

# Note

## On the Use of Star-Shaped Genealogies in Inference of Coalescence Times

Noah A. Rosenberg<sup>\*,1</sup> and Aaron E. Hirsh<sup>†</sup>

<sup>\*</sup>Program in Molecular and Computational Biology, University of Southern California, Los Angeles, California 90089 and

<sup>†</sup>Department of Biological Sciences, Stanford University, Stanford, California 94305

Manuscript received August 15, 2002

Accepted for publication April 7, 2003

### ABSTRACT

Genealogies from rapidly growing populations have approximate “star” shapes. We study the degree to which this approximation holds in the context of estimating the time to the most recent common ancestor ( $T_{\text{MRCA}}$ ) of a set of lineages. In an exponential growth scenario, we find that unless the product of population size ( $N$ ) and growth rate ( $r$ ) is at least  $\sim 10^5$ , the “pairwise comparison estimator” of  $T_{\text{MRCA}}$  that derives from the star genealogy assumption has bias of 10–50%. Thus, the estimator is appropriate only for large populations that have grown very rapidly. The “tree-length estimator” of  $T_{\text{MRCA}}$  is more biased than the pairwise comparison estimator, having low bias only for extremely large values of  $Nr$ .

A fundamental development in population genetics has been the recognition that the pattern of genetic variation in a set of sampled sequences is heavily affected by the particular genealogy of the lineages. In general, however, this underlying genealogy is unknown. To account for the effect of the genealogy in analyses of population genetic data, it is useful to consider “random genealogies” that are consistent with the data and to average over many such genealogies. The coalescent framework provides a natural way to construct these random genealogies under various assumptions about the demography of populations (DONNELLY 1996; NORDBORG 2001).

Use of the coalescent to model unknown genealogies sometimes leads to intensive computations in statistical inference from data (STEPHENS 2001). Thus, for ease of computation, analyses can be conditioned on assumed genealogical shapes that are specified prior to analysis or that are inferred from genetic data. Such methods often ignore uncertainty about the exact shape of the genealogy, producing potentially biased estimates with misleadingly small confidence intervals (SLATKIN and RANNALA 2000; RANNALA and BERTORELLE 2001; ROSENBERG and NORDBORG 2002). When methods based on assumed genealogies are applied, it is important to quantify associated limitations.

The “star-shaped” or “maypole” genealogy (Figure 1A) is perhaps the simplest type of genealogy and the easiest to analyze, as it has the only shape for which all

lineages experience independent evolution. In a star-shaped genealogy, sampled lineages provide independent replicates of the evolutionary process since the time of their most recent common ancestor (MRCA). SLATKIN and HUDSON (1991) found that genealogies of samples taken from populations growing exponentially in size tend to be “star-like,” much more so than genealogies from constant-sized populations (Figure 1; see also DONNELLY 1996; NORDBORG 2001). Because many human populations have experienced rapid population growth, star-shaped genealogies have been explicitly assumed in diverse analyses of human genetic data (RISCH *et al.* 1995; THOMAS *et al.* 1998; MCPEEK and STRAHS 1999; REICH and GOLDSTEIN 1999; LIU *et al.* 2001; STUMPF and GOLDSTEIN 2001, for example). Additionally, star-shaped genealogies are implicit in methods of analysis that treat “unrelated” individuals as independent trials from a population with specified allele frequencies or other parameters.

Here we determine the degree to which the assumption of a star-shaped genealogy is appropriate for a sample taken from an exponentially growing population. Importantly, the error introduced by the assumption depends on the nature of the eventual calculation that will be performed conditional on the star-shaped genealogy. SLATKIN (1996) defined a “stellate index” to quantify the degree to which a given genealogy resembles a star-shaped genealogy and considered properties of this index under various population models. Our goal is different, in that we aim to determine biases of estimators that result from assuming that genealogies of samples taken from exponentially growing populations are star shaped. For this purpose, we find that other quanti-

<sup>1</sup>Corresponding author: Program in Molecular and Computational Biology, 1042 W. 36th Pl., DRB 289, University of Southern California, Los Angeles, CA 90089. E-mail: noahr@usc.edu

