

# The existence and abundance of ghost ancestors in biparental populations



Simon Gravel<sup>a,\*</sup>, Mike Steel<sup>b</sup>

<sup>a</sup> Genome Quebec Innovation Centre, Department of Human Genetics, McGill University, 740 Dr Penfield, Montreal, Quebec, H3A 0G1, Canada

<sup>b</sup> Biomathematics Research Centre, University of Canterbury, New Zealand

## ARTICLE INFO

### Article history:

Received 6 February 2014

Available online 19 February 2015

### Keywords:

Genealogy

Recombination

Genetic drift

Wright–Fisher model

## ABSTRACT

In a randomly-mating biparental population of size  $N$  there are, with high probability, individuals who are genealogical ancestors of every extant individual within approximately  $\log_2(N)$  generations into the past. We use this result of J. Chang to prove a curious corollary under standard models of recombination: there exist, with high probability, individuals within a constant multiple of  $\log_2(N)$  generations into the past who are simultaneously (i) genealogical ancestors of *each* of the individuals at the present, and (ii) genetic ancestors to *none* of the individuals at the present. Such ancestral individuals – ancestors of everyone today that left no genetic trace – represent ‘ghost’ ancestors in a strong sense. In this short note, we use simple analytical argument and simulations to estimate how many such individuals exist in finite Wright–Fisher populations.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

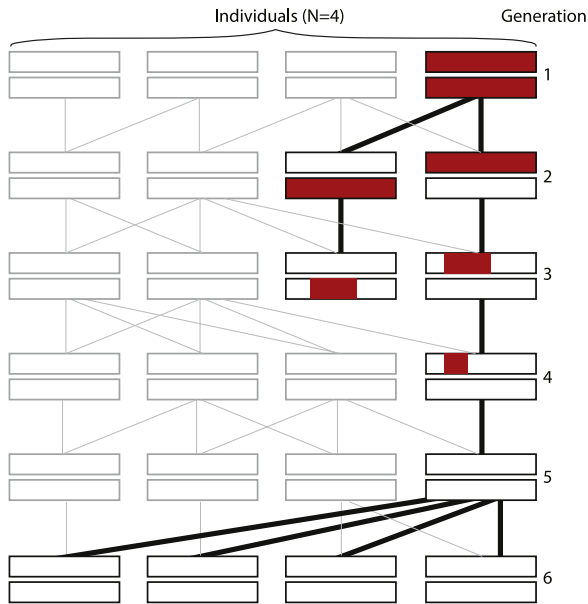
The reproductive success of an individual is traditionally measured by the number of offspring it produces. However, even given a fixed number of descendants, there is a fair amount of variation in the amount of genetic material left by an individual. An ancestor that contributes genetic material to an individual through multiple lineages is expected to leave more genetic material than an ancestor that contributes along a single lineage, for example (Baird et al., 2003; Barton and Etheridge, 2011; Lachance, 2009; Manrubia et al., 2003; Matsen and Evans, 2008). Even if we fix the genealogy, there remains considerable stochasticity in the amount of genetic material left by each ancestor during meiosis. The probability of inheriting genetic material along a given ancestral lineage decreases rapidly with the number of generations. Thus, an individual leaving a single descendant after more than 10 generations is likely to leave no genealogical material in future generations, even if this descendant is wildly prolific. This limits the information about past genealogies that can be recovered from present-day genomic data (Thatte, 2013; Thatte and Steel, 2008). Not only are we limited to individuals that left descendants, but even the genomes of individuals that left offspring may be completely inaccessible.

A result of Chang (Chang, 1999; Donnelly et al., 1999) states that first-generation (or *founding*) individuals of a constant-sized, randomly mating Wright–Fisher diploid population rapidly reach one of two ancestral states: about 80% become ancestral to the complete subsequent population, and 20% leave no descendants. We are interested in the 80% of ‘successful’ individuals, and wonder about the proportion of such individuals that leave no genetic material to any of their large number of descendants. The question of whether genealogical ancestors leave genetic material in an infinite population was discussed in detail in Baird et al. (2003), where a branching approximation to the infinite-population Wright–Fisher model was used to show a logarithmic decrease of survival probability with time. Here we consider the finite population case and obtain analytical and numerical results under the exact Wright–Fisher model. Because the approximations of Baird et al. (2003) are reasonable, we expect that our results should agree in the large- $N$  and long-time limits, but our results will provide additional insight for small population and short times.

We first formally show that, amongst those individuals who are genealogically ancestral to the complete population, the proportion that leaves no genetic material at the present is nonzero in a finite population, and indeed approaches 1 when the number of generations and population size become large. We turn to simulation to estimate their abundance for finite genomes and finite number of generations. We find that the proportion of ghost ancestors grows approximately logarithmically with the number of generations in a constant-size population and that convergence to the large-population limits occurs rapidly at short time-scales.

\* Corresponding author.

E-mail address: [simon.gravel@mcgill.ca](mailto:simon.gravel@mcgill.ca) (S. Gravel).



