

Distributions of time to coalescence under stochastic population growths: application to MRCA dating

Krzysztof A. Cyran¹ and Marek Kimmel²

Keywords: coalescence distributions, MRCA dating, Markov branching processes, mtDNA, HVR

1 Introduction

In the last decade a lot of effort has been spent on inferring human population history from genetic diversity data [6, 7]. The majority of methods were based on the Wright-Fisher (W-F) model of genetic drift which assumes multinomial sampling scheme and thus (for large population) Poisson distribution of the number of progeny for any particular locus. Since this model is not always accurate, the question arises: What is the influence of the departure from W-F model on the distribution of the coalescence time and further analysis of genetic variation? To answer it we performed an extensive (forward direction) simulation study estimating the coalescence distribution for populations evolving according to various stochastic scenarios. We compared coalescence distributions of W-F type models and of the O'Connell (OC) model [4] (corrected in [1]) and the results allowed us to estimate the time to the most recent female common ancestor (MRFCA) of modern humans. For this purpose we used genetic data from HVRI and HVRII of mtDNA of modern humans and Neanderthal fossils.

2 Coalescence distributions in Markov branching processes

We modeled the population trajectories by the slightly supercritical Markov branching process. For such processes, the OC model predicts the asymptotic coalescence distribution to be independent of the type of offspring distribution with given mean and bounded variance. We used the OC model as a standard and checked how well the distributions obtained in various W-F (time-homogeneous) models would match it (Fig. 1a.). We also performed the Kolmogorov-Smirnov test for statistical comparison of distributions obtained with the standard (Table 1, last column), and performed simulations for (time-inhomogeneous) branching processes in variable environment (Fig. 1b.).

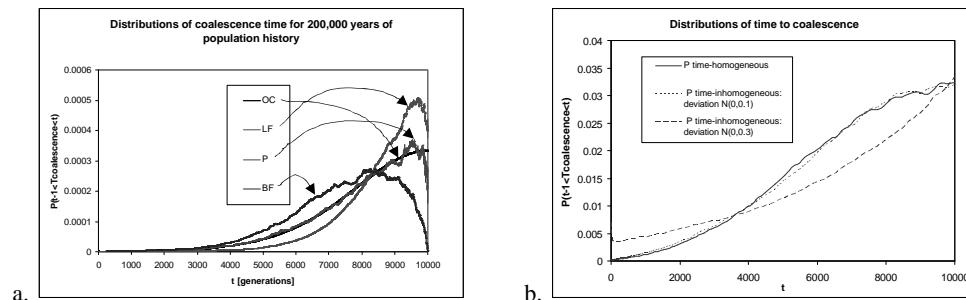


Figure 1: Distributions of time to coalescence for different population histories: a. Distribution for the case of stochastic time homogeneous growth (progeny distributions: P- Poisson, LF-linear fractional, BF- binary fission), b. Coalescence distribution for the cases of stochastic time-homogeneous vs. time-inh. Poisson growth.

¹ (a) Department of Statistics, Rice University, Houston, Texas, USA. E-mail: chrisc1@rice.edu
(b) Institute of Computer Science, Silesian University of Technology, Gliwice, Poland
² Department of Statistics, Rice University, Houston, Texas, USA. E-mail: kimmel@rice.edu

Population trajectory scenario	$g =$	$s^2 =$	s	Final pop. size	Equal to OC distribution?
	$E(T_c/T / N_0=1)$	$Var(T_c/T / N_0=1)$			
OC	0.801	0.0253	0.159	10^7	–
W-F with P	0.802	0.0253	0.159	10^7	Yes
W-F with BF	0.735	0.0289	0.17	0.5×10^7	No
W-F with LF	0.844	0.0243	0.156	2×10^7	No
W-F with P, s_{e1}	0.794	0.0289	0.17	10^7	Not sure
W-F with P, s_{e2}	0.699	0.0724	0.269	2×10^7	No

Table 1: Relative time to coalescence g and its variance s^2 for a pair of alleles for various demographic scenarios starting from a common ancestor. Two last rows are for randomly changing environment with std. dev. $s_{e1} < s_{e2}$.

3 Estimation time to mtEve

In order to be able to date the MRCFA based on data from Table 1, we have to know the average genetic distance between modern humans d_{avg} , divergence rate d and the duration of one human generation λ . After a series of successful sequencings of *H. neanderthalensis* mtDNA [2, 3] dated to live until about 40, 000 years ago [5], d no longer had to be estimated from human-chimpanzee divergence. Assuming the infinite site model, relying on data from [2] and using *H. neanderthalensis* as an outgroup, we estimated d to be about 1.2×10^{-7} . For $d_{avg}=1.8\%$ [2] and $\lambda=20$ years this results in times to mtEve given in Table 2. For all stochastic trajectories we analyzed, the resulting time falls into the 95% confidence interval of the estimate reported in [2]. However, our results, with the average of 193×10^3 years, indicate a systematic shift of 30×10^3 years towards the past. We also showed in this paper that after changing the outgroup from chimpanzee to Neanderthals, genetic models with different assumptions tend to give similar (therefore mutually supporting) predictions. However, interestingly, the estimates in the deterministic growth models are systematically higher than those in the stochastic model. The computer program used for calculations of the coalescence distribution can be downloaded from web: www.stat.rice.edu/~kimmel/software/coalescence.

OC.	W-F time-homogeneous			W-F time-inh. P.		W-F exponential growth			
	P	BF	LF	s_{e1}	s_{e2}	$Z_T=10^9$	$Z_T=10^8$	$Z_T=10^7$	$Z_T=10^6$
187	187	204	178	189	215	223	239	266	311

Table 2: Estimates of the time to mtEve $E(T_0)$ in $[10^3 \text{ years}]$ for various population history scenarios.

References

- [1] Kimmel, M. and Axelrod D. E. 2002. *Branching Processes in Biology*. New York: Springer-Verlag.
- [2] Krings, M., Geisert, H., Schmitz, R., Krainitzki, H. and Pääbo S. 1999. DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen, *Proceedings of the National Academy of Sciences USA*, 96: 5581-5585.
- [3] Krings, M., Capelli, C., Tschentscher, F., Geisert, H., Meyer, S. von Haeseler, A., Grossschmidt, K., Possnert, G., Paunovic, M. and Pääbo, S. 2000. A view of Neandertal genetic diversity, *Nature Genetics*, 26: 144-146.
- [4] O'Connell, N. 1995. The genealogy of branching processes and the age of our most recent common ancestor, *Advances in Applied Probability* 27: 418-442.
- [5] Schmitz, R., Bonani, G., and Smith, F. H. 2002. New research at the Neandertal type site in the Neander Valley of Germany. *Journal of Human Evolution* 42(3):A32-A32.
- [6] Wooding, S. and Rogers A. 2000. A Pleistocene population X-plosion? *Human Biology* 72: 693-695.
- [7] Wooding, S. and Rogers A. 2002. The matrix coalescence and an Application to Human Single-Nucleotide Polymorphisms, *Genetics* 161: 1641-1650.