TAXON 36(3): 584–587. AUGUST 1987

EXPERIMENTS AND HYPOTHESES IN SYSTEMATICS

In her response to Hagen's (1983) discussion of experimental methods in plant taxonomy, Gilmartin (1986) argued that the word experimental should *not* be equated with the use of a particular technique (e.g., electrophoresis), nor with research focused on a particular taxonomic level (e.g., species). Instead, she proposed that "experimental" be equated with hypothesis testing. "Experimental systematics" is distinguished from "descriptive systematics," in which "explicitly stated hypotheses may be absent" (p. 118). A number of studies were cited by Gilmartin to support her claim that "experimental plant systematics, i.e., research that tests hypotheses, is extremely active with populations, species, genera, etc." (p. 118).

In responding to Gilmartin (1986), LaDuke (1987) apparently accepted the idea that "experimental" and "hypothesis testing" should be equated, but argued that most of the studies cited by Gilmartin did not actually involve hypothesis testing. He maintained that hypothesis testing is rare in systematics, and that systematic research would be much improved if hypotheses were stated explicitly and if lower level phylogenetic hypotheses were tested before higher level problems were addressed.

Gilmartin (1986) cited my preliminary phylogenetic analysis of *Viburnum* (Donoghue, 1983) as an example of hypothesis testing and, therefore, of experimental research, but LaDuke (1987) argued that this study involved hypothesis generation, *not* hypothesis testing. I will argue that *both* Gilmartin and LaDuke are mistaken. I feel obliged to comment on this matter since Doyle and I (e.g., Doyle and Donoghue, 1986) may have inadvertently added to the confusion by using the word experimental in describing our analyses of seed plant phylogeny.

Comments on Gilmartin's Argument

I agree with Gilmartin that it is wrong to associate the word experiment with a particular technique or hierarchical level. However, I disagree with her claim that "whenever specific hypotheses are tested, the term experimental is aptly applied" (p. 118). Many scientists and philosophers consider hypothesis testing to be the hallmark of science. Thus, in equating "experimental" with "hypothesis testing" Gilmartin effectively equates "experimental" with "scientific." Under this view, the experiment is the one and only kind of scientific test. This obliterates a distinction that is commonly made (e.g., Hempel, 1966: 19–22) between observation (taking note of something) and experiment (manipulation or perturbation of a system relative to a control) and denies that hypotheses can be tested by observation alone. I think the distinction between observation and manipulative experiment is a useful one, and that hypothesis testing in systematics often involves observation rather than experimentation.

For example, consider the standard phylogenetic hypothesis that two taxa are more closely related to one another than either is to a third (i.e., they share a more recent common ancestor). This can be tested by observing the characters of relevant organisms. In conjunction with auxiliary principles (such as parsimony) these character data can be brought to bear on the phylogenetic hypothesis, with the result that it is either temporarily rejected as inferior or accepted as the best candidate for future tests. Assuming that Gilmartin would consider this to be hypothesis testing, she would also consider it an experiment. I would say it is a test, but *not* an experiment. All experimental tests involve observations, but not all observations or tests involve experiments.

Of course, the distinction I have drawn between observations and experiments is not without its difficulties. In fact, there is a fine line between the two in some instances. For example, in the case of so-called natural experiments, a perturbation occurs and is observed, but the perturbation is not caused by the investigator. It is also important to remember that "observation" is not a trouble-free concept. Indeed, this is a major issue in the philosophical literature (e.g., Brown, 1977), where it is commonly said that observations are low-level hypotheses accepted for the sake of argument and should themselves be tested. I agree, but the idea that observations are hypotheses does not affect my argument since I am distinguishing observation from experiment.

If a distinction between observational and experimental tests is admitted, and if it is agreed that there can be observational tests of phylogenetic hypotheses, we should consider whether there can also be experimental tests of such hypotheses. I suggest that, at least in one sense, there can be and that Doyle and Donoghue (1986) provide an example. In our studies of seed plant phylogeny we wished to test the robustness of the cladograms obtained from parsimony analysis of our data set. This we did by manipulating the data set in a number of ways. Thus, for example, we performed computer experiments in which we eliminated poorly known fossil groups from the data set (also see Doyle and Donoghue, 1987). Through this sort of experimentation with our data, we were able to conclude that some relationships were much more robust than others. Obviously, this is not laboratory or field experimentation; we did not physically eliminate taxa or characters in the real world and observe the effects. Instead, we performed *thought experiments* (analogous to those in some areas of physics), wherein we manipulated our data and assumptions and compared the results to an analysis of a control data set.

