Theoretical Population Biology 115 (2017) 1–12

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Population structure and coalescence in pedigrees: Comparisons to the structured coalescent and a framework for inference

Peter R. Wilton [a,*], Pierre Baduel [a,c], Matthieu M. Landon [b,c], John Wakeley [a]

[a] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, 02138, USA
[b] Department of Systems Biology, Harvard University, Cambridge, MA, 02138, USA
[c] École des Mines de Paris, Mines ParisTech, Paris 75272, France

ARTICLE INFO

Article history:
Received 23 May 2016
Available online 28 January 2017

Keywords:
Pedigree
Coalescent theory
Identity-by-descent
Demographic inference

ABSTRACT

Contrary to what is often assumed in population genetics, independently segregating loci do not have completely independent ancestries, since all loci are inherited through a single, shared population pedigree. Previous work has shown that the non-independence between gene genealogies of independently segregating loci created by the population pedigree is weak in panmictic populations, and predictions made from standard coalescent theory are accurate for populations that are at least moderately sized. Here, we investigate patterns of coalescence in pedigrees of structured populations. We find that the pedigree creates deviations away from the predictions of the structured coalescent that persist on a longer timescale than in the case of panmictic populations. Nevertheless, we find that the structured coalescent provides a reasonable approximation for the coalescent process in structured population pedigrees so long as migration events are moderately frequent and there are no migration events in the recent pedigree of the sample. When there are migration events in the recent sample pedigree, we find that distributions of coalescence in the sample can be modeled as a mixture of distributions from different initial sample configurations. We use this observation to motivate a maximum-likelihood approach for inferring migration rates and mutation rates jointly with features of the pedigree such as recent migrant ancestry and recent relatedness. Using simulation, we show that our inference framework accurately recovers long-term migration rates in the presence of recent migration events in the sample pedigree.

1. Introduction

The coalescent is a stochastic process that describes the distribution of gene genealogies, the tree-like structures that describe relationships among the sampled copies of a gene. Since its introduction (Kingman, 1982a,b; Hudson, 1983; Tajima, 1983), the coalescent has been extended and applied to numerous contexts in population genetics and is now one of the foremost mathematical tools for modeling genetic variation in samples (Hein et al., 2005; Wakeley, 2009).

In a typical application to data from diploid sexual organisms, the coalescent is applied to multiple loci that are assumed to have entirely independent ancestries because they are found on different chromosomes and thus segregate independently, or are far enough apart along a single chromosome that they effectively segregate independently. Even chromosome-scale coalescent-based inference methods that account for linkage and recombination (e.g., Li and Durbin, 2011; Sheehan et al., 2013; Schiffels and Durbin, 2014) multiply probabilities across distinct chromosomes that are assumed to have completely independent histories due to their independent segregation.

In reality, the ancestries of gene copies sampled at independently segregating loci from a fixed set of diploid, sexually reproducing individuals are independent only after conditioning on the population pedigree, i.e., the set of familial relationships between all individuals in the population throughout all time. Fig. 1 illustrates this. Fig. 1(A) depicts coalescence within the framework of standard coalescent theory under a diploid, monoecious Wright–Fisher model with the possibility of selfing. Under this model, from which the coalescent can be derived as a limiting process, the probability that two distinct ancestral lineages coalesce in a given generation is $1/(2N)$, where N is the diploid population size. The averaging over pedigrees in this model is depicted in Fig. 1(A) by each individual having a "parental" relationship to every individual in the previous generation. Thus, two ancestral lineages follow parental relationships to the same individual with

---

* Corresponding author.
E-mail address: pwilton@fas.harvard.edu (P.R. Wilton).

