

MATTERS OF THE RECORD

Extinct meets extant: simple models in paleontology and molecular phylogenetics

Sean Nee

Ashworth Laboratories, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, United Kingdom. E-mail: sean.nee@ed.ac.uk

Accepted: 9 November 2003

Introduction

Paleontologists have a long tradition of the use of mathematical models to assist in describing and understanding patterns of diversification through time (e.g., Raup et al. 1973; Stanley 1975; Sepkoski 1978; Raup 1985; Foote 1988; Gilinsky and Good 1989). This is natural, as the information, phylogenetic and otherwise, that paleontologists work with comes equipped with a temporal dimension, albeit approximate, which endows these phylogenies with information about the tempo of evolution as well as the genealogical relationships among the lineages. Mathematical and statistical modeling are the tools for unlocking the quantitative information in the phylogenies.

Recently, molecular phylogenetics (e.g., Hillis et al. 1996) has created a new source of phylogenies with a temporal dimension, now frequently provided by molecular clocks. Many people have applied mathematical models to these phylogenies as well. Here, I highlight the areas of overlap as well as differences in what the simple models in the two fields have to tell us. To save ink, I will hereafter refer to paleontology as P and molecular phylogenetics as MP.

First I review the use of simple mathematical models to extract information about the tempo of evolution from phylogenies in the fields of P and MP. The same models, or variants thereof—these being the birth, birth-death, and Moran models—are used in the two areas, but there are differences in what they tell us, arising from differences in the nature of the phylogenies themselves. Finally, I address a high-profile assault on this common

framework of understanding that has recently been launched.

The Pure Birth Process

This is one of the two simplest mathematical models used in P and MP (the other is the Moran process—discussed below) and, in its stochastic form, was one of the first stochastic processes to have been studied (Yule 1924; Kendall 1948, 1949; Feller 1957; Bailey 1964). It assumes that clades grow as follows. At each point in time, each lineage (or higher taxon) has the same probability, b , as every other to give birth to a new lineage, and extinction—or lineage death—does not occur. Under this model, average clade size, $N(t)$, grows exponentially: $N(t) = N(0)e^{bt}$. This deterministic result from the pure birth process has been used by Sepkoski (e.g., 1978) and Stanley (e.g., 1975) among others.

It is notable that the statistician who first studied this process, Yule (1924), was inspired to do so by exactly the same sorts of questions that motivate us. The Willis in his paper's title ("A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, FRS") was an expert on angiosperms who had formed what were perceived to be anti-Darwinian views on the sizes and distributions of species and higher taxa. The paper contains a thorough treatment of the pure birth process and even has such contemporary features as the use of the birth process to estimate the rate of cladogenesis of the angiosperms. Happily, the birth process is often called the Yule process, anchoring its origins in our own areas of interest. For a time, a competing name was the Furry process, after the astronomer who stud-