

- phate, 10 Na-phosphocreatine, and 0.6 EGTA (pH 7.2). Neurons were perfused with artificial cerebrospinal fluid (ACSF) of the following composition (in mM): 119 NaCl, 2.5 KCl, 4 CaCl₂, 4 MgCl₂, 26.2 NaHCO₃, 1 NaH₂PO₄, and 11 glucose; then the neurons were gassed with 5% CO₂ and 95% O₂ at 28°C. Responses to glutamate were evoked by adding (in μ M) 40 L-glutamic acid γ -(α -carboxyl-2-nitrobenzyl) ester (Molecular Probes), 1 tetrodotoxin, 100 picrotoxin, 100 APV, and 1 3-((R)-2-carboxypiperazin-4-yl)-propyl-1-phosphoric acid. Ten to 25 ms of flash (100 W of Hg; quartz optics, Zeiss Axiovert 135) was delivered to a spot (\sim 50- μ m diameter) positioned over dendrites.
16. Organotypic slice culture of hippocampus was prepared as described [L. Stoppini *et al.*, *J. Neurosci. Methods* **37**, 173 (1991)]. Hippocampal slices (400 μ m) were cut from postnatal 6- to 8-day-old rats with tissue chopper and cultured in medium as described [W. Musleh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9451 (1997)]. Slices were placed in a recording chamber and perfused with ACSF. Synaptic transmission was evoked (0.1 Hz, 100 μ s, \sim 10 V) with a glass stimulation electrode (1 to 2 megaohms, filled with 3 M NaCl), positioned in CA1 stratum radiatum. Tetanic stimuli consisted of two 100-Hz trains of 1-s duration separated by 20 s.
 17. D. L. Pettit *et al.*, *Science* **266**, 1881 (1994).
 18. R. S. Petralia *et al.*, *J. Comp. Neurol.* **385**, 456 (1997).
 19. A. Miyawaki *et al.*, *Nature* **388**, 882 (1997).
 20. The distribution of GluR1 and GluR1-GFP was measured quantitatively with immuno-gold labeling (78). Sections from dendritic regions \sim 100 μ m from the cell bodies were examined. To identify a cell infected with GluR1-GFP virus, we identified a dendritic segment (\sim 1 μ m in diameter) with immunolabeling. Comparable regions (with dendritic segments \sim 1 μ m in diameter) were chosen in tissue from uninfected P10 animals for analysis of endogenous GluR1. For each section, the label was identified as being in one of the following compartments: A, dendritic cytosol; B, dendritic surface; C, spine, non-PSD; and D, PSD. The total number of grains counted was 515 for endogenous GluR1 and 785 for GluR1-GFP. For background subtraction, 50 presynaptic terminals were randomly chosen, label was counted, and density was calculated (0.17 gold per square micrometer). This background density constituted 10 to 20% of the signal found in dendritic cytosol (0.92 gold per square micrometer) and was subtracted to reach dendritic cytosol value (A, above); this region is the only compartment with sufficient area to be affected by background labeling.
 21. K. Svoboda *et al.*, *Science* **272**, 716 (1996); W. Denk and K. Svoboda, *Neuron* **18**, 351 (1997); Z. F. Mainen *et al.*, *Methods*, in press. Optical stacks (30 μ m by 30 μ m by 50 μ m; step size, 0.5 μ m) were captured about every 15 min with a custom-built two-photon laser scanning microscope [Zeiss, 63X objective; Ti: sapphire laser tuned to wavelength (λ) \sim 900 nm]. Stimulation electrode was placed nearby dendrites expressing GluR1-GFP under visual guidance. Quantification of fluorescence from images was carried out with custom-made programs in IDL (Research Systems). Images were imported into Adobe Photoshop for figure presentation. Each displayed image is generally the average of three to five optical sections. In some cases, an image is the composite of images captured at slightly different (\sim 1 μ m) optical sections; this was necessary to offset tilting of structures over the course of hours of examination.
 22. Single-spine Ca²⁺ imaging studies with the same stimulus electrodes (tip resistance, 1 to 2 megaohms) and parameters indicate that synaptic excitation in this preparation is sparse. Thus, as in previous studies (21, 37), the tip of the stimulating electrode must be placed around 5 to 15 μ m from imaged postsynaptic regions.
 23. Within 3 SD of the noise measured in background regions.
 24. The random fluctuation of GluR1-GFP fluorescence was measured in six spines identified during control conditions. At these spines, the GluR1-GFP intensity decreased 344 ± 259 AU (mean \pm SD) from one observation period to the next (from 1755 ± 613 AU to 1411 ± 478 AU). Thus, the increase in spine GluR1-GFP fluorescence after tetanus differs (by 6 SD for "active" spines and by 7 SD for "empty" spines) from the average changes seen in the absence of tetanus. This indicates that such events are unlikely to occur from chance alone. Furthermore, the density of GluR1-GFP-containing spines detected after tetanus in the analyzed dendrites ($0.93 \pm 0.12 \mu\text{m}^{-1}$ over 10 dendritic segments $7.5 \pm 4.7 \mu\text{m}$ in length) was \sim 16 SD outside the mean density of GluR1-GFP-containing spines detected in the absence of tetanus ($0.021 \pm 0.055 \mu\text{m}^{-1}$, mean \pm SD, measured over 30 randomly chosen dendritic segments).
 25. After selecting a region of interest (about 8 μ m by 8 μ m by 3 μ m), an autocorrelation function, $A(r)$, of fluorescence intensity, $I(x, y, z)$, was computed as a function of distance, r :

