GENE CONVERSION AND THE INFINITE-SITES MODEL

Stanley A. Sawyer
Washington University

Abstract

The infinite-sites model assumes that there have been only simple point mutations since the common ancestor of a sample of aligned DNA sequences, and in particular no recombination or gene conversion. A statistical test for detecting a history of gene conversion within a sample of aligned DNA sequences is applied to various data sets. The results show strong evidence for multiple intragenic conversion events at two loci in the bacterium *E. coli* and in a repeated family in maize. The data suggest that the rate of these short conversion events may be much larger than the neutral mutation rate per base. Thus the basic assumptions of the infinite-sites model may often be violated in natural data sets. The same test applied to a gene in an RNA virus and to a sample of mitochondrial DNA did not show gene conversion.

Introduction. The basic genetic material or DNA of plants and animals can be thought of as a set of strings of letters from an alphabet {T, C, A, G}, with one string for each chromosome. Thus to a geneticist a mouse is the same as a tomato, since both have about the same amount of DNA. A gene or genetic locus is a segment of a chromosome that affects some trait. Experiments beginning in the middle 1960's showed that many enzymes were unexpectedly polymorphic; i.e., their genetic loci had genes of many different types. A protracted and sometimes heated debate arose about the selective significance of these polymorphisms. Some biologists felt that most of these observed enzyme polymorphisms were selectively neutral or nearly neutral, while others asserted that different enzymes were unlikely to be selectively equivalent, and so various forms of balancing selection must be involved (Kimura, 1968; Wills, 1973; see also Hartl, 1989; Hartl and Sawyer, 1991). In an attempt to resolve this controversy, Ewens (1972) derived a formula for the expected distribution of genes among allelic classes in a random sample from a population that is selectively neutral at that genetic locus. The infinite-alleles model in all of its various forms can be considered an extension of Ewens' (1972) work. The vast number of recent papers on the infinite-alleles model by Donnelly, Ethier, Griffiths, Hoppe, Hudson, Kaplan, Kurtz, Tavare, Watterson, and others shows the health and vigor of current mathematical research in this field.

The infinite-alleles model assumes an equilibrium selectively-neutral random-mating population subject to mutation, where each new mutant gene is of a type that is entirely new to the population. In 1972, most data on enzyme poly-