



Number of database sequences matching the 551-ACR subset at various score thresholds. Two different scoring matrices, pam120 (squares) and pam250 (circles), produce similar results. Scores are on a logarithmic scale. The distribution is limited on the right (high scores) by the 551 ACRs matching themselves and on the left (lower score) by the random matching of unrelated sequences. Both curves exhibit a quasi-linear part corresponding to the progressive recruiting of significant matches. An abrupt change in the slope occurs at $S=60$ (pam120, $P \leq 2-5 \times 10^{-2}$) or $S=85$, pam250, $P \leq 2-5 \times 10^{-2}$) where random matches begin to occur. Up to 50% of the entries could be ACR-related.

arthropods and nematodes. For each subset, I compared every sequence (using the local alignment program Blastp³) to the other sequence subsets, recording all matches scoring above a significant threshold (see table).

I then submitted the resulting set of cross-phylum matching sequences (ACR-containing sequences) to self-comparison (again with Blastp and the same score threshold), and partitioned them into clusters of related sequences. For each similarity cluster, I then selected as a representative the member exhibiting the largest conserved region (computed in terms of total length of the cross-phylum matching segment(s)).

This protocol gave an ACR representative set of 551 independent sequences (see table). This result is remarkably insensitive to the details of the procedure (scoring matrices, significance threshold, and so on). Half the 551 ACR representatives correspond to enzymatic activities representing all basic cellular systems: DNA replication, transcription, translation, signal transduction, glycolysis, tricarboxylic acid cycle and oxido-reduction. Structural elements such as ribosomal proteins, histones, cytoskeleton and matrix components are all present. But about 200 ACR representatives correspond to sequences of less obviously central importance, and warrant further analyses.

The 551-representative set constitutes 1,061 distinct contiguous segments with lengths ranging from 18 to 1,497 residues, with an average of 131 residues and a median of 84. Those numbers in part depend on the local nature of the Blastp algorithm, and do not accurately reflect

the domain organization of the ancestral proteins.

It has previously been suggested that proteins are put together from a limited set of 1,000–7,000 exons⁴, or 1,000 distinct folding domains⁵. The coincidence of these numbers is striking. But ACRs are too variable in size simply to be interpreted as structural domains. Furthermore, more than 96% of the database sequences related to the ACR set match a single representative, suggesting that these ancestral segments are rarely independently shuffled in contemporary sequences. Remarkably, 40–60% of the sequences in the database appear related to the 551-ACR representative set (see figure). This small ‘ancestral’ subset thus constitutes a convenient resource for the fast screening and identification of new sequences (for instance numerous cDNA partial sequences) and the definition of motifs. The 551-representative ACR set is available on e-mail request (jmc@ncbi.nlm.nih.gov).

J.-M. Claverie

National Center for Biotechnology Information,
National Library of Medicine,
NIH, Bethesda, Maryland 20894, USA

- Green, P. *et al. Science* **259**, 1711–1716 (1993).
- Bairoch, A. *Nucleic Acids Res.* **20**, 2019–2022 (1992).
- Altschul, S. F. *et al. J. molec. Biol.* **215**, 403–410 (1990).
- Dorit, R. L., Schoenbach, L. & Gilbert, W. *Science* **250**, 1377–1382 (1990).
- Chothia, C. *Nature* **357**, 543–544 (1992).

Species concepts

SIR — The biological species concept (BSC) is a valuable heuristic that rightly focuses attention on the fundamental importance of reproductive and genetic connections among populations. Nevertheless, we disagree with Mayr’s notion¹, discussed by Diamond in *News and Views*², that the BSC applies easily to plants, and argue that those phenomena that provide the most difficulty for the BSC (for example asexual reproduction, hybridization, polyploidy) represent critical aspects of plant evolution rather than rare aberrations in a flora consisting mostly of a ‘good species’. We do not intend to belittle the BSC, nor to suggest an alternative, but argue that the unavoidable limitations of the BSC indicate the need for different species concepts for different taxa, so that the concept does not obscure the variety of biological processes.

For 83 per cent of the species considered by Mayr¹, the typological species concept gave unambiguous and accurate identification. Use of the BSC helps for another 10 per cent, but other species concepts (such as the evolutionary species concept) would probably do just as well for such a broad range of taxa (from

aspen to clubmosses). Indeed, no account is taken of how often ambiguous results might conceivably occur in such a diverse flora (for example by focusing on congeners, for which hybridization is more likely). Comparison of the success of different species concepts would in any case be more valuable than arguing that 93 per cent is an acceptable criterion of success. Unless the usefulness of other species concepts (such as the evolutionary³, phylogenetic⁴ or typological⁵ species concepts) in dealing with these taxa is also assessed, there is no valid comparison on which to base acceptance or rejection of a species concept.

Furthermore, the flora studied by Mayr is a small temperate one that includes few members of the groups that give species concepts the most difficulty, such as orchids. Mayr’s survey also does not address the performance of the BSC in describing geographical variation within a plant taxon.

Use of a particular species concept always depends on the questions considered. Those interested in speciation and reproductive isolation might not abandon the BSC even if another species concept helped resolve more difficulties, because the BSC embodies a rich background of evolutionary theory dealing with barriers to hybridization. Similarly, many plant species are asexual or apomictic, and well over a third have hybrid origins⁶, so botanists are understandably reluctant to abandon consideration of these factors when delimiting species.

The issue is not how universally a species concept applies to plants or animals, but the implications of the obvious difficulties in its application for some taxa. Even a single species that clearly does not fit the BSC leads an evolutionarily trained botanist to wonder if modes of speciation other than the traditional route of allopatry are important. We suggest that those instances yield insights into the speciation process for both plants and animals.

Randall Mitchell, David Bleakly, Robert Cabin, Raymund Chan, Brian Enquist, Ann Evans, Timothy Lowrey, Diane Marshall, Steven Reed, George Stevens
Department of Biology,
University of New Mexico,
Albuquerque, New Mexico 87131, USA
Nickolas Waser
Department of Biology,
University of California,
Riverside, California 92521, USA

- Mayr, E. *Am. J. Bot.* **79**, 222–238 (1992).
- Diamond, J. *Nature* **360**, 627–628 (1992).
- Ashlock, P. D. *Syst. Zool.* **28**, 441–450 (1979).
- de Queiroz, K. & Donoghue, M. *Cladistics* **4**, 317–338 (1988).
- Mayr, E. & Ashlock, P. D. *Principles of Systematic Zoology* (McGraw Hill, New York, 1991).
- Riesberg, L. H. & Soltis, D. E. *Evol. Trends Plants* **5**, 65–81 (1991).