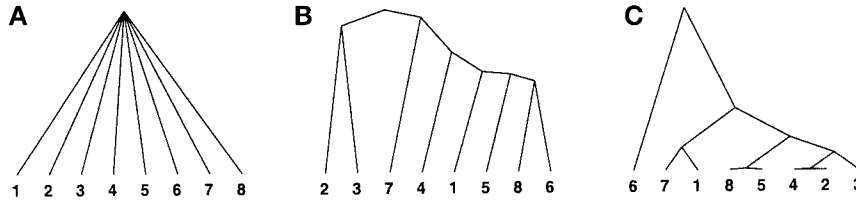


FIGURE 1.—Shapes of genealogies. (A) Perfect star-shaped genealogy or “maypole genealogy.” (B) Example of genealogy simulated using the coalescent process for the exponential growth model with population size 100,000 and growth rate 0.01. (C) Example of genealogy simulated using the coalescent process for the constant population size model. Genealogies are scaled to have the same  $T_{MRCA}$ . In B and C, labels are assigned randomly to lineages.

ties are more natural than the index of SLATKIN (1996). We focus on estimation of the time to the MRCA of a set of sampled lineages.

Consider the genealogy of a set of  $n$  sampled lineages at a nonrecombining locus. Properties of this genealogy include the time to the MRCA of the sample ( $T_n$  or  $T_{MRCA}$ ), the total length of all branches of the genealogy ( $L_n$ ), and the average coalescence time of a pair of sampled lineages ( $P_n$ ). Ratios of these quantities can be used to explore shapes of genealogies under various demographic models (SLATKIN 1996; UYENOYAMA 1997; SCHIERUP and HEIN 2000).

Suppose that an estimate for  $T_n$  is desired. Under the assumption of a star-shaped genealogy,  $P_n$  and  $T_n$  are equivalent. Because unbiased estimates of the coalescence time of a pair of lineages can frequently be obtained (TAJIMA 1983, for example),  $P_n$  is estimated as the average of estimated pairwise coalescence times, over all pairs of lineages. This idea underlies methods summarized by STUMPF and GOLDSTEIN (2001), in which  $P_n$  is estimated under a stepwise mutation model (either by comparing pairs of lineages or by comparing each lineage to a putative ancestral type), and  $\hat{P}_n$  is used as the estimator of  $T_n$ . This “pairwise comparison estimator” of  $T_n$  is unbiased only if the sample has a star-shaped genealogy (or if  $n = 2$ ).

In general, bias of this estimator of  $T_n$  is downward, because the average pairwise coalescence time is always less than or equal to the overall coalescence time. In a constant-sized population of haploid size  $N$ , the bias of the pairwise estimator can be considerable. If we treat  $P_n$  and  $T_n$  as functions of population size  $N$  and growth rate  $r$  and measure them in units of  $N$  generations, then instead of  $P_n/T_n = 1$  as in a star-shaped genealogy, the constant model yields

$$\begin{aligned} \mathbb{E}[P_n(N, 0)/T_n(N, 0)] &\approx \mathbb{E}[P_n(N, 0)]/\mathbb{E}[T_n(N, 0)] \\ &= 1/[2(1 - 1/n)] \\ &= n/(2n - 2) \end{aligned} \tag{1}$$

(using Equation 1 of UYENOYAMA 1997). For a large sample from a constant-sized population,  $P_n$  underestimates  $T_n$  by 50%. Of course, for exponentially growing populations, this bias decreases as the rate of growth increases. To determine the nature of the bias, we simulated the coalescent with exponential growth (SLATKIN and HUDSON 1991; DONNELLY and TAVARÉ 1995; NORD-

BORG 2001), with current population size  $N$ , sample size  $n$ , and exponential growth rate  $r$  (so that  $t$  generations in the past, population size was  $N \exp[-rt]$ ). For each set of parameter values, we computed the mean value of  $P_n/T_n$  over 10,000 simulations.

Figure 2 demonstrates that the bias of the pairwise comparison estimator increases with sample size, decreases with growth rate, and decreases with population size. To explain the perhaps surprising dependence on sample size, note that  $\mathbb{E}[P_n]$  is constant as a function of  $n$  (TAJIMA 1983), whereas  $\mathbb{E}[T_n]$  increases slowly with  $n$  (UYENOYAMA 1997, for example). Thus, the increase in bias with sample size is slow, yet noticeable in the difference between Figure 2A (smaller sample size) and Figure 2B (larger sample size).

