Statistical Traps Lurk in the Fossil Record

A combination of the skewed distribution in the species abundance and the vagaries of fossilization may suggest differences in fossil samples where no real difference exists

When paleontologists analyze a fossil assemblage from a particular locality, they can in effect acquire a glimpse of an ancient ecological community. Comparisons with other fossil assemblages can then produce information on how ecological communities differed through time or between separate geographical areas. But, warns Carl Koch of Old Dominion University, Norfolk, Virginia, paleontologists must be cautious about reading biological significance into differences between fossil assemblages, because of a statistical quirk of nature that may be misleading them.

"It is possible to get as much as a 25% difference in the species represented in two separate samples taken from the same fossil assemblage," Koch estimates, "even if the samples are quite large. Before biological meaning can be ascribed to differences in the fossil record, paleontologists therefore need a way of knowing just how different fossil samples will look when they are in fact sampling the same fossil community."

Koch has been sounding this kind of warning for some time—thereby irritating many professionals in the field, who believe that experienced workers can separate signal from noise in the record—but has now produced a quantitative analysis that, he says, "should be disconcerting to many paleontologists." In an article in the current issue of *Paleobiology*, he concludes that "much of the past work and interpretations based on the measured values [of species distributions] should be reevaluated."

"It's true that many people haven't been sufficiently aware of the sample size effect that Koch talks about," says David Raup of the University of Chicago. "It is not clear to me that the problem is as serious as he suggests. But his point is very well taken."

The source of the problem is a simple fact of biological life that ecologists have long been familiar with: namely, that in any ecological community a few species will occur in abundance, some will occur frequently, but most are rare and will occur only infrequently. Add to this another simple fact of biological life—that the odds of any individual animal becoming fossilized after death, and thus entering the paleontological record, are vanishingly small—and you are immediately faced with potential statistical traps through vagaries of sampling.

For instance, no single fossil sample is going to pick up all the rare species in an assemblage. And if this sample were to be compared with another from higher in the stratigraphic sequence (which represents a later slice of time) then an overlapping, but different, set of rare species will be recorded. "Species that are found in the lower sample but not in the upper might erroneously be inferred to have gone extinct," says Koch. "Similarly, species that appear in the later sample but not the earlier might wrongly be thought to have originated there." In an earlier paper with Martin Buzas and others, of the Smithsonian Institution, Washington D.C., Koch concluded that "The list of examples where rare species could lead to erroneous conclusions is limited only by the imagination of the reader."

Clearly, the origination and extinction of species through the fossil record is often extremely important in building up an accurate picture of the change of ecological communities through time. "The more samples you are able to get from the two stratigraphic levels, the smaller will be the potential problem," says Koch, "but it will always be a problem to some degree."

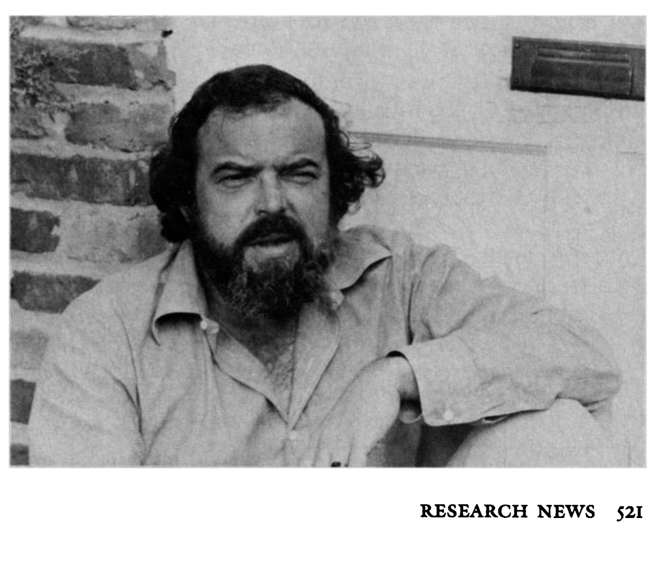
In his recent study Koch analyzed the species distribution in ten fossil data sets, each of which differed in the number of locations from which the fossils were collected, the total number of fossils in the data set, and the number of species represented. All the data sets were from marine ecosystems, which varied from open ocean to restricted basins. Koch found that each of the collections fitted the typical ecological pattern of a few abundant species, some frequently occurring species, and many rare species—in other words, a logarithmic series distribution. As a result, says Koch, "we can predict the sample size effects for many kinds of paleontological study."

Using these statistical distributions as a predictor, Koch carried out a theoretical estimate of the number of species that would occur only once in a data set comprising 200 collections giving a total of 600 species: the answer turned out to be 156. Suppose now that a further 200 collections were taken from the same sediments. According to the calculations, the expectation is that "529 (88%) of the original species would be found." In addition, 71 rare species will be present in this second data set that were absent in the first, even though they essentially come from the same assemblage. This difference of nearly 12% in overall species composition between the two data sets is simply the result of statistical sampling effects, not of biological differences.

Turning to a concrete example from the ten data sets in his study, Koch asked the question, "How many molluscan species found in the East Gulf range geographically into the Atlantic Coastal Plain during the late Cretaceous?" According to the information in the two relevant fossil data sets, the number of species in common is 289, which represents 48% of the total East Gulf assemblage. This leaves a difference between the two assemblages of 52%, which looks sub-

Carl Koch

"The list of examples where rare species could lead to erroneous conclusions is limited only by the imagination of the reader."



stantial, even against the rule of thumb of 20 to 25% that paleontologists use for denoting different faunal ranges.

