the ratio of the scanned areas of two known geographic regions to the ratio of their actual areas. From this comparison, we calculated a projection error as $1 - k$, where k is a constant such that

$$\left\{\frac{\text{scanned area of region1}}{\text{actual area of region 1}}\right\} = k \times \left\{\frac{\text{scanned area of region 2}}{\text{actual area of region 2}}\right\}$$

For consistency, the regions were arranged in this formula (assigned as region 1 or region 2) so that k was always less than 1. To be conservative, we tried to choose reference areas as different in latitude as possible but covering a large latitudinal range.

15. From projections of estimated species richness for these groups (*17*), arthropods should account for 94% of the estimate, versus 3.4% for plants, 2.1% for mollusks, and 0.5% for all chordates combined. Using these exact proportions changes the arthropod estimate by +0.02 populations in 10,000 km^2.
16. P. H. Raven, *The Futurist* **19**, 8 (1985).
17. P. M. Hammond, in *Global Biodiversity Assessment*, V. H. Heywood, Ed. (Cambridge Univ. Press, Cambridge, 1995), pp. 113–138.
18. T. L. Erwin, *Coleopt. Bull.* **36**, 74 (1982).
19. J. T. Legge, R. Roush, R. Desalle, A. P. Vogler, B. May, *Conserv. Biol.* **10**, 85 (1996).
20. J. Thomas and N. Webb, *Butterflies of Dorset* (Dorset Natural History and Archeological Society, Dorchester, UK, 1984).
21. I. Brown and P. R. Ehrlich, *Oecologia* **47**, 239 (1980).
22. K. J. Gaston, *Rarity* (Chapman and Hall, London, 1994).
23. P. H. Raven, *Bull. ESA* **Spring**, 4 (1983).
24. E. H. Rapoport, *Aerography: Geographical Strategies of Species* (Pergamon, Oxford, UK, 1982); M. D. Pagel, R. M. May, A. R. Collie, *Am. Nat.* **137**, 791 (1991); R. France, *ibid.* **139**, 342 (1992).
25. D. L. Hawksworth, Ed., *The Biodiversity of Microorganisms and Invertebrates: Its Role in Sustainable Agriculture* (CAB International, Wallington, UK, 1991).
26. S. L. Pimm, G. J. Russell, J. L. Gittleman, T. M. Brooks, *Science* **269**, 347 (1995); S. L. Pimm and R. A. Askins, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9343 (1995).
27. Food and Agriculture Organization of the United Nations, *Forest Resources Assessment: Tropical Countries* (Rome, 1990). This estimate of the rate of tropical deforestation may be conservative; the actual rate may be as high as 2% per year.
28. The model used for this estimate is $S = cA^z$, where S is the number of species, A is the area where the species are found, and c and z are constants. Empirical studies of a variety of taxa reveal a rough range of values of z from 0.15 to 0.35 [R. H. MacArthur and E. O. Wilson, *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, NJ, 1967); M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995)].
29. We thank D. Ackerley, C. Boggs, G. Ceballos, M. Feldman, J. Hellmann, M. Lachmann, H. Mooney, J. Pritchard, T. Ricketts, M. Tanaka, P. Vitousek, and W. Watt for comments on earlier drafts of the manuscript; and S. Daily, K. Freeman, L. Light, and V. Tubbesing for help with data collection. This research was supported by Peter and Helen Bing, the Pew Charitable Trusts, the Winslow Foundation, and the late LuEsther Mertz.

19 June 1997; accepted 7 August 1997

Extinction and the Loss of Evolutionary History

Sean Nee* and Robert M. May

Extinction episodes, such as the anthropogenic one currently under way, result in a pruned tree of life. But what fraction of the underlying evolutionary history survives when k of n species in a taxon are lost? This is relevant both to how species loss has translated into a loss of evolutionary history and to assigning conservation priorities. Here it is shown that approximately 80 percent of the underlying tree of life can survive even when approximately 95 percent of species are lost, and that algorithms that maximize the amount of evolutionary history preserved are not much better than choosing the survivors at random. Given the political, economic, and social realities constraining conservation biology, these findings may be helpful.

