two such paths coalesce with probability one when they meet. For these loci, there is only one gene genealogy within the pedigree.

Under Wright–Fisher reproduction, parents are chosen at random uniformly from among all possible parents. This determines the structure of the pedigree in that generation. Assume that there are N_f karyotypic females and N_m karyotypic males in every generation. For autosomal loci, the familiar effective population size $N_e = 4N_f N_m / (N_f + N_m)$ from classical forward-time analysis (Wright 1931) and its backward-time counterpart $1/(2N_e)$ for the pairwise coalescence probability (Möhle 1998a, 1998b) come from averaging over the possible outcomes of reproduction in a single generation. Sections 6.1 and 6.2 in Wakeley (2009) give a detailed illustration. For uniparentally inherited loci, this averaging yields $1/N_f$ and $1/N_m$ for the pairwise coalescence probabilities. In the diploid monoecious Wright–Fisher model or by setting $N_f = N_m = N/2$, these average probabilities of coalescence become $2/N$ for uniparentally inherited loci and $1/(2N)$ for autosomal loci. For simplicity in this work, we will focus on the diploid monoecious Wright–Fisher model.

Averaging over pedigrees is what leads to the effective population size, N_e, being the primary determinant of forward-time and backward-time dynamics in neutral population genetic models. For very large populations, N_e becomes the only parameter of the Wright–Fisher diffusion (Ewens 2004) and the standard neutral or Kingman coalescent process (Sjödin et al. 2005). In particular, N_e sets the timescale over which mutation acts to produce genetic variation. Such averaging removes the pedigree as a possible latent variable which could be important for structuring genetic variation. As a result, from the perspective of the standard neutral coalescent, information about the (marginal) gene genealogical process together with the mutation process is all we can hope to infer from genetic data (Sjödin et al. 2005).

The situation in which it makes the most sense to use this marginal process of coalescence is when the only data available come from a single nonrecombining locus. In fact, the initial applications of ancestral inference to single-locus data, namely to restriction fragment length polymorphisms in human mitochondrial DNA (mtDNA) (Brown 1980; Cann et al. 1987) then to sequences of the hypervariable control region (Vigilant et al. 1989, 1991; Ward et al. 1991; Di Rienzo and Wilson 1991), did not even use of the statistical machinery of population genetics. They instead took the gene genealogy and times to common ancestry to be fixed, and estimated them using traditional phylogenetic methods (Felsenstein 2004). But this in turn spurred the development of likelihood-based methods of ancestral inference using coalescent prior distributions for gene genealogies (Lundstrom et al. 1992; Griffiths and Tavaré 1994; Kuhner et al. 1995). We note that in the interim it has also become common to treat phylogenies as random variables using a wide variety of prior models (Ronquist et al. 2012; Suchard et al. 2018; Bouckaert et al. 2019).

The desirability of accounting for variation in gene genealogies became especially clear when the first sample DNA sequences of the human ZFY gene was obtained and was completely monomorphic (Dorit et al. 1995). The mutation rate is lower on the Y chromosome than in the hypervariable region of mtDNA but it is not equal to zero (Brown et al. 1979; Wilson et al. 1985; Ingman et al. 2000; The 1000 Genomes Project Consortium 2015). Using coalescent priors it was shown that the complete lack of variation in that first sample at ZFY was consistent with a wide range of times to common ancestry for the Y chromosome (Dorit et al. 1995; Donnelly et al. 1996; Fu and Li 1996; Weiss and von Haeseler 1996).

If instead data come from multiple loci, it is impossible to ignore variation in gene genealogies regardless of whether one thinks of the pedigree as fixed or random. Variation in gene genealogies across the genome is, for example, what coalescent hidden Markov models use to estimate trajectories of population sizes. The simplest illustrative case is when the loci are on different chromosomes or far enough apart on the same chromosome that they assort independently into gametes, and when within each locus there is no recombination. The gene genealogies of such loci will vary due to the particular outcomes of Mendelian segregation. They will also be independent due to Mendelian assortment, but only given the pedigree. Mendel's law of independent assortment is a law of conditional independence. It applies once relationships have been specified.

However, throughout much of the history of population genetics, it was assumed that independently assorting loci would have completely independent evolutionary histories. In coalescent theory, this means independent gene genealogies. As Charlesworth (2022) recently noted, Fisher (1922, 1930) and Wright (1931) intended their results on allele frequency spectra and probability density functions of allele frequencies at stationarity to be descriptions of the behavior of large numbers of independently assorting loci in the same genome. This is evident in their application of these distributions to the multiple Mendelian factors of Fisher's infinitesimal model (Fisher 1918) in their arguments about the Dominance Ratio (Fisher 1922; Charlesworth 2022).

An early application to multilocus data was made by Cavalli-Sforza and Edwards (1967) and Felsenstein (1973) who developed likelihood-based methods to infer trees of populations within species from multilocus allele-frequency data, specifically human blood group data, by modeling the forward-time process of random genetic drift independently at each locus conditional on the population tree. Felsenstein (1981) further developed and

applied these methods to gel electrophoretic data. Today's methods of inferring admixture from single nucleotide polymorphism (SNP) data using *F*-statistics are based on the same notion of independence (Patterson *et al.* 2012).

Like the population size itself, demographic features such as the splitting of populations have mostly been treated as fixed in population genetics. Cavalli-Sforza and Edwards (1967) and Felsenstein (1973) did discuss but did not implement prior models for trees of populations, specifically as outcomes of birth–death processes. More recently, Heled and Drummond (2009) did implement this in a coalescent framework for multilocus sequence data, using the prior distribution of Gernhard (2008); see also Lambert and Stadler (2013). Yang (2002) and Rannala and Yang (2003) took a different approach, using gamma-distributed pseudopriors for times in trees.

Previous work on pedigrees

Although the underlying assumption that unlinked loci have completely independent evolutionary histories is mistaken because it would require them having independent pedigrees, most theoretical work has followed the lead of Fisher (1922, 1930) and Wright (1931). Examples in which this is made explicit include Karlin and McGregor (1967), Kimura (1969), Ewens (1974), and Ewens and Maruyama (1975). Multiplying likelihoods across loci in applications to genetic data subsequently became common practice (Watterson 1985; Padmadisastra 1988; Sawyer and Hartl 1992; Wakeley 1999; Nielsen 2000; Wooding and Rogers 2002; Adams and Hudson 2004). It is built into current inference packages, including ∂a∂i (Gutenkunst *et al.* 2009), momi2 (Kamm *et al.* 2020), and *fastsimcoal*2 (Excoffier *et al.* 2013, 2021).

As it happens, this conceptual mistake has almost no practical ramifications if the population is large and well mixed, and the variance of offspring numbers among individuals is not too large. Ball *et al.* (1990) were the first to address the question of gene genealogies within pedigrees. They used simulations to show that the distribution of pairwise coalescence times among loci on a single pedigree do not differ substantially from their distributions among loci which have independent pedigrees. Their population model was similar to the Wright–Fisher model with population size $N = 100$: Poisson offspring numbers with strong density regulation to a carrying capacity of 100. Their results were based on simulations of 50 gene genealogies for each of 50 pedigrees and samples of size $n = 100$, in which a single gene copy was taken at random at each locus within each individual. They also showed that the distribution of coalescence times among pairs of individuals on a single pedigree are very similar to the prediction obtained by averaging over pedigrees.

Wakeley *et al.* (2012) confirmed these results and related them to coalescent theory using simulations of 10^8 gene genealogies for $n = 2$ for each of 10^4 pedigrees and population sizes up to 10^5, together with more limited treatments of larger samples $n = 20$ and 100. Pedigrees were constructed in three different ways: assuming Wright–Fisher reproduction, using empirically derived human family structures, and under a model in which the outcome of a single generation of Wright–Fisher reproduction was repeated over time, resulting in a so-called cyclical pedigree. These simulations showed that times to common ancestry conditional on the pedigree conform well to the probability law underlying coalescent theory, with a constant coalescence probability $1/(2N_e) = 1/(2N)$ each generation under the Wright–Fisher model with $N_f = N_m = N/2$, except for in the recent past where they differ greatly and depend on the pedigree. But they also showed that as long as N is large these idiosyncrasies in the short-time

behavior of the ancestral process have little effect on the overall distribution of coalescence times given the pedigree, whether it is among independent loci in the same individuals or among independently sampled pairs of individuals.

Here "recent" means proportional to $\log_2(N)$ generations, which is the timescale for the first occurrence of a common ancestor of all present-day individuals (Chang 1999) and for the complete overlap of all individuals' ancestries in a well-mixed biparental population (Chang 1999; Derrida *et al.* 1999, 2000a, 2000b; Barton and Etheridge 2011; Coron and Le Jan 2022). This is much shorter than the *N*-generations timescale required for common ancestry of uniparental genetic lineages (Chang 1999; Donnelly *et al.* 1999). Additional work on these properties of pedigrees include Rohde *et al.* (2004) and Lachance (2009) who showed that population structure and inbreeding do not strongly affect the time to the first occurrence of a common ancestor of all individuals. Blath *et al.* (2014) proved that the ancestries of the great majority of individuals overlap even in cyclical pedigrees as $N \to \infty$. Matsen and Evans (2008) and Gravel and Steel (2015) showed that ancestral genetic lineages pass through only a small minority of the shared pedigree ancestors. See Agranat-Tamir *et al.* (2024) for further developments and an extension to admixed populations. Sainudiin *et al.* (2016) constructed a model with recombination which interpolates between uniparental common ancestry on the *N*-generations timescale and biparental common ancestry on the $\log_2(N)$-generations timescale.

Tyukin (2015) proved what was implied by the simulations of Ball *et al.* (1990) and Wakeley *et al.* (2012), specifically that when the population is large and well mixed the pedigree-averaged coalescent process is a good substitute for the actual coalescent process conditional on the pedigree. Questions of this sort have a long history in mathematical physics and probability theory, where "quenched" and "annealed" are often used to refer to conditional as opposed to averaged processes. Molchanov (1994) and Bolthausen and Sznitman (2002b) provide background and developments in the classical context of random walks in random environments. What Tyukin (2015) proved is that the quenched coalescent process conditional on the pedigree converges to the pedigree-averaged standard neutral or Kingman coalescent process in the limit $N \to \infty$. Tyukin (2015) did this under a broader set of reproduction models with mating analogous to Wright–Fisher but with a general exchangeable distribution of offspring numbers (Cannings 1974) in the domain of attraction of Kingman's coalescent (Möhle and Sagitov 2001; Sagitov 2003).

Since time in the Kingman coalescent process is measured in units proportional to *N* generations, the result of Tyukin (2015) provides insight into the role of the pedigree in the recent ancestry of the sample ($O(\log_2(N))$ generations) under the Cannings and Wright–Fisher models. Specifically, the chance of any events in the recent past which would dramatically alter the rate of coalescence must be negligible as $N \to \infty$. Intuitively we might surmise that (1) individuals randomly sampled from a large well-mixed population are unlikely to be closely related, and barring coalescence for some small number of generations until their ancestries overlap does not affect the limit and (2) by the time their ancestries do overlap in the pedigree, their numbers of ancestors are approaching the population size, making the chance of coalescence of order $1/N$.

Two cases have been identified using simulations where quenched and annealed results are noticeably different. The first is population subdivision, especially with limited migration. Wollenberg and Avise (1998) showed that as the migration distance decreases in a linear habitat, fewer independent loci are needed to accurately measure pairwise coefficients of coancestry

on the pedigree. Wilton *et al.* (2017) described increasingly strong pedigree effects as the migration rate decreased in a two-subpopulation model, specifically spikes in the distribution of pairwise coalescence times corresponding to the particular series of individual migration events that occurred in the ancestry. These results illustrate how even single gene genealogies may contain information about events in the ancestry of geographically structured populations, via the pedigree. Thus, they are relevant for applications of ancestral inference to single-locus data, such as mtDNA, as well as to the broader field of intraspecific phylogeography (Avise *et al.* 1987; Avise 1989, 2000). For recent empirical studies of spatiotemporally structured pedigrees and their effects on local patterns of genetic variation, see Aguillon *et al.* (2017) and Anderson-Trocmé *et al.* (2023).

The second situation in which pedigrees have a strong effect on coalescence times and gene genealogies is when there is a high variance of offspring numbers among individuals. This variance is comparatively low in the Wright–Fisher model, which has a multinomial distribution of offspring numbers (becoming Poisson as $N \to \infty$). In deriving the standard neutral coalescent process, Kingman (1982) started with the general exchangeable model of Cannings (1974) then assumed that the variance of offspring numbers was finite as $N \to \infty$. Without this assumption, the ancestral limit process is not the Kingman coalescent process but rather a coalescent process with multiple mergers (see below). In addition, in this situation, simulations have shown that the pedigree has a marked effect on genetic ancestries.

Wakeley *et al.* (2016) simulated pedigrees in which a single individual had a very large number of offspring in some past generation and otherwise there was Wright–Fisher reproduction. This large reproduction event greatly increased the probability of coalescence in the generation in which it occurred, causing a spike in the distribution of pairwise coalescence times and altering the allele frequency spectrum. A strong selective sweep at one locus gave similar effects at unlinked loci via the pedigree (Wakeley *et al.* 2016). Similar deviations from standard neutral coalescent predictions are produced by cultural transmission of reproductive success (Guez *et al.* 2023).

Plan of the present work

Here, we present a new quenched limit result for coalescent processes in fixed pedigrees under a modified Wright–Fisher model which allows for large reproduction events. Wright–Fisher reproduction on its own produces various kinds of large reproduction events but these are all extremely rare. Our model adds big families with two parents and numbers of offspring proportional to the population size. These are inserted into the pedigree either on the same N-generations timescale as coalescent events in the Wright–Fisher background model or much faster so that they completely dominate the ancestral process. In both cases, the limiting ancestral process conditional on the pedigree is different than the limiting ancestral process which averages over pedigrees. For simplicity, we focus on samples of size two. Consistent with the results of Tyukin (2015), our result reduces to the Kingman coalescent with $n = 2$ in the case where there are no big families.

Note that the corresponding averaged process is not the Kingman coalescent but rather a coalescent process with multiple mergers; see Tellier and Lemaire (2014) for an overview of these models in the context of population genetics. Multiple-merger coalescent processes arise as $N \to \infty$ limits when the variance of offspring numbers is large, and so may be applicable to a broad range of species with the capacity for high fecundity (Eldon 2020). They also arise from recurrent selective sweeps, when differences in offspring numbers are determined by individuals' genotypes

(Durrett and Schweinsberg 2004, 2005; Schweinsberg and Durrett 2005). Whereas the Kingman coalescent includes only binary mergers of ancestral genetic lineages, these more general processes allow mergers of any size. At issue here is how these models should be interpreted and applied.

By averaging over the process of reproduction, two kinds of multiple-merger coalescent processes have been described: Λ-coalescents which have asynchronous multiple mergers (Donnelly and Kurtz 1999; Pitman 1999; Sagitov 1999) and Ξ-coalescents which have simultaneous multiple mergers (Schweinsberg 2000; Möhle and Sagitov 2001; Sagitov 2003). Multiple-merger processes for diploid organisms are always Ξ-coalescents with the possibility of an even number simultaneous mergers (Birkner *et al.* 2018). Our quenched limit result brings into question what seems like a natural extension from applications of the standard neutral coalescent model, namely to assume that multiple-merger models may be applied independently to independent loci as has been done both in theoretical explorations (Der and Plotkin 2014; Eldon *et al.* 2015; Spence *et al.* 2016; Matuszewski *et al.* 2018) and in analyses of SNP data (Birkner *et al.* 2013a; Blath *et al.* 2016; Árnason *et al.* 2023; Freund *et al.* 2023).

To establish the quenched limit process, we adapt the method that Birkner *et al.* (2013c) used for a quenched limit of a random walk in a random environment. See also the earlier work of Bolthausen and Sznitman (2002a). In this approach, the problem of convergence in distribution is addressed by analyzing a pair of conditionally independent processes, here corresponding to the ancestries of samples at two independently assorting loci on the pedigree. As Koskela (2018) has pointed out, positive correlations of coalescence times for pairs of unlinked loci are a hallmark of (pedigree-averaged) multiple-merger coalescent models. Our result frames this in terms of pedigrees, in which big families are the only elements that persist as $N \to \infty$. If a big family has occurred in a particular generation, the probability of coalescence is greatly increased in that generation for all loci. All other aspects of the pedigree, that is to say the outcomes of ordinary Wright–Fisher reproduction, "average out" such that the Kingman coalescent process describes the ancestral process during the times between big families.

Theory and results

In this section, we present the population model considered in this paper, the mathematical statement of our main result and its proof. This result (Theorem 1) is stated as a convergence of the *conditional* distribution of the coalescence time of a pair of gene copies, given that we know the pedigree and which individuals were sampled. We assume that the pedigree is the outcome of a random process of reproduction, the population model described in the following section, and that the two individuals are sampled without replacement from the current generation. To connect with known results and highlight the effect of conditioning, we first state and prove the corresponding result (Lemma 1) for the *un*conditional distribution of the coalescence time. This corresponds to fixing the sampled individuals and averaging over the pedigree. We close this section with simulations illustrating multilocus genetic ancestry and further analysis showing how nonzero correlations of coalescence times at unlinked loci result from averaging over the pedigree.

The population model

We consider a diploid, monoecious, biparental, panmictic population of constant fixed size $N \in \mathbb{N}$ with discrete, nonoverlapping generations. Implicitly there is no selection, but we do not in fact model mutation or genetic variation, only the generation of the pedigree

and coalescence within it. There are two different types of reproduction. With high probability, reproduction follows the diploid biparental Wright–Fisher model. With small probability α_N each generation, there is a highly reproductive pair whose offspring comprise a proportion $\psi \in [0, 1]$ of the population. Note that ψ is a fixed deterministic constant. More precisely, for each positive integer g, the reproductive dynamics between the parent generation $g + 1$ and the offspring generation g is given as follows:

1. With probability $1 - \alpha_N$, each individual in the next generation is formed by choosing two parents at random, uniformly with replacement from the N adults of the current generation. Genetically, each offspring is produced according to Mendel's laws which means each of the two gene copies in a parent is equally likely to be the one transmitted to the offspring. In this case, we call g a "Wright–Fisher generation." An example of this standard reproduction dynamics between the parent generation $g + 1$ and the offspring generation g is depicted below for a population of size $N = 7$.
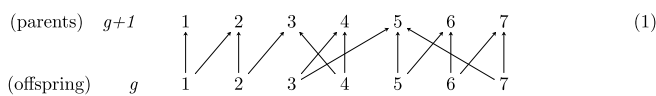
(parents) $g+1$
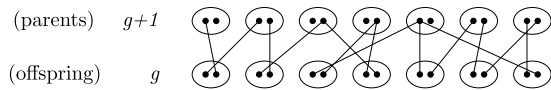
(offspring) g

2. With probability α_N, a pair of adults is chosen uniformly without replacement to have a very large number of offspring, $[\psi N]$ where $\psi \in [0, 1]$ is a fixed fraction of the population. The other $N - [\psi N]$ offspring are produced as above according to the Wright–Fisher model. In this case, we call g a "generation with a big family." An example of this special reproduction dynamics is depicted below for $N = 7$ and $\psi = 0.72$ in which the highly reproductive pair $(I_1, I_2) = (4, 5)$ in generation $g + 1$ has $[\psi N] = [0.72 \cdot 7] = 5$ offspring in generation g.

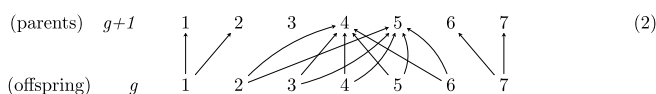(parents) $g+1$

(offspring) g
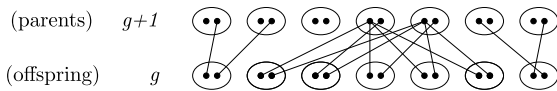
These two possibilities happen independently for all generations $g \in \mathbb{Z}_{\geq 0}$. The classical Wright–Fisher model corresponds to the case when $\alpha_N = 0$. In this case, every $g \in \mathbb{Z}_{\geq 0}$ is a Wright–Fisher generation. Note, we allow selfing with probability $1/N$ for all offspring produced by Wright–Fisher reproduction but we assume that the $[\psi N]$ offspring of big families have two distinct parents.

The *parent assignment* between (parental) generation $g + 1$ and (offspring) generation g is the collection of edges connecting the offspring with their parents. The diagram in equation (1) below shows the parent assignment corresponding to the example above in which g is a Wright–Fisher generation.

(parents) $g+1$ 1 2 3 4 5 6 7   (1)

(offspring) g 1 2 3 4 5 6 7

On the other hand, the diagram in equation (2) below shows the parent assignment corresponding to the example above in which g is a generation with a big family.

(parents) $g+1$ 1 2 3 4 5 6 7   (2)

(offspring) g 1 2 3 4 5 6 7

Pedigree

The collection of all the parent assignments among all pairs of consecutive generations is called the *pedigree* and it is denoted as $\mathcal{A}^{(N)}$. The pedigree models the set of all family relationships among the members of the population for all generations. The pedigree is shared among all loci. It is the structure through which genetic lineages are transmitted. Patterns of ancestry, or gene genealogies are outcomes of Mendelian inheritance in this single shared pedigree.

Frequency of big families

Recall that α_N denotes the probability of a big family to appear in a generation. We set

$$\alpha_N = \frac{\lambda}{N^\theta}, \qquad (3)$$

where $\theta \in (0, 1]$ and $\lambda \in \mathbb{R}_{\geq 0}$ is a fixed parameter which determines the relative frequency of big families on the timescale of N^θ generations.

Timescale

Suppose two individuals are sampled uniformly *without* replacement among the N individuals of the current generation $g = 0$ and we sample one gene copy from each. Let $\tau^{(N,2)}$ be the pairwise coalescence time, that is, the number of generations in the past until the two sampled gene copies coalesce. How long is the pairwise coalescence time $\tau^{(N,2)}$? This will depend on N and also on θ owing to our assumption (3). In considering the limiting ancestral process for the sample, we rescale time so that it is measured in units of N^θ generations. We study the distribution of the rescaled pairwise coalescence time, $\tau^{(N,2)}/N^\theta$, with different results depending on whether $\theta \in (0, 1)$ or $\theta = 1$. In the latter case, our timescale is N generations, which we note is $1/2$ the usual coalescent timescale for diploids. In the former case, where we may infer from equation (3) that big families will dominate the ancestral process, the timescale is accordingly much shorter than the usual coalescent timescale. Coalescence times in both cases also depend on a combined parameter $\psi^2/4$ which is the limiting probability of coalescence when a big family occurs.

Limiting process by averaging over the pedigree

For reference and to illustrate our choice of timescale, we begin with a Kingman coalescent approximation for the pairwise coalescence time in the classical Wright–Fisher model, here the special case $\theta = 1$ and $\lambda = 0$ or $\alpha_N = 0$. Averaging over the process of reproduction in a single generation gives a coalescence probability of $1/(2N)$. With $\theta = 1$, we measure time in units of N generations. To parallel the derivation of our main result, we consider the probability that the coalescence time $\tau^{(N,2)}$ is more than $[tN]$ generations. The limiting ancestral process is obtained as

$$\mathbb{P}^{(N)}(\tau^{(N,2)} > [tN]) = \left(1 - \frac{1}{2N}\right)^{[tN]} \to e^{-t/2} \quad \text{as } N \to \infty. \qquad (4)$$

In words, the rescaled coalescence time $\frac{\tau^{(N,2)}}{N}$ converges in distribution to an exponential random variable with rate parameter $1/2$.

Before stating our main result, we first prove Lemma 1 below, generalizing (4) to our population model, in the sense that for $\theta = 1$ with $\psi = 0$ or $\lambda = 0$ in Lemma 1 we recover (4).

Lemma 1. *Let $\lambda \in \mathbb{R}_{\geq 0}$, $\theta \in (0, 1]$, and set $\alpha_N = \frac{\lambda}{N^\theta}$. The rescaled coalescence time $\frac{\tau^{(N,2)}}{N^\theta}$ converges in distribution to an exponential random variable with rate parameter*

$$\begin{cases} \lambda \frac{\psi^2}{4}, & \text{when } \theta \in (0, 1) \\ \frac{1}{2} + \lambda \frac{\psi^2}{4}, & \text{when } \theta = 1. \end{cases} \tag{5}$$

We note that in Birkner *et al.* (2013b), the full ancestral recombination graph (ARG) for samples of arbitrary size and genomes consisting of arbitrary numbers of linked loci is described for a population model nearly identical to ours here. The ancestral recombination graph (Hudson 1983a; Griffiths and Marjoram 1997), like the Kingman coalescent itself, averages over the pedigree. Lemma 1 describes the marginal ancestral process for a sample of size two at a single locus.

