J. Paleont., 76(4), 2002, pp. 587–589

0022-3360/02/0076-587$03.00

STRATIGRAPHY IN PHYLOGENY RECONSTRUCTION—REPLY TO SMITH (2000)

JOHN ALROY

National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara 93101,
<alroy@nceas.ucsb.edu>

IN A brief commentary, Smith (2000) broadly dismisses several newly developed phylogenetic methods that make use of temporal data. In this comment, I wish to concur with other critics of Smith (Fisher et al., 2002; Wagner, 2002), and in particular expand on an argument of Fisher et al. regarding stratocladistics and the divisibility of paleontological time scales. Smith's strongest and most detailed argument is that "With enough subdivisions of time . . . stratigraphic debt can be made sufficiently large to overturn any phylogeny based on morphology . . . " As pointed out by Fisher et al. (2002), Clyde and Fisher (1997) actually employed a rather conservative algorithm for defining the number of time intervals. Here I argue that stratocladistics fundamentally depends on global temporal durations of fossil taxa, not just on their ranges across individual (or even composite) stratigraphic sections, and that these durations are intrinsically expressed in discrete, not continuous, units. Thus, paleontological time scales cannot be subdivided arbitrarily and are not literally "stratigraphic." I also discuss additional objective criteria for limiting temporal subdivision, which show that a series of stratocladistic studies involving North American mammals actually have underdivided the time scale. Thus, stratocladistics does not allow for infinite temporal subdivision either in principle or in practice.

TIME AND STRATIGRAPHY

Smith (2000) almost casually mentions that there always may be "enough subdivision of time" to create as much "stratigraphic debt" as one might want. Smith nowhere explains why. One might think so because the literature has used "stratigraphy" and "time" as virtual synonyms; local stratigraphic sections are measured in continuous units such as meters; and perhaps, then, time is mathematically continuous and infinitely divisible. Alternatively, in an illustration Smith (2000, Fig. 1) shows subdivisions of "absolute time," and numerical time scales also are necessarily discrete and fundamentally different from measured sections.

The terms "time" and "stratigraphy".—Almost all of the large literature on stratocladistics, and on related "stratigraphic" measures of the quality of the fossil record (e.g., Paul, 1982; Benton, 1994; Huelsenbeck, 1994; Foote and Raup, 1996; Clyde and Fisher, 1997; Wills, 1999), has used the term "stratigraphic" to mean "having to do with time." The very terms "stratocladistics" (Fisher, 1982, 1991) and "stratophenetics" (Gingerich, 1976) are formed by using stratigraphy as a shorthand for temporal data. Indeed, Fox et al. (1999) flatly defined "stratigraphic data" as "the temporal order of specimens in the fossil record." "Stratigraphic" also is overused as a synonym for "temporal" in the larger paleontological geological literature [e.g., the North American "Stratigraphic" Code (NACSN, 1983); the International "Stratigraphic" Guide of the International Subcommission on Stratigraphic Classification (Hedberg, 1976)].

However, "stratigraphy" in a literal sense only has to do with the description and analysis of "stratified rocks" (Webster's New Universal Unabridged Dictionary, 1994), "rock strata" (Jackson, 1997), or "stratified (layered) rocks; the arrangement of strata" (Woodburne, 1987). Meanwhile, Hedberg (1976) uses "chronostratigraphy" to mean the study of "the age of strata" and as a subset of geochronology, not the other way around. Finally, Harland et al. (1990) refer to "chronometric" and "chronostratic" time scales, only tying the latter to stratigraphy. Numerical time scales of the kind used, for example, by Wills (1999) are clearly not "chronostratic" but "chronometric." In sum, all responsible authors should agree that "stratigraphy" and "stratigraphic" concern rocks, not time, so the terms "chronology" and "temporal" should always be preferred in the phylogenetic literature.

Local vs. regional and global scales of stratocladistic analyses.—Putting terminology aside, some stratocladistic studies do focus on a single, local stratigraphic section (e.g., Clyde and Fisher, 1997; Polly, 1997). So it might be logically justified (if uncharitable) to argue that stratocladists might overdivide their time scales just by slicing up their sections too narrowly.

Of course, sections are not arbitrarily and infinitely divisible because the rock record accumulates at variable rates and omits variable amounts of time in between depositional packages. Furthermore, very small stratigraphic intervals would break up sedimentary packages that were laid down in single events. Regardless, most applications of "stratigraphic" methods in phylogenetics (e.g., Huelsenbeck, 1994; Wagner, 1995; Wills, 1999; Bloch et al., 2001; Bodenbender and Fisher, 2001) have relied upon continental or intercontinental paleontological time scales, which are not even simple agglomerations of stratigraphic sections: they are empirical hypotheses about the ordering of real-world biological events.

