Statistics in Molecular Biology
IMS Lecture Notes - Monograph Series (1999) Volume 33

USES OF STATISTICAL PARSIMONY IN HIV ANALYSES

BY KEITH A. CRANDALL

Brigham Young University

Molecular phylogenies have become powerful tools in human epidemiological studies. Because the phylogeny represents the historical relationship of genes through time, it plays an important role in the elucidation of both historical patterns and processes at work on the gene region of interest, and therefore, on the disease associated with that gene region. However, phylogenetically based analyses are only as good as the phylogenies upon which they are based. Two common problems result from the application of phylogenetic techniques to the population genetic level; 1) lack of resolution due to the short divergence times of a population study, and 2) incorrect inference due to the comparison of non-homologous sequence regions resulting from recombination. A population based method for reconstructing historical relationships among gene sequences is statistical parsimony. In this paper, I outline the limitations of traditional methods, outline the advantages and demonstrated superiority of statistical parsimony when divergences among sequences are low. Finally, I demonstrate the multiple applications of this estimation procedure to problems relating to human immunodeficiency virus evolution.

1. Introduction. Recent advances in population genetic theory, especially coalescence theory (Ewens 1990; Hudson 1990; Donnelly and Tavaré 1995), coupled with an expansion of molecular techniques, have allowed detailed phylogenetic information at the population level. Such genealogical relationships are termed gene trees, allele trees, or haplotype trees, in which different haplotypes or alleles are merely unique nucleotide sequences for a specific region of DNA (loosely termed a gene). With these advances, phylogenetic approaches have proven powerful in studying problems in population genetics and human epidemiology. For example, researchers have utilized phylogenies to explore the origin and spread of retroviruses such as HIV-1, HIV-2 and SIV through a population (Hirsch et al. 1989; Gojobori et al. 1990) and to identify transmission events among individuals and between species (Ou et al. 1992; Holmes et al. 1993; Crandall 1995; Sharp et al. 1996). Phylogenetic studies have also been used to examine the population dynamics of viral infections and the associations of host/pathogen (Harvey and Nee 1994; Holmes and Garnett 1994). Phylogenies have played a central role in longitudinal studies examining the diversification of HIV through time (Kuiken et al. 1993; Strunnikova et al. 1995) and how this

Research partially supported by NIH Grant R01-HD34350-01A1 and the Alfred P. Sloan Foundation.

AMS 1991 *subject classifications.* Primary 92D15, 62G10; secondary 92D10, 92D20.

Key words and phrases. Phylogeny, HIV, parsimony, population genetics.