$$A(r) = \frac{\sum_x \sum_y \sum_z \sum_{\theta r} [I(x, y, z) * I(x_{\theta r}, y_{\theta r}, z)]}{[I(x, y, x) + I(x_{\theta r}, y_{\theta r}, z)]^2}$$
 where θr indexes x and y over 72 locations about a circle centered at x, y and of radius r . This function decayed smoothly and monotonically as a function of distance (Fig. 5B), reflecting the nonperiodic distribution of fluorescence. We used the distance at which the autocorrelation function reached 50% decay ($R_{50\%}$) as an index of signal homogeneity.
 26. This decrease is 3.6 SD beyond those changes seen in the absence of stimulation, indicating a small likelihood that these changes occur from chance alone.
 27. J. M. Bekkers and C. F. Stevens, *Nature* **346**, 724 (1990).
 28. K. Harris *et al.*, *J. Neurosci.* **12**, 2685 (1992).
 29. M. Maletic-Savatic *et al.*, *Science* **283**, 1923 (1999).
 30. P.-M. Lledo *et al.*, *ibid.* **279**, 399 (1998).
 31. A. Nishimune *et al.*, *Neuron* **21**, 87 (1998); P. Osten *et al.*, *ibid.*, p. 99; I. Song *et al.*, *ibid.*, p. 393.
 32. J. Spacke and K. M. Harris, *J. Neurosci.* **17**, 190 (1997).
 33. M. Maletic-Savatic and R. Malinow, *ibid.* **18**, 6803 (1998); M. Maletic-Savatic *et al.*, *ibid.*, p. 6814.
 34. Z. Nusser *et al.*, *Neuron* **21**, 545 (1998).
 35. U. Frey and R. G. Morris, *Nature* **385**, 533 (1997).
 36. P. Worley, *Neuron* **21**, 936 (1998).
 37. R. Malinow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8170 (1994); R. Yuste and W. Denk, *Nature* **375**, 682 (1995); H. Takechi *et al.*, *ibid.* **396**, 757 (1998).
 38. We thank S. Schlesinger for Sindbis virus expression system; P. De Camilli for antibody against synapsin I; C. S. Zucker for antibody to GFP; J. Boulter, S. F. Heinemann, and P. H. Seeburg for cDNA clones; N. Dawkins-Pisani, B. Burbach, and P. O'Brian for technical assistance; and Y.-X. Wang for assistance in the immunogold studies. Y.H. is a recipient of research fellowships from Japan Society for the Promotion of Science and Uehara Memorial Foundation. S.H.Z. was a recipient of Wellcome Trust Fellowship. This study was supported by the NIH (R.M. and K.S.), the Mathers Foundation (R.M.), and the Human Frontier Science Program, Pew, and Whitaker Foundations (K.S.).