**Fig. 1.** In this schematic example, the rightmost individual in generation 1 has genealogical descendants until generation 6, but left no genetic material in generations 5 and 6. By generation 5 it has become a ghost; by generation 6 it is a 'super-ghost', in that it is now ancestral to the entire population.

Our work complements recent efforts to study the relationship between genealogical and genetic ancestries. Wakeley et al. (2012) recently proposed an improved approach to model the recent pedigree structure of samples within coalescent theory, but found that this improved modeling had a relatively modest effect on the simulated statistics. Wiuf and Hein (1997) used mathematical modeling and simulations to study the distribution of ancestral material of an extant chromosome, with the aim of addressing two questions: (i) how many ancestors are there to a present human chromosome? and (ii) how many different sequences in an ancestral population can one sample by sequencing extant sequences? More recently, Matsen and Evans (2008) used simulations and probabilistic analysis to investigate relationships between the number of descendant alleles of an ancestor allele and the number of genealogical descendants of the individual who possessed that allele. Here we explore the impact of demography on the short- and long-term probability that an individual with many genealogical descendants leaves no genetic material. We provide a short formal proof for the existence of such individuals, in a probabilistic sense, and study their prevalence under different demographic scenarios with Wright–Fisher reproduction.

## 2. Ghost and super-ghost ancestors

Given a population at the present, an ancestral individual  $I$  is said to be a *ghost* ancestor if:

- (i)  $I$  is the genealogical ancestor of at least one individual at the present, and
- (ii)  $I$  contributes nothing genetically to any individual at the present.

A *super-ghost* replaces condition (i) with the stronger condition:

- (i')  $I$  is the genealogical ancestor of *all* individuals at the present.

A schematic example of a super-ghost is shown in Fig. 1.

### 2.1. The existence of super-ancestors: Chang's result

Consider the genealogical ancestry of a randomly-mating biparental population under the neutral Wright–Fisher model. In this model, generations are discrete, and each individual selects two parents uniformly at random from the previous generation, and this process is extended back in time independently from generation to generation. This generates a random genealogy for the population.

Because each individual has on average two offspring per generation, we would expect individuals to leave a large number of offspring if their descending lineage survives the stochasticity of the first few generations. In fact, the probability of eventually becoming ancestral to the *complete* population in a randomly mating population of size  $N$  converges to  $1 - \rho \approx 0.7968$  for large  $N$ , where  $\rho$  satisfies  $\rho = e^{-2+2\rho}$  (Chang, 1999) (see also Derrida et al., 1999, 2000a,b). Because the expected number of descendants of successful individuals initially grows close to exponentially, we would expect that it must take at least  $\log_2(N)$  generations to become ancestral to the complete population. In fact, Chang (1999) established the following two tight asymptotic results that are key to this paper:

- (a) the number of generations  $G_1$  required to find the first super-ancestor approaches  $\log_2(N)$ , in the sense that  $\frac{G_1}{\log_2(N)}$  converges in probability to 1 as  $N$  grows; and
- (b) the number of generations  $G_a$  until *all* individuals that left descendants are super-ancestors approaches  $(1 + \zeta) \log_2(N)$ , where  $\zeta \approx 0.7698$ , in the sense that  $\frac{G_a}{(1+\zeta)\log_2(N)}$  converges in probability to 1 as  $N$  grows.

### 2.2. The existence of super-ghosts

We consider again the genealogical model from Chang (1999), and model the transmission of genetic material. Given a genealogy, each individual receives one copy of each chromosome from each parent. An individual transmits chromosomes by recombining the two chromosome copies it has inherited. We model recombination as a Markovian copying process along the chromosome length: starting from one edge, the transmitted chromosome is generated by copying one of the two inherited chromosomal copies, with a transition rate between the two of 1 per Morgan.

**Proposition 2.1.** Under the random biparental genealogical model with discrete generations of size  $N$ , the probability that there is a super-ghost ancestor within  $c \cdot \log_2(N)$  generations from the present converges to 1 as  $N$  grows, for any  $c > 1$ .

The formal proof of this proposition is provided in the Appendix, and here we provide an informal argument for the existence of super-ghosts and logarithmic dependence on  $N$ . The proof has two steps: first, we show that ghosts exist, i.e., that there are individuals that leave genealogical descendants but no genetic material. Second, we show that the descendants are likely to become ancestral to the entire population, ensuring that the ancestral individual is now a super-ghost. To estimate the timing it is convenient to proceed back in time. In a Wright–Fisher model, the probability that an individual is the unique descendant to an individual living  $G$  generations is nonzero. It can be easily calculated through dynamic programming, and is trivially bounded below by  $P(1)^G$ , where  $P(1)$  is the probability of being an only child. In this particular case of a single descending lineage, we can also easily show that the probability of leaving no genetic material becomes nonzero in any reasonable model of recombination: the descendant haploid genome received genetic contributions along  $2^G$  lineages (some of

Download English Version:

<https://daneshyari.com/en/article/6372329>

Download Persian Version:

<https://daneshyari.com/article/6372329>

[Daneshyari.com](https://daneshyari.com)