Phylogenetic assessment of total biodiversity

R. H. Crozier,[1] K. Pedersen[2] and P-M. Agapow[1,3]

[1]Centre for Conservation Genetics and Department of Genetics, La Trobe University, Bundoora, Victoria 3083

[2]Institute of Cell and Molecular Biology, Department of General and Marine Microbiology, University of Göteborg, Medicinaregatan 9 C, S-413 19 Göteborg, Sweden

[3]Present address: Department of Biology, Imperial College, Silwood Park, Ascot, Berkshire, United Kingdom

Conservation of the other 99% is usually best served by the preservation of whole communities rather than of treasured species. Setting priorities for habitat preservation requires development of theory beyond anthropocentrism. A starting point comes from treating the coding DNA of the world's organisms as the ultimate measure of biodiversity. The numbers of genes in different organisms are important, and phylogenetic relationships are crucial. Groups of localities preserving more of the heritable information content are favoured over groups preserving less. Especially for bacteria and soil-dwelling organisms (such as protists and small invertebrates generally), the process should lend itself to automation, via the detection of unseen, even unseeable, organisms through universal primers for the amplification and sequence analysis of ribosomal DNA genes. The approach is illustrated by an analysis of bacterial communities from groundwaters near the sole remaining natural fission reactor, near Oklo, Gabon, showing the importance of statistical sufficiency in biodiversity estimation, the statistical assessment of species numbers, the shortcut potential of Higher Taxon Richness, and the interdependence of sites in determining their conservation worth.

INTRODUCTION

The conservation of biodiversity generally seems to fall into two approaches, both benefitting from genetical insights and technology. The most prominent of these is the effort to preserve species gravely threatened by small population size. Ideally the response to such threats would be to remove the factors which degraded the habitat and led to the demographic crisis, but this is not always possible and hence careful management of the population may be needed to combat inbreeding (Mangel *et al.* 1996). Charismatic large animals (such as koalas, cheetahs, and whooping cranes) tend to dominate such cases of crisis management, and indeed non-vertebrates are not likely to be threatened separately from the danger of total habitat loss, although exceptions occur. An exception is provided by social insect populations, because these are effectively small due to their reproductive division of labor (Pamilo and Crozier 1997).

The conservation approach based on single species tends towards anthropocentrism as its rationale for species selection (Metrick and Weitzman 1996). While any means of gathering support for conservation may be worthwhile, choosing species for preservation on the basis of anthropocentric appeal abandons rationality as the chief basis and hence is likely to lead to poor decisions and the wastage of resources.

The other main approach of conservation is the preservation of habitats. The scientific task is then how to choose localities for preservation, with the goal the preservation of the largest proportion of biodiversity over the long term (Crozier 1997).

A major question is the definition of biodiversity itself. The soundest way to avoid anthropocentrism and whimsy seems to be to develop Wilson's (1992) suggestion of biodiversity as genetic information content, defining this as the information in coding sequences (Crozier and Kusmierski 1994; Crozier 1997). Biodiversity as information content then weights species differentially according to phylogenetic distinctiveness, but also (in the very broad picture) according to the number of genes (Crozier 1997). Given estimates of about 80 000 genes for mammals and 20 000 for insects such as *Drosophila*, other things being equal a mammal species should be given no more than four times the conservation priority of an insect species. Given that the additional genes present in mammals mostly arise from the duplication of pre-existing genes and represent "fine-tuning" rather than major metabolic innovation (Loomis 1988; Miklos and Rubin 1996), a mammal species probably represents less than four times the genetical information content of a fruit-fly.

Comparisons of biodiversity estimators

Species diversity is determined by the number of species present and the evenness of their abundances. Because we are interested in the long-term preservation of genetic

Pp. 30–33 in *The Other 99%. The Conservation and Biodiversity of Invertebrates* ed by Winston Ponder and Daniel Lunney, 1999. Transactions of the Royal Zoological Society of New South Wales, Mosman 2088.

information content the evenness information is not pertinent. Consequently, simple species richness, the number of species present, is a better measure. However, species richness lacks phylogenetic information. A subsidiary problem is that species richness, in treating each species as equal no matter how similar it is to any other, is vulnerable to problems of species definition (although when very large numbers are being counted one hopes for stability against taxonomic uncertainties in particular groups).