1 February 1999; accepted 4 May 1999

Reconstructing Phylogeny with and without Temporal Data

David L. Fox,* Daniel C. Fisher, Lindsey R. Leighton

Conventional cladistic methods of inferring evolutionary relationships exclude temporal data from the initial search for optimal hypotheses, but stratocladistics includes such data. A comparison of the ability of these methods to recover known, simulated evolutionary histories given the same, evolved character data shows that stratocladistics recovers the true phylogeny in over twice as many cases as cladistics (42 versus 18 percent). The comparison involved 550 unique taxon-by-character matrices, representing 15 evolutionary models and fossil records ranging from 100 to 10 percent complete.

Phylogenetic analysis seeks to identify the pattern of historical relationships among organisms. However, controversy persists over not only what constitutes an appropriate inference strategy, but also what categories of data are acceptable as evidence (1). One recent debate focuses on the use of stratigraphic data, or the

temporal order of specimens in the fossil record, as evidence for inferring relationships (2). Proponents of cladistic methods often argue that temporal data may be misleading as indicators of relationship (3). Arguments supporting this position are essentially a priori and do not address the relative efficacy of

methods that do and do not use temporal data. We explore the efficacy of using temporal data through simulations of evolutionary histories and associated character data, assessing the relative performance of two phylogenetic methods, cladistics and stratocladistics.

Both conventional cladistic analysis and stratocladistics rely on parsimony, in the sense of minimizing ad hoc auxiliary hypotheses, to evaluate alternative interpretations of relationship. Cladistic analysis selects interpretations that minimize hypotheses of homoplasy, or shared traits that do not result from common ancestry (4, 5). Hypotheses of homoplasy are counted individually and, lacking contravening evidence, equally.

Stratocladistics (6) incorporates stratigraphic data into the logic of cladistic hypothesis choice by selecting interpretations of relationship that minimize hypotheses of homoplasy and nonpreservation of lineages through intervals that preserve other lineages under analysis, giving neither category of

data priority. Like cladistics, stratocladistics judges hypotheses in terms of consistency with data in hand; any choice may be tested with more data, whether from new specimens, new characters, or stratigraphic range extensions or refinements. Although not necessary, explicit hypotheses of ancestry may emerge from stratocladistic analysis. By incorporating time and allowing ancestors, stratocladistics operates at the level of phylogenetic trees, not cladograms, and thus yields hypotheses that are more specific and more easily refuted with additional data (7).

The basic measure of fit between data and hypothesis in stratocladistics is "parsimony debt," or the total number of ad hoc hypotheses required to explain a given body of character and stratigraphic data; more debt indicates less parsimonious solutions. Character debt easily converts to treelength, the conventional measure of fit in cladistics, and adding stratigraphic parsimony debt yields "stratigraphically augmented treelength." In Fig. 1A, the phylogenetic tree is consistent with the character data and hence has 0 character debt. However, the lineage leading to A from the hypothetical ancestor Y must pass through two intervals without representation by fossils, even though B and C are preserved in those intervals. This requires two ad hoc hypotheses of nonpreservation, and hence two units of stratigraphic debt. In contrast, the tree in Fig. 1B predicts the order of occurrence of taxa in the fossil record, but requires two character reversals to explain the resemblance between the traits of taxon A and the conditions taken as primitive. Thus, this hypothesis has no stratigraphic debt, but two units of debt for character data. Both trees are equally parsimonious overall; additional data, either character or stratigraphic, are needed to select between them. Like equal weighting of characters, equal weighting of character and stratigraphic debt is not required, but is chosen unless evidence indicates otherwise.

Our simulation program is a time-homogeneous branching model that tracks the evolution of 50 binary "morphological" characters through discrete stratigraphic intervals. The program has five parameters. Speciation (λ) and extinction (μ) rates control the branching behavior of the model and essentially determine the stratigraphic scope of simulations in model units. Character state transition probabilities (F and B) control character evolution. Forward transitions (F) are from primitive to derived; back transitions (B) are reversals. In each interval, characters are checked within each lineage for evolution, accommodating change during both anagenesis and cladogenesis. Values selected for these first four parameters

constitute an evolutionary model. The final parameter, OTU-loss, converts a complete simulated history to an incomplete fossil record. Our operational taxonomic units (OTUs) are lineage segments, the representative of each lineage in each stratigraphic interval. OTU-loss specifies a percentage of OTUs that are chosen at random for deletion (8).