Doyle and I wished to call attention to the fact that we had performed extensive manipulation of our data set, and it was for that purpose that we chose the word experimental. It is less clear why Gilmartin wishes to extend the word experimental to cover all hypothesis testing. However, since the word experimental tends to have favorable connotations, it may seem advantageous to consider systematists to be experimentalists. The temptation to extend the meaning of "experimental" in this way should be resisted—this maneuver distorts the usual meaning of the word without substantial benefit.

I am also troubled by Gilmartin's distinction between experimental systematics and descriptive systematics, noted above. This is a distinction based on what one says rather than on what one actually does. If I state a hypothesis explicitly, I am to be considered an experimental systematist. If I conduct the same study but the hypothesis is not stated explicitly (perhaps only because I think it will be obvious to the reader), I am then considered a descriptive systematist. The critical thing, in my view, is whether or not there is a hypothesis being tested; it is less important whether or not the hypothesis is explicitly stated.

Comments on LaDuke's Argument

LaDuke (1987) claims that hypothesis testing is woefully rare in plant systematics and that instead many systematists simply generate hypotheses. In considering the studies referred to by Gilmartin (1986), he concluded that Donoghue (1983) was an example of hypothesis generation as opposed to testing, apparently because I used information in the literature in delimiting terminal taxa ("cladistic units" or "CUs") used in the analysis. I suggest that in my analyses of *Viburnum* phylogeny I *simultaneously* tested previous hypotheses of relationships and generated new hypotheses to be tested in subsequent studies. In particular, I considered my analysis to be a test of Rehder's (1908, 1940) sectional classification of the genus. For this reason the correspondence between the terminal taxa in my analysis and Rehder's nine sections was presented in a table, and the largest section of the paper was devoted to a discussion of each of Rehder's sections in relation to the results obtained in the parsimony analysis. By this test I concluded that there is evidence that some of Rehder's sections are monophyletic, whereas others appear to be paraphyletic. The latter, I would argue, should be abandoned.

At the same time that I tested a previous hypothesis of relationships, I generated new hypotheses. Specifically, I concluded that some sets of terminal taxa consistently emerged as monophyletic, even when I modified the data set by subtraction of some characters and taxa. These robust groups represent concrete hypotheses of relationship to be tested in subsequent analyses. Indeed, I have now tested them by more carefully examining pollen and leaf characters (Donoghue, 1985; Donoghue and Levin, 1986), and for the most part they have been corroborated. I also note that in these subsequent studies some of the terminal taxa used in the preliminary analysis were divided into smaller units based on additional information, especially on Asian species. Using smaller groups in the analysis of leaf characters allowed a test of the monophyly of the terminal taxa used previously. These were found to be monophyletic, or their relations were unresolved but not inconsistent with monophyly. Obviously, the monophyly of these and other terminal taxa will be further tested using additional character information.

I conclude that I simultaneously tested and generated phylogenetic hypotheses, and I submit that this is often the case in systematics. Therefore, it is not surprising that LaDuke had difficulty in identifying particular analyses as either one or the other. He seemed aware of this problem when he noted that "in some instances, it is difficult to tease the mental acts of hypothesis generation away from hypothesis testing, as they can be occurring easily in rapid succession" (p. 62). I suggest only that they can occur at the same time.

In addition to *describing* systematic research, LaDuke was also interested in *prescribing* a temporal sequence of analysis—one thing should be done before another. Thus, in discussing my paper on *Viburnum* (Donoghue, 1983) he says that "it would be better to test the monophyly of each CU first, . . . then do cladistic analysis among the CUs . . ." (p. 61). This same kind of argument is repeated in his discussion of Wanntorp (1983). The implication is that we should always start at the bottom and work upward, that is, first we should determine what species exist, then we should assemble these into monophyletic groups of species, and so on up the line. In this case, Doyle and I should not have carried out an analysis of seed plant phylogeny until we had first established that all of the terminal taxa we used were monophyletic groups of species. Obviously, if LaDuke's prescription were adopted it would be quite a long time before we reached the higher levels, and we would be unable in the meantime to address issues such as the evolution of the seed or of the angiosperms.

