

LINEAR ESTIMATORS FOR THE EVOLUTION OF TRANSPOSABLE ELEMENTS

BY PAUL JOYCE¹, LINETTE FOX, N. CAROL CASAVANT, AND
HOLLY A. WICHMAN

*Department of Mathematics, Division of Statistics, University of Idaho,
Department of Zoology and Genetics, Iowa State University and Department of
Biological Sciences, University of Idaho*

Pairwise differences and segregating sites are two measures of sequence divergence often used to estimate the rate of evolution. There are other measures of sequence divergence. Which method is most appropriate? Motivated by a study of the evolution of transposable elements, we develop a new and more precise method for estimating the rate of evolution. We apply our method to LINE-1 data from Casavant *et al.* (1996).

1. Introduction. The motivation for this paper grew out of a very curious discovery made while analyzing some DNA sequence data. See Fox (1997). Under certain model assumptions for the evolution of transposable elements (mobile repetitive DNA found dispersed throughout the genome), we considered two simple measures of sequence divergence to estimate the rate of evolution. One estimate was based on the number of pairwise differences and the other was based on the number of segregating sites. We showed that the estimator based on the number of pairwise differences was inconsistent (variance of the estimator does not go to zero), while the segregating sites method was consistent. This is not at all surprising. The same story is true for the well studied neutral coalescent model, see Watterson (1975) and Tajima (1983). However, not only was it demonstrated that the segregating sites estimator is consistent under our model assumptions, but the variance of the estimator goes to zero like $1/n$, where n is the sample size. For evolutionary models involving DNA data it is unusual for estimators to have such good asymptotic properties.

The most curious discovery came when we applied each method to data. We found that in most cases of biological relevance, the pairwise difference estimator actually outperformed the segregating sites estimator. In this paper we resolve this apparent paradox.

We begin with a brief description of the relevant biology followed by the assumptions of the single master model used to analyze the data. We then

¹This research is supported by the National Science Foundation grant 96-26764.

AMS 1991 subject classifications. Primary 62F05; secondary 60G42, 60F15.

Key words and phrases. Coalescent, transposable elements, best linear unbiased estimator (BLUE).