The branching model begins with a single lineage, with all 50 characters showing the primitive condition. This primitive taxon is always preserved and identified as such in all analyses so that it may act as an outgroup (9). Outgroups play a pivotal role for cladistics, but not for stratocladistics, so their inclusion favors cladistics. The branching model tracks evolution of the characters through anagenesis and cladogenesis, until at least seven lineages evolve (10). With a threshold greater than seven lineages, the number of OTUs in complete histories often leads to more cladograms than can be saved with the current version of PAUP (11). The model runs to a maximum of 26 intervals, a limitation imposed by the maximum number of states in the stratigraphic character in MacClade, the software used for stratocladistic analysis (12). Each run of the model generates a single simulated history from which 10 taxon-by-character matrices are extracted by incrementing OTU-loss in 10% steps from 0% OTU-loss (a "complete" fossil record) to 90% OTU-loss (a greatly impoverished fossil record). Incrementing OTU-loss in this manner allows examination of the effects of the completeness of the fossil record on the relative performance of stratocladistics and cladistics. This percentage value is different from per-individual preservation probability (which might be vanishingly small for the real fossil record). Each lineage, in each unit of model-time, corresponds to many individual organisms, of which only one needs to be preserved and studied to count the lineage in that interval. Estimates of real familial, generic, and species-level completeness for a variety of invertebrates and vertebrates range from 60 to 90% (13).

We explored multiple models to determine the relative behavior of stratocladistics and cladistics in parts of the parameter space. We have analyzed two sets of experiments, both with $\lambda = 0.15$ and $\mu = 0.08$. These values were chosen, like the threshold of seven lineages, to yield complete histories that would not exceed software limits. The first set includes 10 evolutionary models, each with $F = 0.1$, whereas B increased from 0.01 to 0.1 in steps of 0.01. We replicated each of these models five times, yielding 50 unique histories and 500 taxon-by-character matrices. In the second set, $F = 2B$, and B had values of 0.025, 0.05, 0.1, 0.15, 0.2, and 0.25. Of these six evolutionary models, one ($F = 0.1, B = 0.05$) was included in the first set of experiments; we retained all five replicates of this model in the second set, but did not replicate the other models in the second set. Thus,

the second set of experiments includes 10 unique histories and 100 data matrices. Transition probabilities were equal for anagenetic and cladogenetic change. We presume that models with low transition probabilities (≤ 0.1) are more realistic than those with higher values, but "real" values of these parameters are unknown. The 550 unique data matrices encompass phylogenetic problems with ingroups that range in size from 4 to 61 OTUs (mean = 23) and simulated stratigraphic ranges from 6 to 26 intervals long (mean = 15).

For each simulation, we analyzed the matrices representing the 10 fossil records cladistically and stratocladistically. Cladistic analyses of the equally weighted morphological characters were done with the tree bisection-reconnection heuristic search algorithm in PAUP; we replicated each search 50 times using random seeds, swapping only on minimal trees (14). Stratocladistic analyses were done with the same morphological characters used in PAUP and the stratigraphic character (15) and "make ancestor" tool in MacClade. Currently, no computer algorithm performs complete stratocladistic analyses, partly because of the increased number of hypotheses involved when considering phylogenetic trees rather than just cladograms. Thus, our stratocladistic analyses are not exhaustive searches of all possible combinations of topologies and ancestors. Our method was first to sort the set of cladograms from the PAUP analysis by stratigraphically augmented treelength and use the overall most parsimonious cladistic topologies among the cladistic results as starting points for manual branch-swapping in search of new, stratocladistically parsimonious topologies rejected by cladistics on the basis of character data alone. Whether or not new topologies were identified, the stratocladistic hypotheses were then