The dependence on  $r$  and  $N$  can be understood as follows. In the constant population size model, denote the random time to the coalescence of  $k$  to  $k - 1$  lineages by  $W_k$ , measured in units of  $N$  generations. For  $k = 1, 2, \dots, n - 1$ , let  $X_k$  (also in units of  $N$  generations) denote the total time elapsed in the coalescence of  $n$  lineages to  $k$  lineages, so that  $X_k = W_n + W_{n-1} + \dots + W_{k+1}$ . The corresponding time that it takes for  $n$  lineages to coalesce to  $k$  lineages in the exponential growth model is obtained from  $g^{-1}(X_k)$ , where

$$g^{-1}(X) = \frac{\ln(1 + NrX)}{r} \tag{2}$$

(NORDBORG 2001, Equation 8). For any genealogy,  $P_n(N, r)$  is a linear combination of the coalescence times for that genealogy. Thus, for the exponential growth model, the numerator of  $P_n(N, r)/T_n(N, r)$  is a linear combination of  $g^{-1}(X_{n-1}), g^{-1}(X_{n-2}), \dots, g^{-1}(X_1)$ , and the denominator is  $g^{-1}(X_1)$ . All terms in both numerator and denominator have an  $r^{-1}$  coefficient, which can therefore be removed. Thus, we have

$$\frac{P_n(N, r)}{T_n(N, r)} = \frac{\sum_{k=1}^{n-1} c_k \ln(1 + NrX_k)}{\ln(1 + NrX_1)} \tag{3}$$

for some collection of nonnegative constants  $c_k$  (with  $c_k > 0$  for at least one value of  $k$ ). By demonstrating that its derivative with respect to  $Nr$  is positive, it is easily shown that  $P_n(N, r)/T_n(N, r)$  is an increasing function of  $Nr$ . Consequently,  $\mathbb{E}[P_n(N, r)/T_n(N, r)]$  is increasing in  $Nr$ , and increases in either  $N$  or  $r$  decrease the bias of the pairwise comparison estimator. Equation 3 im-