Quantitative biochronology and stratigraphy.—One still could try to translate these discrete, broad-based time scales into continuous stratigraphic units by means of quantitative correlation methods. Indeed, a large body of literature on graphic correlation (Shaw, 1964) seeks to do just this. However, graphic correlation actually seeks to find best-fit hypothetical arrangements of discrete biological events, such as first and last appearance events (not "appearance datums," which pertain only to individual, local stratigraphic sections: Alroy, 1994, 1996; Walsh, 1998). As stated by Shaw (1964, p. 104), "the only objective temporal facts that a fossil species can supply—the only epochs it can mark in the rocks—are the beginning and end of its range."

More importantly, numerous rival methods do not generate continuous master stratigraphic sections (archaeological seriation—Burroughs and Brower, 1982; probabilistic biostratigraphy—Hay, 1972; Gradstein and Agterberg, 1982; unitary associations—Guex, 1977, 1991; appearance event ordination—Alroy, 1992, 1994). Their common currency is discrete appearance event sequences, which need not correspond perfectly to any one local stratigraphy (e.g., see Guex, 1991), and typically are computable even from taxonomic inventories of fossil collections that have unknown stratigraphic interrelationships. Thus, objective, quantitative paleontological time scales are not simply stratigraphy writ large.

Admittedly, most of these methods are said to concern quantitative "biostratigraphy," not time per se. However, "biostratigraphy" itself only has to do with vertical distributions of fossils in rocks per se, and with characterizations of strata on the basis of fossils (Jackson, 1997). Thus, methods like unitary associations and appearance event ordination that are independent of superposition are not strictly "biostratigraphic," and we need another term—i.e., biochronology—for the combination of biostratigraphic data and inventory data (Guex, 1991; Alroy, 1992, 1994).

Global time scales and stratigraphy.—Of course, most real-world paleontological time scales still are constructed using qualitative, non-statistical criteria. However, traditional criteria are themselves directly translatable into the currency of first and last appearances. All biostratigraphic time scales reduce to interval, assemblage, or abundance zones (NACSN, 1983). Any interval or assemblage zone is defined by the combinations of overlapping ranges, which are just sets of appearance events. Abundance zones are not routinely used and have been rejected by most quantitative workers (e.g., Shaw, 1964; Guex, 1977; Gradstein and Agterberg, 1982; Alroy, 1992). Some workers tie names of paleontological time intervals to "reference localities" (e.g., Mein, 1975; Alroy, 2000), but the temporal correlations underpinning such time scales always can be expressed in terms of appearance events.

In summary, paleontological time scales always reduce to discrete historical events, so stratocladistics, related phylogenetic methods, and the so-called stratigraphic fit measures all have to work with event-based time scales whenever they stray from local stratigraphic sections. Algorithmically, these methods only make sense in the first place if they do at some level correspond with global biochronology—as recognized by Clyde and Fisher (1997) and Bodenbender and Fisher (2001), who required that each of their "stratigraphic" intervals be bounded on each end by at least one event. Perhaps, then, stratocladistics is better called "chronocladistics."

CONSTRAINTS ON SUBDIVIDING TIME SCALES

Smith (2000) also has failed to indicate that numerous logical and quantitative criteria may be used to limit the subdivision of time scales, and that there are specific reasons to favor the level of precision actually used in stratocladistic analyses of the mammalian record.

Logical limits.—Both quantitative methods and formal nomenclatural rules define time intervals directly on the basis of appearance events, and each taxon may have only one first and one last appearance. Thus, paleontological time scales never can break out more than twice as many intervals as taxa, and the maximum number of intervals is necessarily finite. Furthermore, each time interval must uniquely include at least one compositionally unique fossil collection. Otherwise, it would not be possible to tie at least one appearance event to the time interval.

Stratocladistics and all related methods are of course subject to these two constraints. Similarly, Clyde and Fisher (1997) and Bodenbender and Fisher (2001) counted intervals only if they included fossils pinning down appearances of ingroup taxa. Thus, intervals were required to sample fossiliferous horizons. Thanks to emphasizing only fossils of ingroup taxa, this criterion is even more conservative than the two mentioned here.

Limits based on diachrony.—The divisibility of time scales is further limited by the precision of paleontological correlations. Some authors equate precision with the average diachrony of individual appearance events seen in multiple regions, a figure often given in studies of Cenozoic deep sea microfossils (e.g., Spencer-Cervato et al., 1994). Importantly, the average interregional diachrony of North American mammalian species is 1.4 m.y. (Alroy, 1998a).

Of course, if time intervals are actually based on appearances of a few taxa selected specifically because they exhibit low diachrony, then these general, average estimates will be too liberal, at least when it comes to interval boundaries. However, paleontologists routinely fail to identify low-diachrony taxa (Alroy, 1998a). Thus, average diachronies do provide a rough estimate of the point at which further subdivision of a time scale would become reckless.

