Does stress increase ecosystem diversity?

SIR — Peter D. Moore (*Nature* **306**, 17; 1984) exposes the general state of confusion about diversity of communities. He reports various authors, concluding that a certain degree of stress may be a factor increasing diversity of a community and quotes several examples, and he implies that stress can take the form of a physical limitation to overall productivity or of disturbance. However, in every case, stress is used to mean an increase or decrease of habitat heterogeneity or of the community itself by the addition or elimination of other species, or physical factors, or by simulating their action. In a quoted experiment on diversity of algae (A) grazed upon by coral reef fish species (B) which in turn are hunted by predatory fish (C), manipulation of both B and C affects diversity of A. If C is absent, then overgrazing reduces A's diversity. If B and C are absent, then A diversity is higher than in the absence of C only, but lower than in the presence of both B and C. (In the last instance, results may be inconclusive as a downward trend in algae diversity was observed at the end of the experiment).

It becomes clear that the real system being studied is not an algae assemblage but a complex structure composed of non-equivalent elements of A, B and C which are linked by a number of functions of which trophic appears to be the most important. The whole structure has its own diversity, or species richness, and by removing some of the species, particularly those functionally most important, we decrease its overall diversity which affects other parts of the system in a variety of patterns. The reason for that, most probably, is that competition acquires the sole role in shaping species composition and abundance pattern. It is not, however, the fact of impact of predator on algal diversity that I question but the term "stress". The problem immediately arises if we try to estimate what is the stress of the combined system of A, B and C in which A manifested the highest diversity. And what would happen if one more trophic level is added? Perhaps the real application of stress would be removal of B and/or C from the complete system A,B,C. Moreover, if B is absent, then there would be no stress on A and maximum stress on C (hunger in this case). What seems to be the case is that the concept of stress is a derivative of the currently applied notion of community. Depending on one's view, the stress is either present or absent. As the community unit is as subjective in most studies as in this discussion, too great a relativity of the stress makes it irrelevant as a theoretical concept. It leads to paradoxes where sometimes a keystone species increases diversity and sometimes it is necessary to speak of a "reverse keystone species effect" (Hixon, M.A. & Brostoff, W.N. *Science* **220**, 511; 1983).

In short, the "stress concept" says that diversity increases (decreases) if overall diversity of the system increases (decreases) by adding species: grazers and predators in this case. The new wording touches upon a circular argument. This mistake is more difficult to notice when a non-functional fragment of community is analysed in separation from the functional whole.

It seems that this, partly semantic, confusion results from our imperfect concepts of ecological systems (communities) stemming from our lack of a coherent theory of ecosystem organization and it leads to a misinterpretation of observed phenomena (Kolasa, J. & Biesiadka, E. *Acta biotheoretica*; submitted).

JERZY KOLASA

Yale Arbovirus Research Unit,
Yale University School of Medicine,
New Haven, Connecticut 06510, USA

How do you spell DNA?

SIR — In a provocative column, you discussed the theory and practice of artificial intelligence[1], using spelling correction as an example of a task in which automated computer procedures should be of great value. In the past ten years, a number of tools have been developed to solve similar problems in molecular biology.

The usual connection between molecular biology and language is made by drawing parallels between the bases of DNA and the letters of the alphabet, between codons and words, and between proteins and sentences. Expanding on this analogy, mutations in DNA sequences can be thought of as spelling errors, creating new sequences from those already in existence. The processes of evolution, whether of DNA or language, select which of these new sequences or words are to survive.

What then can computer programs tell us about these molecular spelling errors? One approach, pioneered by Needleman and Wunsch and elaborated by Sankoff, Sellers and others, maximizes matches, subtracting a weighted sum of mismatches and insertion/deletions, between two DNA or protein sequences. For example, AATCAG and ATTCG might be related by

A A T C A G
A T T C * G

where there is one point mutation and one insertion/deletion. Two sequences of length N have of the order of 2^{2N} possible relationships. For $N = 1,000$, this number is approximately 10^{600} so that an exhaustive search is impossible. Nonetheless, rigorous algorithms and programs exist to locate rapidly the optimal relationships for sequences of 1,000 or more bases. Many modifications and improvements exist. Long insertions/deletions can be weighted according to length. Other mutational events, such as inversions, are being incorporated into these rapid comparison algorithms.

These algorithms can be extended in a novel way to predict RNA secondary structure; the base-paired regions correspond to matches in the sequence comparison algorithms. Indeed, these basic methods have been adapted and applied to speech recognition, geological strata, handwriting recognition, bird song and gas chromatography. Many such applications and associated theory appear in a recent book edited by D. Sankoff and J.B. Kruskal[2]. In addition to spelling correction, the analogue to a dictionary search for a word is the comparison of a new sequence to an existing DNA or protein data base. These searches locate library sequences with regions of similarity to the new sequence, combining the concepts of dictionary and imperfect spelling. New tests are being devised to estimate the statistical significance of the similarities. Methods to perform these large searches have been developed, and were successfully applied in the recent discovery of sequence similarity between the transforming protein of a primate sarcoma virus and a platelet-derived growth factor[3,4]. Another recent computer finding indicates that an oncogene product appears to have arisen as a result of recombination of two unrelated cellular genes[5].

One of the most intriguing areas of DNA sequence analysis parallels the concept of language interpretation. Patterns such as repetitive DNA are frequently noticed before their meaning is understood. An important case is the search for promoter sequences in *Escherichia coli*[6]. DNA upstream from *E. coli* coding regions is presumed to contain base sequences that spell "begin transcription". These are searches for patterns of unknown composition and length which we know must occur, however imperfectly, within specific regions of DNA. Text editors, even those equipped to search for regular expression patterns[7], are not adequate to this task. If useful and rigorous algorithms are developed for these tasks of locating imprecise words of unknown spelling, new and nontrivial insights could result.

New techniques of pattern recognition in DNA and protein sequences are resulting from creating and applying concepts of mathematics, statistics and computer science appropriate to specific questions of molecular biology. As often happens in science, these methods may turn out to have broad applicability.

MICHAEL S. WATERMAN*

Departments of Mathematics and of Biological Sciences,
University of Southern California,
Los Angeles, California 90089, USA

1. *Nature* **306**, 637 (1983).
2. Sankoff, D. & Kruskal, J.B. (eds) *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* (Addison-Wesley, London, 1983).
3. Weiss, R. *Nature* **304**, 12 (1983).
4. Doolittle, R.F. *et al. Science* **221**, 275 (1983).
5. Naharro, G., Robbins, K.C. & Reddy, E.P. *Science* **223**, 67 (1984).
6. Hawley, D.K. & McClure, W.R. *Nucleic Acids Res.* **11**, 2237 (1983).
7. Aho, A.V., Hopcroft, J.E. & Ullman, J.D. *The Design and Analysis of Computer Algorithms* (Addison-Wesley, London, 1974).