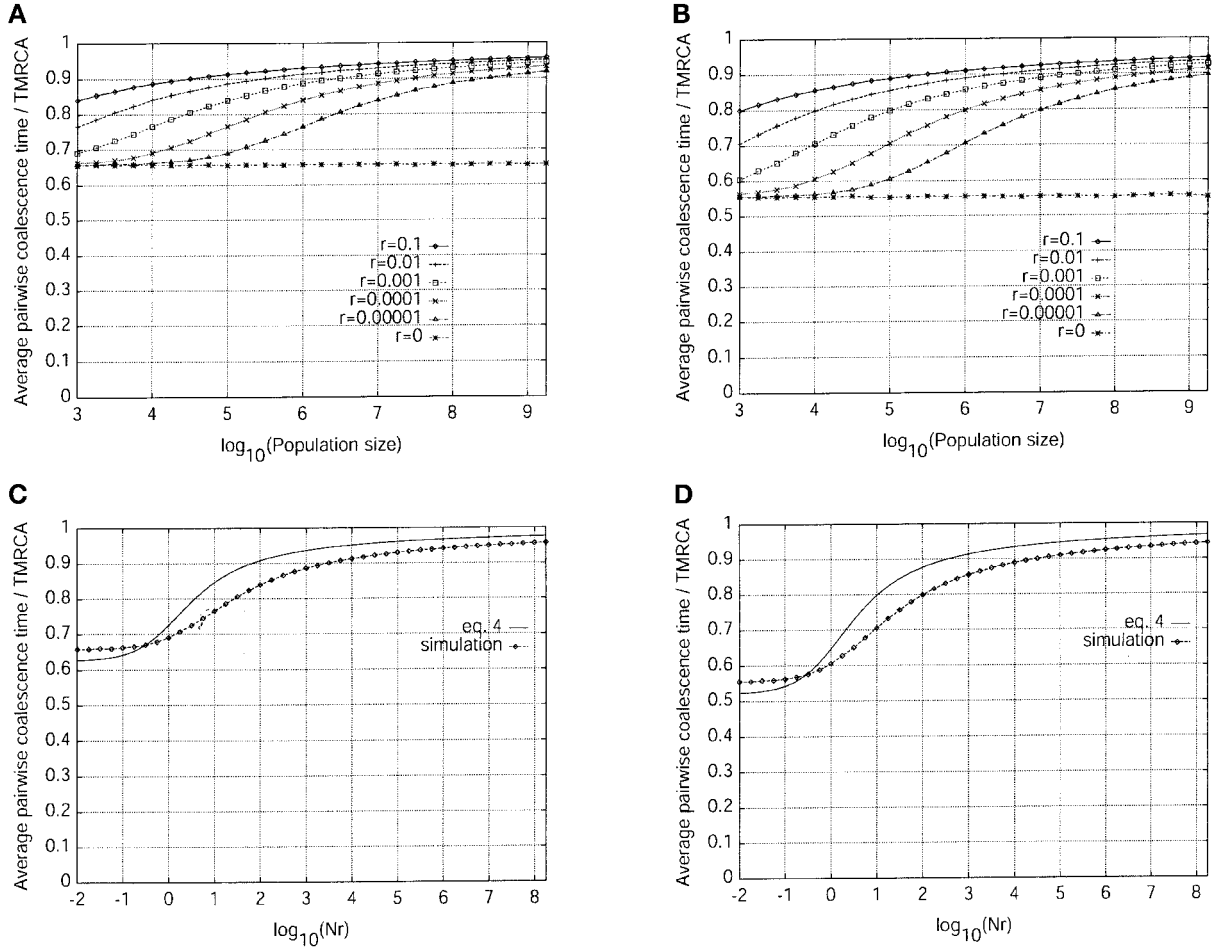


FIGURE 2.—Expected values of the ratio of the average pairwise coalescence time to the time to the most recent common ancestor, or  $\mathbb{E}[P_n/T_n]$ . (A) Simulations with samples of size 5. (B) Simulations with samples of size 25. (C) Comparison of simulations with closed-form approximation (4),  $n = 5$ . (D) Comparison of simulations with closed-form approximation (4),  $n = 25$ . Each point in A and B is based on 10,000 realizations of the coalescent process with exponential growth; the simulation results in C and D derive from averaging  $P_n/T_n$  across all appropriate simulations among those that are shown in A and B.

plies that the parameters  $N$  and  $r$  are coupled in the compound parameter  $Nr$ : for example, the parameter sets  $(N, r) = (10^4, 0.001), (10^5, 0.0001), (10^6, 0.00001)$ , all of which have  $Nr = 10$ , produce the same value of  $\mathbb{E}[P_n(N, r)/T_n(N, r)]$  (Figure 2).

As in the constant population size case,  $\mathbb{E}[P_n(N, r)/T_n(N, r)]$  can be approximated under exponential growth:

$$\begin{aligned} \mathbb{E}\left[\frac{P_n(N, r)}{T_n(N, r)}\right] &\approx \frac{\mathbb{E}[P_n(n, r)]}{\mathbb{E}[T_n(N, r)]} \\ &\approx \frac{g^{-1}(\mathbb{E}[P_n(N, 0)])}{g^{-1}(\mathbb{E}[T_n(N, 0)])} \\ &= \frac{g^{-1}(1)}{g^{-1}[(2n - 2)/n]} \\ &= \frac{\ln(1 + Nr)}{\ln[1 + 2Nr(n - 1/n)]}. \end{aligned} \tag{4}$$

Note that (1) can be obtained from (4) by taking the limit as  $r \rightarrow 0$ . The closed-form expression in (4) gives

a reasonable approximation to  $\mathbb{E}[P_n/T_n]$  (Figure 2, C and D).

For rapidly growing large populations, the bias of the pairwise comparison estimator may be 10% or less (Figure 2). Estimated growth rates for periods of exponential growth of various human populations range from 0.001 to 0.02 per generation (PRITCHARD *et al.* 1999; THOMSON *et al.* 2000). Thus, for groups with sufficiently large  $N$ , the star-shaped genealogy might lead to nearly unbiased estimation of  $T_n$ . However, under the exponential growth model,  $N$  equals the current census population size only if the variance of reproductive success equals 1 (other properties of populations are also incorporated into the parameter  $N$ —see NORDBORG 2001; NORDBORG and KRONE 2002). Estimates of  $N$  for human populations under exponential growth models are considerably smaller than census sizes (PRITCHARD *et al.* 1999; THOMSON *et al.* 2000, for example). For human groups, it is questionable whether  $N$  is large enough for star-shaped genealogies to be applied to estimation of  $T_n$ .