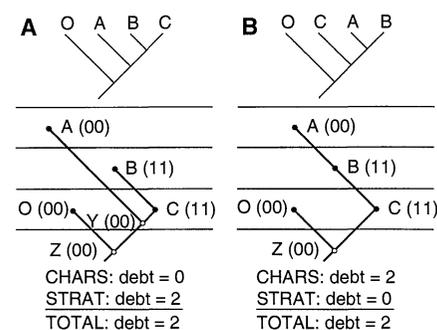


Fig. 1. (A and B) Examples illustrating character, stratigraphic, and total parsimony debt for three ingroup taxa (A, B, and C), one outgroup (O), and hypothetical ancestors (Y and Z). At the top are alternative cladograms and below each is an associated phylogenetic tree. The two characters shown in parentheses occur in primitive (0) and derived (1) states in O, Y, and Z and in B and C, respectively. The hypotheses differ in how they interpret the 0 states in taxon A. Vertical position of taxa indicates superpositional relations.

Museum of Paleontology, University of Michigan, 1109 Geddes Road, Ann Arbor, MI 48109-1079, USA.

*To whom correspondence should be addressed. E-mail: dlfox@umich.edu

identified from the overall most parsimonious cladistic topologies with the "make ancestor" tool in MacClade. The stratocladistic hypotheses are the shortest hypotheses after testing each OTU as an ancestor and leaving as ancestral those that do not increase stratigraphically augmented treelength. True phylogenies for each simulation were reconstructed from model output after completion of phylogenetic analyses.

Comparing cladistic (that is, not stratigraphically augmented) treelengths for our results, the true cladogram, which cannot be shorter than the cladistic result, is typically longer than the arrangement of taxa preferred by cladistics. For all 550 analyses, the median treelengths without stratigraphy or ancestors for cladistics, stratocladistics, and true phylogenies are 111, 127, and 144 steps, respectively. This implies that cladistics underestimates the length of the true phylogeny and that stratocladistics often selects topologies rejected by cladistics.

Stratocladistics produced far fewer hypotheses than cladistics from the same character data. The mean and median numbers of cladistic topologies were 298 and 12, with a maximum of 18,900 (for an analysis that yielded a single stratocladistic topology); the mean and median numbers of stratocladistic topologies were 2.4 and 1, with a maximum of only 44. To check whether our simulated character data

might be preserving less phylogenetic information than would real data, we compared retention indices (RIs) of our cladistic results with published values for 80 data matrices that include fossils (16). The published matrices had a mean RI of 0.76 ± 0.12 , whereas the mean RI was 0.85 ± 0.06 for our first set of experiments ($F = 0.1, n = 500$), 0.57 ± 0.17 for our second set ($F = 2B, n = 50$), and 0.73 ± 0.11 for all 550 simulations. Paired *t*, Mann-Whitney, and two-sample Kolmogorov-Smirnov tests show that these differences in RI are statistically significant. Characters in our first set of experiments thus preserve more phylogenetic signal than real examples, the second set preserves less, and the combined experiments differ least from published studies.

To compare the performance of each method, we calculated all pairwise consensus fork indices (CFIs) between the topology of the true cladogram and each of the topologies in the cladistic and stratocladistic result-sets. CFI expresses the proportion of subclades shared between cladograms (17). Only the cladistic topology of the stratocladistic hypotheses was considered. CFI is unbiased by tree shape and ranges from 0, if compared cladograms share no subclades, to 1, for identity. It is a conser-

vative metric, as the position of a mismatch within a topology, basal or crownward, does not affect its value. The maximum, average, and minimum CFIs for both methods for all data matrices indicate that stratocladistics does better than cladistics at estimating the true phylogeny (Fig. 2). The maximum CFI (CFI_{max}) is a method's best approximation to the true cladogram, and the minimum is that method's worst guess at the true cladogram. Focusing on CFI_{max} , the stratocladistic result included a hypothesis closer to the truth than the closest cladistic hypothesis in over half of the 550 data matrices (Fig. 2A). Stratocladistics recovered the true phylogeny in 42% of the simulations; in 28%, stratocladistics recovered the true phylogeny and cladistics did not. Cladistics recovered the true phylogeny in only 18% of the simulations; in only 4% did cladistics recover the true phylogeny whereas stratocladistics did not. Considering both recovery of the true phylogeny and numbers of topologies generated, the 52% of all data sets in which stratocladistics outperformed cladistics yielded a mean of 3.3 stratocladistic topologies, compared with 316.0 cladistic topologies. In the 33% of data sets for which the methods tied on CFI_{max} , a mean of 1.6 stratocladistic topologies opposed 301.3