Proof of Lemma 1

The lineage dynamics of our model can be analyzed using a Markov chain. In any generation g in the past, the ancestral lineages of a pair of gene copies must be in one of the three states $\{\xi_0, \xi_1, \xi_2\}$, where

$\xi_0 = (\bullet)(\bullet)$ represents two ancestral lineages in two distinct individuals,

$\xi_1 = (\bullet\bullet)$ represents two ancestral lineages on different chromosomes in the same individual,

$\xi_2 = (\bullet)$ represents that the ancestral lineages have coalesced.

The diploid ancestral process for a pair of gene copies can thus be represented as a Markov chain $(M_g)_{g \in \mathbb{Z}_{\geq 0}}$ with state space $\{\xi_0, \xi_1, \xi_2\}$, where M_g is the state of the two lineages g generations in the past. Its one-step transition matrix Π_N is given by

$$\Pi_N := (1 - \alpha_N)\Pi_N^{\text{WF}} + \alpha_N \Pi_N^{\text{BF}}, \tag{6}$$

where

$$\Pi_N^{\text{WF}} = \begin{array}{c} \\ \xi_0 \\ \xi_1 \\ \xi_2 \end{array} \begin{array}{ccc} \xi_0 & \xi_1 & \xi_2 \\ \begin{bmatrix} 1 - \frac{1}{N} & \frac{1}{2N} & \frac{1}{2N} \\ 1 - \frac{1}{N} & \frac{1}{2N} & \frac{1}{2N} \\ 0 & 0 & 1 \end{bmatrix} \end{array} \tag{7}$$

and

$$\Pi_N^{\text{BF}} = \begin{array}{c} \\ \xi_0 \\ \xi_1 \\ \xi_2 \end{array} \begin{array}{ccc} \xi_0 & \xi_1 & \xi_2 \\ \begin{bmatrix} 1 - \frac{\psi^2}{2} & \frac{\psi^2}{4} & \frac{\psi^2}{4} \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} + O\left(\frac{1}{N}\right). \tag{8}$$

The matrix Π_N^{WF} in equation (7) is the transition matrix for a Wright–Fisher generation, whereas Π_N^{BF} in equation (8) is for a generation with a big family. The entries of Π_N^{BF} in (8) are derived by conditioning on the parent assignment(s) for the individual(s) containing the ancestral lineages, with respect to the highly reproductive pair. For instance, for ancestral lineages currently in two distinct individuals, the coalescence probability is 1/4 if both individuals are members of the big family and 1/(2N) otherwise. Thus, we have

$$\mathbb{P}(M_{g+1} = \xi_2 \mid M_g = \xi_0, \text{ BF}) = \frac{[\psi N]([\psi N] - 1)}{N(N-1)} \frac{1}{4}$$
$$+ \left(1 - \frac{[\psi N]([\psi N] - 1)}{N(N-1)}\right) \frac{1}{2N} = \frac{\psi^2}{4} + O\left(\frac{1}{N}\right)$$

for the transition $\xi_0 \to \xi_2$ in Π_N^{BF}, and where we have also specified that this contribution to the overall probability in equation (6) is conditional on the occurrence of a big family.

The rest of the proof is a straightforward application of Möhle (1998a, Lemma 1). This is a separation-of-timescales result. To see how it works, using equation (3) we can rewrite equation (6) as

$$\Pi_N := A + \frac{1}{N^\theta} B_N + O\left(\frac{1}{N^{\theta+1}}\right), \tag{9}$$

where

$$A = \begin{array}{c} \\ \xi_0 \\ \xi_1 \\ \xi_2 \end{array} \begin{array}{ccc} \xi_0 & \xi_1 & \xi_2 \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \tag{10}$$

and

$$B_N = \begin{array}{c} \\ \xi_0 \\ \xi_1 \\ \xi_2 \end{array} \begin{array}{ccc} \xi_0 & \xi_1 & \xi_2 \\ \begin{bmatrix} -\frac{N^\theta}{N} - \lambda\frac{\psi^2}{2} & \frac{N^\theta}{N}\frac{1}{2} + \lambda\frac{\psi^2}{4} & \frac{N^\theta}{N}\frac{1}{2} + \lambda\frac{\psi^2}{4} \\ -\frac{N^\theta}{N} & \frac{N^\theta}{N}\frac{1}{2} & \frac{N^\theta}{N}\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \end{array}. \tag{11}$$

The matrix A contains the fastest parts of the process. The matrix B_N contains the next-fastest parts of the process, specifically those occurring on the timescale of N^θ generations.

Möhle's result depends on the existence of equilibrium stochastic matrix $P := \lim_{k \to \infty} A^k$ which in this case is equal to A. Möhle's result also requires the existence of the limiting infinitesimal generator $G := \lim_{N \to \infty} P B_N P$. Note that in our application $B := \lim_{N \to \infty} B_N$ itself converges. From equation (11), it is clear that the limiting result will differ depending on whether $\theta \in (0, 1)$ or $\theta = 1$. If $\theta \in (0, 1)$, the contribution of Wright–Fisher generations to the coalescence rate shrinks to zero in the limit. If $\theta = 1$, the contribution of Wright–Fisher generations, which is 1/2 on our timescale, remains comparable to the contribution of generations with big families in the limit.

Applying Möhle (1998a, Lemma 1) to compute our probability of interest,

$$\mathbb{P}^{(N)}(\tau^{(N,2)} > [tN^\theta]) = (1, 0, 0)\Pi_N^{[tN^\theta]}(1, 1, 0)^T \tag{12}$$

$$= (1, 0, 0)\left(A + \frac{1}{N^\theta}B_N + O\left(\frac{1}{N^{\theta+1}}\right)\right)^{[tN^\theta]}(1, 1, 0)^T \tag{13}$$

$$\to (1, 0, 0)Pe^{tG}(1, 1, 0)^T \quad \text{as } N \to \infty. \tag{14}$$

The initial vector $(1, 0, 0)$ enforces our assumed starting state, ξ_0. The end vector $(1, 1, 0)^T$ enforces the requirement that the lineages remain distinct at generation $[tN^\theta]$, i.e. that the Markov chain $(M_g)_{g \in \mathbb{Z}_{\geq 0}}$ has not reached state ξ_2. Möhle's result Pe^{tG} is in the middle. Recall that P, which here is equal to A, instantaneously adjusts the sample so that the effective starting state is ξ_0 even if the sample state is ξ_1. The lineages then enter the continuous-time process with rate matrix G. Overall we have

$$Pe^{tG}(1, 1, 0)^T = \begin{cases} (e^{-t\lambda\frac{\psi^2}{4}}, e^{-t\lambda\frac{\psi^2}{4}}, 0)^T, & \text{if } \theta \in (0, 1), \\ (e^{-t(\frac{1}{2} + \lambda\frac{\psi^2}{4})}, e^{-t(\frac{1}{2} + \lambda\frac{\psi^2}{4})}, 0)^T, & \text{if } \theta = 1. \end{cases} \tag{15}$$

The right-hand side of equation (14) is equal to equation (5), and the proof of Lemma 1 is complete. $\square$

Remark 1 (Robustness against perturbation of initial condition). The form of P shows that the limiting result in Lemma 1 holds regardless of whether the sample begins in state ξ_0, as we have assumed, or in state ξ_1. So, other sampling schemes could be considered. In fact Lemma 1 still holds if the initial distribution lies in the set $\mathcal{I} := \{(c, 1-c, 0) \in [0, 1]^3 : c \in [0, 1]\}$. This can be seen clearly in equation (15).

Limiting process by conditioning on the pedigree

Our main result is about the conditional distribution. We let

$$F_N(t, \mathcal{A}^{(N,2)}) := \mathbb{P}^{(N)}(\tau^{(N,2)} > [tN^\theta] \,|\, \mathcal{A}^{(N,2)}) \tag{16}$$

be the conditional probability of the event $\{\tau^{(N,2)} > [tN^\theta]\}$ given the (random) pedigree and the sampled pair of individuals. Mathematically, $\mathcal{A}^{(N,2)}$ is the sigma-field (all information) generated by the outcome of the random reproduction of the population and the knowledge which pair of individuals was sampled.

Theorem 1. *Let $\lambda \in \mathbb{R}_{\geq 0}$, $\theta \in (0, 1]$, and set $\alpha_N = \frac{\lambda}{N^\theta}$. For all $t \in (0, \infty)$, we have the following convergence in distribution as $N \to \infty$*

$$F_N(t, \mathcal{A}^{(N,2)}) \to \begin{cases} \left(1 - \frac{\psi^2}{4}\right)^{Y(t)}, & \text{when } \theta \in (0, 1), \\ e^{-t/2}\left(1 - \frac{\psi^2}{4}\right)^{Y(t)}, & \text{when } \theta = 1, \end{cases} \tag{17}$$

where $Y(t)$ is Poisson process with rate λ. In fact, the convergence in equation (17) holds jointly for all $t > 0$, see the discussion in Remark A1 in the Strengthening the convergence in Theorem 1 section of Appendix A for details.

Theorem 1 offers a description of the conditional distribution of the coalescence time $\tau^{(N,2)}$ for a sample of two genes in a population of size N given the pedigree. It says that the law of $\frac{\tau^{(N,2)}}{N^\theta}$, under the conditional probability $\mathbb{P}(\cdot \,|\, \mathcal{A}^{(N,2)})$, converges weakly as $N \to \infty$ to the law of a random variable (call it T) under a probability measure $\mathbb{P}_Y$ that depends on the Poisson process Y with rate λ. Furthermore, the survival function $\mathbb{P}_Y(T > t)$ is equal to the right-hand side of equation (17). In what follows, we will refer to $F_N(t, \mathcal{A}^{(N,2)})$ defined in equation (16) as the discrete survival function.

Theorem 1 has an intuitive interpretation. Taking the case $\theta = 1$, the $e^{-t/2}$ represents the probability that the two lineages have not coalesced by time t due to ordinary Wright–Fisher/Kingman coalescence. Against this smooth backdrop there are $Y(t)$ points, representing essentially instantaneous events in which a big family occurs and the lineages have a large probability, $\psi^2/4$, of coalescing. Thus, there is an additional factor in the survival function representing the probability that the pair does not coalesce in any of these extreme events. The case $\theta \in (0, 1)$ is analogous except the timescale is so short that there is no chance of an ordinary Wright–Fisher/Kingman coalescent event.

Note that when $\lambda = 0$, there are no large reproduction events and $Y(t) \equiv 0$. Then for $\theta \in (0, 1)$, the right-hand side of equation (17) is 1, i.e. there is no coalescence with probability 1. For $\theta = 1$, the right-hand side of equation (17) is $e^{-t/2}$ which is expected from the cumulative distribution function (CDF) of the Kingman coalescent for a sample of size 2, with our timescale. The degenerate case $\lambda > 0$ but $\psi = 0$ effectively gives these same results for any $Y(t)$.

Proof of Theorem 1

Recall that each $g \in \mathbb{Z}_{\geq 0}$ is a Wright–Fisher generation (resp. a generation with a big family) with probability $1 - \alpha_N$ (resp. α_N), independently for all $g \in \mathbb{Z}_{\geq 0}$. The number of generations with big

families in $\{0, 1, \ldots, G-1\}$, denoted by $H_N(G)$, therefore has the binomial distribution $\text{Bin}(G, \alpha_N)$.

We begin by addressing the technical point that we cannot actually know just by looking at the pedigree whether g is a generation with a big family, the way we have defined these as occurring only in special generations. Even in the classical Wright–Fisher model, every individual has the capacity to produce a large number of offspring. But reproductive outcomes as extreme as our big families are exceedingly rare under ordinary Wright–Fisher reproduction when N is large.

To illustrate, consider the event that, spanning generations $g + 1$ and g, there exists a pair of parents with at least $[\psi N]$ offspring. In our population model, this is guaranteed to occur in generations with big families. Note that the two parents of a big family have an additional $\sim \text{Poisson}(2(1 - \psi))$ offspring because the other $N - [\psi N]$ offspring are produced according to the Wright–Fisher model. The event that a pair of parents with at least $[\psi N]$ offspring can also occur randomly in Wright–Fisher generations, but only with small probability

$$\epsilon_N \leq \binom{N}{2}\binom{N}{[\psi N]}\left(\frac{1}{\binom{N}{2}}\right)^{[\psi N]} \leq \frac{2^{[\psi N]-1}}{N^{[\psi N]-2}}.$$

Let $Q_N(G)$ be the number of generations $g \in \{0, 1, \ldots, G-1\}$ in which such an event occurs between $g + 1$ and g. Then $Q_N(G)$ is extremely close to the binomial variable $H_N(G) \sim \text{Bin}(G, \alpha_N)$ because

$$H_N(G) \leq Q_N(G) \quad \text{and} \quad Q_N(G) \leq \text{Bin}(G, \alpha_N + \epsilon_N), \tag{18}$$

where the first inequality holds almost surely and the second is a stochastic dominance. Since $\alpha_N = \frac{\lambda}{N^\theta}$, for each $t \in (0, \infty)$ and $\theta \in (0, 1]$ we have convergence in distribution

$$Q_N([tN^\theta]) \to Y(t) \quad \text{as } N \to \infty \tag{19}$$

which is identical to the limiting result for $H_N([tN^\theta])$. In other words, ϵ_N is so small for any sizeable N, that we are safe in assuming that such extreme events in the pedigree reliably signify generations with big families as defined under our model.

Indeed, from the discussion above we have

$$\lim_{N \to \infty} \mathbb{E}^{(N)}\left[\left|\left(1 - \frac{\psi^2}{4}\right)^{H_N([tN])} - \left(1 - \frac{\psi^2}{4}\right)^{Q_N([tN])}\right|^2\right] = 0 \tag{20}$$

so that we can (and will) in the following computations replace $H_N([tN])$ by $Q_N([tN])$ without changing any limit as $N \to \infty$.

Proof of equation (17) when $\theta = 1$

In this case, it suffices to show that

$$\lim_{N \to \infty} \mathbb{E}^{(N)}\left[\left|F_N(t, \mathcal{A}^{(N,2)}) - e^{-t/2}\left(1 - \frac{\psi^2}{4}\right)^{Q_N([tN])}\right|^2\right] = 0. \tag{21}$$

Expanding the square in equation (21) gives

$$\mathbb{E}^{(N)}\left[F_N^2(t, \mathcal{A}^{(N,2)}) - 2e^{-t/2}F_N(t, \mathcal{A}^{(N,2)})\left(1 - \frac{\psi^2}{4}\right)^{Q_N([tN])}\right.$$
$$\left. + e^{-t}\left(1 - \frac{\psi^2}{4}\right)^{2Q_N([tN])}\right], \tag{22}$$

which requires the computation of three expectations. The first is the expectation of the square of the discrete survival function,

$$\mathbb{E}^{(N)}[F_N^2(t, \mathcal{A}^{(N,2)})]. \tag{23}$$

The second is the expectation of the discrete survival function times the probability that a single pair of lineages does not coalesce in any of the generations with big families in the pedigree up to time t,

$$\mathbb{E}^{(N)}\left[F_N(t, \mathcal{A}^{(N,2)})\left(1-\frac{\psi^2}{4}\right)^{Q_N([tN])}\right], \tag{24}$$

The third is the expectation of the square of the same, latter probability that a single pair of lineages does not coalesce in any of the generations with big families in the pedigree up to time t,

$$\mathbb{E}^{(N)}\left[\left(1-\frac{\psi^2}{4}\right)^{2\,Q_N([tN])}\right]. \tag{25}$$

First term in equation (22)

The expectation in equation (23) can be computed by considering two samples of size 2 whose lineage dynamics are conditionally independent given $\mathcal{A}^{(N,2)}$. Genetically, this corresponds to the ancestral processes of two unlinked loci given the pedigree and the two sampled individuals, and where one gene copy has been sampled at each locus from each of the individuals. Let τ and τ' be the coalescence times of these two pairs of sampled gene copies. Due to the conditional independence of these coalescence times, for all $g \in \mathbb{Z}_+$ we have

$$\mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau > g, \; \tau' > g) = \mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau > g)\mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau' > g) \tag{26}$$

in which $\mathbb{P}_{\mathcal{A}^{(N,2)}}(\cdot)$ is shorthand for $\mathbb{P}(\cdot \,|\, \mathcal{A}^{(N,2)})$ in equation (16). Setting $g = [tN]$ and taking expectations on both sides of equation (26) gives

$$\begin{aligned}
\mathbb{E}^{(N)}[F_N^2(t, \mathcal{A}^{(N,2)})] &= \mathbb{E}^{(N)}[\mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau > [tN])\,\mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau' > [tN])] \\
&= \mathbb{E}^{(N)}[\mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau > [tN], \; \tau' > [tN])] \\
&= \mathbb{P}^{(N)}(\tau > [tN], \; \tau' > [tN]).
\end{aligned} \tag{27}$$

In order to compute the limit as $N \to \infty$ in equation (27), we introduce the ancestral process of two conditionally independent samples given the pedigree.

Joint diploid ancestral process

The stochastic dynamics of the two conditionally independent, given the pedigree, pairs of lineages are described by the *joint diploid ancestral process* $\widetilde{M} := (\widetilde{M}_g)_{g \in \mathbb{Z}_{\geq 0}}$. This is a Markov chain with state space $\mathcal{S} = \{\xi_{00}^{(4)}, \xi_{00}^{(3)}, \ldots, \xi_\Delta\}$ described below, where $\widetilde{M}_g$ is the state of the two pairs of lineages in a common pedigree g generations backwards in time. Denote by $\widetilde{\Pi}_N$ its transition matrix, the derivation of its entries is available in the *The joint diploid ancestral process with 10 states is Markovian* section of Appendix A and its entries are available in the *Transition matrix for a generation with a big family* section of Appendix A for a generation with a big family and in the *Two extremes for the sizes of big families* section of Appendix A for a Wright–Fisher generation.

Similarly to the proof of Lemma 1, denote by $\bullet$ an ancestral lineage of a gene copy in the first pair and by $\star$ the same for the second pair. Parentheses are used to denote individuals. More precisely, consider the following 10 states:

$$\begin{aligned}
\xi_{00}^{(4)} &= (\bullet)(\bullet)(\star)(\star) \\
\xi_{00}^{(3)} &= (\bullet)(\bullet\star)(\star) \\
\xi_{00}^{(2)} &= (\bullet\star)(\bullet\star) \\
\xi_{10}^{(2)} &= (\bullet\bullet\star)(\star) \\
\xi_{01}^{(2)} &= (\star\star\bullet)(\bullet) \\
\xi_{10}^{(3)} &= (\bullet\bullet)(\star)(\star) \\
\xi_{01}^{(3)} &= (\star\star)(\bullet)(\bullet) \\
\xi_{11}^{(2)} &= (\bullet\bullet)(\star\star) \\
\xi_{11}^{(1)} &= (\bullet\bullet\,\star\star) \\
\xi_\Delta &= \text{coal}.
\end{aligned}$$

The superscript indicates the total number of individuals in which the 4 ancestral lineages reside. The two subscripts tell us the states of the two pairs respectively: 0 means a pair of lineages in state ξ_0 and 1 means a pair of lineages in state ξ_1, with these as defined in the proof of Lemma 1. For example, $\xi_{10}^{(3)}$ involves 3 individuals in which the first pair of lineages are in the the same individual and the second pair of lineages is in different individuals. Finally, the state ξ_Δ is an absorbing state which represents the event that at least one of the two pairs has coalesced. The order of the states is arbitrary, based first on the subscripts then on the superscripts.

By definition, the two pairs of gene copies are drawn from the same pair of individuals at the present generation $g = 0$, where for each pair one gene copy is picked from each of the individuals. Hence the initial state $\widetilde{M}_0$ must be $\xi_{00}^{(2)}$. In other words, the distribution $\vec{p}_0$ of $\widetilde{M}_0$ is given by

$$\vec{p}_0 = (0, \, 0, \, 1, \, \ldots, \, 0). \tag{28}$$

It follows from Lemma A1 in the *Limiting behavior of the joint diploid ancestral process* section of Appendix A that

$$\lim_{N \to \infty} \mathbb{P}^{(N)}(\tau > [tN], \; \tau' > [tN]) = \lim_{N \to \infty} \vec{p}_0 \cdot \widetilde{\Pi}_N^{[tN]}(1, \, \ldots, \, 1, \, 0)^T \tag{29}$$

$$= (0, \, 0, \, 1, \, 0, \, \ldots, \, 0)\,\widetilde{P}e^{t\widetilde{G}}(1, \, \ldots, \, 1, \, 0)^T \tag{30}$$

$$= e^{-t}e^{-\lambda t(\frac{\psi^2}{2} - \frac{\psi^4}{16})}, \tag{31}$$

where equation (29) follows from the definition of $\widetilde{M}$ and $\vec{p}_0$, equation (30) from Möhle (1998a, Lemma 1) as explained in the *Limiting behavior of the joint diploid ancestral process* section of Appendix A and equation (31) by Lemma A1. Note that the vector $(1, \, \ldots, \, 1, \, 0)^T$ in equations (29) and (30) amounts to the Markov chain $(\widetilde{M}_g)_{g \in \mathbb{Z}_{\geq 0}}$ not reaching state ξ_Δ, i.e. that neither pair has coalesced.

Limiting behavior of the joint diploid ancestral process

Remark 2 (Robustness of joint process to initial condition). Our assumed initial state $\xi_{00}^{(2)}$ is the usual way multilocus data are sampled in population genetics. But Lemma A1 and Theorem 1 both hold for any initial state $\vec{p}_0$ whose last coordinate is zero. This is because the sample will undergo an instantaneous adjustment by $\widetilde{P}$ given in equation (A20), so that the effective starting state is always $\xi_{00}^{(4)}$. Whatever idiosyncrasies $\mathcal{A}^{(N,2)}$ may possess,

especially in the recent past, sensu Chang (1999), meaning the most recent $\log_2(N)$ generations, these matter less and less as N grows. In the limit, the lineages of the sample immediately disperse to different individuals without undergoing any coalescent events. Similarly, the factors of $\widetilde{P}$ in $\widetilde{G}$ guarantee that the lineages will remain in state $\zeta_{00}^{(4)}$ throughout the ancestral process, except for instants in which they have a chance to coalesce. This robustness against initial condition is analogous to equation (14).

Second term in equation (22)

We now show that equation (24) converges to $e^{-t/2}e^{-\lambda t(\frac{\psi^2}{2} - \frac{\psi^4}{16})}$ as $N \to \infty$. Through the use of the law of total expectation, equation (24) is equal to

$$\sum_{k=0}^{[tN]} \mathbb{E}^{(N)}[\mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau > [tN]) \mid Q_N([tN]) = k] \left(1 - \frac{\psi^2}{4}\right)^k \mathbb{P}^{(N)}(Q_N([tN]) = k). \quad (32)$$

By the fact that $Q_N([tN])$ is known given the pedigree and an application of the tower property, the conditional expectation in equation (32) is equal to

$$\mathbb{P}^{(N)}(\tau > [tN] \mid Q_N([tN]) = k),$$

which is approximately equal to

$$\mathbb{P}^{(N)}(\tau > [tN] \mid H_N([tN]) = k),$$

by equation (18). That is to say

$$\mathbb{E}^{(N)}\left[\mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau > [tN]) \left(1 - \frac{\psi^2}{4}\right)^{Q_N([tN])}\right]$$
$$\approx \mathbb{E}^{(N)}\left[\mathbb{P}_{H_N}^{(N)}(\tau > [tN]) \left(1 - \frac{\psi^2}{4}\right)^{H_N([tN])}\right],$$

in the sense that equation (20) holds. By Lemma A2 in Appendix A, for $g = [tN]$, it follows that for each $N \geq 2$ and $t \in (0, \infty)$,

$$\mathbb{E}^{(N)}\left[\mathbb{P}_{H_N}^{(N)}(\tau > [tN]) \left(1 - \frac{\psi^2}{4}\right)^{H_N([tN])}\right]$$
$$= (1, 0, 0)\left(\Pi_N^{\mathrm{mid}}\right)^{[tN]}(1, 1, 0)^T, \quad (33)$$

where Π_N^{mid} is defined as

$$\Pi_N^{\mathrm{mid}} := \alpha_N \left(1 - \frac{\psi^2}{4}\right)\Pi_N^{\mathrm{BF}} + (1 - \alpha_N)\Pi_N^{\mathrm{WF}}. \quad (34)$$

It now follows by Lemma A3 that the right-hand side of equation (33) converges to $e^{-t/2}e^{-\lambda t(\frac{\psi^2}{2} - \frac{\psi^4}{16})}$.

Third term in equation (22)

Finally, equation (25) is computed by first noticing that the number of big families up to generation T, $Q_N(T)$, is (almost) binomially distributed according to $\mathrm{Bin}([T], \alpha_N)$ for all $T \geq 1$, as observed by equation (18). Using the probability generating function of $Q_N(T)$ we get that the third term in equation (22) is equal to $e^{-\lambda t(\frac{\psi^2}{2} - \frac{\psi^4}{16})}$.