Limits based on calibration error.—Researchers also have tried to estimate precision directly by calibrating paleontological time scales against independent geochronological age estimates, most famously including Harland et al. (1982, 1990). Quantitative analyses of fossil mammal biochronology provide numerous estimates (e.g., Alroy, 1992, 1996, 1998b, 2000 for North America; Alroy, 1994, for Africa; Azanza et al., 1997, for Europe), and average 95 percent confidence intervals typically are as low as 1.4 m.y. (Alroy, 2000).

Limits based on the precision/accuracy tradeoff.—An even more direct approach (Alroy, 1996, 1998b) is to examine the tradeoff between precision and accuracy as time scales are subdivided by comparing observed appearances with "true," fully-resolved appearance dates estimated by adding 50 percent confidence intervals (Paul, 1982; Strauss and Sadler, 1989; Marshall, 1994). The inferred optimal interval length for the North American mammalian time scale is about 0.7 m.y. (Alroy, 1996, 1998b).

Stratocladistic precision in practice.—The 0.7 m.y. precision figure is quite close to the average interval lengths actually used in key stratocladistic studies: Clyde and Fisher (1997) subdivided the Tiffanian through early Bridgerian mammal ages into eight intervals averaging about 1.5 m.y. (see Alroy, 2000); Polly (1997) subdivided the Clarkforkian and Wasatchian into 10 intervals averaging 0.65 m.y.; and Bloch et al. (2001) subdivided the Paleocene into eight intervals averaging 1.4 m.y. The one major stratocladistic study not involving mammals (Bodenbender and Fisher, 2001) defined just 18 intervals for almost the entire Paleozoic.

In his brief comments on the subject, Smith (2000) made it appear that the entire problem of time scale resolution is subjective and arbitrary. His argument only could have resulted from a misreading of recent literature on both stratocladistics and quantitative biochronology: multiple objective criteria do exist, and they all suggest that Clyde and Fisher (1997), Polly (1997), and Bloch et al. (2001) used entirely conservative temporal schemes.

ACKNOWLEDGMENTS

I thank D. Fisher, M. Foote, C. Marshall, and P. Wagner for helpful comments and words of caution. This work was funded by the NSF Division of Environmental Biology (grant DEB-0083893), and hosted by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0072909), the University of California, and the Santa Barbara campus. This is Paleobiology Database publication #8.

REFERENCES

Alroy, J. 1992. Conjunction among taxonomic distributions and the Miocene mammalian biochronology of the Great Plains. Paleobiology, 18:326–343.

Alroy, J. 1994. Appearance event ordination: a new biochronologic method. Paleobiology, 20:191–207.

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. Palaeogeography, Palaeoclimatology, Palaeoecology, 127:285–311.

Alroy, J. 1998a. Diachrony of mammalian appearance events: implications for biochronology. Geology, 26:23–27.

Alroy, J. 1998b. Equilibrial diversity dynamics in North American mammals, p. 232–287. *In* M. L. McKinney and J. Drake (eds.), Biodiversity Dynamics: Turnover of Populations, Taxa and Communities. Columbia University Press, New York.

Alroy, J. 2000. New methods for quantifying macroevolutionary patterns and processes. Paleobiology, 26:707–733.

Azanza, B., M. T. Alberdi, E. Cerdeño, and J. L. Prado. 1997.

Biochronology from latest Miocene to middle Pleistocene in the Western Mediterranean area: a multivariate approach. Memoires et Travaux E.P.H.E., Institut de Montpellier, 21:567–574.

BENTON, M. J. 1994. Palaeontological data and identifying mass extinctions. Trends in Ecology and Evolution, 9:181–185.

BLOCH, J. L., D. C. FISHER, K. D. ROSE, AND P. D. GINGERICH. 2001. Stratocladistic analysis of Paleocene Caroplestidae (Mammalia, Plesiadapiformes) with description of a new Late Tiffanian genus. Journal of Vertebrate Paleontology, 21:119–131.

BODENBENDER, B. E., AND D. C. FISHER. 2001. Stratocladistic analysis of blastoid phylogeny. Journal of Paleontology, 75:351–369.

BURROUGHS, W. A., AND J. C. BROWER. 1982. SER, a FORTRAN program for the seriation of biostratigraphic data. Computers & Geosciences, 8:137–148.

CLYDE, W. C., AND D. C. FISHER. 1997. Comparing the fit of stratigraphic and morphologic data in phylogenetic analysis. Paleobiology, 23:1–19.

FISHER, D. C. 1982. Phylogenetic and macroevolutionary patterns within the Xiphosurida. Proceedings of the Third North American Paleontological Convention, 1:175–780.

FISHER, D. C. 1991. Phylogenetic analysis and its implications in evolutionary paleobiology, p. 103–122. *In* N. L. Gilinsky and P. W. Signor (eds.), Analytical Paleobiology. Paleontological Society, Knoxville, Tennessee.