Phylogenetic measures of species diversity (reviewed by Crozier 1997) present advantages, including the objective use of molecular information, but also face difficulties in large-scale implementation because the appropriate instrumentation (e.g., automated connection of the DNA-extraction and sequencing operations) is still under development. Higher taxon richness (Balmford *et al.* 1996) was conceived to estimate species richness from the number of higher categories (genera, tribes etc.) present. However, given that such higher categories represent the systematist's judgement of evolutionary distinctiveness, this measure is promising but not proven as an indicator of the more important phylogenetic distinctiveness of habitats. Calibration (e.g., Gaston and Blackburn 1996) of the distinctiveness of higher taxa is desirable before higher taxon richness can be used as a surrogate for complete phylogenetic analysis.

METHODS

The statistical methods for biodiversity assessment are susceptible to much improvement. There is a tendency to extrapolate from repeated samples to derive estimates of species richness, despite the availability of sample-coverage estimators which can yield confidence limits (Chao and Lee 1992; Bunge and Fitzpatrick 1993; Bunge *et al.* 1995; Chao *et al.* 1996). Confidence limits for biodiversity in terms of information content can also be estimated, perhaps less satisfactorily than the methods available for species richness, using bootstrap subsampling of the molecular dataset and the program CONSERVE (Crozier 1997) in conjunction with Felsenstein's PHYLIP package (Felsenstein 1993). The approach using CONSERVE is to compare possible sets of reserves in terms of their genetical information content and to test whether the various sets of reserves are significantly different from each other in the amount of genetical biodiversity which would be preserved.

The ability to place confidence limits on estimates of biodiversity is likely to be of increasing importance if the scientific recommendations of conservationists as to the relative conservation worth of alternative potential reserve systems are to be taken seriously. A recommendation stating that one set of reserves is statistically significantly better than another will carry considerably more force than one lacking statistical support.

APPLICATION

Molecular methods are applicable to all living things, but are particularly finding application in the study of microorganisms. We therefore illustrate the methods outlined above using a bacterial dataset, but stress that they are applicable in principle to any soil cryptobiota and ultimately to organisms generally.

Most species of bacteria cannot be cultured and hence are virtually undetectable except by the polymerase chain reaction (PCR) amplification of their genes. The number of unseen species is large: the number of bacterial species estimated to occur in a gram of Norwegian forest soil was 4 000 based on DNA-hybridization curves (Torsvik *et al.* 1990). New species have been reported from marine and terrestrial habitats on the basis of gene sequences plucked from samples using "universal" PCR primers for rRNA genes (reviewed by Crozier 1997). It is uncertain how accurately PCR-based methods reflect the unseen bacterial communities present, but the wealth of information they present is obviously better than none and, provided that any biases are consistent across localities, they will provide valid comparative information.

Systematic efforts to compare localities even in terms of the bacterial communities present are rare. Subterranean bacterial communities, however, present a case where species numbers are sufficiently low that a study should give a reasonable picture of the phylogenetic distinctiveness of different sites (Pedersen 1993).

A comparison of rDNA sequences from five subterranean sites from the Oklo natural reactor region of Gabon (Pedersen *et al.* 1996) provided an opportunity for testing the methods sketched above. The sequences derived from groundwater samples taken from boreholes, one per site, and each sequence results from a single bacterial cell. Bacteria with identical rDNA sequences were termed clones (although their sequences may differ in regions of the DNA other than that sequenced). In addition to the phylogenetic distinctiveness of the sites, the data allowed estimation of sample coverage — the

estimated proportion of the total numbers of clones and bacterial genera actually present that were sequenced. The number of clones sequenced was estimated to represent 16%–62% of those actually present (95% confidence limits). The coverage of genera was estimated to be 48%–93% or 16%–68%, depending on the genetical criterion used for assignment of clones to genera. These estimates indicate that coverage was quite high, although not complete. Sites differed greatly in the proportions of the major groups of bacteria present and this led to large differences between sites in their potential to preserve biodiversity. Combinations of sites generally but not always differed significantly in terms of biodiversity present: even the most biodiverse site by itself would preserve only about 80% of the total. Further, assignment of clones to genera showed that the number of these higher categories was highly predictive of the amount of biodiversity preserved (if the sites are protected), indicating that the genus level might be appropriate for application of the higher taxon richness approach in this case.