Putting everything together

As $N \to \infty$, equation (22) is equal to 0 since equation (25) multiplied by e^{-t} and equation (23) add up to $2e^{-t}e^{-\lambda(\frac{\psi^2}{2} - \frac{\psi^4}{16})t}$ which cancel

out with (24) multiplied by $-2e^{-t}$. This gives (21) which concludes the proof of Theorem 1 in the case of $\theta = 1$.

Convergence (17) when $\theta \in (0, 1)$

The proof is similar to the case of $\theta = 1$. In all of the above, substitute $[tN]$ by $[tN^\theta]$, and show instead of equation (22) that

$$\lim_{N \to \infty} \mathbb{E}^{(N)}\left[\left|F_N(t, \mathcal{A}^{(N,2)}) - \left(1 - \frac{\psi^2}{4}\right)^{Q_N([tN^\theta])}\right|^2\right] = 0. \quad (35)$$

Expanding equation (35) gives the same three terms as in equations (23)–(25). In this faster timescale, as $N \to \infty$, equation (23) is now equal to

$$\lim_{N \to \infty} \mathbb{E}^{(N)}[F_N^2(t, \mathcal{A}^{(N,2)})] = e^{-\lambda t(\frac{\psi^2}{2} - \frac{\psi^4}{16})}, \quad (36)$$

as available in Lemma A1. The limiting behavior of equation (24) is the same as before, that is

$$\lim_{N \to \infty} \mathbb{E}^{(N)}\left[F_N(t, \mathcal{A}^{(N,2)})\left(1 - \frac{\psi^2}{4}\right)^{Q_N([tN])}\right] = e^{-\lambda t(\frac{\psi^2}{2} - \frac{\psi^4}{16})} \quad (37)$$

and

$$\lim_{N \to \infty} \mathbb{E}^{(N)}\left[\left(1 - \frac{\psi^2}{4}\right)^{2Q_N([tN^\theta])}\right] = e^{-\lambda t(\frac{\psi^2}{2} - \frac{\psi^4}{16})}. \quad (38)$$

Multiplying equation (37) by -2 and summing it up with equations (36) and (37) concludes the proof in the case of $\theta \in (0, 1)$.

The proof of Theorem 1 is complete.

Remark 3 (Only big families matter). Let $\vec{G}^{(N)} = (G_1^{(N)}, G_2^{(N)}, \ldots)$, where $0 \leq G_1^{(N)} < G_2^{(N)} < \cdots$, be the generations with big families that have (randomly) occurred. Then $\vec{G}^{(N)}$ is known if we know the pedigree. Similar to equation (16), we let

$$F_N(t, \vec{G}^{(N)}) := \mathbb{P}^{(N)}(\tau^{(N,2)} > [tN^\theta] \mid \vec{G}^{(N)}) \quad (39)$$

be the conditional probability of the event $\{\tau^{(N,2)} > [tN^\theta]\}$ given the (random) generations $\vec{G}^{(N)}$. Hence, here we condition on less information than on the left-hand side of equation (17). We can show that Theorem 1 still holds (i.e. the weak convergence in equation (17) still holds) if we replace $F_N(t, \mathcal{A}^{(N,2)})$ by $F_N(t, \vec{G}^{(N)})$. For a proof sketch, see *Proof sketch for Remark 3* section of Appendix A.

Coalescence times, gene genealogies, and correlations

Here, we briefly recap then provide three illustrations of our results. Our main result is Theorem 1 which describes two limiting distributions of coalescence times conditional on the pedigree. As the number of unlinked loci examined in the sampled individuals increases, the empirical distribution of their coalescence times should converge to Theorem 1. In this case, conditional on the pedigree, the probability of coalescence in a generation depends on whether that particular generation includes a big family. For background and comparison, Lemma 1 presents the corresponding two limiting distributions obtained by the usual method of averaging over pedigrees, i.e. over all possible outcomes of reproduction in a single generation, including the possibility of a big

family. In this case, the probability of coalescence is the same in every generation.

Time is rescaled in all of these limiting ancestral processes. It is measured in units of N^θ generations for some $\theta \in (0, 1]$. When $\theta \in (0, 1)$, the timescale for big families to occur is much shorter than the usual Wright–Fisher coalescent timescale of N generations. When $\theta = 1$, the timescales for big families and for ordinary Wright–Fisher coalescence are the same. Big families occur at rate λ in rescaled time, and their offspring comprise a fraction $\psi \in [0, 1]$ of the population in that generation. Underpinning our results is the fact that as $N \to \infty$ ancestral genetic lineages spend the overwhelming majority of their time in separate individuals, i.e. in state ξ_0 for a pair of lineages at the same locus (cf. Lemma 1) or state $\xi_{00}^{(4)}$ for two pairs of lineages at two unlinked loci (cf. Theorem 1 and Remark 2). Thus, when a big family occurs, each lineage independently: (1) is among the offspring of the highly reproductive pair with probability ψ and (2) if so, is equally likely to descend from each of the four copies of the corresponding locus in the two parents. A pair of lineages at the same locus coalesces in the big family with probability $\psi^2/4$. Pairs of lineages at different, unlinked loci do this independently.

Our first illustration compares our limiting results to the CDF (i.e. one minus the survival function) of pairwise coalescence times in the discrete model. Figure 1a displays CDFs for five simulated pedigrees for $N = 500$, assuming that the probability of a big family is equal to the expected pairwise coalescence probability,

$1/(2N) = 0.001$, and the offspring make up the entire population in that generation. This corresponds to the limiting process in Theorem 1 with $\theta = 1$, $\lambda = 1/2$ and $\psi = 1$. This makes the coalescence probability ($\psi^2/4$) equal to 1/4 in each generation with a big family. We computed coalescence probabilities on each pedigree in each generation starting from a pair of randomly sampled individuals using the method in Wakeley *et al.* (2012). The corresponding "expected" CDF of the pedigree-averaged process from Lemma 1, i.e. of an exponential random variable with rate parameter 5/8, is shown for comparison.

The left panel of Fig. 1a illustrates that the ancestral process conditional on the pedigree is quite close to limiting result in Theorem 1, even when $N = 500$. The CDFs make discrete jumps whenever big families occur. In this case, with $\psi = 1$ the magnitude of a jump is always 1/4 of the remaining distance to 1. Between jumps the CDFs show a steady increase in the cumulative coalescence probability, in line with the limiting prediction with its rate of 1/2. In contrast, the pedigree-averaged process in Lemma 1 predicts a faster rate of increase of the CDF and no jumps.

The right panel of Fig. 1a details the short-time behavior of the ancestral process conditional on the pedigree, displaying these same CDFs only over the most recent 40 generations. The scale on the vertical axis is such that the diagonal corresponds approximately to the prediction of the background Wright–Fisher model (not shown) and a line with slope 1.25 corresponds approximately
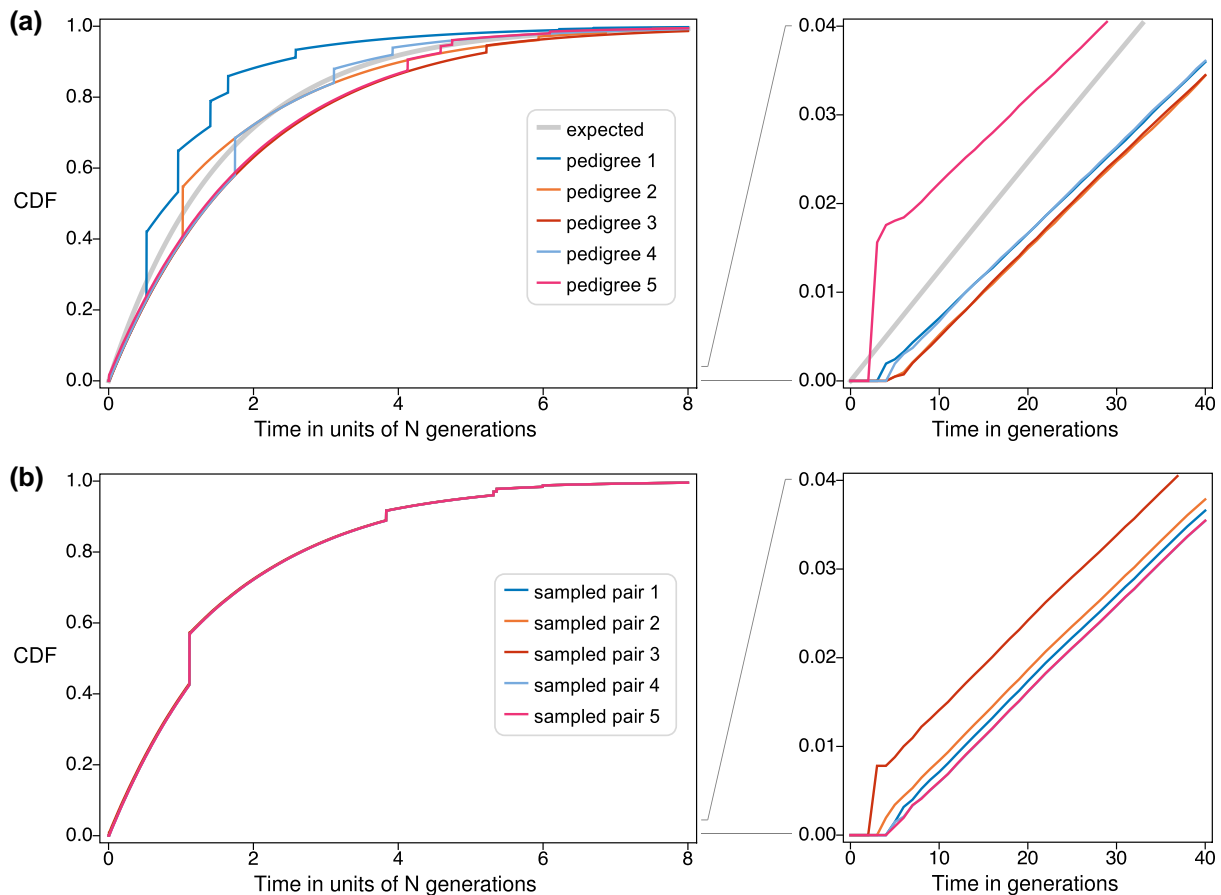

Fig. 1. CDFs of pairwise coalescence times for $\theta = 1$ and $\lambda = 1/2$. a) (left panel) CDFs for five simulated pedigrees for populations of size $N = 500$ together with the corresponding expected CDF from Lemma 1. (Right panel) The same five CDFs and the corresponding expectation from Lemma 1, only plotted over the most recent 40 generations. b) Corresponding results for a single pedigree for a population of size $N = 500$ but five different pairs of individuals, each sampled independently without replacement from the population.

to prediction of Lemma 1 which is shown. After a small number of generations, which from Chang (1999) should be of order $\log_2(N)$, the CDFs for the five pedigrees start to show the predicted Wright–Fisher slope of one. However, they start at different places depending on the particular ancestries of the sampled individuals, specifically whether there are very recent shared ancestors as in pedigree 5 or more likely there are no very recent shared ancestors as in pedigrees 1 through 4; cf. also Wakeley *et al.* (2012). These differences are barely visible on the timescale of the left panel of Fig. 1a, and it is implicit in Theorem 1 that they become negligible as $N \to \infty$.

As stated in equation (26), the predictions for each of the five pedigrees in Fig. 1a apply equally and independently to every locus in the sampled individuals. These five, like five instances of Theorem 1, are again predictions for the empirical distributions of coalescence times among unlinked loci. Different instances of $\mathcal{A}^{(N,2)}$ will have different times of big families (Fig. 1a, left panel) and different patterns of recent common ancestry of the samples (Fig. 1a, right panel). For comparison, Fig. 1b shows the same two graphs for five independently sampled pairs of individuals on a single pedigree. Again, each sample has its own pattern of recent common ancestry, producing visible differences on the scale of the right panel. But now all five samples access the same shared set of big families, resulting in the five closely overlapping CDFs in the left panel of Fig. 1b.

Next we illustrate the effects that big families have on the gene genealogies of larger samples, in particular the sharing of identical coalescence times at unlinked loci. Rather than simulating pedigrees for finite populations, we use the limiting model directly so that big families are the only possible cause of shared coalescence times. We set $\psi = 1$ as before, and for simplicity assume that big families drive the ancestral process, i.e. $\theta \in (0, 1)$. We set $\lambda = 1$ without loss of generality, as λ is arbitrary except when $\theta = 1$.

Based on Theorem 1, we model gene genealogies by generating a series of exponential waiting times between big families and, since $\theta \in (0, 1)$, disallowing coalescence between them. When the n ancestral lineages of the sample reach the first big family, their distribution among the four parental gene copies will be multinomial with parameters n and $(1/4, 1/4, 1/4, 1/4)$. Anywhere from one to four simultaneous multiple mergers will occur. The number of ancestral lineages which emerge is also at most four. If more than one lineage emerges, the same process is repeated until a single lineage remains which is the most recent common ancestor of the entire sample. The only aspects of the pedigree which persist in the limit are the big families (cf. Remark 3). Thus, independent runs of this multinomial coalescent process using the same series of exponential waiting times correspond to gene genealogies of unlinked loci conditional on the pedigree.

Figure 2a displays the gene genealogies of seven unlinked loci for a sample of size 16, assuming in this way that all loci share the same pedigree. The trees are oriented with the present-day samples at the bottom. Solid lines trace (unlabeled) ancestral lineages up into the past. Thin dotted lines show the times of the big families. All seven gene genealogies have multiple mergers at the most recent big family in the past, and five have common ancestor events at the second one. In the more distant past when there are small numbers of ancestral lineages, there is less sharing of coalescence times among gene genealogies. This is expected; for example, the final two lineages only coalesce with probability 1/4 each time they encounter a big family.

Figure 2b shows seven gene genealogies, again for samples of size 16, but now assuming that each locus has its own pedigree. These are equivalent to seven gene genealogies sampled from seven independent populations, each with its own series of exponential waiting times between big families as in Fig. 2a (not displayed in Fig. 2b). These gene genealogies differ from the
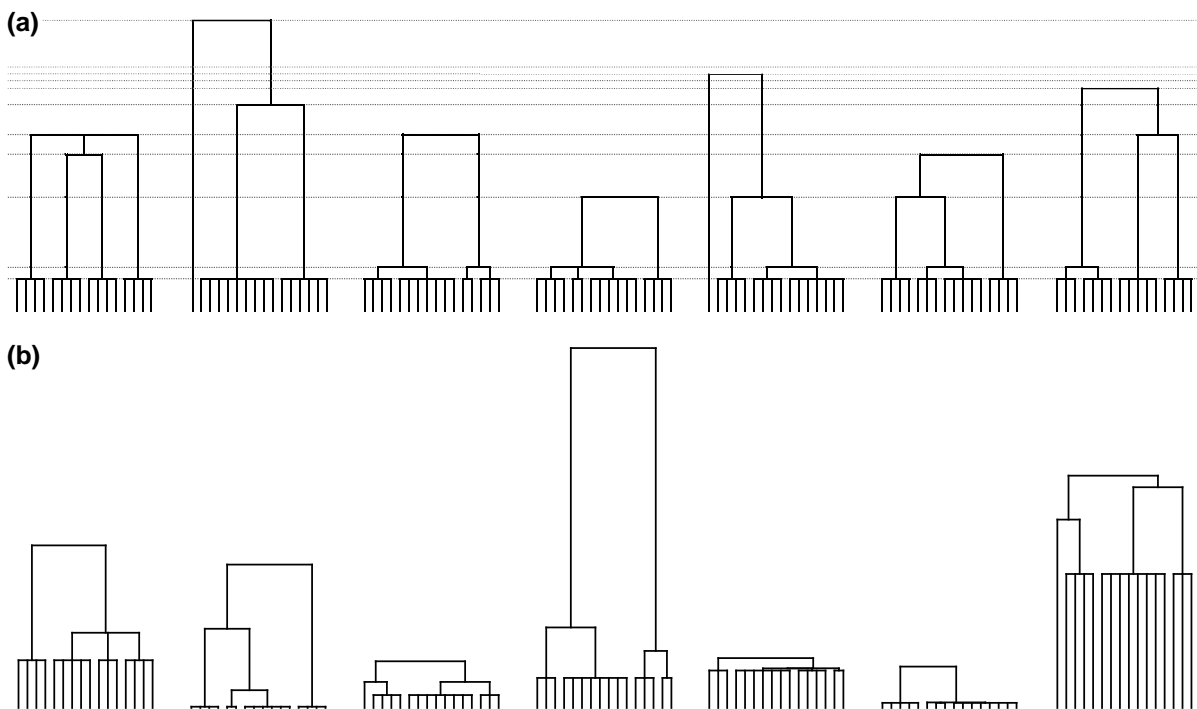

Fig. 2. Simulated gene genealogies for seven independently assorting loci when all seven share the same pedigree a) vs when each locus has its own independently generated pedigree b). The sample size is $n = 16$ for every locus. Gene genealogies were generated as described in the text for the limiting model with $\theta \in (0, 1)$ and $\lambda = \psi = 1$. Thin dotted lines in the top row show the particular series of times of big families in that population.

ones in the top row, most obviously in the different timings of their first common ancestor events. Clearly, the distribution of gene genealogies produced in this way will not be close to the distribution of gene genealogies of unlinked loci in the same genome which perforce come from the same population.

Finally, we illustrate how averaging over pedigrees as in Lemma 1 results in positive correlations of coalescence times between unlinked loci. Explicitly modeling pedigrees as in Theorem 1 predicts these to be zero as might be expected for independently assorting loci. Based on the property of ancestral lineages spending the overwhelming majority of their time in separate individuals, cf. Remarks 1 and 2, we consider the two-locus analog of Lemma 1 with reduced state space

$$
\begin{aligned}
\xi_{00} &= (\bullet)(\bullet)(\star)(\star) \\
\xi_{10} &= (\bullet)(\star)(\star) \\
\xi_{01} &= (\bullet)(\bullet)(\star) \\
\xi_{11} &= (\bullet)(\star),
\end{aligned}
$$

where now the first and second subscripts are indicators of whether locus 1 or locus 2 has coalesced. By extension from Lemma 1, the limiting ancestral process for two unlinked loci has transition rate matrix

$$
Q = \begin{cases} \lambda Q^{\mathrm{BF}}, & \text{if } \theta \in (0,\ 1) \\ Q^{\mathrm{WF}} + \lambda Q^{\mathrm{BF}}, & \text{if } \theta = 1, \end{cases} \tag{40}
$$

where

$$
Q^{\mathrm{BF}} = \begin{array}{c} \\ \xi_{00} \\ \xi_{10} \\ \xi_{01} \\ \xi_{11} \end{array} \overset{\displaystyle \begin{array}{cccc} \xi_{00} & \xi_{10} & \xi_{01} & \xi_{11} \end{array}}{\begin{bmatrix} -\frac{\psi^2}{4}\left(2-\frac{\psi^2}{4}\right) & \frac{\psi^2}{4}\left(1-\frac{\psi^2}{4}\right) & \frac{\psi^2}{4}\left(1-\frac{\psi^2}{4}\right) & \frac{\psi^4}{16} \\ 0 & \frac{-\psi^2}{4} & 0 & \frac{\psi^2}{4} \\ 0 & 0 & -\frac{\psi^2}{4} & \frac{\psi^2}{4} \\ 0 & 0 & 0 & 0 \end{bmatrix}} \tag{41}
$$

and

$$
Q^{\mathrm{WF}} = \begin{array}{c} \\ \xi_{00} \\ \xi_{10} \\ \xi_{01} \\ \xi_{11} \end{array} \overset{\displaystyle \begin{array}{cccc} \xi_{00} & \xi_{10} & \xi_{01} & \xi_{11} \end{array}}{\begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}}. \tag{42}
$$

Focusing on the case $\theta = 1$, the rate matrix Q is the sum of a Wright–Fisher (or Kingman coalescent) component and a big-family component. We have factored the tuning parameter λ out of the latter to emphasize that, conditional on the occurrence of a big family, samples at the two loci coalesce or do not coalesce independently of each other.

Let T_1 and T_2 be the coalescence times at the two loci. These correspond to the limiting random variables τ/N^θ and τ'/N^θ in Lemma A1. Here individually they are the times to state ξ_{11} starting from states ξ_{01} and ξ_{10}, respectively. From the rate matrix Q in equation (40) or from Lemma 1 directly, T_1 and T_2 are identically distributed. In particular,

$$
T_1 \sim \begin{cases} \text{exponential}\left(\lambda \frac{\psi^2}{4}\right), & \text{if } \theta \in (0,\ 1) \\ \text{exponential}\left(\frac{1}{2} + \lambda \frac{\psi^2}{4}\right), & \text{if } \theta = 1. \end{cases} \tag{43}
$$

However, T_1 and T_2 are not necessarily independent. Lemma A1 accounts for this nonindependence in the proof of Theorem 1,

and we note that equation (40) also gives equation (A22). Here, we use first-step analysis to compute the correlation coefficient, $\mathrm{Corr}[T_1,\ T_2]$. Let W be the waiting time to the first event in the ancestry of the two loci starting from state ξ_{00}, and T_1^* and T_2^* be the additional times to coalescence at each locus following the first event. In this formulation,

$$
T_i = W + T_i^* \tag{44}
$$

for $i \in \{1,\ 2\}$. From equation (40), we have

$$
W \sim \begin{cases} \text{exponential}\left(\lambda \frac{\psi^2}{4}\left(2-\frac{\psi^2}{4}\right)\right), & \text{if } \theta \in (0,\ 1) \\ \text{exponential}\left(1 + \lambda \frac{\psi^2}{4}\left(2-\frac{\psi^2}{4}\right)\right), & \text{if } \theta = 1 \end{cases} \tag{45}
$$

and we point out that, since W is exponentially distributed,

$$
\mathbb{E}[W^2] = 2\mathbb{E}[W]^2. \tag{46}
$$

Conditioning on the first step from state ξ_{00} and simplifying,

$$
\mathbb{E}[T_1 T_2] = \mathbb{E}[W^2] + \mathbb{E}[W]\mathbb{E}[T_1^*] + \mathbb{E}[W]\mathbb{E}[T_2^*] + \mathbb{E}[T_1^* T_2^*] \tag{47}
$$

$$
= 2\mathbb{E}[W]\mathbb{E}[T_1]. \tag{48}
$$

Going from equations (47) to (48) uses equations (46), (44), $\mathbb{E}[T_1] = \mathbb{E}[T_2]$, and the fact that either T_1^* or T_2^* or both are equal to zero following the first event. Then for the correlation coefficient, we have simply

$$
\mathrm{Corr}[T_1,\ T_2] = \frac{2\mathbb{E}[W]\mathbb{E}[T_1] - \mathbb{E}[T_1]^2}{\mathrm{Var}[T_1]}, \tag{49}
$$

which, using equations (43) and (45), becomes

$$
\mathrm{Corr}[T_1,\ T_2] = \begin{cases} \frac{\psi^2}{8-\psi^2}, & \text{if } \theta \in (0,\ 1) & (50\mathrm{a}) \\ \frac{\lambda \psi^4}{16+\lambda \psi^2(8-\psi^2)}, & \text{if } \theta = 1. & (50\mathrm{b}) \end{cases}
$$

Even though the loci assort independently, their ancestries in the pedigree-averaging model jointly depend on the random process that generates big families in the population. As a result, their coalescence times are positively correlated.

The correlation coefficient (50b), obtained here under the assumption that the loci are unlinked, corresponds to equation (31) in Birkner *et al.* (2013b, p. 266), obtained there by modeling recombination explicitly and then taking the limit as the rescaled recombination parameter tends to infinity. The timescales in these two works differ by a factor of two. Our equation (50b) becomes identical to equation (31) in Birkner *et al.* (2013b) by putting $\lambda = c/2$.

For a given value of ψ, the correlation coefficient is smaller when $\theta = 1$ than when $\theta \in (0,\ 1)$. When coalescence can be due to either big families or ordinary Wright–Fisher reproduction ($\theta = 1$), the correlation tends to zero as λ tends to zero. As λ grows, equation (50b) grows until it approaches (50a). Thus, the occurrence of big families may be said to be the source of positive correlations in coalescence times at unlinked loci. In a similar vein, $\mathrm{Corr}[T_1,\ T_2]$ tends to zero as the fraction of the population replaced by each big family, ψ, tends to zero. This is true even if $\theta \in (0,\ 1)$, i.e. when there is no Wright–Fisher/Kingman component in the limit process. At the other extreme, as $\psi \to 1$, $\mathrm{Corr}[T_1,\ T_2] \to 1/7$ which is considerably less than one. Even

when all coalescence happens in big families and the offspring of each big family replace the entire population, there are still two diploid parents and the loci will generally have different coalescence times.

The following alternate derivation of equation (50a) shows how these positive correlations arise. In short it is because T_1 and T_2 have a shared dependence on the times between big families in the pedigree. Implicitly, Lemma 1 averages over these times whereas Theorem 1 retains them.