We approach questions about pruning the tree of life and the calculus of biodiversity (*1*), so forcefully raised by the current extinction crisis (*2*), in the context of theoretical clades that either have been growing exponentially throughout their history or have been of constant size, such that each time a new lineage has appeared by speciation another lineage has gone extinct. These extremes bracket the plausible dynamical histories of real clades. The radiations of both the New World and Old World monkeys are consistent with the exponential growth model (*3*), whereas the history of the Plethodontid salamanders is consistent with the constant size model (*4*). Logistic growth, in which diversity rises to some maximum, is a convenient model for macroevolutionary clade expansion as well as population growth (*5*). In this framework, exponential growth is the early phase of logistic growth, and the constant size model describes a clade that has been at its maximum size for some time. From the data of marine families compiled by Benton (*6*), to which the logistic model has been fitted (*5*), the number of families appears to have been roughly constant for about 200 million years before the Late Permian mass extinction.

Suppose k species are saved from a total of n. This may be done in many ways. At one extreme, the species may be picked at random with respect to their phylogenetic relationships—the "field of bullets" scenario (*7*); at another extreme, useful for comparison, the species may be chosen according to the following algorithm, which maximizes the amount of evolutionary history preserved. The $k - 1$ lowest nodes in a tree (counting from the root) are selected. These define k clades. One species from each clade is picked; if a clade has more than one species in it, then one is picked at random. Figure 1 illustrates the relation between species loss and the loss of evolutionary history and shows that this algorithm optimizes the amount of evolutionary history preserved.

If k species out of a total of n are saved, it is natural to express the amount of history preserved as a fraction of the total amount that could have been preserved if all n species had been saved. How can this "amount of evolutionary history" be measured? For many purposes, it may be best simply to count species as such. But, as emphasized by Vane-Wright and others (*1*, *8*), it is often useful to measure the loss at a more fundamental level; ultimately, it would be best to assess this loss at the genetic level, by some measure of underlying information molecularly coded in DNA. Proximally, we work here with the tree structure. The above algorithm clearly works whether the actual "lengths" of the branches are known, or merely the branching order of the nodes (although firmer estimates of the fraction saved can be made in the former case). Also, note that we assume all branch tips are equidistant from the root; more details of molecular evolution could give a picture in which such lengths varied, although it seems likely that our general conclusions will remain valid in these more general circumstances.

We now present approximate equations for the average fractional amount of evolutionary history preserved, $f(k,n)$, when we save k of the original n species, under various assumptions about the history of the clade (*9*). For a random set of species from a clade that has been of constant size (indicated by the subscript r, const.), the equation for $f(k,n)_{\text{r, const.}}$ is

$$f(k,n)_{\text{r, const.}} \approx \frac{\ln(k-1) + C}{\ln(n-1) + C} \qquad (1)$$

where C is Euler's constant, with a value of ~0.577. This is obviously only meaningful for $k > 1$. Numerical simulations show that this analytical approximation performs very well for $k > 3$.

For a random set of species from a clade

Department of Zoology, Oxford University, South Parks Road, Oxford, OX1 3PS, UK.

*To whom correspondence should be addressed. E-mail: sean.nee@zoo.ox.ac.uk

that has been growing exponentially (indicated by r, expon.),

$$f(k,n)_{\text{r, expon.}} \approx \frac{k\ln\left(\frac{2n}{k(k-1)}\right) + \sum_{j=2}^{k-1} j\ln\left(\frac{j+1}{j-1}\right)}{n-1} \quad (2)$$

Unlike the other three equations we present, Eq. 2 is only a good approximation for $k << n$ (*10*).