FISHER, D. C., M. FOOTE, D. L. FOX, AND L. R. LEIGHTON. 2002. Stratigraphy in phylogeny reconstruction—comment on Smith (2000). Journal of Paleontology, 76:585–586.

FOOTE, M., AND D. M. RAUP. 1996. Fossil preservation and the stratigraphic ranges of taxa. Paleobiology, 22:121–140.

FOX, D. L., D. C. FISHER, AND L. R. LEIGHTON. 1999. Reconstructing phylogeny with and without temporal data. Science, 284:1816–1819.

GINGERICH, P. D. 1976. Cranial anatomy and evolution of early Tertiary Plesiadapidae (Mammalia, Primates). University of Michigan Museum of Paleontology Papers on Paleontology, 15:1–141.

GRADSTEIN, F. M., AND F. P. AGTERBERG. 1982. Models of Cenozoic foraminiferal stratigraphy, northwestern Atlantic margin, p. 119–173. *In* J. M. Cubitt and R. A. Reyment (eds.), Quantitative Stratigraphic Correlation. John Wiley & Sons, Cichester.

GUEX, J. 1991. Biochronological Correlations. Springer-Verlag, Berlin, 252 p.

GUEX, J. 1997. Une nouvelle méthode d'analyse biochronologique: note preliminaire. Bulletin de la Societé vaudoise des sciences naturelles, 73:309–322.

HARLAND, W. B., R. L. ARMSTRONG, A. V. COX, L. E. CRAIG, A. G. SMITH, AND D. G. SMITH. 1990. A Geologic Time Scale. Cambridge University Press, Cambridge, 263 p.

HARLAND, W. B., A. V. COX, P. G. LLEWELLYN, C. A. G. PICKTON, A. G. SMITH, AND R. WALTERS. 1982. A Geologic Time Scale. Cambridge University Press, Cambridge, 131 p.

HAY, W. W. 1972. Probabilistic stratigraphy. Eclogae Geologicae Helveticae, 65:255–266.

HEDBERG, H. D. 1976. International Stratigraphic Guide. John Wiley & Sons, New York, 200 p.

HUELSENBECK, J. P. 1994. Comparing the stratigraphic record to estimates of phylogeny. Paleobiology, 20:470–483.

JACKSON, J. A. 1997. Glossary of Geology. American Geological Institute, Alexandria, Virginia, 769 p.

MARSHALL, C. R. 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. Paleobiology, 20:459–469.

MEIN, P. 1975. Résultats du groupe de travail des vertébrés: biozonation du Néogène méditerranéean à partir des mammifères, p. 78–81. *In* J. Senes (ed.), Report on Activity of the RCMNS Working Groups (1971–1975), Bratislava.

NORTH AMERICAN COMMISSION ON STRATIGRAPHIC NOMENCLATURE. 1983. North American stratigraphic code. American Association of Petroleum Geologists Bulletin, 67:841–875.

PAUL, C. R. C. 1982. The adequacy of the fossil record, p. 75–117. *In* K. A. Joysey and A. E. Friday (eds.), Problems of Phylogenetic Reconstruction. Academic Press, London.

POLLY, P. D. 1997. Ancestry and species definitions: a stratocladistic analysis of Viverravidae (Carnivory, Mammalia) from Wyoming. Contributions from the Museum of Paleontology, University of Michigan, 30:1–53.

SHAW, A. B. 1964. Time in Stratigraphy. McGraw-Hill, New York, 365 p.

SMITH, A. B. 2000. Stratigraphy in phylogeny reconstruction. Journal of Paleontology, 74:763–766.

SPENCER-CERVATTO, C., H. R. THIERSTEIN, D. B. LAZARUS, AND J. P. BECKMANN. 1994. How synchronous are Neogene marine plankton events? Paleoceanography, 9:739–763.

STRAUSS, D., AND P. M. SADLER. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology, 21:411–421.

WAGNER, P. J. 1995. Stratigraphic tests of cladistic hypotheses. Paleobiology, 21:153–178.

WAGNER, P. J. 2002. Stratigraphy in phylogeny reconstruction—comment on Smith (2000). Journal of Paleontology. 76:590–593.

WALSH, S. L. 1998. Fossil datum and paleobiological event terms, paleontostratigraphy, chronostratigraphy, and the definition of Land Mammal "Age" boundaries. Journal of Vertebrate Paleontology, 18: 150–179.

WILLS, M. A. 1999. Congruence between phylogeny and stratigraphy: randomization tests and the gap excess ratio. Systematic Biology, 48: 559–580.

WOODBURNE, M. O. 1987. Cenozoic Mammals of North America: Geochronology and Biostratigraphy. University of California Press, Berkeley, 336 p.

ACCEPTED 29 NOVEMBER 2001