When $\theta \in (0, 1)$, coalescence can only happen when a big family occurs. Let K_1 and K_2 be the numbers of such events it takes for locus 1 and locus 2 to coalesce, respectively. These do not depend on the times between big families when $\theta \in (0, 1)$. Further, K_1 and K_2 are independent because the loci are unlinked. They are geometric random variables with parameter $\psi^2/4$. Let X_i, $i \in \mathbb{Z}_{\geq 0}$, be the time from the $(i-1)$th to the ith big family backward in time, with $X_0 \equiv 0$. In the context of Theorem 1, these times are independent and identically distributed exponential random variables with rate parameter λ. Under this formulation,

$$T_i = \sum_{j=1}^{K_i} X_j \qquad (51)$$

for $i \in \{1, 2\}$. There are two sources of variation in T_i: variation in K_i and variation in the lengths of the intervals, X_j, $j \in \{1, \ldots, K_i\}$. Starting with equation (51), it is straightforward to confirm that the distribution of T_i is exponential with rate parameter $\lambda\psi^2/4$ as in equation (43) or Lemma 1.

From equation (51) and the fact that X_i and X_j are independent for $i \neq j$, it is also clear that intervals in a common to T_1 and T_2 are a key source of their covariation. For given values of K_1 and K_2, they are the only source. The first interval is always shared, as are all subsequent intervals until one or the other locus coalesces. Let K_{12} be the number of these shared intervals and

$$T_{12} = \sum_{i=1}^{K_{12}} X_i \qquad (52)$$

be the corresponding total length of time. By definition $K_{12} = \min(K_1, K_2)$. The more ancient $K_1 - K_2$ or $K_2 - K_1$ intervals are only ancestral to one of the loci.

Applying the conditional covariance formula, or law of total covariance, we have

$$
\begin{aligned}
\mathrm{Cov}[T_1, T_2] &= \mathbb{E}[\mathrm{Cov}[T_1, T_2|K_1, K_2]] + \mathrm{Cov}[\mathbb{E}[T_1|K_1, K_2], \mathbb{E}[T_2|K_1, K_2]] \\
&= \mathbb{E}[\mathrm{Var}[T_{12}|K_{12}]] + \mathrm{Cov}[\mathbb{E}[T_1|K_1], \mathbb{E}[T_2|K_2]].
\end{aligned}
\qquad (53)
$$

The outer expectation and covariance are with respect to the joint distribution of K_1 and K_2. Note that K_{12} is a marginal property of this distribution. The inner variance (or covariance) and expectations are with respect to the joint distributions of the X_i which are the only parts of T_1 and T_2 that vary conditional on K_1 and K_2.

At this point, in equation (53), we have not applied the fundamental property that K_1 and K_2 are independent since the loci are unlinked, nor have we assumed any particular distribution(s) for the X_i. We have only used the definitions of T_1 and T_2 as sums of random variables and the assumption that X_i and X_j are independent for $i \neq j$. So we may consider that the interval times are fixed numbers: $X_i \equiv x_i$, $i \in \mathbb{Z}_{\geq 0}$. They could be the outcomes of the exponential random times implicit in Theorem 1. Fixing

the X_i means fixing the only aspects of the pedigree that persist in the limiting model. Conditioning on the pedigree, T_1 and T_2 are independent even in the limiting model; cf. equation (26). The point we wish to emphasize here is that fixing the X_i removes one particular source of covariation of T_1 and T_2. It makes $\mathrm{Var}[T_{12}|K_{12}] = 0$.

Continuing from equation (53) and assuming that X_i, $i \in \mathbb{Z}_{\geq 0}$, are independent and identically distributed

$$
\begin{aligned}
\mathrm{Cov}[T_1, T_2] &= \mathbb{E}[K_{12}] \, \mathrm{Var}[X_i] + \mathbb{E}[X_i]^2 \, \mathrm{Cov}[K_1, K_2] \\
&= \mathbb{E}[K_{12}] \, \mathrm{Var}[X_i],
\end{aligned}
\qquad (54)
$$

the latter following from the independence of K_1 and K_2. Again, $X_i \sim$ exponential(λ), and from the definition of K_{12} as the number of big-family events it takes for one locus or the other to coalesce,

$$K_{12} \sim \mathrm{geometric}\left(1 - \left(1 - \frac{\psi^2}{4}\right)\right). \qquad (55)$$

Putting the required quantities in equation (54) and simplifying gives

$$\mathrm{Cov}[T_1, T_2] = \frac{16}{\lambda^2 \psi^2 (8 - \psi^2)} \qquad (56)$$

which is exactly the covariance needed to produce the correlation coefficient (50a). In sum, the model of Lemma 1 predicts a positive correlation of coalescence times at unlinked loci because it averages over the distributions of the intervals X_i. Starting instead with the model of Theorem 1 shows that the particular quantity controlling these positive correlations is $\mathrm{Var}[X_i]$.

Discussion

The use of random models to describe past events raises many questions in population genetics. Everything in the past has already occurred, including all instances and timings of reproduction and genetic transmission. For empirical work this may be a truism. But population genetics has always been concerned with evolutionary processes. How do mutation, recombination, selection, random genetic drift, nonrandom mating, limited dispersal, etc., conspire to produce observable patterns of variation? By emphasizing the fixed nature of the past, we highlight the subjectivity of theoretical work, specifically when the goal is to interpret data from natural populations. Ultimately, the choices one makes about modeling the past may be application-dependent.

Motivated by applications to multilocus data, we singled out the pedigree as a key feature of the past and obtained a result (Theorem 1) concerning the application of neutral coalescent models in sexually reproducing species. We have the following sampling structure in mind. Processes of survival and reproduction result in a pedigree. Genetic transmission, including mutation and recombination across the entire genome, occurs within the pedigree. A number of individuals are sampled from the population and some or all of their genomes are sequenced. We modeled the single-locus coalescent process conditional on the pedigree. Our results specify the distribution of coalescence times given the pedigree and the sampled individuals. This distribution can be interpreted either as a prior for a single locus or as a prediction about the distribution of coalescence times among unlinked loci. We contrasted our results conditional on

the pedigree with results obtained by averaging over pedigrees, noting that the latter is the tradition of theoretical population genetics. We did not model mutation or recombination, but our fundamental conclusion—that some population processes cause the quenched and averaged processes to be very different—should be as important for genetic variation as it is for coalescence times.

We can compare our framework with that of Ralph (2019). The two have a lot in common. Ralph (2019) takes the pedigree and the outcomes of genetic transmission, including recombination across the entire genome, to be fixed. The latter is referred to as the ancestral recombination graph (ARG), which we note differs slightly from the corresponding objects in Hudson (1983a) and Griffiths and Marjoram (1997) because it is embedded in the fixed pedigree. Without specifying a generative model for the pedigree, Ralph (2019) focuses on the ARG as the fixed but unknown object of interest in empirical population genetics. A sample is taken and some stretch of the genome is sequenced. Its ancestry is a collection of gene genealogies, a subset of the ARG. Implicitly, it is the outcome of the random process of genetic transmission within the fixed pedigree, but this too is not modeled.

The only randomness is in how the collection of gene genealogies of the sample is revealed by mutation. Ralph (2019) assumes the infinite-sites mutation process and uses this to show that predictions about summary statistics of DNA sequence variation, such as the average number of pairwise nucleotide differences or the *F*-statistics of Patterson *et al.* (2012), can be expressed in terms of the fixed branch lengths in the sampled subset of the ARG. This is the empirical version of what Slatkin (1991), Griffiths and Tavaré (1998), Nielsen (2000), McVean (2002), and Peter (2016) had done in the context of the standard neutral coalescent, where instead the moments of summary statistics can be expressed in terms of corresponding moments of branch lengths. Ralph *et al.* (2020) describe a hybrid approach, with the ARG conceived as in Ralph (2019) and with times of events in the ARG for data from humans (The 1000 Genomes Project Consortium 2015) estimated with the aid of the standard neutral coalescent (Speidel *et al.* 2019).

Whereas we model the production of the pedigree and the process of coalescence within it but do not model mutation, Ralph (2019) models only mutation on the fixed ARG. Consider a species in which recurrent selective sweeps across the genome have structured the ARG. An empirical estimate of the ARG would find regions of the genome with reduced variation due to reduced times to common ancestry. In order to relate these observations to an evolutionary process within the empirical framework of Ralph (2019), for example to describe them in terms of recurrent selective sweeps as in Durrett and Schweinsberg (2005), additional modeling would be needed. In contrast, in a theoretical approach such as ours here, recurrent sweeps would be included in the model at the outset, and this in turn would facilitate the interpretation of patterns in the data. Under our model, it is important to keep in mind that the ARG is in fact a fixed object and that the process of coalescence within the pedigree models the sampling of a locus in the ARG.

Today detailed estimates of the ARG for large samples of human genomes are available (Wohns *et al.* 2022; Zhang *et al.* 2023). These have been obtained, like other recent estimates (Kelleher *et al.* 2019; Speidel *et al.* 2019; Albers and McVean 2020), using the standard neutral coalescent as a prior for gene genealogies and times to common ancestry. Our results and those of Tyukin (2015) help to justify using such a prior despite the fact that the pedigree is fixed, so long as the processes which laid

down the pedigree are not too different from the Wright–Fisher or Cannings models with relatively low variation of offspring numbers. The empirically oriented interpretations in these works, for example in Wohns *et al.* (2022), connect features of the ARG with major events in human history, such as the out-of-Africa event which has been studied genetically since the first mtDNA discoveries (Cann *et al.* 1987; Vigilant *et al.* 1991) and the novel finding of an accumulation of ancestry in Papua New Guinea more than 100-thousand years ago. This is intraspecific phylogeography (Avise *et al.* 1987; Avise 1989, 2000) at genome scale.

Our model for generating the pedigree includes the possibility of special generations in which a big family is guaranteed to occur. We obtained different coalescent processes as $N \to \infty$, depending on the relative rate of these big families in the limit and whether the ancestral process is conditional on the pedigree (Theorem 1) or not (Lemma 1). This essentially negative result, that the averaged process cannot be used in place of the conditional process, includes the positive finding that the Kingman coalescent can be used between big families in the case that both occur on the same timescale (Theorem 1 with $\theta = 1$). The numbers and timings of big families are all that is left of the pedigree in the limit (cf. Remark 3). Needing to keep track of just these is much less daunting than the prospect of including entire pedigrees in all of our population-genetic models. There may be other circumstances in which aspects of the pedigree are important, but so far the only other instance identified is when subpopulations are connected by limited migration (Wilton *et al.* 2017).

Limiting coalescent processes for our model generally involve simultaneous multiple mergers. Yet the familiar extensions of the Kingman coalescent to include multiple mergers have been derived by averaging over the pedigree, not by conditioning on it. They begin with single-generation marginal probabilities of coalescence, whereas in truth the individuals in the sample either have or do not have common ancestors in any preceding generation and this is what determines probabilities of coalescence. Without big families, our results and those of Tyukin (2015) provide belated justification for the early uses of the Kingman coalescent process as a prior model for the gene genealogy of a single locus (Lundstrom *et al.* 1992; Griffiths and Tavaré 1994; Kuhner *et al.* 1995). Our work also clarifies what is involved in using pedigree-averaged ancestral processes as single-locus priors in cases where big families can occur.

If the only data available were from a single locus without recombination, one could model the gene genealogy using the pedigree-averaged ancestral process. The logic would be that a single locus has one unknown random pedigree and one unknown random gene genealogy within that pedigree, and that the gene genealogies from multiple-merger coalescent models are marginal predictions over both of these unknowns. For example, with $n = 2$, a single draw of a coalescence time from the appropriate exponential distribution in Lemma 1 accounts for both sources of variation. Implicit in this accounting is that repeated samples would each have their own pedigree and conditional gene genealogy. This 2-fold sampling structure is precisely what Theorem 1 describes. It is straightforward to show that repeated sampling under Theorem 1 (each time drawing a new pedigree) gives the same exponential distributions as in Lemma 1. Yet even for a single locus, it may be preferable to record the additional information about big families as in Theorem 1.

Applying this type of repeated sampling (i.e. including resampling the pedigree) to multiple loci is another matter. Population-genetic models should not allow the pedigree to vary among loci. Theorem 1 is a simple initial example of the kind of

coalescent modeling required for multilocus data generally but especially when multiple-merger processes are implicated. In cases where big families may occur with some frequency, it is crucial to retain the information about the pedigree which matters for the gene genealogies at all loci. All multiple-merger coalescent models so far described, which implicitly average over the pedigree, are inadequate in this sense. The broader implication of Lemma 1 and Theorem 1 is that there exists a collection of quenched limits conditional on the pedigree which await description and are the appropriate models for multilocus data.

The diploid exchangeable population models in Birkner *et al.* (2018) are a natural starting point for the description of general quenched-pedigree Ξ-coalescent models. Alternatively, parameterized models could be considered, controlling for example rates of monogamy and the distribution of offspring numbers as in the program SLiM 3 (Haller and Messer 2019) or the Pólya urn scheme of Gasbarra *et al.* (2005). The latter was used for the prior in the Bayesian inference methods of Gasbarra *et al.* (2007a, 2007b) and Ko and Nielsen (2019) for estimating the recent few generations of the pedigree from sequence data. Selfing in the production of big families, which we assumed does not occur, could also be considered. Nonexchangeable models for generating pedigrees are possible, for example with recurrent selective sweeps (Durrett and Schweinsberg 2005) or cultural transmission of reproductive success (Guez *et al.* 2023). It could also be of interest to describe these coalescent models directly in terms of the properties of pedigrees as directed graphs, and here we note the study of Blath *et al.* (2014) as a start in this direction.

The model underlying Lemma 1 and Theorem 1 is very simple. Only one type of big family is allowed, these are distributed in time according to a Poisson process, and we only considered a sample of size two. We hypothesize that the basic principles of Theorem 1 will be robust to all of these. For example, other ways of generating big families should be possible, such that $Y(t) \sim \text{Poisson}(\lambda t)$ would be replaced by some other distribution. Extensions to larger sample sizes and to variation in the numbers and types of big families seem straightforward in principle, though they will require a lot more bookkeeping. Such generalization of big families will need to include sufficient details that Mendel's laws can be applied. For example, if four parents have $[\psi N]$ offspring, it will matter whether they form two monogamous pairs or comprise one big family with four parents, and in either case just how many offspring each pair has. In a more general model, such details will need to be specified for each big-family event.

The implications of our results for inference can be sketched as follows. Consider a general situation in which there may be special events like our big families in a well mixed population which has possibly changed in size over time. Assume that data are available for L unlinked loci and there is no intra-locus recombination. Let $\mathcal{D}$, $\mathcal{G}$, $\mathcal{A}$, Θ_m, and Θ_c represent the data, the collection of gene genealogies at the loci, the pedigree, the parameters of a mutation model and the parameters of a coalescent model, specifically a trajectory of relative population sizes over time in a Kingman coalescent with variable size. Let $\mathcal{D}_i$ and $\mathcal{G}_i$ be the data and the gene genealogy at the ith locus. Consider the likelihood, which is key to any sort of statistical inference. Traditional coalescent-based inference disregards $\mathcal{A}$ and computes the likelihood $\mathbb{P}(\mathcal{D}; \Theta_m, \Theta_c)$ in which we use "; " to indicate that Θ_m and Θ_c are treated as fixed parameters. The traditional computation proceeds by conditioning on $\mathcal{G}$, treating this as a random variable, but does so under the erroneous assumption that $\mathbb{P}(\mathcal{G}; \Theta_c)$ is equal to the product of $\mathbb{P}(\mathcal{G}_i; \Theta_c)$ across loci.

Instead, because a shared pedigree has been fixed by past events, a better approach would be to use $\mathcal{A}$ as the parameter in place of Θ_c and to compute the likelihood

$$
\begin{aligned}
\mathbb{P}(\mathcal{D}; \Theta_m, \mathcal{A}) &= \sum_{\mathcal{G}} \mathbb{P}(\mathcal{D} \,|\, \mathcal{G}; \Theta_m) \mathbb{P}(\mathcal{G}; \mathcal{A}) \\
&= \sum_{\mathcal{G}} \prod_{i=1}^{L} \mathbb{P}(\mathcal{D}_i \,|\, \mathcal{G}_i; \Theta_m) \mathbb{P}(\mathcal{G}_i; \mathcal{A}),
\end{aligned}
\tag{57}
$$

where now in equation (57) the independence assumption of gene genealogies (given the pedigree) is correct. Theorem 1 is a simple example of what we expect will be possible under a variety of population models. Intuitively, $\mathcal{A}$ can be replaced by the pair $\{\mathcal{Y}, \mathcal{A} \setminus \mathcal{Y}\}$, where $\mathcal{Y}$ is a list of special events and $\mathcal{A} \setminus \mathcal{Y}$ is the remainder of the pedigree. In the limiting ancestral process, $\mathcal{Y}$ may need to be preserved while $\mathcal{A} \setminus \mathcal{Y}$ can be replaced by a coalescent model with parameters Θ_c. For our model, $\mathcal{Y}$ would be the times and sizes (ψ) of big families, and the coalescent model would be the Kingman coalescent. Thus, our results suggest the simplification

$$
\begin{aligned}
\mathbb{P}(\mathcal{D}; \Theta_m, \mathcal{A}) &\approx \mathbb{P}(\mathcal{D}; \Theta_m, \mathcal{Y}, \Theta_c) \\
&= \sum_{\mathcal{G}} \prod_{i=1}^{L} \mathbb{P}(\mathcal{D}_i \,|\, \mathcal{G}_i; \Theta_m) \mathbb{P}(\mathcal{G}_i; \mathcal{Y}, \Theta_c),
\end{aligned}
\tag{58}
$$

where the approximation is for large N. In the present work, equation (58) has the probabilistic interpretation in Theorem 1, where $\mathcal{A}$ and the limiting object $\mathcal{Y}$ are random outcomes of a population process. Then $\mathbb{P}(\mathcal{Y})$ could also serve as the prior for Bayesian inference of $\mathcal{Y}$, using equation (58) but with "|" not "; " for conditioning on $\mathcal{A}$ and $\mathcal{Y}$. In any case, the pair $\{\mathcal{Y}, \Theta_c\}$ is a much more manageable variable than $\mathcal{A}$. For many species, it will not be necessary to record special events in the limiting model. Without $\mathcal{Y}$, equation (58) reduces to traditional coalescent-based inference.

The issues we raise here about pedigrees parallel those in recent work on population bottlenecks. Like the trajectories of population sizes through time in coalescent hidden Markov models, bottlenecks have traditionally been considered fixed events of the past. But models of recurrent bottlenecks have recently been considered. A bottleneck is the event that a population ordinarily of size N_0 has size $N_B < N_0$ for a period of time. In Birkner *et al.* (2009, Section 6), it was shown that a Ξ-coalescent describes the limiting gene-genealogical process for a model with recurrent severe bottlenecks, specifically with the bottleneck duration going to zero and $N_B/N_0 \to 0$ as both N_B and N_0 go to infinity. González Casanova *et al.* (2022) used a similar framework but allowed that N_B could be finite. They described a new class of Ξ-coalescents they called the symmetric coalescent. A model like ours with $\psi = 1$ and random selfing between the parents of the big family would give one of these, being identical to a short drastic bottleneck (González Casanova *et al.* 2022, Definition 3) with $N_B = 4$ and our θ and λ corresponding to their α and $k^{(N)}$. But as noted in Birkner *et al.* (2009), Ξ-coalescent models are only obtained for recurrent bottlenecks by averaging over the exponential process which generates them. When the times and severities of bottlenecks are fixed, the result will depend on these and will not be a time-homogeneous Markov process. Against this backdrop of similarities, a small but notable difference is that the bottleneck models in Birkner *et al.* (2009) and González Casanova *et al.* (2022) are haploid rather than diploid.

The proof of Theorem 1 uses the idea of two independent copies of the coalescent process on the same pedigree. Genetically, these

are the gene-genealogies of two independently assorting loci conditional on their shared pedigree. Our results suggest a reinterpretation of population-genetic models which predict nonzero correlations or positive covariances of coalescence times at unlinked loci. Eldon and Wakeley (2008) found that the correlation could be positive in a model of recombination and the haploid (or gametic) equivalent of big families. Birkner *et al.* (2013b) extended this finding to a diploid model of recombination with big families similar to the ones we studied here. It appears that such correlations result from averaging over the pedigree. In the simple model we considered, they arise from averaging over the times of big families, cf. equations (53) and (54). On any fixed pedigree the correlation of coalescence times at unlinked loci must be zero.

The comparison with recurrent bottlenecks is apt here as well. Schaper *et al.* (2012) constructed a recurrent-bottleneck model for recombination and coalescence at two loci with recombination. They note that what is relevant for data is the covariance of coalescence times conditional on the series bottleneck events in the ancestry of the population, not the unconditional covariance which averages over these. They showed that the conditional covariance goes to zero as the recombination parameter goes to infinity. They also showed, in the Ξ-coalescent limit of Birkner *et al.* (2009) that the covariance could be positive even as the recombination parameter goes to infinity. We note an analogous finding for yet another model in Wakeley and Lessard (2003), in which nonzero correlations of coalescence times at unlinked resulted from taking the number of subpopulations to infinity in an island migration model, even though for any finite number of subpopulations the correlation goes to zero as the recombination parameter tends to infinity.

In sum, for a century it has been common practice in population genetics to compute probabilities of past events by averaging over an assumed process of reproduction. What we have shown is that when big families occur with some frequency, or more generally when the descendants of a small number of individuals take over a sizable fraction of the population in a short period of time, this averaging is not justified and can produce spurious results. Instead, such extreme outcomes of reproduction should be viewed as fixed, and probabilities of coalescence and other events conditioned upon them. In light of this, existing multiple-merger coalescent models must be reassessed and most likely replaced with conditional or quenched models. A comparison with how population size has been treated as fixed is of some interest because it too is an outcome of reproduction. In both cases, it is when population-genetic models are applied to explain variation among loci that the importance of conditioning on past events is most readily apparent.

Data availability

Mathematica notebooks containing some calculations, detailed in Appendix A, are available at: https://github.com/diamantidisdimitris/Bursts-of-coalescence.

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Adams AM, Hudson RR. 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. Genetics. 168(3): 1699–1712. doi:10.1534/genetics.104.030171.

Agranat-Tamir L, Mooney JA, Rosenberg NA. 2024. Counting the genetic ancestors from source populations in members of an admixed population. Genetics. iyae011. doi:10.1093/genetics/iyae011.

Aguillon SM, Fitzpatrick JW, Bowman R, Schoech SJ, Clark AG, Coop G, Chen N. 2017. Deconstructing isolation-by-distance: the genomic consequences of limited dispersal. PLoS Genet. 13(8): e1006911. doi:10.1371/journal.pgen.1006911.

Albers PK, McVean G. 2020. Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS Biol. 18(1): e3000586. doi:10.1371/journal.pbio.3000586.

Anderson-Trocmé L, Nelson D, Zabad S, Diaz-Papkovich A, Kryukov I, Baya N, Touvier M, Jeffery B, Dina C, Vézina H, *et al.* 2023. On the genes, genealogies, and geographies of Quebec. Science, 380(6647):849–855. doi:10.1126/science.add5300.

Árnason E, Koskela J, Halldórsdóttir K, Eldon B. 2023. Sweepstakes reproductive success via pervasive and recurrent selective sweeps. eLife. 12:e80781. doi:10.7554/eLife.80781.

Avise JC. 1989. Gene trees and organismal histories: a phylogenetic approach to population biology. Evolution. 43(6):1192–1208. doi:10.1111/j.1558-5646.1989.tb02568.x.

Avise JC. 2000. Phylogeography: The History and Formation of Species. Cambridge (MA): Harvard University Press.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst. 18(1):489–522. doi:10.1146/annurev.es.18.110187.002421.

Avise JC, Wollenberg K. 1997. Phylogenetics and the origin of species. Proc Natl Acad Sci U S A. 94(15):7748–7755. doi:10.1073/pnas.94.15.7748.

Ball RM, Neigel JE, Avise JC. 1990. Gene genealogies within the organismal pedigrees of random-mating populations. Evolution. 44(2): 360–370. doi:10.1111/j.1558-5646.1990.tb05205.x.

Barton NH, Etheridge AM. 2011. The relation between reproductive value and genetic contribution. Genetics. 188(4):953–973. doi:10.1534/genetics.111.127555.

Barton NH, Etheridge AM, Véber A. 2017. The infinitesimal model: definition, derivation, and implications. Theor Popul Biol. 118: 50–73. doi:10.1016/j.tpb.2017.06.001.

Birkner M, Blath J, Eldon B. 2013a. Statistical properties of the site-frequency spectrum associated with λ-coalescents. Genetics. 195(3):1037–1053. doi:10.1534/genetics.113.156612.

Birkner M, Blath J, Eldon B. 2013b. An ancestral recombination graph for diploid populations with skewed offspring distribution. Genetics. 193(1):255–290. doi:10.1534/genetics.112.144329.