If, instead of choosing species randomly, we apply the maximizing algorithm, then, on average, the fractional history preserved by saving the best set from a clade that has been of constant size (indicated by m, const.) is

$$f(k,n)_{\text{m, const.}} \approx \frac{\ln(k-1) + C + k\ln\left(\frac{k(n-1)}{n(k-1)}\right)}{\ln(n-1) + C} \quad (3)$$

The corresponding expression for an exponentially growing clade is

$$f(k,n)_{\text{m, expon.}} \approx \frac{k\ln(en/k)}{n-1} \quad (4)$$

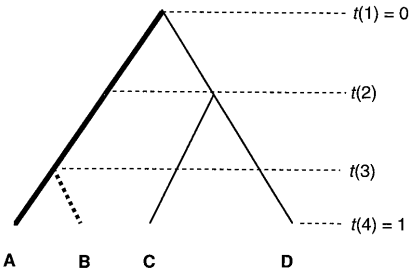

Fig. 1. The loss of species B results in the loss of the evolutionary history indicated by the dashed bold line, whereas if both A and B are lost, the entire history indicated by the bold line is lost. We chose a time scale such that the time from the first bifurcation to the present is 1, and $t(i)$ denotes the time of the ith node [when the $i + 1$ species appears; thus, $t(n-1)$ is the time at which the nth species appeared, $1 - t(n-1)$ time units ago]. For a tree with n species, $t(n)$ is the present and $t(n) \equiv 1$. The total branch lengths of a tree with n species, that is, the total amount of independent evolutionary history, is thus

$$\sum_{i=1}^{n} i[t(i) - t(i-1)] = n - \sum_{i=1}^{n-1} t(i)$$

regardless of the tree topology. A subset of k species defines a pruned tree with $k - 1$ nodes. If we save nodes $i, j, \ldots$, then the total saved evolutionary history is $k - t(i) - t(j) - \ldots$. Hence, the algorithm described in the text optimizes the amount of evolutionary history preserved. Note that only "living tissue" is of interest, so we ignored extinct lineages.

The average fraction of history preserved as a function of the number of species saved from clades of various sizes under our two growth scenarios is illustrated in Fig. 2. On average, for clades of size 50 and 500, about half the history is preserved by saving 20% of the species. In both cases, the increase in the fraction saved by the optimizing algorithm is, at best, about 10%. This result, that a large amount of evolutionary history can survive an extinction episode, is closely related to the result that a large amount of genetic variation can remain in a population immediately after a severe decline (*11*), when this latter result is considered from the viewpoint of the genealogies of genes (*12*).

The diminishing returns from saving more species, in terms of evolutionary history saved, is especially striking in larger clades that have been of constant size. For a clade size of 5×10^6 under the constant size model, choosing 5% of the species to be saved retains, on average, 81% of the total evolutionary history, in the sense defined above. This choice of a quantitative example is inspired by the greatest mass extinction to date, which occurred in the Late Permian 251 million years ago. [Although it was suggested that up to 95% of marine species (*7, 13*) perished in this extinction, this is probably an overestimate (*7, 14*).] Erwin and colleagues (*15*) compared the increase in diversity of marine animal taxa in the period after the Precambrian-Cambrian transition and in the period after the Late Permian mass extinction: species richness was very low at the start of both periods. They found a striking contrast. The higher taxa—phyla, classes, and orders—appeared in much greater numbers during the first period compared with the second. They argued that this may have been a result of the second radiation being seeded by species with a large variety of body plans already widely scattered among adaptive zones; hence, there were more limited evolutionary opportunities. Our contention—that a substantial proportion of the tree of life could survive even such a large extinction as occurred in the Late Permian—is entirely consistent with this interpretation.