Koch notes, however, that the data set for the East Gulf area is four times the size of the Atlantic Coast data set, which difference warns of potential statistical traps. Using the appropriate curves, Koch calculates that, given this disparity in sample size, there would be an apparent difference of some 35 to 40% in species composition between the East Gulf region and the Atlantic Coast, even if the two had been ecologically identical. "I conclude that most of the measured faunal difference results from differences in the amount of data between the two areas," says Koch.

In another example, in which the data sets were of similar size, a species count showed a 31% difference between two locations. But Koch shows that under these circumstances a difference of 12 to 14% would be expected from sampling errors alone, even if the two areas were in fact ecologically identical. The 31% therefore falls to 17 to 19%, which is a real but much less marked difference. And so it goes.

Although Koch is not the first person to remove the statistical traps resulting from sample size effects, his technique is very straightforward and simple. Raup, who with Rex Crick developed in 1979 a "prohibitively expensive" computer program to address the problem, describes Koch's work as "a giant shortcut." Koch's method, says Raup, "produces a very good approximation and is therefore potentially very helpful."

Koch suggests that one way to avoid the sample size problem is simply to eliminate the rare species from the data sets. David Jablonski, of the University of Chicago, is not convinced. "Koch believes that our samples are hopelessly distorted," he says. "But if you look through a whole series of horizons you can reduce the problem and come close to getting a real biological signal out of the data." Jablonski does agree, however, that "There's no question that taking the fossil record at face value can be horribly misleading. Koch is right in saying that although ecologists are aware of the skewed distribution of species in ecosystems, paleontologists in general have not been. We have to be careful in evaluating fossil assemblages. You have to try to compare what you see in the fossil record with what you see in living ecosystems. When you get a good match you know you are looking at a real biological signal." ■ ROGER LEWIN

ADDITIONAL READING

C. F. Koch, "Prediction of sample size effects on the measured temporal and geographic distribution patterns of species," *Paleobiology* **13**, 100 (1987).

Supernova 1987A: Notes from All Over

As Supernova 1987A in the Large Magellanic Cloud continues on its inexorable course—it has currently reached magnitude 3.3, with perhaps another magnitude of brightening to go in the next few months before it begins its long, slow fade into oblivion—astronomers are following its every step with avid interest. Some recent developments:

A New Limit on the Mass of the Neutrino?

As the first reasonably close supernova in modern times, Supernova 1987A also has the distinction of being the first such outburst ever observed via neutrino radiation. The pulse of 11 neutrino events seen in Japan's Kamiokande II proton decay detector, together with the 8 events seen simultaneously in the IMB detector near Cleveland, have not only pinpointed the exact time of the explosion—7:35:41 a.m. Universal Time on 23 February 1987—but have gone a long way toward validating researchers' understanding of the physics of the explosion itself.

Just as important, as more than one theorist was quick to realize, the events can also be used to place an upper bound on the mass of the neutrino (or more precisely, the electron neutrino, which dominates the Kamiokande II and IMB signal). One such analysis was recently presented in *Nature* by John N. Bahcall of the Institute for Advanced Study in Princeton and Nobel laureate Sheldon L. Glashow of Harvard University.

If neutrinos actually have zero mass, they point out, then relativity theory says that the particles must move at precisely the speed of light, which means in turn that they will take a certain amount of time—about 170,000 years—to make the trip from the Large Magellanic Cloud. If neutrinos have a nonzero mass, however, then relativity says they must travel more slowly than light, which means that they will take somewhat longer to make the trip. For any given particle, the actual time delay depends on its ratio of mass to energy; thus, say Bahcall and Glashow, if one could assume that Supernova 1987A had emitted all its neutrinos at once, it would be a simple matter to determine the particles' mass: just plot the energy versus arrival time for all the detected events and read off the answer.

Intriguingly enough, the Kamiokande II and IMB events are spread out over several seconds, which is roughly what one would expect for a neutrino mass of a few electron volts (eV)—precisely the mass range that theorists have favored on the basis of cosmological arguments. Unfortunately, say Bahcall and Glashow, the neutrinos were most assuredly *not* emitted at the same instant. For one thing, the energies and arrival times of the observed events show no consistent pattern. For another, the best theoretical models of supernova explosions suggest that neutrinos were actually emitted over the course of several seconds, as the core of the supernova's massive progenitor star collapsed to form a neutron star.

So from a strictly logical standpoint, say the authors, there is no way to disentangle the effect of an unknown neutrino mass from the effect of an unknown spread in emission times. However, they argue that one can set an upper limit on the mass through a plausibility argument: unless we have been extremely unlucky, the various time delays have probably not conspired to make a very long neutrino pulse look like a very short pulse.

Consider the Kamiokande II record, in which the first eight events are tightly bunched within the first 2 seconds. (The remaining three come almost 10 seconds later.) If one assumes that the mass effect has not compressed this pulse during its flight by more than a factor of 2—a crude and admittedly arbitrary cutoff, but one that Bahcall and Glashow believe is conservative—then one can derive an upper bound for the neutrino mass: 11 eV, or considerably better than the best limits set in terrestrial laboratories (18 eV).

It should be said that not everyone agrees with this analysis. Well before 1987A, for example, University of Chicago physicist David Schramm and his colleagues had already done detailed calculations of supernova neutrino signals and had found that a neutrino mass would produce exactly the kind of pulse compression that Bahcall and Glashow reject. "You can set a reasonable limit of 30 eV," says Schramm. But he is reluctant to push the available data any further.

On the other hand, it is also worth noting that a mass limit as low as 11 eV, if true, would have large cosmological conse-