Birkner M, Blath J, Möhle M, Steinrücken M, Tams J. 2009. A modified lookdown construction for the Xi-Fleming-Viot process with

mutation and populations with recurrent bottlenecks. ALEA Lat Am J Probab Math Stat. 6:35–61.

Birkner M, Černý J, Depperschmidt A, Gantert N. 2013c. Directed random walk on the backbone of an oriented percolation cluster. Electron J Probab. 18:1–35. doi:10.1214/EJP.v18-2302.

Birkner M, Liu H, Sturm A. 2018. Coalescent results for diploid exchangeable population models. Electron J Probab. 23:1–44. doi:10.1214/18-EJP175.

Blath J, Cronjäger MC, Eldon B, Hammer M. 2016. The site-frequency spectrum associated with xi-coalescents. Theor Popul Biol. 110:36–50. doi:10.1016/j.tpb.2016.04.002.

Blath J, Kadow S, Ortgiese M. 2014. The largest strongly connected component in the cyclical pedigree model of Wakeley et al. Theor Popul Biol. 98:28–37. doi:10.1016/j.tpb.2014.10.001.

Bolthausen E, Sznitman A. 2002a. On the static and dynamic points of view for certain random walks in random environment. Methods Appl Anal. 9(3):345–376. doi:10.4310/MAA.2002.v9.n3.a4.

Bolthausen E, Sznitman A. 2002b. Ten lectures on random media. In: DMV-Seminar. Vol. 32. Basel: Birkhäuser (Oberwolfach Seminars).

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 15(4):e1006650. doi:10.1371/journal.pcbi.1006650.

Brown WM. 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci U S A. 77(6):3605–3609. doi:10.1073/pnas.77.6.3605.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A. 76(4):1967–1971. doi:10.1073/pnas.76.4.1967.

Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. Nature. 325(6099):31–36. doi:10.1038/325031a0.

Cannings C. 1974. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Adv Appl Probab. 6(2):260–290. doi:10.2307/1426293.

Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. Evolution. 21(3):550–570. Also published as: Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet 19(3 Pt 1):233–257. doi:10.1111/j.1558-5646.1967.tb03411.x.

Chang JT. 1999. Recent common ancestors of all present-day individuals. Adv Appl Probab. 31(4):1002–1026. doi:10.1239/aap/1029955256.

Charlesworth B. 2022. Fisher's historic 1922 paper On the dominance ratio. Genetics. 220(3):iyac006. doi:10.1093/genetics/iyac006.

Coron C, Le Jan Y. 2022. Pedigree in the biparental Moran model. J Math Biol. 84(6):51. doi:10.1007/s00285-022-01752-0.

Der R, Plotkin JB. 2014. The equilibrium allele frequency distribution for a population with reproductive skew. Genetics. 196(4):1199–1216. doi:10.1534/genetics.114.161422.

Derrida B, Manrubia SC, Zanette DH. 1999. Statistical properties of genealogical trees. Phys Rev Lett. 82:1987–1990. doi:10.1103/PhysRevLett.82.1987.

Derrida B, Manrubia SC, Zanette DH. 2000a. Distribution of repetitions of ancestors in genealogical trees. Physica A. 281(1–4):1–16. doi:10.1016/S0378-4371(00)00031-5.

Derrida B, Manrubia SC, Zannette DH. 2000b. On the genealogy of a population of biparental individuals. J Theor Biol. 203(3):303–315. doi:10.1006/jtbi.2000.1095.

Di Rienzo A, Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci U S A. 88(5):1597–1601. doi:10.1073/pnas.88.5.1597.

Donnelly P, Kurtz TG. 1999. Particle representations for measure-valued population models. Ann Probab. 27(1):166–205. doi:10.1214/aop/1022677258.

Donnelly P, Tavaré S, Balding DJ, Griffiths RC. 1996. Estimating the age of the common ancestor of men from the ZFY intron. Science. 272(5266):1357–1359. doi:10.1126/science.272.5266.1357.

Donnelly P, Wiuf C, Hein J, Slatkin M, Ewens WJ, Kingman JFC. 1999. Discussion: recent common ancestors of all present-day individuals. Adv Appl Probab. 31(4):1027–1035. doi:10.1239/aap/1029955257.

Dorit RL, Akashi H, Gilbert W. 1995. Absence of polymorphism at the ZFY locus on the human Y chromosome. Science. 268(5214):1183–1185. doi:10.1126/science.7761836.

Durrett R, Schweinsberg J. 2004. Approximating selective sweeps. Theor Popul Biol. 66(2):129–138. doi:10.1016/j.tpb.2004.04.002.

Durrett R, Schweinsberg J. 2005. A coalescent model for the effect of advantageous mutations on the genealogy of a population. Stoch Process Their Appl. 115(10):1628–1657. doi:10.1016/j.spa.2005.04.009.

Eldon B. 2020. Evolutionary genomics of high fecundity. Annu Rev Genet. 54(1):213–236. doi:10.1146/annurev-genet-021920-095932.

Eldon B, Birkner M, Blath J, Freund F. 2015. Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents? Genetics. 199(3):841–856. doi:10.1534/genetics.114.173807.

Eldon B, Wakeley J. 2008. Linkage disequilibrium under skewed offspring distribution among individuals in a population. Genetics. 178(3):1517–1532. doi:10.1534/genetics.107.075200.

Ewens WJ. 1974. A note on the sampling theory for infinite alleles and infinite sites models. Theor Popul Biol. 6(2):143–148. doi:10.1016/0040-5809(74)90020-3.

Ewens WJ. 1990. Population genetics theory—the past and the future. In: Lessard S, editor. Mathematical and Statistical Developments of Evolutionary Theory. Amsterdam: Kluwer Academic Publishers. p. 177–227.

Ewens WJ. 2004. Mathematical Population Genetics, Volume I: Theoretical Foundations. Berlin: Springer-Verlag.

Ewens WJ, Maruyama T. 1975. A note on the variance of the number of loci having a given gene frequency. Genetics. 80(1):221–222. doi:10.1093/genetics/80.1.221.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9(10):e1003905-1–e1003905-17. doi:10.1371/journal.pgen.1003905.

Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics. 37(24):4882–4885. doi:10.1093/bioinformatics/btab468.

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am J Hum Genet. 25(5):471–492.

Felsenstein J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution. 35(6):1229–1242. doi:10.1111/j.1558-5646.1981.tb04991.x.

Felsenstein J. 2004. Inferring Phylogenies. Sunderland (MA): Sinauer Associates, Inc.

Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 52:399–433. doi:10.1017/S0080456800012163.

Fisher RA. 1922. On the dominance ratio. Proc R Soc Edinb. 42:321–341. doi:10.1017/S0370164600023993.

Fisher RA. 1930. The distribution of gene ratios for rare mutations. Proc R Soc Edinb. 50:205–220.

Freund F, Kerdoncuff E, Matuszewski S, Lapierre M, Hildebrandt M, Jensen JD, Ferretti L, Lambert A, Sackton TB, Achaz G. 2023.

Interpreting the pervasive observation of U-shaped site frequency spectra. PLoS Genet. 19(3):e1010677-1–e1010677-18. doi:10.1371/journal.pgen.1010677.

Fu Y, Li W. 1996. Estimating the age of the common ancestor of men from the ZFY intron. Science. 272(5266):1356–1357. doi:10.1126/science.272.5266.1356.

Gasbarra D, Pirinen M, Sillanpää MJ, Arjas E. 2007a. Estimating genealogies from linked marker data: a Bayesian approach. BMC Bioinformatics. 8(1):411. doi:10.1186/1471-2105-8-411.

Gasbarra D, Pirinen M, Sillanpää MJ, Salmela E, Arjas E. 2007b. Estimating genealogies from unlinked marker data: a Bayesian approach. Theor Popul Biol. 72(3):305–322. doi:10.1016/j.tpb.2007.06.004.

Gasbarra D, Sillanpää MJ, Arjas E. 2005. Backward simulation of ancestors of sampled individuals. Theor Popul Biol. 67(2):75–83. doi:10.1016/j.tpb.2004.08.003.

Gernhard T. 2008. The conditioned reconstructed process. J Theor Biol. 253(4):769–778. doi:10.1016/j.jtbi.2008.04.005.

González Casanova A, Miró Pina V, Siri-Jégousse A. 2022. The symmetric coalescent and Wright–Fisher models with bottlenecks. Ann Appl Probab. 32(1):235–268. doi:10.1214/21-AAP1676.

Gravel S, Steel M. 2015. The existence and abundance of ghost ancestors in biparental populations. Theor Popul Biol. 101:47–53. doi:10.1016/j.tpb.2015.02.002.

Griffiths RC, Marjoram P. 1997. An ancestral recombination graph. In: Donnelly P, Tavaré S, editors. Progress in Population Genetics and Human Evolution. New York: Springer-Verlag (IMA Volumes in Mathematics and its Applications, vol. 87). p. 257–270.

Griffiths RC, Tavaré S. 1994. Ancestral inference in population genetics. Stat Sci. 9(3):307–319. doi:10.1214/ss/1177010378.

Griffiths RC, Tavaré S. 1998. The age of a mutation in a general coalescent tree. Commun Stat Stoch Models. 14(1–2):273–295. doi:10.1080/15326349808807471.

Guez J, Achaz G, Bienvenu F, Cury J, Toupance B, Austerlitz F. 2023. Cultural transmission of reproductive success impacts genomic diversity, coalescent tree topologies, and demographic inferences. Genetics. 223(4):iyad007. doi:10.1093/genetics/iyad007.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5(10):e1000695-1–e1000695-11. doi:10.1371/journal.pgen.1000695.

Haller BC, Messer PW. 2019. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Mol Biol Evol. 36(3):632–637. doi:10.1093/molbev/msy228.

Heled J, Drummond AJ. 2009. Bayesian inference of species trees from multilocus data. Mol Biol Evol. 27(3):570–580. doi:10.1093/molbev/msp274.

Hudson RR. 1983a. Properties of a neutral allele model with intragenic recombination. Theor Popul Biol. 23(2):183–201. doi:10.1016/0040-5809(83)90013-8.

Hudson RR. 1983b. Testing the consta neutral allele model with protein sequence data. Evolution. 37(1):203–217. doi:10.1111/j.1558-5646.1983.tb05528.x.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature. 408(6813):708–713. doi:10.1038/35047064.

Kamm J, Terhorst J, Durbin R, Song YS. 2020. Efficiently inferring the demographic history of many populations with allele count data. J Am Stat Assoc. 115(531):1472–1487. doi:10.1080/01621459.2019.1635482.

Karlin S, McGregor J. 1967. The number of mutant forms maintained in a population. In: Le Cam LM, Neyman J, editors. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Held at the Statistical Laboratory; June 21–July 18, 1965 and December 27, 1965–January 7, 1966. Berkeley (CA): University of California Press. p. 415–538.

Kelleher J, Wong Y, Wohns AW, Fadil C, Albers PK, McVean G. 2019. Inferring whole-genome histories in large population datasets. Nat Genet. 51(9):1330–1338. doi:10.1038/s41588-019-0483-y.

Kimura M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to the steady flux of mutations. Genetics. 61(4):893–903. doi:10.1093/genetics/61.4.893.

Kingman JFC. 1982. On the genealogy of large populations. J Appl Probab. 19(A):27–43. doi:10.2307/3213548.

Ko A, Nielsen R. 2019. Joint estimation of pedigrees and effective population size using Markov chain Monte Carlo. Genetics. 212(3):855–868. doi:10.1534/genetics.119.302280.

Koskela J. 2018. Multi-locus data distinguishes between population growth and multiple merger coalescents. Stat Appl Genet Mol Biol. 17(3):20170011. doi:10.1515/sagmb-2017-0011.

Kuhner MK, Yamato J, Felsenstein J. 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics. 140(4):1421–1430. doi:10.1093/genetics/140.4.1421.

Lachance J. 2009. Inbreeding, pedigree size, and the most recent common ancestor of humanity. J Theor Biol. 261(2):238–247. doi:10.1016/j.jtbi.2009.08.006.

Lambert A, Stadler T. 2013. Birth–death models and coalescent point processes: the shape and probability of reconstructed phylogenies. Theor Popul Biol. 90:113–128. doi:10.1016/j.tpb.2013.10.002.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature. 475(7357):493–496. doi:10.1038/nature10231.

Lundstrom R, Tavaré S, Ward RH. 1992. Estimating substitution rates from molecular data using the coalescent. Proc Natl Acad Sci U S A. 89(13):5961–5965. doi:10.1073/pnas.89.13.5961.

Malécot G. 1941. Etude mathématique des populations Mendélienne. Ann Univ Lyon Sci A. 4:45–60.

Malécot G. 1946. La consanguinité dans une population limitée. C R Acad Sci, Paris. 222:841–843.

Malécot G. 1948. Les Mathématiques de l'Hérédité. Masson, Paris. Available from https://wellcomecollection.org/works/msfaxgkw. Extended translation: The Mathematics of Heredity. W.H. Freeman, San Francisco (1969).

Matsen FA, Evans SN. 2008. To what extent does genealogical ancestry imply genetic ancestry? Theor Popul Biol. 74(2):182–190. doi:10.1016/j.tpb.2008.06.003.

Matuszewski S, Hildebrandt ME, Achaz G, Jensen JD. 2018. Coalescent processes with skewed offspring distributions and nonequilibrium demography. Genetics. 208(1):323–338. doi:10.1534/genetics.117.300499.

McVean GAT. 2002. A genealogical interpretation of linkage disequilibrium. Genetics. 162(2):987–991. doi:10.1093/genetics/162.2.987.

Möhle M. 1998a. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. Adv Appl Probab. 30(2):493–512. doi:10.1239/aap/1035228080.

Möhle M. 1998b. Coalescent results for two-sex population models. Adv Appl Probab. 30(2):513–520. doi:10.1239/aap/1035228081.

Möhle M. 1999. The concept of duality and applications to Markov processes arising in neutral population genetics models. Bernoulli. 5:761–777. doi:10.2307/3318443.

Möhle M, Sagitov S. 2001. A classification of coalescent processes for haploid exchangeable population models. Ann Probab. 29(4):1547–1562. doi:10.1214/aop/1015345761.

Molchanov SA. 1994. Lectures on random media. In: Bernard P, editor. École d'Été de Probabilités de Saint-Flour—1992. Berlin: Springer-Verlag (Lectures Notes in Mathematics, vol. 1581). p. 242–411. Available from https://doi.org/10.1007/BFb0073871.

Nielsen R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics. 154(2):931–942. doi:10.1093/genetics/154.2.931.

Padmadisastra S. 1988. Estimating divergence times. Theor Popul Biol. 34(3):297–319. doi:10.1016/0040-5809(88)90026-3.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. Genetics. 192(3):1065–1093. doi:10.1534/genetics.112.145037.

Peter BM. 2016. Admixture, population structure, and F-statistics. Genetics. 202(4):1485–1501. doi:10.1534/genetics.115.183913.

Pitman J. 1999. Coalescents with multiple collisions. Ann Probab. 27(4):1870–1902. doi:10.1214/aop/1022874819.

Ralph P, Thornton K, Kelleher J. 2020. Efficiently summarizing relationships in large samples: a general duality between statistics of genealogies and genomes. Genetics. 215(3):779–797. doi:10.1534/genetics.120.303253.

Ralph PL. 2019. An empirical approach to demographic inference with genomic data. Theor Popul Biol. 127:91–101. doi:10.1016/j.tpb.2019.03.005.

Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics. 164(4):1645–1656. doi:10.1093/genetics/164.4.1645.

Rohde DLT, Olson S, Chang JT. 2004. Modelling the recent common ancestry of all living humans. Nature. 431(7008):562–566. doi:10.1038/nature02842.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542. doi:10.1093/sysbio/sys029.

Sagitov S. 1999. The general coalescent with asynchronous mergers of ancestral lines. J Appl Probab. 36(4):1116–1125. doi:10.1239/jap/1032374759.

Sagitov S. 2003. Convergence to the coalescent with simultaneous multiple mergers. J Appl Probab. 40(4):839–854. doi:10.1239/jap/1067436085.

Sainudiin R, Thatte B, Véber A. 2016. Ancestries of a recombining diploid population. J Math Biol. 72(1–2):363–408. doi:10.1007/s00285-015-0886-z.

Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. Genetics. 132(4):1161–1176. doi:10.1093/genetics/132.4.1161.

Schaper E, Eriksson A, Rafajlovic M, Sagitov S, Mehlig B. 2012. Linkage disequilibrium under recurrent bottlenecks. Genetics. 190(1):217–229. doi:10.1534/genetics.111.134437.

Schweiger R, Durbin R. 2023. Ultrafast genome-wide inference of pairwise coalescence times. Genome Res. 33:1023–1031. doi:10.1101/gr.277665.123.

Schweinsberg J. 2000. Coalescents with simultaneous multiple collisions. Electron J Probab. 5:1–50. doi:10.1214/EJP.v5-68.

Schweinsberg J, Durrett R. 2005. Random partitions approximating the coalescence of lineages during a selective sweep. Ann Appl Probab. 15(3):1591–1651. doi:10.1214/105051605000000430.

Sheehan S, Harris K, Song YS. 2013. Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. Genetics. 194(3):647–662. doi:10.1534/genetics.112.149096.

Sjödin P, Kaj I, Krone S, Lascoux M, Nordborg M. 2005. On the meaning and existence of an effective population size. Genetics. 169(2):1061–1070. doi:10.1534/genetics.104.026799.

Slatkin M. 1991. Inbreeding coefficients and coalescence times. Genet Res. 58(2):167–175. doi:10.1017/S0016672300029827.

Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation for thousands of samples. Nat Genet. 51(9):1321–1329. doi:10.1038/s41588-019-0484-x.

Spence JP, Kamm JA, Song YS. 2016. The site frequency spectrum for general coalescents. Genetics. 202(4):1549–1561. doi:10.1534/genetics.115.184101.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4(1):vey016. doi:10.1093/ve/vey016.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105(2):437–460. doi:10.1093/genetics/105.2.437.

Tellier A, Lemaire C. 2014. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. Mol Ecol. 23(11):2637–2652. doi:10.1111/mec.12755.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. Nature. 526(7571):68–74. doi:10.1038/nature15393.

Tyukin A. 2015. Quenched limits of coalescents in fixed pedigrees [master's thesis]. [Mainz (Germany)]: Johannes-Gutenberg-Universität.

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci U S A. 86(23):9350–9354. doi:10.1073/pnas.86.23.9350.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science. 253(5027):1503–1507. doi:10.1126/science.1840702.

Wakeley J. 1999. Nonequilibrium migration in human history. Genetics. 153(4):1863–1871. doi:10.1093/genetics/153.4.1863.

Wakeley J. 2009. Coalescent Theory: An Introduction. Greenwood Village (CO): Roberts & Company Publishers. Current publisher: New York (NY): Macmillan Learning.

Wakeley J, King L, Low BS, Ramachandran S. 2012. Gene genealogies within a fixed pedigree, and the robustness of Kingman's coalescent. Genetics. 190(4):1433–1445. doi:10.1534/genetics.111.135574.

Wakeley J, King L, Wilton PR. 2016. Effects of the population pedigree on genetic signatures of historical demographic events. Proc Natl Acad Sci U S A. 113(29):7994–8001. doi:10.1073/pnas.1601080113.

Wakeley J, Lessard S. 2003. Theory of the effects of population structure and sampling on patterns of linkage disequilibrium applied to genomic data from humans. Genetics. 164(3):1043–1053. doi:10.1093/genetics/164.3.1043.

Wang K, Mathieson I, O'Connell J, Schiffels S. 2020. Tracking human population structure through time from whole genome sequences. PLoS Genet. 16(3):e1008552-1–e1008552-24. doi:10.1371/journal.pgen.1008552.

Ward RH, Frazier BL, Dew-Jager K, Pääbo S. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci U S A. 88(19):8720–8724. doi:10.1073/pnas.88.19.8720.

Watterson GA. 1985. Estimating species divergence times using multi-locus data. In: Ohta T, Aoki K, editors. Population Genetics and Molecular Evolution: Papers Marking the Sixtieth Birthday of Motoo Kimura. Tokyo: Japan Scientific Societies Press; Berlin: Springer-Verlag. p. 163–183.

Weiss G, von Haeseler A. 1996. Estimating the age of the common ancestor of men from the ZFY intron. Science. 272(5266):1359–1360. doi:10.1126/science.272.5266.1359.

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, *et al.* 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc. 26(4):375–400. doi:10.1111/j.1095-8312.1985.tb02048.x.

Wilton PR, Baduel P, Landon MM, Wakeley J. 2017. Population structure and coalescence in pedigrees: comparisons to the structured coalescent and a framework for inference. Theor Popul Biol. 115: 1–12. doi:10.1016/j.tpb.2017.01.004.

Wohns AW, Wong Y, Ben J, Akbari A, Mallick S, Pinhasi R, Patterson N, Reich D, Kelleher J, McVean G. 2022. A unified genealogy of modern and ancient genomes. Science. 375(6583):eabi8264. doi: 10.1126/science.abi8264.

Wolfram Research, Inc. 2022. Mathematica, Version 13.1.

Wollenberg K, Avise JC. 1998. Sampling properties of genealogical pathways underlying population pedigrees. Evolution. 52(4): 957–966. doi:10.1111/j.1558-5646.1998.tb01825.x.

Wooding S, Rogers A. 2002. The matrix coalescent and an application to human single-nucleotide polymorphisms. Genetics. 161(4): 1641–1650. doi:10.1093/genetics/161.4.1641.

Wright S. 1921a. Systems of mating. I. The biometric relations between parent and offspring. Genetics. 6(2):111–123. doi:10.1093/genetics/6.2.111.

Wright S. 1921b. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. Genetics. 6(2):124–143. doi:10.1093/genetics/6.2.124.

Wright S. 1921c. Systems of mating. III. Assortative mating based on somatic resemblance. Genetics. 6(2):144–161. doi:10.1093/genetics/6.2.144.

Wright S. 1921d. Systems of mating. IV. The effects of selection. Genetics. 6(2):162–166. doi:10.1093/genetics/6.2.162.

Wright S. 1921e. Systems of mating. V. General considerations. Genetics. 6(2):167–178. doi:10.1093/genetics/6.2.167.

Wright S. 1922. Coefficients of inbreeding and relationship. Am Nat. 56(645):330–338. doi:10.1086/279872.

Wright S. 1931. Evolution in Mendelian populations. Genetics. 16(2): 97–159. doi:10.1093/genetics/16.2.97.

Yang Z. 2002. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. Genetics. 162(4):1811–1823. doi:10.1093/genetics/162.4.1811.

Zhang BC, Biddanda A, Palamara PF. 2023. Biobank-scale inference of ancestral recombination graphs enables genealogical analysis of complex traits. Nat Genet. 55(5):768–776. doi:10.1038/s41588-023-01379-x.

Appendix A
Joint diploid ancestral process

Recall the joint ancestral process $\widetilde{M}$ describes the lineage dynamics of two conditionally independent samples of size 2. In the *The joint diploid ancestral process with 10 states is Markovian* section we establish its Markov property. This property should hold for any population model in which the outcomes of reproduction in different generations are independent. In the *Transition matrix for a generation with a big family* section, we provide its transition matrix for a generation with a big family and in the *Two extremes for the sizes of big families* section, we study two extreme cases, $\psi = 0$ and $\psi = 1$. In the former case, we recover the transition matrix of the joint diploid ancestral process for a Wright–Fisher generation and in the latter we consider the case when the highly reproductive couple replaces the entire population.

The joint diploid ancestral process with 10 states is Markovian

Recall the state space of the process $\widetilde{M} = (\widetilde{M}_g)_{g \in \mathbb{Z}_{\geq 0}}$ is $\mathcal{S} = \{\xi_{00}^{(4)}, \xi_{00}^{(3)}, \ldots, \xi_\Delta\}$ which has 10 states including the state ξ_Δ where one, the other or both pairs of lineages have coalesced.

Fix a pair of states $(\xi, \eta) \in \mathcal{S}^2$. It holds that

$$
\mathbb{P}^{(N)}(\widetilde{M}_{g+1} = \eta \mid \widetilde{M}_g = \xi, \ \{\widetilde{M}_k\}_{k=0}^{g-1}) = \sum_{\alpha^{(g)}} \mathbb{P}^{(N)}(\widetilde{M}_{g+1} = \eta \mid
$$
$$
\widetilde{M}_g = \xi, \ \alpha^{(g)}, \ \{\widetilde{M}_k\}_{k=0}^{g-1}) \mathbb{P}^{(N)}(\alpha^{(g)} \mid \widetilde{M}_g = \xi, \ \{\widetilde{M}_k\}_{k=0}^{g-1})
$$
(A1)

$$
= \sum_{\alpha^{(g)} \in A_{\xi,\eta}} \mathbb{P}^{(N)}(\widetilde{M}_{g+1} = \eta \mid \widetilde{M}_g = \xi, \ \alpha^{(g)}) \mathbb{P}^{(N)}(\alpha^{(g)}),
$$
(A2)

where the last sum $\sum_{\alpha^{(g)} \in A_{\xi,\eta}}$ is over all possible parent assignments $\alpha^{(g)}$ that belong to the pattern $A_{\xi,\eta}$ described in Tables A1–A3. These are exactly the parent assignments for which transition to η is *possible* in the sense that $\mathbb{P}^{(N)}(\widetilde{M}_{g+1} = \eta \mid \widetilde{M}_g = \xi, \ \alpha^{(g)}) > 0$ if and only if $\alpha^{(g)} \in A_{\xi,\eta}$.