We must now consider the nature of the variation around the averages we have been discussing. There are many sources of natural variation: variation in the times between nodes, variation in the tree topology, and, given a particular tree, variation as a result of the random sampling of species to be saved. It is this latter source of variation that we explore on the grounds that what we are ultimately interested in are the trees that actually exist in nature. We repeatedly sampled 12 species from a clade consisting of 64 species (about one-fifth of the species), which has grown exponentially to that size. The motivation for these choices is to provide a large scope for variation: our qualitative conclusions are unaffected. We imposed two extreme topologies, "comb" and "bush," and the resulting frequency histograms for 1000 random samples are illustrated in Fig. 3. The main features of these distributions are as follows. As is intuitively to be expected from the differences in tree topologies, random samples from bush topologies, on average, preserve more history than samples from comb topologies; the mean of the bush distribution is 0.418, whereas the mean of the comb distribution is 0.235. The mean for random samples of 12 from random clades of size 64 is 0.38. For both of the trees studied, the amount saved by the maximizing algorithm is 0.487. The amount saved by the maximizing algorithm is always independent of topology, depending only on the node locations. Samples from comb topologies exhibit greater variability in the amount of history preserved; the standard deviation

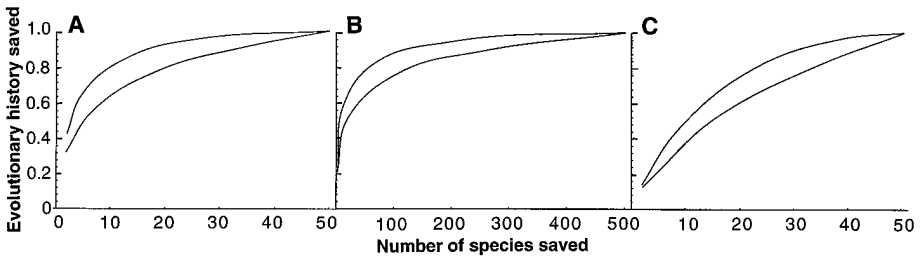

Fig. 2. The average amount of evolutionary history saved as a function of the number of species saved. The bottom curve in each panel is for a random sample of species, and the top curve is for a set of species chosen according to the optimizing algorithm. (**A**) A clade that has been of constant size, 50, throughout its history. (**B**) A clade that has been of constant size, 500, throughout its history. (**C**) A clade that has been growing exponentially to a present size of 50. The corresponding figure for a present size of 500 is visually indistinguishable, except that the x axis runs to 500 [as in (B)] and the left limits of the curves correspondingly are extrapolated in the direction of zero; this dependence on proportion, and independence of number, is evident in Eq. 4.

of the bush distribution is 0.027, whereas the standard deviation of the comb distribution is 0.041 (*16*). Studies of the topologies of contemporary phylogenetic trees (*17*) acquire a new significance in light of these results: they have been motivated up to now by interest in how the topology may provide information on the processes that produced the tree—now they can provide information about the remains of the tree as species depart.

Our macroscale approach to evolutionary history has its microscale counterpart in the emphasis conservation biologists place on "evolutionarily significant units" (*18*) and "phylogenetic species" (*19*); essentially, separately evolving lineages. There is considerable resistance to adopting the phylogenetic species concept: it would result in a doubling of the number of bird species, for example (*20*). The nomenclature debate aside, the promotion of the concept reflects a broad concern with independent evolutionary history per se.

At an intermediate scale of study, several authors have proposed a variety of methods for incorporating phylogenetic information into the weighting of species for conservation priority (*1, 8*). From one viewpoint, for example, the demise of the sole survivor of an ancient lineage (tuatara, for example) is a greater loss than the demise of a member of a rich species flock (any one grass snake species, for example). From another viewpoint (*21*), the sole survivor is a dead end and the species flock should be the target of concern, because its vigorous speciation potential will be needed to restock the world after the deluge. As Krajewski (*22*) put it, the argument is whether to focus on the branches or the twigs of the tree of life. We suspect that such debates about how to optimize choices may have limited applicability in practice, given the many economic and social factors that are likely to cut across conclusions drawn from academic "glass bead games" (*23*).

