Times to a most recent common ancestor in a coalescent simulation model with age structure and non-Poisson variation in family size among individuals

Brett Cotler[1], Earl Zedd[2], and Matthew B Hamilton[3]

[1]Brown University , [2]Cal State Long Beach[3], Georgetown University Dept. Of Biology

INTRODUCTION

- Genetic Drift is the random change in allele frequencies in a population over time due to sampling.
- Estimates of genetic drift in age structured populations can be made using either individuals of random ages (corresponds to N_e) or uniform age individuals such as newborns (corresponds to N_b).
- Estimates of the strength of genetic drift from age structured species can differ greatly depending on whether N_e or N_b has been estimated. This has lead to the general question of whether these two measures are really based on different processes of genetic drift in age structured populations.
- Based on a model of sampling variation, Waples et al. 2013 have suggested that differences in life tables can explain how these two methods of sampling yield different strengths of genetic drift. Understanding the relationship between N_e and N_b can inform future empirical studies.
- Analytical time-backward coalescent models suggest that different sampling schemes with age structure will produce the same strength of drift. We built a stochastic coalescent simulation to compare N_e and N_b.

An Age Structured Coalescent Model

We wrote Matlab code for an age structured coalescent model based on an algorithm with several steps:

1. Input a life table (see Table 1), number of generations, initial total population size, and a number of lineages to sample.
2. Convert the life table to a Leslie matrix, and construct a population with the census number of individuals in each age class over the number of generations. We assumed population size and life table parameters were constant over time.
3. Sample k lineages (k = 2 in Results shown here) in the present generation.
4. Using age cohort population sizes, life table probabilities, and family size probability distributions, sample lineages backwards and when age one (newborn) lineages randomly sample an ancestor (Fig. 1).
5. Lineages continue to sample back in time until they reach an MRCA. Record total time to the MRCA and number of newborn to parent sampling events.
6. Repeat steps 3-5 for many iterations to obtain distributions of time to MRCA.

Simulations used life tables reported by Waples et al. (2013).

Table 1: Example life table for a mouse. Column 1 is age class (x), column 2 is age-specific survival (S_x), column 3 is age-specific fecundity (B_x).

Mouse		
Age	Sx	Bx
1	0.951	0.85
2	0.432	1.91
3	0.341	1.3
4	0.341	1.3
5	0.341	1
6	0.000	1

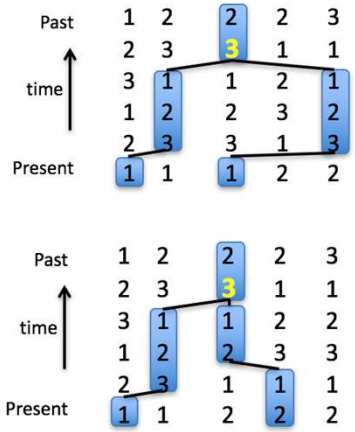

Figure 1. Example age structured genealogies that coalesce to an MRCA (yellow number) where numbers indicate lineage age. Two individuals in the present generation age back in time and sample ancestors in the past. Sampling age = 1 individuals measures N_b (top), while sampling random age lineages measures N_e (bottom).

RESULTS

A desert tortoise life table with 37 age classes and k = 2 was used in Figures 3-6.

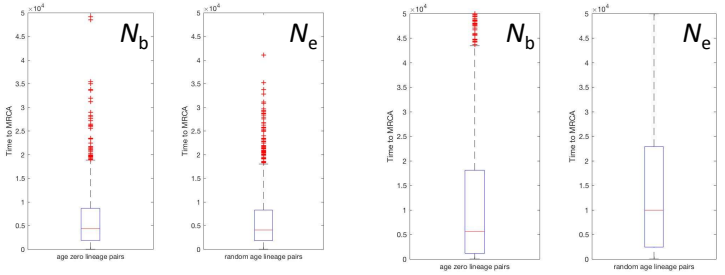

Figure 2 (far left): Times to MRCA for 1000 iterations where reproduction starts at age 6 and age-specific fecundity has lower variation among age classes.

Figure 3 (left): Times to MRCA for a modified life table where only the last three age classes are reproductively mature, making variance in reproduction among age classes very high.

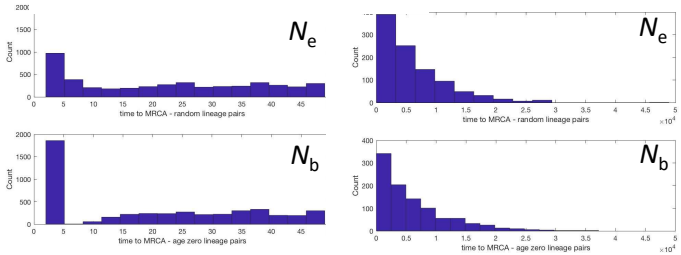

Figure 4 (right): Times to MRCA for those iterations where coalescence occurred within 50 generations (~5% of cases).

Figure 5 (far right): Times to MRCA for all iterations where coalescence occurred within 50,000 generations (~99% of cases).

- Figures 2 and 3 show that N_e and N_b times to MRCA are identical unless there is highly skewed fecundity among age classes. Even in the latter case, time to MRCA distributions overlap extensively for N_e and N_b.
- Figures 4 and 5 show a difference in time to MRCA for N_e and N_b when coalescence occurs very rapidly, a difference that disappears over more generations. This is a product of the small proportion of cases where same age lineages coalesce immediately, an outcome not possible for different age lineages.

CONCLUSIONS

- We found no differences in time to MRCA for N_e and N_b using empirical life tables.
- Median time to MRCA for N_e and N_b differs when life tables exhibit extreme variance in fecundity among age classes. Yet the distributions broadly overlap because of the random variation in coalescence time.
- Preliminary simulations using much greater than Poisson variation in family size within age classes resulted in small differences in time to MRCA for N_e and N_b (results not shown).
- Our results contrast with those of Waples et. al (2013) who found stronger differences between N_e and N_b. We believe the difference is a consequence of their model assumptions that do not capture the stochastic nature of genetic drift as seen in our coalescent model.

Acknowledgements

BC and EZ were supported by the Georgetown University Environmental Science and Policy in the Nation's Capitol NSF REU site program. We thank Professor Sivan Leviyang for discussion.

Literature

Donovan, T. M. and C. Welden. 2002. Spreadsheet exercises in ecology and evolution. Sinauer Associates, Inc. Sunderland, MA, USA.

Hamilton, MB. 2009. *Population Genetics.* Wiley-Blackwell.

Waples RS et al. 2011. Calculating N_e and N_a/N in age-structured populations: a hybrid Felsenstein-Hill approach. Ecology 92(7): 1513-1522.

Waples RS et al. 2013. Simple life-history traits explain key effective population size ratios across diverse taxa. *Proc R Soc Bio Sci 280(1768):20131339.*