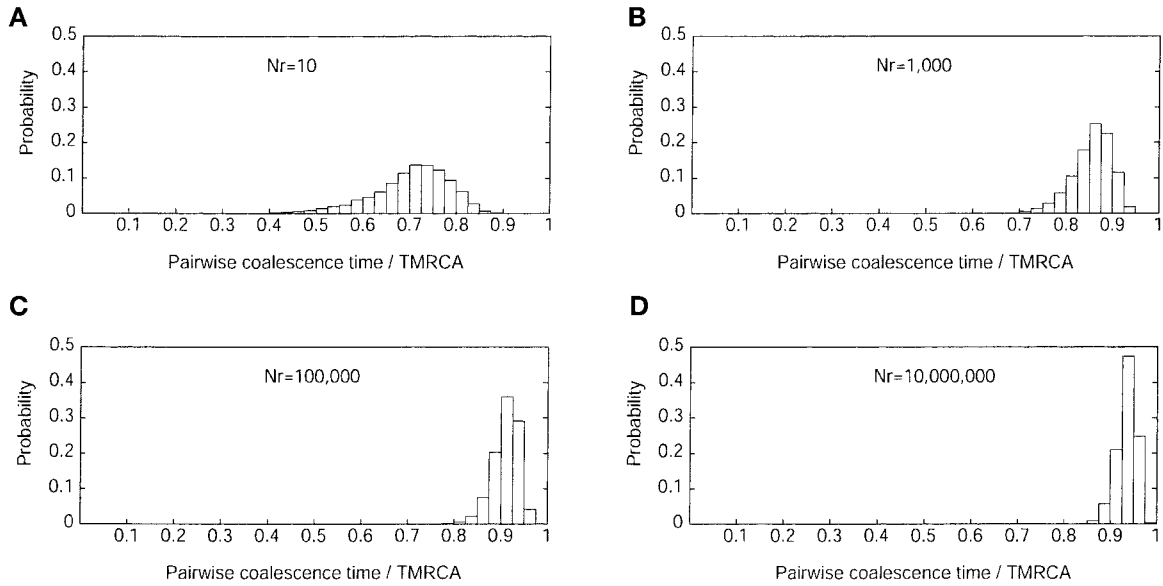


FIGURE 3.—Distribution of the ratio of the average pairwise coalescence time to the time to the most recent common ancestor ( $P_n/T_n$ ). Each distribution is based on 10,000 realizations of the coalescent process with exponential growth, using a sample of size 25.

For small populations the bias of the pairwise comparison estimator is particularly large. For  $Nr < 100$  the bias for a sample of reasonable size will be considerable,  $>20\%$ . Thus, in small populations, even if they have expanded exponentially, the star-shaped genealogy assumption cannot substitute for genealogical modeling of the data; schemes that explicitly account for uncertainty in the genealogy (for reviews, see ROSENBERG and FELDMAN 2002; TANG *et al.* 2002) are likely more appropriate. For estimating  $T_{\text{MRCA}}$ , population sizes and growth rates of relatively small groups such as Jewish priests (THOMAS *et al.* 1998) may be too small to produce approximate star-shaped genealogies. In these groups it is probable that the pairwise comparison estimator underestimates coalescence times, and use of the estimator should be accompanied by quantification of its bias.

A further problem with this estimation procedure is that on the assumption of a star-shaped genealogy, the variance of the pairwise comparison estimator is typically underestimated, as its calculation ignores uncertainty associated with not knowing the genealogy. Under exponential growth,  $P_n/T_n$  can be quite variable (Figure 3), compared to its constant value of 1 in the star genealogy model. By assuming that  $P_n$  and  $T_n$  are equal, the pairwise comparison estimator ignores inherent variation in the relationship between these two quantities, which exists even if  $N$  and  $r$  are known exactly. Of course, all model-based procedures experience problems similar to this limitation of the star genealogy model. The variance of estimators is typically evaluated conditional on a model, such as the star genealogy model or the constant population size model; uncertainty associated

with not knowing the model is difficult to incorporate into the calculation of confidence intervals.

The variance of  $P_n/T_n$  is larger for smaller values of  $Nr$ . Thus, as  $Nr$  decreases, not only does  $P_n$  move farther away from  $T_n$ , but also  $T_n$  becomes harder to predict from  $P_n$  (Figure 3). Only for  $Nr > \sim 10^5$  is it nearly certain that  $T_n < 1.25P_n$  (Figure 3C).