We have shown that much of the tree of life may survive even vigorous pruning and, perhaps more important, that the proportion surviving is relatively insensitive, on average, to whether the saved species are chosen randomly or optimally. However, community composition is less likely to survive such pruning; the Late Permian extinction was followed by the most pronounced biotic reorganization between the Cambrian explosion and the present (*13*). Furthermore, the very survival of the tree itself probably depends more on geophysiology (*24*) than on how vigorously it is pruned, and here questions of ecological services became paramount (*25*). Finally, we note that conservation concern will often be focused on individual species as such—on their behavior, potential usefulness, or unique role in an ecosystem—rather than simply on overall measures of evolutionary history. To make this personal, we note that the extinction of *Homo sapiens* could be seen as resulting merely in the loss of 5 million years of evolutionary history: this clearly illustrates that the conservation importance we assign to a species is not necessarily proportional to the amount of evolutionary history it represents.

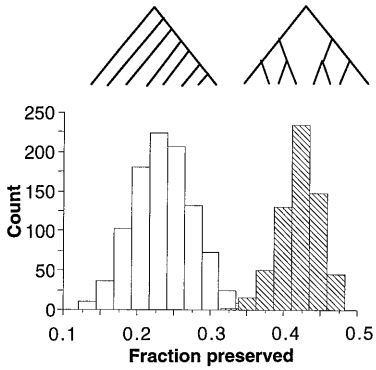

Fig. 3. Histograms of the fraction of evolutionary history saved by random samples of 12 species from a comb topology (left) and a bush topology (right) containing a total of 64 species. In the simulation study the times between nodes are the same for both topologies: the time between the *i*th and $(i + 1)$ node is $1/(i + 1)$, that is, the expected time interval under a pure birth process.

REFERENCES AND NOTES

1. The phrase "calculus of biodiversity," coined in R. M. May [*Nature* **347**, 129 (1990)], refers to any explicit quantitative procedure for weighting species for conservation priority; in particular, procedures that exploit the hierarchically structured relationships between species.
2. S. L. Pimm, G. J. Russell, J. L. Gittleman, T. M. Brooks, *Science* **269**, 347 (1995); J. H. Lawton and R. M. May, Eds. *Extinction Rates* (Oxford Univ. Press, Oxford, 1995).
3. A. Purvis, S. Nee, P. H. Harvey, *Proc. R. Soc. London Ser. B* **260**, 329 (1993).
4. S. Nee *et al*., in *Extinction Rates*, J. L. Lawton and R. M. May, Eds. (Oxford Univ. Press, Oxford, 1994), pp. 164–182.
5. V. Courtillot and Y. Gaudemer, *Nature* **381**, 146 (1996).
6. M. J. Benton, *Science* **268**, 52 (1995).
7. D. M. Raup, *Extinction* (Oxford Univ. Press, Oxford, 1993).
8. R. I. Vane-Wright, C. J. Humphries, P. H. Williams, *Biol. Conserv.* **55**, 235 (1991); D. P. Faith, *ibid.* **68**, 69 (1994).
9. All four equations are derived by fixing the times between nodes at the mean times for the appropriate probability distributions. The random sampling models exploit coalescence theory of gene genealogies from population genetics to provide the probability distributions of the times between nodes in the tree. For a random sample from a tree that has been growing exponentially, this probability distribution is implicit in a simulation algorithm presented by M. Slatkin and R. Hudson [*Genetics* **129**, 555 (1991)] and made explicit by S. Nee, R. M. May, and P. H. Harvey [*Philos. Trans. R. Soc. London Ser. B* **349**, 25 (1994)]. The expression for the maximizing algorithm treats the exponentially growing population as a birth process. The expressions for the constant size models are derived in a natural way by using the well-known distribution of coalescence times from population genetics, for example, J. Felsenstein, *Genet. Res. Cambridge* **59**, 139 (1992), which is discussed in a macroevolutionary context, most relevant for our purposes, in (*4*), and by J. Hey [*Evolution* **46**, 627 (1992)]. Hence, Eqs. 1 and 3 are derived as follows. For both the entire tree with n species and for the pruned tree of k randomly chosen species, the amount of time between nodes i and $i + 1$ is proportional to $1/i(i + 1)$. (The constants of proportionality all cancel so are not made explicit.) As there are $i + 1$ lineages between these nodes, this interval contributes an amount $(i + 1)/i(i + 1)$ to the total evolutionary history. So the total amount of evolutionary history of the pruned tree is