In the last equality above, we noted that a parent assignment $\alpha^{(h)}$ between generations g and $g+1$ is independent of $\{\widetilde{M}_k\}_{k=0}^{g}$ and so $\mathbb{P}^{(N)}(\alpha^{(g)} \mid \widetilde{M}_g = \xi, \ \{\widetilde{M}_k\}_{k=0}^{g-1}) = \mathbb{P}^{(N)}(\alpha^{(g)})$. Also, $\mathbb{P}^{(N)}(\widetilde{M}_{g+1} = \eta \mid \widetilde{M}_g = \xi, \ \alpha^{(g)}, \ \{\widetilde{M}_k\}_{k=0}^{g-1}) = \mathbb{P}^{(N)}(\widetilde{M}_{g+1} = \eta \mid \widetilde{M}_g = \xi, \ \alpha^{(g)})$ because it is determined entirely by Mendelian randomness and not by the population model (whether it is Wright–Fisher or something else). This conditional probability is computed in Tables A4–A10, and we only need to compute this for those $\alpha^{(g)}$ in patterns whose probability is of order $O(\frac{1}{N})$ or larger in view of Möhle (1998a, Lemma 1).

In particular, we have shown that the number $\widetilde{\pi}_{\xi,\eta}^{(N)} := \mathbb{P}^{(N)}(\widetilde{M}_{g+1} = \eta \mid \widetilde{M}_g = \xi, \ \{\widetilde{M}_k\}_{k=0}^{g-1})$ does not depend on the "history" $\{\widetilde{M}_k\}_{k=0}^{g-1}$. Hence, the process $\widetilde{M}$ is Markovian.

Transition matrix for a generation with a big family

For any $\xi \in \mathcal{S}$, let $\vec{r}_\xi = (\widetilde{\pi}_{\xi,\eta}^{(N)})_{\eta \in \mathcal{S}}$ be the ξth row of $\widetilde{\Pi}_N$. By equation (A2), each row of $\widetilde{\Pi}_N$ is computed as

$$
\vec{r}_\xi = A_\xi \cdot \vec{p}_\xi,
$$

where A_ξ is given by the appropriate matrix in Tables A4–A10 corresponding to ξ, and $\vec{p}_\xi$ contains the probabilities of the patterns from Tables A1–A3 used in all the transitions from ξ to $\eta \in \mathcal{S}$. In a generation with a big family, $\widetilde{\Pi}_N$ is given as

$$
\widetilde{\Pi}_N^{\mathrm{BF}} = \begin{pmatrix} r_{\xi_{00}^{(4)}}^{\mathrm{BF}} \\ \vdots \\ r_{\xi_\Delta}^{\mathrm{BF}} \end{pmatrix} + O(N^{-2}),
$$
(A3)

where r_ξ^{BF} in equation (A3) is the row of $\widetilde{\Pi}_N^{\mathrm{BF}}$ that corresponds to initial state ξ up to an error of order $O(N^{-2})$, given in equations (A4)–(A8). For example, $r_{\xi_{11}^{(1)}}^{\mathrm{BF}}$ is derived by multiplying the matrix $A_{\xi_{11}^{(1)}}$ given in Table A4 by the vector column $\vec{p} := (\frac{1-\psi}{N}, \ \psi + (1-\psi) \times (1 - \frac{1}{N}))^T$ on the right. The calculations for each row are available at https://github.com/diamantidisdimitris/Bursts-of-coalescence in the PiBF_row_derivation_print file and can be reproduced in the corresponding Mathematica notebook PiBF_row_derivation_print_notebook, where Mathematica version 13.1.0 (Wolfram Research, Inc. 2022) was used. Its rows are given as

$$
r^{\text{BF}}_{\xi^{(1)}_{11}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(1)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}0\\0\\\frac{\psi-1}{N}+1\\0\\0\\0\\0\\0\\\frac{1-\psi}{4N}\\\frac{3-3\psi}{4N}\end{pmatrix}^{T}, \quad r^{\text{BF}}_{\xi^{(2)}_{11}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(2)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}\frac{4\psi^2+2\psi-6}{N}-\psi^2+1\\\frac{4-4\psi}{N}\\\psi^2\\\frac{1-\psi^2}{2N}\\0\\\frac{1-\psi^2}{2N}\\0\\0\\0\\\frac{-3\psi^2+2\psi+1}{N}\end{pmatrix}^{T}, \tag{A4}
$$

$$
r^{\text{BF}}_{\xi^{(2)}_{10}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(1)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}0\\\frac{3\psi^2}{N}-\frac{3}{N}-\psi^2+1\\-\frac{\psi}{N}+\frac{1}{N}+\frac{\psi^2}{2}\\0\\-\frac{\psi}{N}+\frac{1}{N}+\frac{\psi^2}{4}\\0\\\frac{1}{2N}-\frac{\psi^2}{2N}\\0\\0\\-\frac{5\psi^2}{2N}+\frac{2\psi}{N}+\frac{1}{2N}+\frac{\psi^2}{4}\end{pmatrix}^{T}, \quad r^{\text{BF}}_{\xi^{(3)}_{10}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(1)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}-\frac{45\psi^3}{2N}+\frac{83\psi^2}{2N}-\frac{13\psi}{N}-\frac{6}{N}+\frac{3\psi^3}{2}-\frac{5\psi^2}{2}+1\\\frac{5\psi}{N}-\frac{\psi^2}{N}-\frac{8\psi}{N}+\frac{4}{N}-2\psi^3+2\psi^2\\-\frac{\psi^3}{N}+\frac{\psi^2}{N}+\frac{\psi^3}{2}\\\frac{\psi^3}{2N}-\frac{\psi}{N}+\frac{1}{2N}-\frac{\psi^3}{4}+\frac{\psi^2}{4}\\-\frac{\psi^3}{4N}+\frac{\psi^2}{4N}+\frac{\psi^3}{4}\\\frac{\psi^3}{4N}-\frac{\psi^2}{4N}-\frac{\psi}{2N}+\frac{1}{2N}\\0\\\frac{\psi^2}{4N}-\frac{\psi^3}{4N}\\0\\\frac{73\psi^3}{4N}-\frac{167\psi^2}{4N}+\frac{45\psi}{2N}+\frac{1}{N}+\frac{\psi^2}{4}\end{pmatrix}^{T}, \tag{A5}
$$

$$
r^{\text{BF}}_{\xi^{(2)}_{01}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(1)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}0\\\frac{3\psi^2}{N}-\frac{3}{N}-\psi^2+1\\-\frac{\psi}{N}+\frac{1}{N}+\frac{\psi^2}{2}\\0\\\frac{1}{2N}-\frac{\psi^2}{2N}\\0\\-\frac{\psi}{N}+\frac{1}{N}+\frac{\psi^2}{4}\\0\\0\\-\frac{5\psi^2}{2N}+\frac{2\psi}{N}+\frac{1}{2N}+\frac{\psi^2}{4}\end{pmatrix}^{T}, \quad r^{\text{BF}}_{\xi^{(3)}_{01}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(1)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}-\frac{45\psi^3}{2N}+\frac{83\psi^2}{2N}-\frac{13\psi}{N}-\frac{6}{N}+\frac{3\psi^3}{2}-\frac{5\psi^2}{2}+1\\\frac{5\psi}{N}-\frac{\psi^2}{N}-\frac{8\psi}{N}+\frac{4}{N}-2\psi^3+2\psi^2\\-\frac{\psi^3}{N}+\frac{\psi^2}{N}+\frac{\psi^3}{2}\\\frac{\psi^3}{2N}-\frac{\psi}{N}+\frac{1}{2N}-\frac{\psi^3}{4}+\frac{\psi^2}{4}\\0\\\frac{\psi^3}{4N}-\frac{\psi^2}{4N}-\frac{\psi}{2N}+\frac{1}{2N}\\-\frac{\psi^3}{4N}-\frac{\psi^2}{4N}+\frac{\psi^3}{4}\\\frac{\psi^2}{4N}-\frac{\psi^3}{4N}\\0\\\frac{73\psi^3}{4N}-\frac{167\psi^2}{4N}+\frac{45\psi}{2N}+\frac{1}{N}+\frac{\psi^2}{4}\end{pmatrix}^{T}, \tag{A6}
$$

$$
r^{\text{BF}}_{\xi^{(2)}_{00}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(1)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}\frac{\psi^2}{N}+\frac{\psi}{2N}-\frac{3}{2N}-\frac{\psi^2}{4}+\frac{1}{4}\\\frac{\psi^2}{N}-\frac{\psi}{2N}-\frac{1}{2N}-\frac{\psi^2}{2}+\frac{1}{2}\\-\frac{\psi}{4N}+\frac{1}{4N}+\frac{1}{4}\\\frac{1}{8N}-\frac{\psi}{8N}\\-\frac{\psi}{4N}+\frac{1}{4N}+\frac{\psi^2}{8}\\\frac{1}{8N}-\frac{\psi}{8N}\\-\frac{\psi}{4N}+\frac{1}{4N}+\frac{\psi^2}{8}\\\frac{\psi^2}{32}\\-\frac{\psi}{16N}+\frac{1}{16N}+\frac{\psi^2}{32}\\-\frac{2\psi^2}{N}+\frac{17\psi}{16N}+\frac{15}{16N}+\frac{7\psi^2}{16}\end{pmatrix}^{T}, \quad r^{\text{BF}}_{\xi^{(3)}_{00}} = \begin{array}{c}\xi^{(4)}_{00}\\\xi^{(3)}_{00}\\\xi^{(2)}_{00}\\\xi^{(3)}_{01}\\\xi^{(2)}_{01}\\\xi^{(3)}_{10}\\\xi^{(2)}_{10}\\\xi^{(2)}_{11}\\\xi^{(1)}_{11}\\\xi_{\Delta}\end{array}\begin{pmatrix}-\frac{25\psi^3}{2N}+\frac{22\psi^2}{N}-\frac{13\psi}{2N}-\frac{3}{N}+\psi^3-\frac{3\psi^2}{2}+\frac{1}{2}\\\frac{-29\psi^3}{4N}+\frac{77\psi^2}{4N}-\frac{25\psi}{2N}+\frac{1}{2N}-\frac{\psi^3}{2}+\frac{1}{2}\\\frac{3\psi^3}{4N}-\frac{3\psi^2}{4N}-\frac{\psi^3}{4}+\frac{\psi^2}{2}\\\frac{15\psi^3}{16N}+\frac{3\psi^2}{16N}-\frac{\psi}{N}+\frac{1}{2N}-\frac{\psi^3}{8}+\frac{\psi^2}{8}\\\frac{\psi^2}{2N}-\frac{\psi}{N}+\frac{1}{2N}+\frac{\psi^2}{8}\\\frac{15\psi^3}{16N}+\frac{3\psi^2}{16N}-\frac{\psi}{N}+\frac{1}{2N}-\frac{\psi^3}{8}+\frac{\psi^2}{8}\\-\frac{\psi^3}{8N}+\frac{5\psi^2}{8N}-\frac{\psi}{N}+\frac{1}{2N}+\frac{\psi^2}{8}\\-\frac{\psi^3}{16N}+\frac{\psi^2}{16N}+\frac{\psi^2}{32}\\-\frac{3\psi^3}{32N}+\frac{3\psi^2}{32N}+\frac{\psi^3}{32}\\\frac{597\psi^3}{32N}-\frac{1349\psi^2}{32N}+\frac{23\psi}{N}+\frac{1}{2N}-\frac{\psi^3}{16}+\frac{\psi^2}{2}\end{pmatrix}^{T}, \tag{A7}
$$

and

$$
r^{\mathrm{BF}}_{\xi^{(4)}_{00}} =
\begin{array}{c}
\xi^{(4)}_{00} \\
\xi^{(3)}_{00} \\
\xi^{(2)}_{00} \\
\xi^{(3)}_{01} \\
\xi^{(2)}_{01} \\
\xi^{(3)}_{10} \\
\xi^{(2)}_{10} \\
\xi^{(2)}_{11} \\
\xi^{(1)}_{11} \\
\xi_{\Delta}
\end{array}
\left(
\begin{array}{c}
\frac{(239\psi^2-350\psi-24)(\psi-1)^2}{4N} + (2\psi+1)(\psi-1)^2 \\
\frac{(59\psi^4+121\psi^3-18\psi^2-60\psi+16)}{4N} - (2\psi-1)\psi^2 \\
\frac{\psi^2-\psi^4}{8N} + \frac{\psi^4}{4} \\
\frac{-11\psi^4+23\psi^3-\psi^2-15\psi+4}{8N} - \frac{1}{4}(\psi-1)\psi^2 \\
\frac{7(\psi-1)\psi^3}{8N} - \frac{1}{8}(\psi-2)\psi^3 \\
\frac{-11\psi^4+23\psi^3-\psi^2-15\psi+4}{8N} - \frac{1}{4}(\psi-1)\psi^2 \\
\frac{7(\psi-1)\psi^3}{8N} - \frac{1}{8}(\psi-2)\psi^3 \\
\frac{(\psi^2-3\psi+2)\psi^2}{16N} + \frac{\psi^4}{32} \\
\frac{\psi^4}{32} - \frac{(\psi-1)\psi^3}{16N} \\
\frac{-351\psi^4+1383\psi^3-1794\psi^2+754\psi+8}{8N} - \frac{1}{16}\psi^2(\psi^2-8)
\end{array}
\right)^T .
\tag{A8}
$$

Two extremes for the sizes of big families

Degenerate case of no offspring, $\psi = 0$. When a Wright–Fisher generation is taking place, denote $\widetilde{\Pi}_N$ by $\widetilde{\Pi}^{\mathrm{WF}}_N$. The matrix $\widetilde{\Pi}^{\mathrm{WF}}_N$ can be recovered from $\widetilde{\Pi}^{\mathrm{BF}}_N$ by setting $\psi = 0$ up to a term of order $O(N^{-2})$. The additional $O(N^{-2})$ term represents cases in which both pairs of lineages coalesce and which become negligible in the limit. Note there is a single case of nonnegligible double coalescence in the $O(N^{-1})$ transition $\xi^{(2)}_{00} \to \xi_{\Delta}$. The matrix $\widetilde{\Pi}^{\mathrm{WF}}_N$ is given as

$$
\widetilde{\Pi}^{\mathrm{WF}}_N =
\begin{array}{c}
\xi^{(4)}_{00} \\
\xi^{(3)}_{00} \\
\xi^{(2)}_{00} \\
\xi^{(3)}_{10} \\
\xi^{(2)}_{10} \\
\xi^{(3)}_{01} \\
\xi^{(2)}_{01} \\
\xi^{(2)}_{11} \\
\xi^{(1)}_{11} \\
\xi_{\Delta}
\end{array}
\begin{array}{cccccccccc}
S & \xi^{(4)}_{00} & \xi^{(3)}_{00} & \xi^{(2)}_{00} & \xi^{(3)}_{10} & \xi^{(2)}_{01} & \xi^{(3)}_{01} & \xi^{(2)}_{01} & \xi^{(2)}_{11} & \xi^{(1)}_{11} & \xi_{\Delta}
\end{array}
\left[
\begin{array}{cccccccccc}
1-\frac{6}{N} & \frac{4}{N} & 0 & \frac{1}{2N} & 0 & \frac{1}{2N} & 0 & 0 & 0 & \frac{1}{N} \\
\frac{1}{2}-\frac{3}{N} & \frac{1}{2}+\frac{1}{2N} & 0 & \frac{1}{2N} & \frac{1}{2N} & \frac{1}{2N} & \frac{1}{2N} & 0 & 0 & \frac{1}{2N} \\
\frac{1}{4}-\frac{3}{2N} & \frac{1}{2}-\frac{1}{2N} & \frac{1}{4}+\frac{1}{4N} & \frac{1}{8N} & \frac{1}{4N} & \frac{1}{8N} & \frac{1}{4N} & 0 & \frac{1}{16N} & \frac{15}{16N} \\
1-\frac{6}{N} & \frac{4}{N} & 0 & \frac{1}{2N} & 0 & \frac{1}{2N} & 0 & 0 & 0 & \frac{1}{N} \\
0 & 1-\frac{3}{N} & \frac{1}{N} & 0 & \frac{1}{2N} & 0 & \frac{1}{N} & 0 & 0 & \frac{1}{2N} \\
1-\frac{6}{N} & \frac{4}{N} & 0 & \frac{1}{2N} & 0 & \frac{1}{2N} & 0 & 0 & 0 & \frac{1}{N} \\
0 & 1-\frac{3}{N} & \frac{1}{N} & 0 & \frac{1}{N} & 0 & \frac{1}{2N} & 0 & 0 & \frac{1}{2N} \\
1-\frac{6}{N} & \frac{4}{N} & 0 & \frac{1}{2N} & 0 & \frac{1}{2N} & 0 & 0 & 0 & \frac{1}{N} \\
0 & 0 & 1-\frac{1}{N} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4N} & \frac{3}{4N} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
\right]
+ O\!\left(\frac{1}{N^2}\right).
\tag{A9}
$$

Offspring make up the whole population, $\psi = 1$

When $\psi = 1$, the highly reproductive pair gives birth to the entire next population. This is the case of pure Mendelian randomness within a single family with two parents. The matrix $\widetilde{\Pi}^{\mathrm{BF}}_N$ is then equal to

$$
\begin{array}{c}
\xi^{(4)}_{00} \\
\xi^{(3)}_{00} \\
\xi^{(2)}_{00} \\
\xi^{(3)}_{01} \\
\xi^{(2)}_{01} \\
\xi^{(3)}_{10} \\
\xi^{(2)}_{10} \\
\xi^{(2)}_{11} \\
\xi^{(1)}_{11} \\
\xi_{\Delta}
\end{array}
\begin{array}{cccccccccc}
S & \xi^{(4)}_{00} & \xi^{(3)}_{00} & \xi^{(2)}_{00} & \xi^{(3)}_{10} & \xi^{(2)}_{10} & \xi^{(3)}_{01} & \xi^{(2)}_{01} & \xi^{(2)}_{11} & \xi^{(1)}_{11} & \xi_{\Delta}
\end{array}
\left[
\begin{array}{cccccccccc}
0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{32} & \frac{1}{32} & \frac{7}{16} \\
0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{32} & \frac{1}{32} & \frac{7}{16} \\
0 & 0 & \frac{1}{4} & 0 & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{32} & \frac{1}{32} & \frac{7}{16} \\
0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\
0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\
0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{4} \\
0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{4} \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
\right].
\tag{A10}
$$

The entries of the matrix above can be explained in terms of Mendel's laws of independent segregation and independent assortment, given that all individuals have the same two parents. For example, in the first row the initial state is $\xi^{(4)}_{00}$: all four lineages in distinct individuals. The transition to state $\xi^{(1)}_{11}$, where all four lineages are in the same individual but have not coalesced, takes place with probability 1/32 and is computed as follows. Each lineage chooses a parent with probability 1/2, then chooses a gene copy within that parent also with probability 1/2. All four lineages choose the same parent (which can be either of the two) with probability 1/8. Then each pair independently chooses distinct gene copies (i.e. does not coalesce) with probability 1/2.

Limiting behavior of the joint diploid ancestral process

The joint diploid ancestral process has transition matrix

$$
\widetilde{\Pi}_N := (1-\alpha_N)\widetilde{\Pi}^{\mathrm{WF}}_N + \alpha_N \widetilde{\Pi}^{\mathrm{BF}}_N.
\tag{A11}
$$

One can write

$$
\widetilde{\Pi}^{\mathrm{WF}}_N = \widetilde{A}^{\mathrm{WF}} + \frac{1}{N}\widetilde{B}^{\mathrm{WF}} + o\!\left(\frac{1}{N}\right),
\tag{A12}
$$

where

$$
\widetilde{A}^{\mathrm{WF}} :=
\begin{array}{c}
\xi^{(4)}_{00} \\
\xi^{(3)}_{00} \\
\xi^{(2)}_{00} \\
\xi^{(3)}_{10} \\
\xi^{(2)}_{10} \\
\xi^{(3)}_{01} \\
\xi^{(2)}_{01} \\
\xi^{(2)}_{11} \\
\xi^{(1)}_{11} \\
\xi_{\Delta}
\end{array}
\begin{array}{cccccccccc}
S & \xi^{(4)}_{00} & \xi^{(3)}_{00} & \xi^{(2)}_{00} & \xi^{(3)}_{10} & \xi^{(2)}_{10} & \xi^{(3)}_{01} & \xi^{(2)}_{01} & \xi^{(2)}_{11} & \xi^{(1)}_{11} & \xi_{\Delta}
\end{array}
\left[
\begin{array}{cccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
\right].
\tag{A13}
$$

and

$$
\widetilde{B}^{\mathrm{WF}} :=
\begin{array}{c}
\xi^{(4)}_{00} \\
\xi^{(3)}_{00} \\
\xi^{(2)}_{00} \\
\xi^{(3)}_{10} \\
\xi^{(2)}_{10} \\
\xi^{(3)}_{01} \\
\xi^{(2)}_{01} \\
\xi^{(2)}_{11} \\
\xi^{(1)}_{11} \\
\xi_{\Delta}
\end{array}
\begin{array}{cccccccccc}
S & \xi^{(4)}_{00} & \xi^{(3)}_{00} & \xi^{(2)}_{00} & \xi^{(3)}_{10} & \xi^{(2)}_{10} & \xi^{(3)}_{01} & \xi^{(2)}_{01} & \xi^{(2)}_{11} & \xi^{(1)}_{11} & \xi_{\Delta}
\end{array}
\left[
\begin{array}{cccccccccc}
-6 & 4 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 1 \\
-3 & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\
-\frac{3}{2} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & 0 & \frac{1}{16} & \frac{15}{16} \\
-6 & 4 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 1 \\
0 & -3 & 1 & 0 & \frac{1}{2} & 0 & 1 & 0 & 0 & \frac{1}{2} \\
-6 & 4 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 1 \\
0 & -3 & 1 & 0 & 1 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\
-6 & 4 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 1 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
\right].
\tag{A14}
$$

Similarly, we split

$$
\widetilde{\Pi}^{\mathrm{BF}}_N = \widetilde{A}^{\mathrm{BF}} + \frac{1}{N}\widetilde{B}^{\mathrm{BF}} + o\!\left(\frac{1}{N}\right)
\tag{A15}
$$

with

$$
\widetilde{A}^{\mathrm{BF}} = (r^{(1)}_{\xi^{(4)}_{00}}, \ldots, r^{(1)}_{\xi_{\Delta}})^T, \qquad \widetilde{B}^{\mathrm{BF}} = \frac{1}{N}(r^{(2)}_{\xi^{(4)}_{00}}, \ldots, r^{(2)}_{\xi_{\Delta}})^T
\tag{A16}
$$

defined as $r_\xi^{(1)} := \lim_{N\to\infty} r_\xi^{\mathrm{BF}}$ and $r_\xi^{(2)} := \lim_{N\to\infty} N \cdot (r_\xi^{\mathrm{BF}} - r_\xi^{(1)})$ for all $\xi \in \mathcal{S}$. We provide only the rows of $\widetilde{A}^{\mathrm{BF}}$ since those are the only ones we will need for the subsequent computations. In particular those are given as

$$
r_{\varsigma_{11}^{(1)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}^T\!\!, \quad
r_{\varsigma_{11}^{(2)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} 1-\psi^2 \\ 0 \\ \psi^2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}^T\!\!, \quad
r_{\varsigma_{10}^{(2)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} 0 \\ 1-\psi^2 \\ \frac{\psi^2}{2} \\ 0 \\ \frac{\psi^2}{4} \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{\psi^2}{4} \end{pmatrix}^T\!\!, \quad
r_{\varsigma_{10}^{(3)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} \frac{3\psi^3}{2}-\frac{5\psi^2}{2}+1 \\ 2\psi^2(1-\psi) \\ \frac{\psi^3}{2} \\ \frac{\psi^2-\psi^3}{4} \\ \frac{\psi^3}{4} \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{\psi^2}{4} \end{pmatrix}^T, \tag{A17}
$$

$$
r_{\varsigma_{01}^{(2)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} 0 \\ 1-\psi^2 \\ \frac{\psi^2}{2} \\ 0 \\ 0 \\ 0 \\ \frac{\psi^2}{4} \\ 0 \\ 0 \\ \frac{\psi^2}{4} \end{pmatrix}^T\!\!, \quad
r_{\varsigma_{01}^{(3)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} \frac{3\psi^3}{2}-\frac{5\psi^2}{2}+1 \\ 2\psi^2(1-\psi) \\ \frac{\psi^3}{2} \\ \frac{\psi^2-\psi^3}{4} \\ 0 \\ 0 \\ \frac{\psi^3}{4} \\ 0 \\ 0 \\ \frac{\psi^2}{4} \end{pmatrix}^T\!\!, \quad
r_{\varsigma_{00}^{(2)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} \frac{1}{4}-\frac{\psi^2}{4} \\ \frac{1}{2}-\frac{\psi^2}{2} \\ \frac{1}{4} \\ 0 \\ \frac{\psi^2}{8} \\ 0 \\ \frac{\psi^2}{8} \\ \frac{\psi^2}{32} \\ \frac{\psi^2}{32} \\ \frac{7\psi^2}{16} \end{pmatrix}^T\!\!, \tag{A18}
$$

$$
r_{\varsigma_{00}^{(3)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} \psi^3 - \frac{3\psi^2}{2} + \frac{1}{2} \\ \frac{1}{2} - \frac{\psi^3}{2} \\ \frac{\psi^2}{2} - \frac{\psi^3}{4} \\ \frac{\psi^2}{8} - \frac{\psi^3}{8} \\ \frac{\psi^2}{8} \\ \frac{\psi^2}{8} - \frac{\psi^3}{8} \\ \frac{\psi^2}{8} \\ \frac{\psi^3}{32} \\ \frac{\psi^3}{322} \\ \frac{\psi^2}{2} - \frac{\psi^3}{16} \end{pmatrix}^T\!\!, \quad \text{and} \quad
r_{\varsigma_{00}^{(4)}}^{(1)} = \begin{array}{l} \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}\!\!\begin{pmatrix} (2\psi+1)(\psi-1)^2 \\ -2(\psi-1)\psi^2 \\ \frac{\psi^4}{4} \\ \frac{(1-\psi)\psi^2}{4} \\ \frac{(2-\psi)\psi^3}{8} \\ \frac{(1-\psi)\psi^2}{4} \\ \frac{(2-\psi)\psi^3}{8} \\ \frac{\psi^4}{32} \\ \frac{\psi^4}{32} \\ \frac{\psi^2(8-\psi^2)}{16} \end{pmatrix}^T. \tag{A19}
$$