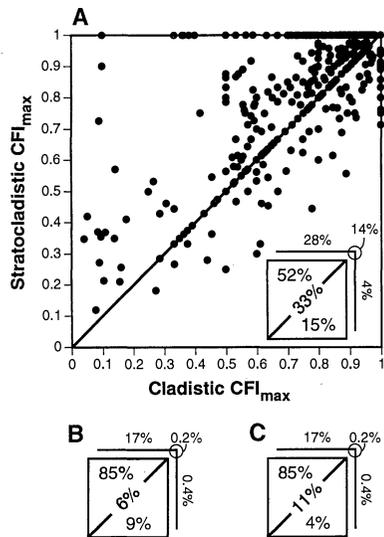


Fig. 2. (A) Maximum consensus fork index for cladistics versus the maximum consensus fork index for stratocladistics for all data. (Inset) Distribution of data within (A). Larger font indicates percentage of data above, below, and on the diagonal of equal CFIs. Smaller font along upper and right margins indicates percentage of total cases in which the methods identified the true cladogram (stratocladistics at top, cladistics on right, both methods at upper right corner). (B) Distribution of CFI_{mean} . (C) Distribution of CFI_{min} . Cases with $CFI_{min} = 1$ necessarily have $CFI_{mean} = 1$ also. $n = 550$ in all plots.

Table 1. Mean CFI_{max} and results of sign tests by percent OTU-loss. $n = 55$ in all cases. NS, not significant.

Percent OTU-loss	Mean CFI_{max} (stratocladistics)	Mean CFI_{max} (cladistics)	No. of cases		Sign
			S > C	C > S	
0	0.945	0.879	40	1	<0.001
10	0.935	0.862	40	2	<0.001
20	0.921	0.848	39	6	<0.001
30	0.898	0.838	34	5	<0.001
40	0.878	0.814	31	9	<0.001
50	0.842	0.766	28	10	<0.001
60	0.834	0.779	25	8	<0.001
70	0.796	0.761	16	11	NS
80	0.753	0.719	16	16	NS
90	0.713	0.699	16	12	NS

Table 2. Mean CFI_{max} and results of sign tests by evolutionary model.

F	B	N	Mean CFI_{max} (stratocladistics)	Mean CFI_{max} (cladistics)	No. of cases		Sign
					S > C	C > S	
0.05	0.025	10	0.944	0.957	0	2	NS
0.10	0.010	50	0.828	0.850	11	17	NS
0.10	0.020	50	0.891	0.855	19	2	<0.001
0.10	0.030	50	0.853	0.853	24	15	NS
0.10	0.040	50	0.950	0.889	22	4	<0.001
0.10	0.050	50	0.894	0.869	24	5	<0.001
0.10	0.060	50	0.813	0.801	16	14	NS
0.10	0.070	50	0.924	0.870	33	8	<0.001
0.10	0.080	50	0.864	0.726	41	2	<0.001
0.10	0.090	50	0.862	0.792	29	3	<0.001
0.10	0.100	50	0.874	0.728	42	3	<0.001
0.20	0.100	10	0.841	0.749	6	1	NS
0.30	0.150	10	0.537	0.569	0	3	NS
0.40	0.200	10	0.322	0.244	8	1	<0.001
0.50	0.250	10	0.446	0.121	10	0	<0.001

cladistic hypotheses. Even in the 15% of cases in which cladistics outperformed stratocladistics, the closest topology was, on average, hidden among 227.5 topologies, compared with 1.4 for stratocladistics. Comparisons of average and minimum CFIs (Fig. 2, B and C) favor stratocladistics even more strongly.