$$\sum_{i=1}^{k-1} \frac{i+1}{i(i+1)} = \sum_{i=1}^{k-1} \frac{1}{i} \approx \ln(k-1) + C$$

In exactly the same fashion, we find that the total amount of evolutionary history of the entire tree is $\ln(n - 1) + C$. Equation 1 in the text now follows immediately. For the derivation of Eq. 3, recall that the optimizing algorithm selects the first $k - 1$ nodes of the tree, defining the k species to be saved. The amount of evolutionary history contained in the pruned tree consists of the sum of (i) the amount of evolutionary history contained in the tree from its root up to the appearance of the $k + 1$ lineage and (ii) k multiplied by the amount of time between this event and the present. That is,

$$\ln(k-1) + C + k\left\{\sum_{i=1}^{n-1} \frac{1}{i(i+1)} - \sum_{i=1}^{k-1} \frac{1}{i(i+1)}\right\} \approx \ln(k-1) + C + k\ln\left(\frac{k(n-1)}{n(k-1)}\right)$$

The expression in the text now follows immediately.
10. S. Tavaré [*Theor. Popul. Biol.* **26**, 119 (1984)] provides a more explicit discussion of the validity of approximations of this sort.
11. M. Nei, T. Maruyama, R. Chakraborty, *Evolution* **29**, 1 (1975).
12. R. R. Hudson, *Oxford Surv. Evol. Biol.* **7**, 1 (1990).
13. D. H. Erwin, *Nature* **367**, 231 (1994).
14. S. M. Stanley and X. Yang, *Science* **266**, 1340 (1994).
15. _____, J. W. Valentine, J. J. Sepkopski, *Evolution* **41**, 1177 (1987).
16. As the size of the tree increases, keeping the fraction of species saved constant, the variation in the fraction of evolutionary history preserved *decreases*. If 300 out of 1600 species arranged in a comb topology and 96 out of 512 arranged in a bush topology are saved, the mean of the comb distribution is 0.191 with a standard deviation of 0.01 and the mean of the bush distribution is 0.427, also with a standard deviation of 0.01.
17. C. Guyer and J. B. Slowinski, *Evolution* **47**, 253 (1993); A. O. Mooers and S. B. Heard, *Q. Rev. Biol.* **72**, 31 (1997).
18. C. Moritz, *Trends Ecol. Evol.* **9**, 373 (1994).
19. M. J. Donohue, *Bryologist* **88**, 172 (1985); D. S. Hibbett and M. J. Donohue, *Conserv. Biol.* **10**, 1321 (1996).
20. G. Martin, *Nature* **380**, 666 (1996).
21. T. L. Erwin, *Science* **253**, 750 (1991).
22. C. Krajewski, *ibid.* **254**, 918 (1991).
23. H. Hesse, *The Glass Bead Game* (Pan Books, London, 1987).
24. J. E. Lovelock, *Philos. Trans. R. Soc. London Ser. B* **352**, 143 (1997).
25. J. H. Lawton, *Oikos* **71**, 1 (1994); S. L. Pimm, *Nature* **387**, 231 (1997); R. Costanza *et al*., *ibid*., p. 253.

26. S.N. acknowledges the Biotechnology and Biological Sciences Research Council (grant G04872). R.M.M. acknowledges the Royal Society of London. We thank D. Erwin for his stimulating comments.

20 May 1997; accepted 28 August 1997