Contribution to the discussion of the paper by Chang, 'Recent Common Ancestors of All Present-Day Individuals'.

Carsten Wiuf and Jotun Hein

Institute of Biological Sciences
University of Aarhus
DK-8000 Aarhus, Denmark

We have read the paper by Chang with great interest and enjoyed his work very much. Understanding the ancestral history of humans and of populations in general has been devoted a lot of work within the last decades. The paper by Chang relates to this work. We found that the questions he adresses are as much motivated from a basic curiosity point of view as from a mathematical point of view. Our attention will be on a few questions that emerge by confronting Chang's work with the 'The Coalescent'.

In the model by Chang, an individual has two ancestors in each generation with certainty. However, in some non-human populations (e.g. some bacterial and plant populations) each individual will not necessarily have two ancestors, because each individual has the opportunity to reproduce sexually as well as asexually. Imagine that an individual at random chooses two ancestors in the previous generation with probability $1-p$, and with probability p chooses just one. On average each individual will have $2\cdot(1-p)+1\cdot p = 2-p$ ancestors. We conjecture that the time $\mathcal{T}_p(n)$ until a most recent common ancestor (MRCA) to the entire population will be about $G_p(n)=\log_{2-p}(n)$ generations (where $\log_q$ denotes base-q logarithm,) $0\leq p<1$, in line with the result by Chang ($p=0, G_0(n)=\log_2(n)$). This we have confirmed by simulations. For an entirely asexual population, i.e. $p=1$, we retrieve the standard coalescent result that a MRCA is found $2n$ generations in mean back in time.

The convergence of $\mathcal{T}_p(n)/G_p(n)$, $0\leq p<1$ in probability towards one cannot happen uniformly in n as a function of p because we have $G_p(n)=\log_{2-p}(n)>2n$ for large p but $E\mathcal{T}_p(n)<E\mathcal{T}_1(n)=2n$ for all $p<1$. Moreover, $\mathcal{T}_1(n)/2n$ does not converge in probability towards one for n tending to infinity, but towards a stochastic variable with mean one and variance 0.29. The latter is well known from standard coalescent theory.

In the proof given by Chang, we can see that Lemma 6 breaks down for $p=1$. The process $M_t=Y_t\,(2-p)^{-t}$ is a martingale for all p, but the proof needs $P\{M_t< b\,(2-p)^{-t}\ \text{ infinitely often}\}<\text{constant}<1$ with b an increasing function of n. The validity of this statement is not guaranteed for $p=1$. We have checked other lemmas and propositions, and they all seem recoverable for p's less than one.

Furthermore, we could imagine a situation where each individual chooses k, $k\geq 2$, ancestors at random in the previous generation. Such a population could either consists of a hypothetical extra-terrestrial life-form or for example in the case $k=4$ be a human population where each individual chooses four grand-parents (or two parents and two parents-in-law). One generation in the grand-parent setting is almost two generations in Chang's setting. The probabilities that an individual has a, $a = 1, 2, 3$, or 4, different ancestors are in Chang's setting $o(1/n), 1/n, 6/n$, and $1 - 7/n$, whereas in the grand-parent setting these are $o(1/n), o(1/n), 6/n$, and $1 - 6/n$. However, the expected number of descendends to a given number of individuals are asymptotically the same in the two settings, and we will expect that the time $\mathcal{T}_n(k)$ until a MRCA fulfill $\mathcal{T}_n(k)/\log_k(n) \to 1$ in probability.

Concerning the second result by Chang, Theorem 2, we also here expect to find a version of the statement similar to Chang's for $0 \leq p < 1$. It is interesting to note that for $p = 1$ we have $\mathcal{T}_n(1) = \mathcal{U}_n$ ($\mathcal{U}_n$ as in Chang's notation), i.e. when the present-day population finds a MRCA in generation $\mathcal{T}_n(1)$ this ancestor is the only individual ancestral to the present-day population.

The way to choose parents defines the reproductive structure; each individual has a binomially distributed number of offspring with mean $\approx 2 - p$ (or k). It is not clear to us how Chang's results are changed if the reproductive structure is different, i.e. if another exchangeable model is assumed to reproduce the population. Only in the case $p = 1$ the result is known; the time $\mathcal{T}_n(1)$ until a MRCA is scaled with the variance, σ^2, of the number of offspring to an individual, $E\mathcal{T}_n(1) = 2n/\sigma^2$.

Let us now turn towards some genetical aspects of Chang's results. Relative to the time-scale of the Wright-Fisher coalescent process (units of n generations) the time until a common biological ancestor (CBA) of all present-day individuals is just a glimpse of time ($\log_2(n)$ compared to $2n$). If we follow two individuals (or in general a small finite sample) from the present generation and about $\log_2(n)$ generations back in time, the chance that some specific part of the genome of the two individuals have found a common ancestor that relates them genetically is negligiable. So, the most recent CBA will in general not be a common genetical ancestor (CGA) to a specific part of the genome of two individuals.

Moreover, a single individual will not necessarily be related genetically to the most recent CBA. This depends on the size of the genome considered and the reproductive model involved. For example, each individual has two copies of each gene, so two of the individuals grand-parents are not grand-parents in the genetic sense. If we consider one gene at each chromosome, we would have to go 6 generations (assuming all $2^6 = 64$ ancestors are different) back in time to find a biological ancestor who (with certainty) is not a genetical ancestor to any of the genes.

Consider the number of biological ancestors, A_t^i, t generations back in time to a single individual, i. It can easily be seen that $f(x) = E(A_{t+1}^i | A_t^i = nx) \approx n(1 - \exp(-2x))$ for n large. Assuming that the process A_t^i at stationarity does not fluctuate very much, we find from solving $f(x) \approx x$ that about 80% of the population is biologically related to individual i. Thus, a biological ancestor will on average be a genetical ancestor to about $g/0.8n$ genes (g denotes the number of genes in the genome) of a single individual.

The paper was very elevating for us to read, and we find that it would be interesting to pursue these matters further.