To examine the relation between the variables in our study and values of CFI_{max} , we grouped the data in Fig. 2A by percent OTU-loss and by F and B (Tables 1 and 2). Each group of results was subjected to a sign test to determine if CFI_{max} was significantly higher for one method or the other (18). In most cases, mean CFI_{max} is higher for stratocladistics and the sign tests are highly significant ($P < 0.001$) in favor of stratocladistics. In three cases, mean CFI_{max} is higher for cladistics, but the differences from CFI_{max} for stratocladistics are not significant. Of the remaining six sign tests that were not significant, only one (OTU-loss = 80%) did not have more cases with higher stratocladistic CFI_{max} than cladistic CFI_{max} . Thus, stratocladistics generally outperforms cladistics to a statistically significant degree, and this outcome is largely independent of the evolutionary model and completeness of the fossil record assumed.

Our results lay to rest the notion that an incomplete fossil record yields no clues for inferring phylogeny. The stratigraphic order of taxa preserved in the fossil record is a complex function of presence and absence controlled by many physical, chemical, and biological factors. To ignore this pattern or dismiss this class of data when it does not agree with a phylogenetic hypothesis makes an unwarranted "covering assumption" (5) that masks the real weight-of-evidence and encourages less than even-handed treatment of data. Stratocladistic hypotheses attempt to explain both the distribution of characters among taxa and the distribution of taxa through time. By doing this, stratocladistic hypotheses explain more features of the natural world and hence have greater explanatory power than purely cladistic hypotheses.

References and Notes

1. D. L. Hull, *Science as a Process* (Univ. of Chicago Press, Chicago, IL, 1988); E. Sober, *Reconstructing the Past: Parsimony, Evolution and Inference* (MIT Press, Cambridge, MA, 1988).
2. A. B. Smith, *Systematics and the Fossil Record* (Blackwell, London, 1984). See also *Nature Debates* [online] (19 November 1998) at <http://helix.nature.com/debates/index.html>.
3. B. Schaeffer, M. K. Hecht, N. Eldredge, in *Evolutionary Biology*, T. Dobzhansky, M. Hecht, W. C. Steere, Eds. (Appleton-Century-Crofts, New York, 1972), vol. 6, chap. 2; G. Nelson, *Syst. Zool.* **27**, 324 (1978).
4. E. O. Wiley, *Phylogenetics: The Theory and Practice of Phylogenetic Systematics* (Wiley, New York, 1981).
5. J. S. Farris, in *Advances in Cladistics*, N. J. Platnick and V. A. Funk, Eds. (Columbia Univ. Press, New York, 1983), vol. 2, pp. 7–36.
6. D. C. Fisher, in W. P. Maddison and D. R. Maddison, *MacClade*, ver. 3.0 (Sinauer, Sunderland, MA, 1992), pp. 124–129; D. C. Fisher, in *Interpreting the Hierarchy of Nature*, L. Grande and O. Rieppel, Eds. (Aca-

- dem Press, New York, 1994), pp. 133–172; P. D. Polly, *Contrib. Mus. Paleontol. Univ. Mich.* **30**, 1 (1997).
7. Cladograms represent only the relative recency of ancestry of the ingroup taxa and not explicit phylogenetic relationships. A given cladogram is consistent with multiple phylogenetic trees, which express explicit relationships between ingroup taxa, either as sister taxa with a hypothetical ancestor or as ancestor-descendant pairs [N. Eldredge and J. Cracraft, *Phylogenetic Patterns and the Evolutionary Process* (Columbia Univ. Press, New York, 1980); G. Nelson and N. I. Platnick, *Systematics and Biogeography: Cladistics and Vicariance* (Columbia Univ. Press, New York, 1981)].
8. Fractional OTUs are not deleted; for example, 70% of 31 OTUs is 21.7 OTUs, but only 21 are deleted.
9. Outgroups implicitly polarize character change, rooting the unrooted networks selected by parsimony [W. P. Maddison, M. J. Donoghue, D. R. Maddison, *Syst. Zool.* **33**, 83 (1984); P. H. Weston, in *Models in Phylogeny Reconstruction*, R. W. Scotland, D. J. Siebert, D. M. Williams, Eds. (Clarendon, Oxford, 1994), pp. 125–155].
10. Most simulated lineages persist through multiple time intervals. Our complete, seven (plus) lineage histories thus include more than seven lineage segments, or OTUs.
11. D. L. Swofford, *Phylogenetic Analysis Using Parsimony (PAUP)*, ver. 3.1.1 (Illinois Natural History Survey, Champaign, IL, 1991).
12. W. P. Maddison and D. R. Maddison, *MacClade*, ver. 3.0 (Sinauer, Sunderland, MA, 1992).
13. M. Foote and D. M. Raup, *Paleobiology* **22**, 121 (1996); C. R. C. Paul, in *The Adequacy of the Fossil Record*, S. K. Donovan and C. R. C. Paul, Eds. (Wiley, Chichester, UK, 1998), pp. 1–22.
14. Occasional branch and bound checks on matrices