Consider now the matrices $\widetilde{A} := \lim_{N\to\infty} \widetilde{\Pi}_N = \widetilde{A}^{\mathrm{WF}}$ and $\widetilde{B}_N^{(\theta)} := N^\theta(\widetilde{\Pi}_N - \widetilde{A})$. We see from equation (A13) that

$$
\tilde{P} := \lim_{m\to\infty} \widetilde{A}^m = \begin{array}{r} \\ \varsigma_{00}^{(4)} \\ \varsigma_{00}^{(3)} \\ \varsigma_{00}^{(2)} \\ \varsigma_{10}^{(3)} \\ \varsigma_{10}^{(2)} \\ \varsigma_{01}^{(3)} \\ \varsigma_{01}^{(2)} \\ \varsigma_{11}^{(2)} \\ \varsigma_{11}^{(1)} \\ \varsigma_\Delta \end{array}
\begin{array}{c} S \quad \varsigma_{00}^{(4)}\ \varsigma_{00}^{(3)}\ \varsigma_{00}^{(2)}\ \varsigma_{10}^{(3)}\ \varsigma_{10}^{(2)}\ \varsigma_{01}^{(3)}\ \varsigma_{01}^{(2)}\ \varsigma_{11}^{(2)}\ \varsigma_{11}^{(1)}\ \varsigma_\Delta \\
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}. \tag{A20}
$$

In fact, $\widetilde{A}$ is the transition matrix of a Markov chain on $\mathcal{S} = \{\xi_{00}^{(4)}, \xi_{00}^{(3)}, \ldots, \xi_\Delta\}$ with two absorbing states $\xi_{00}^{(4)}$ and ξ_Δ; furthermore, $\xi_{00}^{(4)}$ is reached from every $\xi \in \mathcal{S} \setminus \{\xi_\Delta\}$. Denoting $\widetilde{B}_N^{(\theta)} := N(\widetilde{\Pi}_N - \widetilde{A})$ we can write $\widetilde{\Pi}_N = \widetilde{A} + \frac{1}{N}\widetilde{B}_N^{(\theta)}$, for $\alpha_N = \lambda/N^\theta$ and $\theta \in (0, 1]$. Also let

$$\widetilde{B}^{(\theta)} := \lim_{N\to\infty} \widetilde{B}_N^{(\theta)} = \begin{cases} \lambda(\widetilde{A}^{\mathrm{BF}} - \widetilde{A}^{\mathrm{WF}}), & \text{if } \theta \in (0, 1) \\ \lambda(\widetilde{A}^{\mathrm{BF}} - \widetilde{A}^{\mathrm{WF}}) + \widetilde{B}^{\mathrm{WF}}, & \text{if } \theta = 1, \end{cases}$$

and observe that the first nine rows of

$$\widetilde{G}^{(\theta)} := \lim_{N\to\infty} \widetilde{P}\widetilde{B}_N^{(\theta)}\widetilde{P} = \widetilde{P}\widetilde{B}^{(\theta)}\widetilde{P} \tag{A21}$$

are identical and equal to

$$\left(-1 - \lambda\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right), \ 0, \ 0, \ \ldots, \ 0, \ 1 + \lambda\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)\right),$$

when $\theta = 1$, and equal to

$$\left(-\lambda\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right), \ 0, \ 0, \ \ldots, \ 0, \ \lambda\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)\right),$$

when $\theta \in (0, 1)$. The last row of $G^{(\theta)}$ is equal to $(0, 0, \ldots, 0)$ for all $\theta \in (0, 1]$. Note that for computations involving $\widetilde{G}^{(\theta)}$, we can effectively collapse the state space $\mathcal{S}$ into two points: ξ_Δ and $\mathcal{S} \setminus \{\xi_\Delta\}$. Now we have all the required terms to invoke Möhle (1998a, Lemma 1) for the proof of Lemma A1. We note that $\sup_N \|\widetilde{B}_N^{(\theta)}\| < \infty$, where for any matrix $K = (k_{ij})_{i,j}$, the norm $\|K\| := \max_i \sum_j |k_{ij}|$ is used. The calculations for Lemma A1 are available at https://github.com/diamantidisdimitris/Bursts-of-coalescence in the Lemma_2_second_moment_print file. The computations can be reproduced using the Lemma_2_second_moment_notebook Mathematica notebook, version 13.1.0 (Wolfram Research, Inc. 2022).

Lemma A1 *Let τ and τ' be the coalescent times of the two pairs of genes that are conditionally independent given $\mathcal{A}^{(N,2)}$. Suppose the initial distribution $\vec{p}_0$ of $\widetilde{M}_0$ puts zero mass on state ξ_Δ. Then for all $0 < t \le t'$, as $N \to \infty$,*

$$\mathbb{P}^{(N)}(\tau > [tN^\theta], \ \tau' > [t'N^\theta]) \to$$
$$\begin{cases} \exp\left\{-\lambda t\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right) - \lambda\frac{\psi^2}{4}(t' - t)\right\}, & \text{if } \theta \in (0, 1), \\ \exp\left\{-\left(1 + \lambda\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)\right)t - \left(\frac{1}{2} + \lambda\frac{\psi^2}{4}\right)(t' - t)\right\}, & \text{if } \theta = 1. \end{cases} \tag{A22}$$

Proof. We first establish the result for the case $t = t'$, and then strengthen it to the general case $t \le t'$ by using the Markov property.

Let $\vec{p}_0 = (0, 0, 1, 0, \ldots, 0)$ be the 1×13 row vector representing the distribution of $\widetilde{M}_0$, which corresponds to the initial distribution of the Markov chain. Then for each $N \in \mathbb{N}$ and each generation $g \in \mathbb{Z}_{\ge 0}$,

$$\mathbb{P}^{(N)}(\tau > g, \ \tau' > g) = \vec{p}_0 \widetilde{\Pi}_N^g (1, 1, 1, 1, 1, 1, 1, 1, 1, 0)^T, \tag{A23}$$

where $\widetilde{\Pi}_N$ was defined in equation (A11). Setting $g = [tN^\theta]$ in equation (A23), an application of Möhle (1998a, Lemma 1) gives

$$\lim_{N\to\infty} \vec{p}_0 \widetilde{\Pi}_N^{[tN^\theta]} (1, 1, 1, 1, 1, 1, 1, 1, 1, 0)^T$$
$$= \vec{p}_0 \widetilde{P}e^{t\widetilde{G}} (1, 1, 1, 1, 1, 1, 1, 1, 1, 0)^T. \tag{A24}$$

Substituting $\widetilde{P}$ from equation (A20) and $\widetilde{G}^{(\theta)}$ from equation (A21), the right-hand side of (A24) is equal to

$$\vec{p}_0 \widetilde{P}e^{t\widetilde{G}} (1, 1, 1, 1, 1, 1, 1, 1, 1, 0)^T \tag{A25}$$

$$= \begin{cases} e^{-\lambda t\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)}, & \text{if } \theta \in (0, 1), \\ e^{-t}e^{-\lambda t\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)}, & \text{if } \theta = 1, \end{cases} \tag{A26}$$

for any $\vec{p}_0$ in the set $\mathcal{I} := \{(x_1, \ldots, x_{13}) \in [0, 1]^{13} : \sum_{i=1}^{12} x_i = 1, \ x_{13} = 0\}$. We have shown that

$$\lim_{N\to\infty} \mathbb{P}^{(N)}(\tau > [tN^\theta], \ \tau' > [tN^\theta]) = \begin{cases} e^{-\lambda t\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)}, & \text{if } \theta \in (0, 1), \\ e^{-t}e^{-\lambda t\left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)}, & \text{if } \theta = 1. \end{cases} \tag{A27}$$

For each $N \in \mathbb{N}$ and generations $0 \le g \le g'$, by the Markov property,

$$\mathbb{P}^{(N)}(\tau > g, \ \tau' > g') = \mathbb{E}_{\widetilde{M}_0}^{(N)}[1_{\{\tau > g\}} 1_{\{\tau' > g'\}}]$$
$$= \mathbb{E}_{\widetilde{M}_0}^{(N)}[1_{\{\tau > g\}} 1_{\{\tau' > g\}} \mathbb{P}_{\widetilde{M}_g}^{(N)}(\tau > g' - g)]. \tag{A28}$$

Set $g = [tN^\theta]$ and $g' = [t'N^\theta]$. Note that $\widetilde{M}_g \in \mathcal{I} := \{(c, 1 - c, 0) \in [0, 1]^3 : c \in [0, 1]\}$ when $\tau > g$. Hence, by Remark 1 and Lemma 1,

$$\lim_{N\to\infty} \mathbb{P}_{\widetilde{M}_g}^{(N)}(\tau > g' - g) = \begin{cases} e^{-\lambda\frac{\psi^2}{4}(t' - t)} & \text{when } \theta \in (0, 1) \\ e^{-\left(\frac{1}{2} + \lambda\frac{\psi^2}{4}\right)(t' - t)} & \text{when } \theta = 1. \end{cases}$$

Putting this into equation (A28) and using equation (A27), we obtain the desired equation (A22). $\square$

Supporting lemmas

The second step in the proof of Theorem 1 was the computation of equation (24) for $\theta = 1$ and of (37) for $\theta \in (0, 1)$. An important step in those calculations was to establish equation (33). Lemma A2 below establishes (33). Similarly to equation (16), for each $g \in \mathbb{Z}_{\ge 0}$ we let $F_N(g) := \mathbb{P}_{\mathcal{A}^{(N,2)}}^{(N)}(\tau > g)$ and $G_N(g) := \mathbb{P}_{H_N}^{(N)}(\tau > g)$.

Lemma A2 *Recall the definition of Π_N^{mid} in equation (34). For all $g \in \mathbb{N}$, $\psi \in (0, 1]$ and $N \ge 2$ and initial distribution $\vec{v}_0$ (represented as a 1×3 row vector),*

$$\mathbb{E}^{(N)}\left[G_N(g)\left(1 - \frac{\psi^2}{4}\right)^{H_N(g)}\right] = \vec{v}_0(\Pi_N^{\mathrm{mid}})^g(1, 1, 0)^T.$$

Proof. By the law of total expectation,

$$\mathbb{E}^{(N)}\left[G_N(g)\left(1 - \frac{\psi^2}{4}\right)^{H_N(g)}\right] = \mathbb{E}^{(N)}\left[\mathbb{P}_{H_N}^{(N)}(\tau > g)\left(1 - \frac{\psi^2}{4}\right)^{H_N(g)}\right]$$
$$= \sum_{k=0}^{g} \mathbb{P}^{(N)}(\tau > g \mid H_N(g) = k)\left(1 - \frac{\psi^2}{4}\right)^k$$
$$\mathbb{P}^{(N)}(H_N(g) = k). \tag{A29}$$

Given g generations, k of which being generations with big families, the total number of arrangements of those events in the pedigree is equal to gk. Given k, the arrangement is uniform on

W_k, the space of all possible arrangements of k big families within g generations. Denote each such arrangement by $\vec{t}$. Let $\Pi_N^{\vec{t}}$ be the product of g terms, each term being either Π_N^{WF} or Π_N^{BF}. The order of the terms (from left to right) is determined by the vector $\vec{t}$ so that whenever a Wright–Fisher generation is taking place the corresponding factor in the product is Π_N^{WF} and Π_N^{BF} otherwise. Hence, the conditional probability in equation (A29) is equal to

$$\mathbb{P}^{(N)}(\tau > g \mid H_N(g) = k) = \frac{1}{g k} \sum_{\vec{t} \in W_k} \mathbb{P}^{(N)}(\tau > g \mid \vec{t}) \qquad (A30)$$

$$= \frac{1}{g k} \sum_{\vec{t} \in W_k} \vec{v}_0 \, \Pi_N^{\vec{t}} (1, \ 1, \ 0)^T, \qquad (A31)$$

where equation (A30) is true by the conditional law of total probability and (A31) by the definition of conditional probability. Plugging equation (A31) back into equation (A29), since $H_N(g) \sim \mathrm{Bin}(g, \ \alpha_N)$, it follows

$$\mathbb{E}^{(N)}\left[G_N(g) \left(1 - \frac{\psi^2}{4}\right)^{H_N(g)} \right]$$

$$= \vec{v}_0 \sum_{k=0}^{g} \sum_{\vec{t} \in W_k} \Pi_N^{\vec{t}} \left(\left(1 - \frac{\psi^2}{4}\right) \alpha_N \right)^k (1 - \alpha_N)^{g-k} (1, \ 1, \ 0)^T \qquad (A32)$$

$$= \vec{v}_0 \left(\alpha_N \left(1 - \frac{\psi^2}{4}\right) \Pi_N^{BF} + (1 - \alpha_N) \Pi_N^{WF} \right)^g (1, \ 1, \ 0)^T$$

$$= \vec{v}_0 (\Pi_N^{mid})^g (1, \ 1, \ 0)^T,$$

where the second last equality follows from distributing k terms of α_N and $g - k$ terms of $1 - \alpha_N$ into the product $\Pi_N^{\vec{t}}$. □

Finally, we compute the right-hand side of (33). The calculations for Lemma A3 are available at https://github.com/diamantidisdimitris/Bursts-of-coalescence in the file titled Lemma_4_mixed_moment_print and can be reproduced in the Mathematica notebook Lemma_4_mixed_moment_notebook, where Mathematica version 13.1.0 (Wolfram Research, Inc. 2022) was used.

Lemma A3 *Suppose the distribution $\vec{v}_0$ of the initial state M_0 puts zero mass at the coalescent state ξ_2 (recall 3-states Markov chain M in the proof of Lemma 1). For each $t \in (0, \ \infty)$, it holds that*

$$\lim_{N \to \infty} \mathbb{E}^{(N)}\left[G_N([tN^\theta]) \left(1 - \frac{\psi^2}{4}\right)^{H_N([tN^\theta])} \right] = \begin{cases} e^{-\lambda t \left(\frac{\psi^2}{4} - \frac{\psi^4}{16}\right)}, & \text{if } \theta \in (0, \ 1), \\ e^{-t/2} e^{-\lambda t \left(\frac{\psi^2}{4} - \frac{\psi^4}{16}\right)}, & \text{if } \theta = 1. \end{cases}$$

Proof. Let Π_N^{mid} as in equation (34), by Lemma A2 for $g = [tN^\theta]$ it follows that

$$\mathbb{E}^{(N)}\left[G_N([tN^\theta]) \left(1 - \frac{\psi^2}{4}\right)^{H_N([tN^\theta])} \right] = \vec{v}_0 (\Pi_N^{mid})^{tN^\theta} (1, \ 1, \ 0)^T. \qquad (A33)$$

Notice that Π_N^{mid} is not a stochastic matrix, consider instead

$$\widetilde{\Pi}_N^{mid} := \left(1 - \frac{\lambda \psi^2}{4 N^\theta}\right)^{-1} \Pi_N^{mid},$$

so that $\widetilde{\Pi}_N^{mid}$ is a stochastic matrix. For any $\theta \in (0, \ 1]$, let

$$A := \lim_{N \to \infty} \widetilde{\Pi}_N^{mid} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and

$$B^{(\theta)} := \lim_{N \to \infty} N^\theta (\widetilde{\Pi}_N^{mid} - A) \qquad (A34)$$

$$= \begin{cases} \begin{pmatrix} \frac{\lambda \psi^4}{8} - \frac{\lambda \psi^2}{2} & \frac{\lambda \psi^2}{4} - \frac{\lambda \psi^4}{16} & \frac{\lambda \psi^2}{4} - \frac{\lambda \psi^4}{16} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \text{if } 0 < \theta < 1, \\[2em] \begin{pmatrix} \frac{\lambda \psi^4}{8} - \frac{\lambda \psi^2}{2} - 1 & -\frac{\lambda \psi^4}{16} + \frac{\lambda \psi^2}{4} + \frac{1}{2} & -\frac{\lambda \psi^4}{16} + \frac{\lambda \psi^2}{4} + \frac{1}{2} \\ -1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}, & \text{if } \theta = 1. \end{cases} \qquad (A35)$$

Furthermore, consider the matrix $P := \lim_{k \to \infty} A^k$, then $P = A$ and let $G^{(\theta)} := P B^{(\theta)} P$. Then for $\theta \in (0, \ 1)$, the first two rows of $G^{(\theta)}$ are identical to

$$\left(\frac{1}{16} \lambda \psi^2 (\psi^2 - 4), \ 0, \ -\frac{1}{16} \lambda \psi^2 (\psi^2 - 4) \right),$$

and the last one is equal to $(0, \ 0, \ 0)$. For $\theta = 1$, the first two rows of $G^{(1)}$ are identical to

$$\left(\frac{1}{8} \lambda (\psi^2 - 4) \psi^2 + \frac{1}{16}(8 - \lambda \psi^2 (\psi^2 - 4)) - 1, \ 0, \ \frac{1}{16}(8 - \lambda \psi^2 (\psi^2 - 4)) \right),$$

and the last one is equal to $(0, \ 0, \ 0)$. Writing $\widetilde{\Pi}_N^{mid}$ as

$$\widetilde{\Pi}_N^{mid} = A + \frac{B^{(\theta)}}{N^\theta} + o\left(\frac{1}{N^\theta}\right), \qquad (A36)$$

we can apply Möhle's Lemma (Möhle 1998a, Lemma 1) which gives

$$\lim_{N \to \infty} \vec{v}_0 (\widetilde{\Pi}_N^{mid})^{tN^\theta} (1, \ 1, \ 0)^T = \begin{cases} e^{-\lambda t \left(\frac{\psi^2}{4} - \frac{\psi^4}{16}\right)}, & \text{for } \theta \in (0, \ 1), \\ e^{-t/2} e^{-\lambda t \left(\frac{\psi^2}{4} - \frac{\psi^4}{16}\right)}, & \text{for } \theta = 1. \end{cases} \qquad (A37)$$

Hence, equation (A33), through the use of equation (A37), is equal to

$$\lim_{N \to \infty} \vec{v}_0 (\Pi_N^{mid})^{tN^\theta} (1, \ 1, \ 0)^T = \lim_{N \to \infty} \vec{v}_0 \left(\left(1 - \frac{\lambda \psi^2}{4 N^\theta}\right) \widetilde{\Pi}_N^{mid} \right)^{tN^\theta} (1, \ 1, \ 0)^T$$

$$= e^{-\lambda t \frac{\psi^2}{4}} \lim_{N \to \infty} \vec{v}_0 (\widetilde{\Pi}_N^{mid})^{tN^\theta} (1, \ 1, \ 0)^T$$

$$= \begin{cases} e^{-\lambda t \left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)}, & \text{if } \theta \in (0, \ 1), \\ e^{-t/2} e^{-\lambda t \left(\frac{\psi^2}{2} - \frac{\psi^4}{16}\right)}, & \text{if } \theta = 1. \end{cases}$$

The last equality holds because $\vec{v}_0 = (x, \ 1 - x, \ 0)$ for some $x \in [0, \ 1]$. □

Proof sketch for Remark 3

Let $\vec{G}^{(N)} = (G_1^{(N)}, \ G_2^{(N)}, \ \ldots)$ with $0 \leq G_1^{(N)} < G_2^{(N)} < \cdots$ be the (random) indices of the generations when a big family occurs. By the definition of our model, we can think of generating the random pedigree in a two-step procedure: First, generate $\vec{G}^{(N)}$ by deciding for each generation independently with probability α_N that a big family occurs there. Then given $\vec{G}^{(N)} = \vec{g}^{(N)}$ with $\vec{g}^{(N)} = (g_1^{(N)}, \ g_2^{(N)}, \ \ldots)$ fixed, generate the pedigree between generation g and $g + 1$ by assigning offspring to parents according to the "generation with big family" model if $g \in \{g_1^{(N)}, \ g_2^{(N)}, \ \ldots\}$ and according to the diploid Wright–Fisher model otherwise, independently for different g's.

Thus, $\mathbb{P}^{(N)}(\tau^{(N,2)} > [tN^\theta] \mid \vec{G}^{(N)} = \vec{g}^{(N)})$ can be computed by using a time-inhomogeneous Markov chain with the three states ξ_0, ξ_1 and ξ_2 from Lemma 1, as in equation (12),

$$\mathbb{P}^{(N)}(\tau^{(N,2)} > [tN^\theta] \mid \vec{G}^{(N)} = \vec{g}^{(N)}) = (1,\ 0,\ 0)(\Pi_N^{WF})^{g_1^{(N)}}\Pi_N^{BF}(\Pi_N^{WF})^{g_2^{(N)}-g_1^{(N)}-1}$$

$$\Pi_N^{BF}(\Pi_N^{WF})^{g_3^{(N)}-g_2^{(N)}-1}\Pi_N^{BF}\cdots(\Pi_N^{WF})^{g_{Y^{(N)}([tN^\theta])}^{(N)}-g_{Y^{(N)}([tN^\theta])-1}^{(N)}-1}$$

$$\Pi_N^{BF}(\Pi_N^{WF})^{[tN^\theta]-g_{Y^{(N)}([tN^\theta])}^{(N)}}(1,\ 1,\ 0)^T,$$

where $Y^{(N)}(u) := Y^{(N)}(u; \vec{g}^{(N)}) := \max\{i : g_i^{(N)} \le u\}$.

Now $\vec{G}^{(N)}/N^\theta := (G_1^{(N)}/N^\theta,\ G_2^{(N)}/N^\theta,\ \ldots)$ converges in distribution to $(J_1, J_2, \ldots)$ where $0 < J_1 < J_2 < \cdots$ are the jump times of a Poisson process $(Y(t))_{t\ge0}$ with rate λ (and $Y(t) = \max\{i : J_i \le t\}$).

We see from equation (8) that

$$\Pi^{BF} := \lim_{N\to\infty}\Pi_N^{BF} = \begin{bmatrix} 1-\frac{\psi^2}{2} & \frac{\psi^2}{4} & \frac{\psi^2}{4} \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Furthermore, for $\theta = 1$ we obtain from equation (7) that for any $t > 0$

$$\lim_{N\to\infty}(\Pi_N^{WF})^{[Nt]} = \begin{bmatrix} e^{-t/2} & 0 & 1-e^{-t/2} \\ e^{-t/2} & 0 & 1-e^{-t/2} \\ 0 & 0 & 1 \end{bmatrix} = :\Pi^{WF}(t),$$

whereas for $\theta \in (0,\ 1)$

$$\lim_{N\to\infty}(\Pi_N^{WF})^{[N^\theta t]} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This gives for $\theta = 1$

$$\lim_{N\to\infty}\mathbb{P}^{(N)}(\tau^{(N,2)} > [tN^\theta] \mid \vec{G}^{(N)} = \vec{g}^{(N)}) \stackrel{d}{=} (1,\ 0,\ 0)\Pi^{WF}(J_1)$$

$$\Pi^{BF}\Pi^{WF}(J_2-J_1)\Pi^{BF}\cdots\Pi^{WF}(J_{Y(t)}-J_{Y(t)-1})\Pi^{BF}\Pi^{WF}(t-J_{Y(t)})(1,\ 1,\ 0)^T$$

$$= e^{-t/2}(1-\psi^2/4)^{Y(t)}.$$

The argument for $\theta \in (0,\ 1)$ is analogous.