http://dx.doi.org/10.1016/j.tpb.2017.01.004
0040-5809/

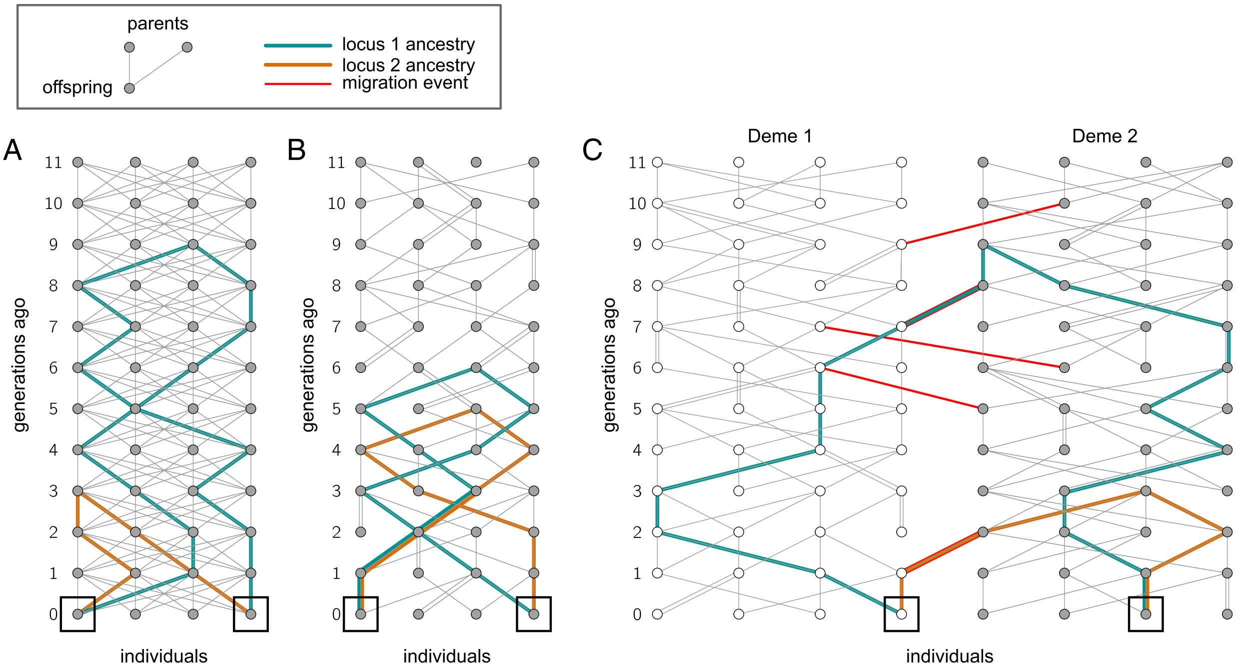

Fig. 1. A conceptual inconsistency in the application of standard coalescent theory to independently segregating loci sampled from a diploid population. Each panel depicts coalescence between two sampled gene copies at a pair of unlinked loci (green and orange lines, respectively) sampled from a pair of individuals in the present generation (black squares). Individuals within a generation are represented by circles arranged horizontally across a row, and generations are arranged vertically, with the present generation at the bottom of the figure and the most ancient generation at the top of the figure. Panel (A) is a conceptual depiction of coalescence at unlinked loci under the assumptions of standard coalescent theory. Under these assumptions, at each unlinked locus, an ancestral lineage has an equal probability of being derived from any individual in the previous generation, and thus it is as if each individual has every potential ancestral individual as a parent. Panels (B) and (C) depict coalescence at two unlinked loci in actual diploid sexual populations, in which the same population pedigree governs the process of coalescence at all loci. In these pedigrees, each individual has exactly two parents (including the possibility of selfing, here), and the distribution of coalescence times amongst unlinked loci depends on both the initially sampled individuals and the particular shape of the pedigree. Panel (B) shows the pedigree of a panmictic population. The probability of coalescence of two ancestral lineages in a given generation depends on the pedigree and generally differs from the $1/(2N)$ that would be assumed after marginalizing over the pedigree. Panel (C) depicts a two-deme population pedigree, with fixed migration events (red lines) constraining the movement of ancestral lineages between the two demes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

probability $1/N$, and in that individual are derived from the same parental chromosome with probability $1/2$. When this probability, $1/(2N)$, is applied independently to unlinked loci, any correlation in ancestry between these loci caused by the pedigree is erased by this averaging over pedigrees.

A conceptually more correct model of coalescence of unlinked loci sampled from individuals in a diploid sexual population is depicted in Fig. 1(B). Here, coalescence occurs within a single diploid population pedigree, with probabilities of coalescence in a given generation determined by the particular shape of the pedigree of the sampled individuals and generally differing from the $1/(2N)$ predicted under the standard interpretation of the diploid Wright–Fisher model. For example, if in a given generation the number of shared parental relationships happens to be greater than usual, the probability of coalescence will tend to be greater, and when the number of shared parental relationships in a given generation is fewer, the probability of coalescence in that generation will tend to be lesser.

These effects of the population pedigree on coalescence in sexual populations were investigated by Wakeley et al. (2012), who studied coalescence in randomly generated population pedigrees of diploid populations reproducing under basic Wright–Fisher-like dynamics, *i.e.*, populations with constant population size, random mating, non-overlapping generations, and no population structure, and in this context, it was found that the shape of the population pedigree affects coalescence probabilities mostly during the first $\sim \log_2(N)$ generations back in time. During these first few generations back in time, there is relatively large variation from pedigree to pedigree in the degree of overlap in the pedigree of the sampled individuals. It may be that for several generations there is zero overlap in the pedigree of the sampled individuals, and consequently the probability of coalescence amongst these individuals will be zero. On the other hand, if by chance the sample contains close relatives (i.e., two individuals possessing a shared pedigree ancestor in these recent generations), the probability of coalescence in the generation of that shared ancestor will tend to exceed the probability of coalescence predicted by standard theory. After the first $\log_2(N)$ generations, the pedigrees of different individuals begin to overlap more completely, and the probabilities of coalescence in these later generations depend less on the pedigree (Wakeley et al., 2012, see also Fig. 2).

This $\log_2(N)$ timescale of convergence or mixing of the pedigree has been studied in other contexts. Chang (1999) found that the number of generations until two individuals share an ancestor (in the biparental pedigree sense) converges to $\log_2(N)$ as the population size grows. Likewise, Derrida et al. (2000) showed that the distribution of the number of repetitions in an individual's pedigree ancestry becomes stationary around $\log_2(N)$ generations in the past. This $\log_2(N)$-generation timescale is the natural timescale of convergence in pedigrees due to the approximate doubling of the number of possible ancestors each generation back in time until the entire ancestral population potentially becomes part of the pedigree.

In these studies it is assumed that the population is panmictic, *i.e.*, that individuals mate with each other uniformly at random. One phenomenon that may alter this convergence in pedigrees is population structure, with migration between subpopulations or demes. In a subdivided population, the exchange of ancestry between demes depends on the history of migration events embedded in the population pedigree (Fig. 1(C)). These fixed past migration events may be infrequent or irregular enough that the generation-by-generation probabilities of coalescence depend on