Other estimators based on the assumption of star-shaped genealogies may suffer from more severe bias than the pairwise comparison estimator of  $T_{\text{MRCA}}$ , because other genealogical ratios decline more rapidly with sample size than does  $P_n/T_n$ . Under the assumption of a star-shaped genealogy, an alternate estimator of  $T_{\text{MRCA}}$  is the “tree-length estimator” or the estimated total branch length of the genealogy divided by  $n$  (KARN *et al.* 2002, for example). Unbiased estimators of the total branch length  $L_n$  can be obtained, for example, under the infinite-sites model, from the number of polymorphic sites observed in a data set divided by the mutation rate. To evaluate the bias of the tree-length estimator, we must consider the ratio of  $L_n/n$  to  $T_n$ , a ratio that equals 1 for a star-shaped genealogy.

Under the constant-sized population model, we have (using Equation 2 of TAVARÉ *et al.* 1997)

$$\begin{aligned} \mathbb{E}\left[\frac{L_n(N, 0)}{nT_n(N, 0)}\right] &\approx \frac{\mathbb{E}[L_n(N, 0)]}{n\mathbb{E}[T_n(N, 0)]} \\ &= \frac{2\sum_{k=1}^{n-1} 1/k}{n[2(1 - 1/n)]} \\ &\approx \frac{\gamma + \ln(n-1)}{n-1}, \end{aligned} \quad (5)$$