In fact, the argument above suggests that the model and the main result can be extended to a scenario where the limiting process $(Y(t))$ is not necessarily a time-homogeneous Poisson process as long as the construction of the pedigree conditional on the sequence of times $\vec{G}^{(N)} = (G_1^{(N)},\ G_2^{(N)},\ \ldots)$ is unchanged.

List of Patterns and Mendelian randomness tables
List of patterns

For Tables A1–A3, let x, y, z, w ∈ [N] represent distinct individuals. Circles represent the parent(s) of these individuals. For any individual, two arrows pointing to the same circle indicate

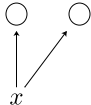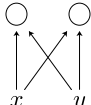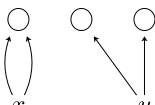**Table A1.** List of patterns when the sampled genes are distributed among one or two individuals.
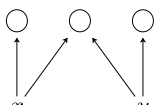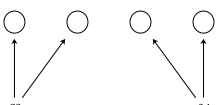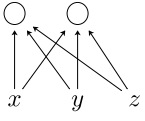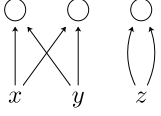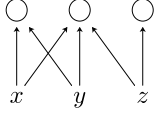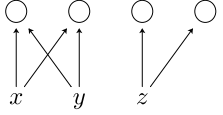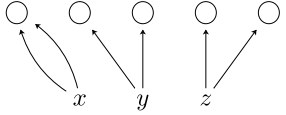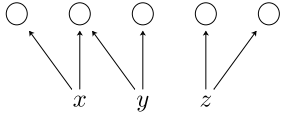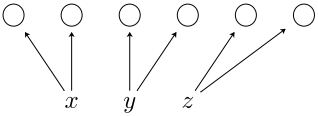
Pattern	Name	Wright–Fisher	Big family
	A_{11}	$\frac{1}{N}$	$\frac{1-\psi}{N}$
	A_{12}	$1-\frac{1}{N}$	$\psi + (1-\psi)\left(1-\frac{1}{N}\right)$
	A_{22}	$O(N^{-2})$	$\psi^2 + O(N^{-2})$
	$A_{23,s}(x,\ y)$	$\frac{(N-1)(N-2)}{N^3}$	$\psi(1-\psi)\frac{1}{N} + (1-\psi)^2\frac{(N-1)(N-2)}{N^3}$
	A_{23}	$\frac{2(N-1)(N-2)}{N^3}$	$2\psi(1-\psi)\frac{2}{N}\left(1-\frac{2}{N}\right) + (1-\psi)^2\frac{2(N-1)(N-2)}{N^3}$
	A_{24}	$\frac{(N-1)(N-2)(N-3)}{N^3}$	$2\psi(1-\psi)\frac{(N-2)(N-3)}{N^2} + (1-\psi)^2\frac{(N-1)(N-2)(N-3)}{N^3}$

Table A2. List of patterns when the sampled genes are distributed among three individuals.
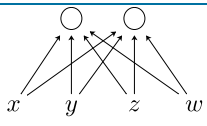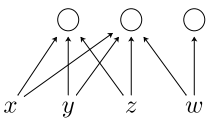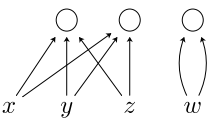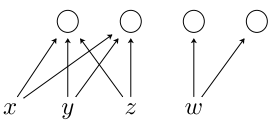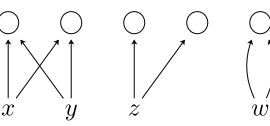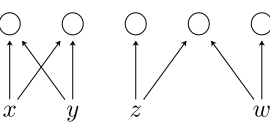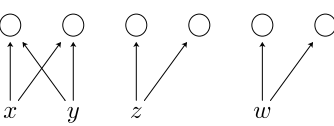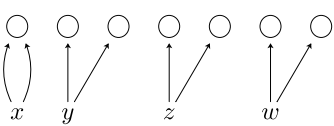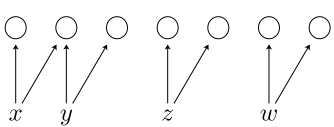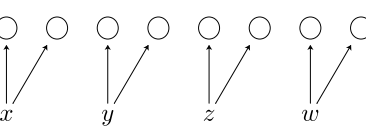
Pattern	Name	Wright–Fisher	Big family
	A_{32}	$O(N^{-2})$	$\psi^3 + O(N^{-2})$
	$A_{33,s}(z)$	$O(N^{-2})$	$\psi^2(1-\psi)(1-\frac{2}{N})\frac{1}{N}$
	$A_{33,2}(z)$	$O(N^{-2})$	$\psi^2(1-\psi)(1-\frac{2}{N})\frac{2}{N}$
	$A_{33}(z)$	$O(N^{-2})$	$\psi^2(1-\psi)(1-\frac{2}{N})(1-\frac{3}{N})$
	$A_{35,s}(x)$	$\frac{(N-1)\cdots(N-4)}{N^5}$	$\psi(1-\psi)^2\frac{(N-2)\cdots(N-4)}{N^4} + (1-\psi)^3\frac{(N-1)\cdots(N-4)}{N^5}$
	$A_{35}(x,\ y)$	$\frac{2(N-1)\cdots(N-4)}{N^5}$	$\psi(1-\psi)^2\frac{2(N-2)\cdots(N-4)}{N^4} + (1-\psi)^3\frac{2(N-1)\cdots(N-4)}{N^5}$
	A_{36}	$\frac{(N-1)\cdots(N-5)}{N^5}$	$3\psi(1-\psi)^2\frac{(N-2)\cdots(N-5)}{N^4} + (1-\psi)^3\frac{(N-1)\cdots(N-5)}{N^5}$

that the individual results from selfing. If two individuals each have one arrow pointing to the same circle, it means that the individuals have exactly one parent in common. The third columns in Tables A1–A3 contain the probabilities of the patterns in a Wright–Fisher generation, while the last column represents a generation with a big family. We only keep track of terms of order $O(1)$ and $O(N^{-1})$.

For example, the probabilities of the patterns A_{11} and A_{12} are explained as follows: In the offspring generation, an individual originates from the highly reproductive pair with probability ψ, signifying two distinct parents. Conversely, the same individual

is *not* among the offspring of the highly reproductive pair with a probability of $1-\psi$. In this scenario, the individual comes from selfing with a probability of $1/N$, and from two distinct parents with probability $1-1/N$. Finally, the probability of pattern A_{22} is explained as follows: Each individual in the offspring generation originates from the highly reproductive pair with probability ψ. By independence, both individuals are offspring of the highly reproductive couple with probability ψ^2. Note that two individuals can share both their parents in various ways. However, in all other scenarios, the probability of such an event is of the order $O(N^{-2})$.

Table A3. List of patterns when the sampled genes are distributed among four individuals.

Pattern	Name	Wright–Fisher	Big family
	A_{42}	$O(N^{-2})$	$\psi^4 + O(N^{-2})$
	$A_{43}(w)$	$O(N^{-2})$	$\psi^3(1-\psi)\frac{2}{N}(1-\frac{2}{N})$
	$A_{43,s}(w)$	$O(N^{-2})$	$\psi^3(1-\psi)(1-\frac{2}{N})\frac{1}{N}$
	$A_{44}(w)$	$O(N^{-2})$	$\psi^3(1-\psi)\frac{(N-2)(N-3)}{N^2}$
	$A_{45,s}(z,\ w)$	$O(N^{-2})$	$\psi^2(1-\psi)^2\frac{(N-2)\cdots(N-4)}{N^4}$
	$A_{45}(z,\ w)$	$O(N^{-2})$	$\psi^2(1-\psi)^2\frac{2(N-2)(N-3)(N-4)}{N^4}$
	$A_{46}(z,\ w)$	$O(N^{-2})$	$\psi^2(1-\psi)^2\frac{(N-2)(N-3)(N-4)(N-5)}{N^4}$
	$A_{47,s}(x)$	$\frac{(N-1)\cdots(N-6)}{N^7}$	$\psi(1-\psi)^3\frac{(N-2)\cdots(N-6)}{N^6}+(1-\psi)^4\frac{(N-1)\cdots(N-6)}{N^7}$
	$A_{47}(x,\ y)$	$\frac{2(N-1)\cdots(N-6)}{N^7}$	$\psi(1-\psi)^3\frac{2(N-2)\cdots(N-6)}{N^6}+(1-\psi)^4\frac{4(N-1)\cdots(N-6)}{N^7}$
	A_{48}	$\frac{(N-1)\cdots(N-7)}{N^7}$	$4\psi(1-\psi)^3\frac{(N-2)\cdots(N-7)}{N^6}+(1-\psi)^4\frac{(N-1)\cdots(N-7)}{N^7}$

Mendelian randomness tables

Tables A4–A10 contain the transition probabilities for every pair of states $(\xi, \eta) \in \mathcal{S}^2$, conditional on the patterns from Tables A1–A3. The derivation of the entries of Tables A4–A10 follow similarly to the entries of equation (A10) in the *Two extremes for the sizes of big families*. For example, in Table A10, the last row of the column labeled A_{47} is explained as follows. Lineages in distinct individuals sharing exactly one parent have descended from different parents with probability 3/4. On the other hand, state $\xi_{00}^{(4)}$ consists of two distinct individuals, hence the pattern A_{47} is used in $42 = 6$ ways. Finally, we note that in Tables A4–A10, the fractional quantities stand for events determined purely by Mendelian randomness (conditional on each pattern) whereas the integer quantities refer to combinatorial terms corresponding to the appropriate use of the patterns.

Table A4. Conditional probabilities given the pattern when the initial state is $\xi_{11}^{(1)}$.

$(\xi_{11}^{(1)}, \eta)$	A_{11}	A_{12}
$(\xi_{11}^{(1)}, \xi_{11}^{(1)})$	$\frac{1}{4}$	0
$(\xi_{11}^{(1)}, \xi_{11}^{(2)})$	0	0
$(\xi_{11}^{(1)}, \xi_{10}^{(2)})$	0	0
$(\xi_{11}^{(1)}, \xi_{10}^{(3)})$	0	0
$(\xi_{11}^{(1)}, \xi_{01}^{(2)})$	0	0
$(\xi_{11}^{(1)}, \xi_{01}^{(3)})$	0	0
$(\xi_{11}^{(1)}, \xi_{00}^{(2)})$	0	$\frac{1}{1}$
$(\xi_{11}^{(1)}, \xi_{00}^{(3)})$	0	0
$(\xi_{11}^{(1)}, \xi_{00}^{(4)})$	0	0

Table A5. Conditional probabilities given the pattern when the initial state is $\xi_{11}^{(2)}$.

$(\xi_{11}^{(2)}, \eta)$	A_{22}	$A_{23,s}$	A_{23}	A_{24}
$(\xi_{11}^{(2)}, \xi_{11}^{(1)})$	0	0	0	0
$(\xi_{11}^{(2)}, \xi_{11}^{(2)})$	0	0	0	0
$(\xi_{11}^{(2)}, \xi_{10}^{(2)})$	0	0	0	0
$(\xi_{11}^{(2)}, \xi_{10}^{(3)})$	0	$\frac{1}{2}$	0	0
$(\xi_{11}^{(2)}, \xi_{01}^{(2)})$	0	0	0	0
$(\xi_{11}^{(2)}, \xi_{01}^{(3)})$	0	$\frac{1}{2}$	0	0
$(\xi_{11}^{(2)}, \xi_{00}^{(2)})$	$\frac{1}{1}$	0	0	0
$(\xi_{11}^{(2)}, \xi_{00}^{(3)})$	0	0	$\frac{1}{1}$	0
$(\xi_{11}^{(2)}, \xi_{00}^{(4)})$	0	0	0	$\frac{1}{1}$

Table A6. Conditional probabilities given the pattern when the initial state is $\xi_{10}^{(2)}$.

$(\xi_{10}^{(2)}, \eta)$	A_{22}	$A_{23,s}$	A_{23}	A_{24}
$(\xi_{10}^{(2)}, \xi_{11}^{(1)})$	0	0	0	0
$(\xi_{10}^{(2)}, \xi_{11}^{(2)})$	0	0	0	0
$(\xi_{10}^{(2)}, \xi_{10}^{(2)})$	0	$\frac{1}{2}$	0	0
$(\xi_{10}^{(2)}, \xi_{10}^{(3)})$	0	0	0	0
$(\xi_{10}^{(2)}, \xi_{01}^{(2)})$	$\frac{1}{4}$	0	$\frac{1}{4}$	0
$(\xi_{10}^{(2)}, \xi_{01}^{(3)})$	0	0	0	0
$(\xi_{10}^{(2)}, \xi_{00}^{(2)})$	$\frac{1}{2}$	0	$\frac{1}{4}$	0
$(\xi_{10}^{(2)}, \xi_{00}^{(3)})$	0	$1 \cdot \frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{1}$
$(\xi_{10}^{(2)}, \xi_{00}^{(4)})$	0	0	0	0

Table A7. Conditional probabilities given the pattern when the initial state is $\xi_{00}^{(2)}$.

$(\xi_{00}^{(2)}, \eta)$	A_{22}	$A_{23,s}$	A_{23}	A_{24}
$(\xi_{00}^{(2)}, \xi_{11}^{(1)})$	$\frac{1}{32}$	0	$\frac{1}{64}$	0
$(\xi_{00}^{(2)}, \xi_{11}^{(2)})$	$\frac{1}{32}$	0	0	0
$(\xi_{00}^{(2)}, \xi_{10}^{(2)})$	$\frac{1}{8}$	0	$\frac{1}{16}$	0
$(\xi_{00}^{(2)}, \xi_{10}^{(3)})$	0	0	$\frac{1}{32}$	0
$(\xi_{00}^{(2)}, \xi_{01}^{(2)})$	$\frac{1}{8}$	0	$\frac{1}{16}$	0
$(\xi_{00}^{(2)}, \xi_{01}^{(3)})$	0	0	$\frac{1}{32}$	0
$(\xi_{00}^{(2)}, \xi_{00}^{(2)})$	$\frac{1}{4}$	$2 \cdot \frac{1}{2}$	$1 \cdot \frac{3}{16}$	$\frac{1}{4}$
$(\xi_{00}^{(2)}, \xi_{00}^{(3)})$	0	$2 \cdot \frac{1}{2}$	$1 \cdot \frac{6}{16}$	$\frac{1}{2}$
$(\xi_{00}^{(2)}, \xi_{00}^{(4)})$	0	0	0	$\frac{1}{4}$

Table A8. Conditional probabilities given the pattern when the initial state is $\xi_{10}^{(3)}$.

$(\xi_{10}^{(3)}, \eta)$	A_{32}	$A_{33,s}$	$A_{33,2}$	A_{33}	$A_{35,s}$	A_{35}	A_{36}
$(\xi_{10}^{(3)}, \xi_{11}^{(1)})$	0	0	0	0	0	0	0
$(\xi_{10}^{(3)}, \xi_{11}^{(2)})$	0	$\frac{1}{4}$	0	0	0	0	0
$(\xi_{10}^{(3)}, \xi_{10}^{(2)})$	0	0	0	0	0	0	0
$(\xi_{10}^{(3)}, \xi_{10}^{(3)})$	0	$\frac{1}{4}$	0	0	$1 \cdot \frac{1}{2}$	0	0
$(\xi_{10}^{(3)}, \xi_{01}^{(2)})$	$\frac{1}{4}$	0	$1 \cdot \frac{1}{8}$	0	0	0	0
$(\xi_{10}^{(3)}, \xi_{01}^{(3)})$	0	0	$3 \cdot \frac{1}{8}$	$1 \cdot \frac{1}{4}$	0	$1 \cdot \frac{1}{8}$	0
$(\xi_{10}^{(3)}, \xi_{00}^{(2)})$	$\frac{1}{2}$	0	$2 \cdot \frac{1}{4}$	0	0	0	0
$(\xi_{10}^{(3)}, \xi_{00}^{(3)})$	0	$2 \cdot 1$	$3 \cdot \frac{1}{2}$	$2 \cdot 1$	0	$2 \cdot \frac{1}{2}$	0
$(\xi_{10}^{(3)}, \xi_{00}^{(4)})$	0	0	0	$1 \cdot \frac{1}{2}$	$2 \cdot 1$	$1 \cdot \frac{3}{4} + 2 \cdot \frac{1}{2}$	$\frac{1}{1}$

Table A9. Conditional probabilities given the pattern when the initial state is $\xi_{00}^{(3)}$.

$(\xi_{00}^{(3)}, \eta)$	A_{32}	$A_{33,s}$	$A_{33,2}$	A_{33}	$A_{35,s}$	A_{35}	A_{36}
$(\xi_{00}^{(3)}, \xi_{11}^{(1)})$	$1 \cdot \frac{1}{32}$	0	$3 \cdot \frac{1}{64}$	0	0	0	0
$(\xi_{00}^{(3)}, \xi_{11}^{(2)})$	$1 \cdot \frac{1}{32}$	0	$2 \cdot \frac{1}{64}$	0	0	0	0
$(\xi_{00}^{(3)}, \xi_{10}^{(2)})$	$1 \cdot \frac{1}{8}$	$1 \cdot \frac{1}{8}$	$2 \cdot \frac{1}{8} + 1 \cdot \frac{1}{16}$	$1 \cdot \frac{1}{8}$	0	$1 \cdot \frac{1}{8}$	0
$(\xi_{00}^{(3)}, \xi_{10}^{(3)})$	0	$1 \cdot \frac{1}{8}$	$1 \cdot \frac{1}{32} + 1 \cdot \frac{1}{16}$	$1 \cdot \frac{1}{8}$	0	$1 \cdot \frac{1}{8}$	0
$(\xi_{00}^{(3)}, \xi_{01}^{(2)})$	$1 \cdot \frac{1}{8}$	$1 \cdot \frac{1}{8}$	$2 \cdot \frac{1}{16} + 1 \cdot \frac{1}{8}$	$1 \cdot \frac{1}{8}$	0	$1 \cdot \frac{1}{8}$	0
$(\xi_{00}^{(3)}, \xi_{01}^{(3)})$	0	$1 \cdot \frac{1}{8}$	$1 \cdot \frac{1}{32} + 1 \cdot \frac{1}{16}$	$1 \cdot \frac{1}{8}$	0	$1 \cdot \frac{1}{8}$	0
$(\xi_{00}^{(3)}, \xi_{00}^{(2)})$	$1 \cdot \frac{1}{4}$	$1 \cdot \frac{1}{2}$	$1 \cdot \frac{3}{8} + 2 \cdot \frac{1}{8}$	$1 \cdot \frac{1}{2}$	0	0	0
$(\xi_{00}^{(3)}, \xi_{00}^{(3)})$	0	$3 \cdot \frac{1}{2}$	$1 \cdot \frac{3}{8} + 2 \cdot \frac{1}{4}$	$3 \cdot \frac{1}{2}$	$1 \cdot 1 + 2 \cdot \frac{1}{2}$	$1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2}$	$1 \cdot \frac{1}{2}$
$(\xi_{00}^{(3)}, \xi_{00}^{(4)})$	0	0	0	0	$2 \cdot \frac{1}{2}$	$1 \cdot \frac{3}{8} + 2 \cdot \frac{1}{4}$	$1 \cdot \frac{1}{2}$

Table A10. Conditional probabilities given the pattern when the initial state is $\zeta_{00}^{(4)}$.

$(\zeta_{00}^{(4)},\ \eta)$	A_{42}	A_{43}	$A_{43,s}$	A_{44}	A_{45}	$A_{45,s}$	A_{46}	$A_{47,s}$	A_{47}	A_{48}
$(\zeta_{00}^{(4)},\ \zeta_{11}^{(1)})$	$1\cdot\frac{1}{32}$	$2\cdot\frac{1}{64}$	0	0	0	0	0	0	0	0
$(\zeta_{00}^{(4)},\ \zeta_{11}^{(2)})$	$1\cdot\frac{1}{32}$	$2\cdot\frac{1}{64}$	0	0	$2\cdot\frac{1}{16}$	0	0	0	0	0
$(\zeta_{00}^{(4)},\ \zeta_{10}^{(2)})$	$1\cdot\frac{1}{8}$	$2\cdot\frac{1}{16}$	$2\cdot\frac{1}{16}$	$2\cdot\frac{1}{16}$	0	0	0	0	0	0
$(\zeta_{00}^{(4)},\ \zeta_{10}^{(3)})$	0	$2\cdot\frac{1}{32}$	$2\cdot\frac{1}{16}$	$2\cdot\frac{1}{16}$	$1\cdot\frac{1}{16}+1\cdot\frac{3}{16}$	$1\cdot\frac{1}{4}$	$1\cdot\frac{1}{4}$	0	$1\cdot\frac{1}{8}$	0
$(\zeta_{00}^{(4)},\ \zeta_{01}^{(2)})$	$1\cdot\frac{1}{8}$	$2\cdot\frac{1}{16}$	$2\cdot\frac{1}{16}$	$2\cdot\frac{1}{16}$	0	0	0	0	0	0
$(\zeta_{00}^{(4)},\ \zeta_{01}^{(3)})$	0	$2\cdot\frac{1}{32}$	$2\cdot\frac{1}{16}$	$2\cdot\frac{1}{16}$	$1\cdot\frac{1}{16}+1\cdot\frac{3}{16}$	$1\cdot\frac{1}{4}$	$1\cdot\frac{1}{4}$	0	$1\cdot\frac{1}{8}$	0
$(\zeta_{00}^{(4)},\ \zeta_{00}^{(2)})$	$1\cdot\frac{1}{4}$	$2\cdot\frac{1}{16}$	0	0	$1\cdot\frac{1}{8}$	0	0	0	0	0
$(\zeta_{00}^{(4)},\ \zeta_{00}^{(3)})$	0	$2\cdot\frac{1}{16}$	$4\cdot\frac{4}{8}$	$4\cdot\frac{1}{2}$	$1\cdot\frac{8}{16}$	$2\cdot\frac{1}{2}$	$4\cdot\frac{1}{2}$	0	$4\cdot\frac{1}{4}$	0
$(\zeta_{00}^{(4)},\ \zeta_{00}^{(4)})$	0	0	0	0	$6\cdot\frac{3}{8}$	$6\cdot\frac{1}{2}$	$6\cdot\frac{1}{2}$	$4\cdot\frac{1}{1}$	$6\cdot\frac{3}{4}$	$1\cdot\frac{1}{1}$

Strengthening the convergence in Theorem 1

Remark A1 The arguments used in the proof of Theorem 1 in fact allow to conclude the (stronger) statement that

$$
(F_N(t,\ \mathcal{A}^{(N,2)}))_{t\geq 0} \to
\begin{cases}
\left(\left(1-\frac{\psi^2}{4}\right)^{Y(t-)}\right)_{t\geq 0}, & \text{when } \theta\in(0,1), \\[2ex]
\left(e^{-t/2}\left(1-\frac{\psi^2}{4}\right)^{Y(t-)}\right)_{t\geq 0}, & \text{when } \theta=1,
\end{cases}
\tag{A38}
$$

in finite-dimensional distributions.

Proof. We will only consider the case $\theta=1$ here, the case $\theta<1$ is similar.

Fix $k\in\mathbb{N}$ and $0<t_1<t_2<\cdots<t_k$. Using equation (21) for $t_1,\ \ldots,\ t_k$, we have

$$
\lim_{N\to\infty}\mathbb{E}^{(N)}\left[\sum_{j=1}^{k}\left|F_N(t_j,\ \mathcal{A}^{(N,2)}) - e^{-t_j/2}\left(1-\frac{\psi^2}{4}\right)^{Q_N([t_jN])}\right|^2\right]=0,
\tag{A39}
$$

which implies as in the proof of Theorem 1 that

$$
(F_N(t_1,\ \mathcal{A}^{(N,2)}),\ \ldots,\ F_N(t_k,\ \mathcal{A}^{(N,2)}))
$$

$$
\to\left(e^{-t_1/2}\left(1-\frac{\psi^2}{4}\right)^{Y(t_1)},\ \ldots,\ e^{-t_k/2}\left(1-\frac{\psi^2}{4}\right)^{Y(t_k)}\right)
$$

$$
\overset{d}{=}\left(e^{-t_1/2}\left(1-\frac{\psi^2}{4}\right)^{Y(t_1-)},\ \ldots,\ e^{-t_k/2}\left(1-\frac{\psi^2}{4}\right)^{Y(t_k-)}\right)
$$

in distribution. This shows convergence of finite-dimensional distributions. □

Furthermore, we conjecture that we have weak convergence in the space of left-continuous functions with right limits, equipped with the Skorohod topology, but we do not pursue this here.

Editor: K. Jain