DISCUSSION

Although for the technical reasons given above application of phylogenetic methods for assessing biodiversity are currently more easily applied to bacterial communities (particularly in deep rocks) than to soil invertebrates, it is possible that the results obtainable will be more meaningful for invertebrates than for bacteria. Bacterial biodiversity is sufficiently poorly known so that, except for those species living in highly restricted and special habitats such as termite guts (Ohkuma *et al.* 1995; Ohkuma *et al.* 1996), it is uncertain that any species show sufficient endemism to be endangered (Fenchel *et al.* 1997). But it is also clear that, although plausible, belief in universally widespread distributions of microbes rests on a lack of evidence and judgement should be withheld until more data are available (Staley 1997). If, contrary to most current opinion, significant evidence of endemism is found for bacteria then results such as those outlined here imply the same kinds of conclusions for the preservation of biodiversity for bacteria as they would for macroorganisms (Crozier 1997). Alternatively if it is indeed true for microorganisms that "everything is everywhere" (Fenchel *et al.* 1997), then the observed differences between sites are still of major conservation interest as indicating the array of bacteria which are actively growing, reflecting site characteristics.

The observation that the numbers of genera were highly predictive of the biodiversity being preserved lends support to the general use of higher taxon richness as a speedy and relatively inexpensive means of estimating biodiversity. However, it is highly desirable to have this finding explicitly tested in eukaryotes. For example, bacterial genera are to some extent defined by sequence difference but we do not know how accurately placement in different genera predicts the evolutionary distinctiveness of eukaryotes.

Random removal of species from simulated phylogenetic trees showed that, depending on tree shape, very high proportions of tree length could be maintained even if most species were removed (Nee and May 1997). Especially noteworthy was the finding that as little as 5% of the species can maintain 80% of the tree length. In the bacterial case examined here, retention of just one site (20% of the total) in most cases led to the preservation of much less than 80% of the total biodiversity. The apparent difference between the analytical and the empirical studies may reflect the fact that the bacterial clones were distributed between sites non-randomly with respect to phylogenetic distinctness. It remains to be seen how general this phenomenon is in ecosystems generally: do different sites tend to have related species, or are species assorted at random to sites?

The statistical framework for biodiversity estimation that incorporates phylogeny requires further development. In particular the relationship between the precision of the present methods and the coverage of communities needs attention; studies involving bootstrap or jackknife subsampling might be helpful. Although the result presented above rests on quite high coverage of the subterranean community, which gives reasonable confidence in the result, cases in which coverage is lower would benefit from improved methods.

ACKNOWLEDGEMENTS

We thank the Australian Research Council (RHC) and the Swedish Nuclear Fuel and Waste Management and the Swedish Natural Science Research Council (KP) for support of this work.

REFERENCES

Balmford, A., Green, M. J. B. and Murray, M. G., 1996. Using higher-taxon richness as a surrogate for species richness: I. Regional tests. *Proceedings of the Royal Society of London B* **263:** 1267–274.

Bunge, J. and Fitzpatrick, M., 1993. Estimating the number of species: A review. *Journal of the American Statistical Association* **88:** 364–73.

Bunge, J., Fitzpatrick, M. and Handley, J., 1995. Comparison of three estimators of the number of species. *Journal of Applied Statistics* **22:** 45–59.

Chao, A. and Lee, S.-M., 1992. Estimating the number of classes via sample coverage. *Journal of the American Statistics Association* **87:** 210–17.

Chao, A., Yip, P. and Lin, H.-S., 1996. Estimating the number of species via a martingale estimating function. *Statistica Sinica* **6:** 403–18.