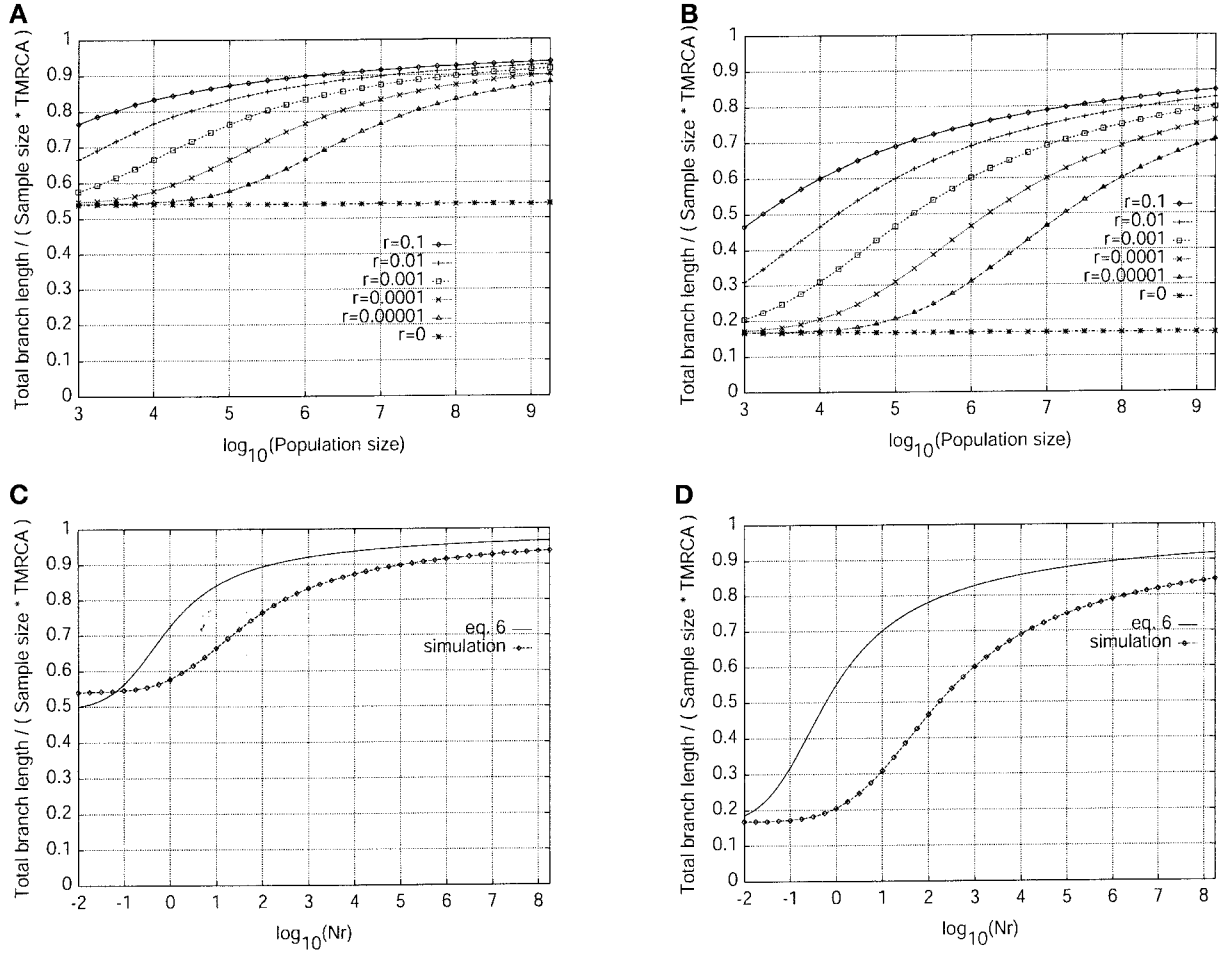


FIGURE 4.—Expected values of the ratio of the total length of the genealogy to the sample size times the time to the most recent common ancestor, or  $\mathbb{E}[L_n/(nT_n)]$ . (A) Simulations with samples of size 5. (B) Simulations with samples of size 25. (C) Comparison of simulations with closed-form approximation (6),  $n = 5$ . (D) Comparison of simulations with closed-form approximation (6),  $n = 25$ . Each point in A and B is based on 10,000 realizations of the coalescent process with exponential growth; the simulation results in C and D derive from averaging  $L_n/(nT_n)$  across all appropriate simulations among those that are shown in A and B.

where  $\gamma \approx 0.5772$  is Euler’s constant. In the exponential growth model the analogous argument to (4) suggests

$$\begin{aligned} \mathbb{E}\left[\frac{L_n(N, r)}{nT_n(N, r)}\right] &\approx \frac{\mathbb{E}[L_n(N, r)]}{\mathbb{E}[nT_n(N, r)]} \\ &\approx \frac{g^{-1}(\mathbb{E}[L_n(N, 0)])}{g^{-1}(n\mathbb{E}[T_n(N, 0)])} \\ &\approx \frac{g^{-1}[2(\gamma + \ln(n - 1))]}{g^{-1}[n(2n - 2)/n]} \\ &= \frac{\ln[1 + 2Nr(\gamma + \ln(n - 1))]}{\ln[1 + 2Nr(n - 1)]}. \end{aligned} \quad (6)$$

The closed-form expression in (6) approximates  $\mathbb{E}[L_n/(nT_n)]$  less accurately than (4) approximates  $\mathbb{E}[P_n/T_n]$  (Figure 4). However,  $L_n/(nT_n)$ , obtained from the same simulated genealogies that underlie Figure 2, is similar to  $P_n/T_n$ , in the ways that it depends on the parameters. The major difference is that the large-sample limit

of  $L_n/(nT_n)$  is zero under constant population size, whereas  $P_n/T_n$  has a large-sample limit of  $1/2$ . Thus, as sample size increases,  $L_n/(nT_n)$  decreases much faster than  $P_n/T_n$ , and the tree-length estimator has bias considerably larger than that of the pairwise comparison estimator.

The potential error of the star genealogy assumption is perhaps greatest when properties of the genealogy itself, such as  $T_n$ , are of interest. If the goal of analysis is to compare genealogies for different loci or populations relative to each other, bias may affect estimates similarly and may have a reduced impact, although differences in sample size and population size should be taken into consideration. Also, if the genealogy is treated as a nuisance parameter, such as in fine mapping of disease susceptibility loci, the assumption might not have severe consequences. The success of methods based on the star genealogy assumption in pinpointing previously identified susceptibility genes (LIU *et al.* 2001, for exam-

ple) suggests that human genealogies may be sufficiently star-like for mapping of some disorders, although modeling of the dependence among lineages can lead to more accurate positional inference (MORRIS *et al.* 2002).

We have seen here that the approximate star-shaped features of genealogies in an exponentially growing population may be insufficient to guarantee low bias in analyses based on the star genealogy assumption, unless the population has grown very rapidly to a very large size. For estimation of  $T_n$ , the numerical results and approximate expressions shown can guide the use of the assumption. Future uses of star-shaped genealogies in population genetic analysis will benefit from demonstration that the assumption is appropriate in the relevant contexts.

We thank H. Innan, P. Marjoram, and M. Nordborg for helpful comments. D. Zulman suggested the term "maypole genealogy." This research was supported by a National Science Foundation Postdoctoral Fellowship in Bioinformatics to N.A.R. and by National Institutes of Health grant GM28016 to M. W. Feldman.