small enough to analyze with that algorithm suggest that for the simulated data, 50 replicates of the heuristic search were adequate. However, given our results, using only 50 replicates is not a liability to cladistics. Because the true cladograms are generally less parsimonious than the associated sets of cladistic results, it is unlikely that even shorter cladograms would be closer topologically to the true cladogram.

15. All intervals are equally weighted, although sedimentological, lithological, and taphonomic data could, in principle, be used to weight the stratigraphic character differentially for specific intervals or lineage segments.
16. RI is a measure of the homoplasy implied by a hypothesis for given character data and ranges from 0 for worst fit to 1 for best [J. S. Farris, *Cladistics* **5**, 417 (1989)]. Published RIs are from W. C. Clyde and D. C. Fisher [*Paleobiology* **23**, 1 (1996)] and all volumes of *Palaeontology*, *Journal of Paleontology*, and *Journal of Vertebrate Paleontology* from 1994 to 1998.
17. Subclades are the proper subsets of a whole cladogram. The whole cladogram does not count as a subclade. The number of subclades is equivalent to the number of internal nodes. The consensus fork index is simply the number of subclades (or internal nodes) shared between cladograms being compared, divided by the total number of subclades.
18. The sign test calculates the probability that the sign of the difference between stratocladistic and cladistic CFI_{max} for each grouping is equally distributed between positive and negative.
19. We thank T. K. Baumiller, J. I. Bloch, W. C. Clyde, P. D. Gingerich, and M. D. Uhen for helpful comments during the development of this project and manuscript.

29 December 1998; accepted 9 March 1999

Two-Dimensional Photonic Band-Gap Defect Mode Laser

O. Painter,¹ R. K. Lee,¹ A. Scherer,^{1*} A. Yariv,¹ J. D. O'Brien,² P. D. Dapkus,² I. Kim²

A laser cavity formed from a single defect in a two-dimensional photonic crystal is demonstrated. The optical microcavity consists of a half wavelength-thick waveguide for vertical confinement and a two-dimensional photonic crystal mirror for lateral localization. A defect in the photonic crystal is introduced to trap photons inside a volume of 2.5 cubic half-wavelengths, approximately 0.03 cubic micrometers. The laser is fabricated in the indium gallium arsenide phosphide material system, and optical gain is provided by strained quantum wells designed for a peak emission wavelength of 1.55 micrometers at room temperature. Pulsed lasing action has been observed at a wavelength of 1.5 micrometers from optically pumped devices with a substrate temperature of 143 kelvin.

In 1946, Edward Purcell (1) first proposed that the spontaneous emission from an excited state of an atom can be significantly altered by placing it in a low-loss cavity with dimensions on the order of the electromagnetic wavelength. More recently, with the

advent of the semiconductor laser and the improvement in crystal growth and fabrication technology, there has been increasing interest in the engineering of optical microcavities in semiconductors for spontaneous emission control (2–4). The vertical cavity surface emitting laser (5) (VCSEL), in which light is confined between two epitaxially grown distributed Bragg reflectors, was one of the first semiconductor cavities with dimensions on the order of the wavelength of light. Another microcavity laser, the micro-disk laser (6), uses total internal reflection at the edge of a high refractive index disk to

¹Departments of Electrical Engineering and Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA. ²Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA.

*To whom correspondence should be addressed. E-mail: etcher@cco.caltech.edu