Crozier, R. H., 1997. Preserving the information content of species: genetic diversity, phylogeny and conservation worth. *Annual Review of Ecology and Systematics* **28:** 243–68.

Crozier, R. H. and Kusmierski, R. M., 1994. Genetic distances and the setting of conservation priorities. Pp. 227–37 in *Conservation Genetics* ed by V. Loeschcke, J. Tomiuk and S. K. Jain. Birkhäuser Verlag: Basel.

Felsenstein, J., 1993. *PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author.* Department of Genetics, University of Washington, Seattle.

Fenchel, T., Esteban, G. F. and Finlay, B. J., 1997. Local versus global diversity of microorganisms: cryptic diversity of ciliated protozoa. *Oikos* **80**(2): 220–25.

Gaston, K. J. and Blackburn, T. M., 1996. The tropics as a museum of biological diversity: analysis of the New World avifauna. *Proceedings of the Royal Society of London. [Biology]* **263**(1366): 63–68.

Loomis, W. H., 1988. *Four billion years. An essay on the evolution of genes and organisms.* Sinauer Associates: Sunderland, Massachusetts.

Mangel, M., Talbot, L. M., Meffe, G. K., Agardy, M. T., Alverson, D. L., Barlow, J., Botkin, D. B., Budowski, G., Clark, T., Cooke, J., Crozier, R. H., Dayton, P. K., Elder, D. L., Fowler, C. W., Funtowicz, S., Giske, J., Hofman, R. J., Holt, S. J., Kellert, S. R., Kimball, L. A., Ludwig, D., Magnusson, K., Malayang, B. S., Mann, S., Norse, E. A., Northridge, S. P., Perrin, W. F., Perrings, C., Peterman, R. M., Rabb, G. B., Regier, H. A., Reynolds, J. E., Sherman, K., Sissenwine, M. P., Smith, T. D., Starfield, A., Taylor, R. J., Tillman, M. F., Toft, C., Twiss, J. R., Wilen, J. and Young, T. P., 1996. Principles for the conservation of wild living resources. *Ecological Applications* **6**(2): 338–62.

Metrick, A. and Weitzman, M. L., 1996. Patterns of behavior in endangered species preservation. *Land Economics* **72**(1): 1–16.

Miklos, G. L. G. and Rubin, G. M., 1996. The role of the genome project in determining gene function: insights from model organisms. *Cell* **86**(4): 521–29.

Nee, S. and May, R. M., 1997. Extinction and the loss of evolutionary history. *Science* **278**(5338): 692–94.

Ohkuma, M., Noda, S., Horikoshi, K. and Kudo, T., 1995. Phylogeny of symbiotic methanogens in the gut of the termite *Reticulitermes speratus*. *FEMS Microbiology Letters* **134**(1): 45–50.

Ohkuma, M., Noda, S., Usami, R., Horikoshi, K. and Kudo, T., 1996. Diversity of nitrogen fixation genes in the symbiotic intestinal microflora of the termite *Reticulitermes speratus*. *Applied Environmental Microbiology* **62**(8): 2747–752.

Pamilo, P. and Crozier, R. H., 1997. Population biology of social insect conservation. *Memiors of the Museum of Victoria* **56:** 411–19.

Pedersen, K., 1993. The deep subterranean biosphere. *Earth-Science Reviews* **34**(4): 243–60.

Pedersen, K., Arlinger, J., Hallbeck, L. and Pettersson, C., 1996. Diversity and distribution of subterranean bacteria in groundwater at Oklo in Gabon, Africa, as determined by 16S rRNA gene sequencing. *Molecular Ecology* **5**(3): 427–36.

Staley, J. T., 1997. Biodiversity: are microbial species threatened? Commentary. *Current Opinion in Biotechnology* **8**(3): 340–45.

Torsvik, V., Goksøyr, J. and Daae, F. L., 1990. High density in DNA of soil bacteria. *Applied Environmental Microbiology* **56:** 782–87.

Wilson, E. O., 1992. *The diversity of life*. Harvard Univeristy Press: Cambridge, MA.