#### LITERATURE CITED

- DONNELLY, P., 1996 Interpreting genetic variability: the effects of shared evolutionary history, pp. 25–40 in *Variation in the Human Genome*. Wiley, Chichester, UK.
- DONNELLY, P., and S. TAVARÉ, 1995 Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.* **29**: 401–421.
- KARN, R. C., A. ORTH, F. BONHOMME and P. BOURSOT, 2002 The complex history of a gene proposed to participate in a sexual isolation mechanism in house mice. *Mol. Biol. Evol.* **19**: 462–471.
- LIU, J. S., C. SABATTI, J. TENG, B. J. B. KEATS and N. RISCH, 2001 Bayesian analysis of haplotypes for linkage disequilibrium mapping. *Genome Res.* **11**: 1716–1724.
- MCPREEK, M. S., and A. STRAHS, 1999 Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *Am. J. Hum. Genet.* **65**: 858–875.
- MORRIS, A. P., J. C. WHITTAKER and D. J. BALDING, 2002 Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies. *Am. J. Hum. Genet.* **70**: 686–707.
- NORDBORG, M., 2001 Coalescent theory, pp. 179–212 in *Handbook of Statistical Genetics*, edited by D. J. BALDING, C. CANNINGS and M. BISHOP. Wiley, Chichester, UK.
- NORDBORG, M., and S. M. KRONE, 2002 Separation of time scales and convergence to the coalescent in structured populations, pp. 194–232 in *Modern Developments in Theoretical Population Genetics*, edited by M. SLATKIN and M. VEUILLE. Oxford University Press, Oxford.
- PRITCHARD, J. K., M. T. SEIELSTAD, A. PÉREZ-LEZAUN and M. W. FELDMAN, 1999 Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**: 1791–1798.
- RANNALA, B., and G. BERTORELLE, 2001 Using linked markers to infer the age of a mutation. *Hum. Mutat.* **18**: 87–100.
- REICH, D. E., and D. B. GOLDSTEIN, 1999 Estimating the age of mutations using variation at linked markers, pp. 129–138 in *Microsatellites: Evolution and Applications*, edited by D. B. GOLDSTEIN and C. SCHLÖTTERER. Oxford University Press, Oxford.
- RISCH, N., D. DE LEON, L. OZELIUS, P. KRAMER, L. ALMASY *et al.*, 1995 Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat. Genet.* **9**: 152–159.
- ROSENBERG, N. A., and M. W. FELDMAN, 2002 The relationship between coalescence times and population divergence times, pp. 130–164 in *Modern Developments in Theoretical Population Genetics*, edited by M. SLATKIN and M. VEUILLE. Oxford University Press, Oxford.
- ROSENBERG, N. A., and M. NORDBORG, 2002 Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* **3**: 380–390.
- SCHIERUP, M. H., and J. HEIN, 2000 Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**: 879–891.
- SLATKIN, M., 1996 Gene genealogies within mutant allelic classes. *Genetics* **143**: 579–587.
- SLATKIN, M., and R. R. HUDSON, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555–562.
- SLATKIN, M., and B. RANNALA, 2000 Estimating allele age. *Annu. Rev. Genomics Hum. Genet.* **1**: 225–249.
- STEPHENS, M., 2001 Inference under the coalescent, pp. 213–238 in *Handbook of Statistical Genetics*, edited by D. J. BALDING, C. CANNINGS and M. BISHOP. Wiley, Chichester, UK.
- STUMPF, M. P. H., and D. B. GOLDSTEIN, 2001 Genealogical and evolutionary inference with the human Y chromosome. *Science* **291**: 1738–1742.
- TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- TANG, H., D. O. SIEGMUND, P. SHEN, P. J. OEFNER and M. W. FELDMAN, 2002 Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. *Genetics* **161**: 447–459.
- TAVARÉ, S., D. J. BALDING, R. C. GRIFFITHS and P. DONNELLY, 1997 Inferring coalescence times from DNA sequence data. *Genetics* **145**: 505–518.
- THOMAS, M. G., K. SKORECKI, H. BEN-AMI, T. PARFITT, N. BRADMAN *et al.*, 1998 Origins of Old Testament priests. *Nature* **394**: 138–140.
- THOMSON, R., J. K. PRITCHARD, P. SHEN, P. J. OEFNER and M. W. FELDMAN, 2000 Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97**: 7360–7365.
- UYENOYAMA, M. K., 1997 Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* **147**: 1389–1400.

Communicating editor: M. K. UYENOYAMA