But is there a logical reason to accept LaDuke's assertion that we should work in one direction, namely from the bottom up? I think not. In some instances it may be highly desirable or even necessary to begin an analysis at a higher level in order to establish a cladistic framework for outgroup polarity assessment at a lower level (Maddison et al., 1984). I can see nothing wrong with making some assumptions about relationships at the outset of a study, since these assumptions can and presumably will be tested in subsequent analyses at other levels. It seems perfectly legitimate to start at a higher level and work down to lower levels or to start in the middle and work in both directions. I agree with Hull (1986: 190) that science is *not* like baseball, wherein the bases must be covered in only one order: "Scientists need not begin their investigations in any particular place. One scientist might begin with a chance observation, another with a fairly particular hypothesis, yet another with a sophisticated theory. Although it is not highly recommended, a scientist might begin with massive accumulation of data. Data gathering, hypothesis formation, testing, reworking old hypotheses, etc. need not proceed in any prescribed temporal order. All that is necessary is that sooner or later all bases get touched."

LaDuke urges that systematists specify the hypotheses they wish to test and then design "experiments" to try to falsify these hypotheses. This will sound very familiar to those who have followed the debates, in *Systematic Zoology* and elsewhere, over Karl Popper's view of science (e.g., Popper, 1968). In a general sense this is good advice; however, it is not entirely obvious how systematics would be improved if we always followed LaDuke's suggestion. In systematics the hypotheses to be tested are usually present in the form of previous treatments of a group under study. In some cases the latest treatment will have been published recently and in other cases it may be hundreds of years old. Older treatments will often have been based on very limited material compared to what is available now, and the previous analysis may have been based on rather different principles than are employed today. These older treatments are, of course, exceedingly important to us, but I wonder if it is always very interesting to succeed in falsifying the hypotheses of relationships proposed in them. How surprising is it, for example, to find that Oersted's hypothesis of relationships in *Viburnum*, published in 1861 and based on relatively little material, is inadequate? Would systematics be better off today if I had set it as my goal to falsify Oersted's hypothesis? I have shown that parts of Oersted's scheme are unparsimonious, but I see little point in dwelling on this accomplishment. In other cases it may be much more interesting to set out to test a previous hypothesis and to focus attention on the comparison. For example, Doyle and I wished to test the phylogenetic hypotheses of Hill and Crane (1982) and Crane (1985). Here the comparison is more meaningful since these are recent analyses based on the same principles that we employed in our studies.

Hopefully it is clear that I am *not* objecting to the idea that hypotheses be stated and tested explicitly. I am only suggesting that there are different degrees to which it is interesting to explicitly state and set out to falsify previous hypotheses in systematics. The quality of research and the significance of the results are more important than having followed a law-and-order procedure. I would value a carefully conducted phylogenetic study that made little mention of previous hypotheses over one that stated and falsified a trivial hypothesis.

Acknowledgments

I am grateful to John LaDuke for calling this matter to my attention by kindly sending me his manuscript. I also thank Jim Doyle for helpful discussion, and David Hull for providing an English translation of a portion of his article (Hull, 1986), originally published in French.

Literature Cited

Brown, H. I. 1977. *Perception, theory, and commitment. The new philosophy of science.* Univ. Chicago Press, Chicago.

Crane, P. R. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. *Ann. Missouri Bot. Gard.* 72: 716–793.

Donoghue, M. J. 1983. A preliminary analysis of phylogenetic relationships in *Viburnum* (Caprifoliaceae s.l.). *Syst. Bot.* 8: 45–58.

———. 1985. Pollen diversity and exine evolution in *Viburnum* and the Caprifoliaceae sensu lato. *J. Arnold Arb.* 66: 421–469.

——— and G. A. Levin. 1986. Leaf morphology and evolution in the genus *Viburnum. Am. J. Bot.* 73: 760.

Doyle, J. A. and M. J. Donoghue. 1986. Seed plant phylogeny and the origin of angiosperms: An experimental cladistic approach. *Bot. Rev.* 52: 321–431.

——— and ———. 1987. The importance of fossils in elucidating seed plant phylogeny and macroevolution. *Rev. Paleobot. Palynol.* 50: 63–95.

Gilmartin, A. J. 1986. Experimental systematics today. *Taxon* 35: 118–119.

Hagen, J. B. 1983. The development of experimental methods in plant taxonomy. *Taxon* 32: 406–416.

Hempel, C. G. 1966. *Philosophy of natural science.* Prentice-Hall, Inc., Englewood Cliffs, N.J.

Hill, C. R. and P. R. Crane. 1982. Evolutionary cladistics and the origin of angiosperms. *In:* K. A. Joysey and A. E. Friday (eds.), *Problems of phylogenetic reconstruction,* pp. 269–361. Syst. Assoc. Spec. Vol. 21. Academic Press, London.

Hull, D. L. 1986. Les fondements épistémologiques de la classification biologique. *In:* P. Tassy (ed.), *L'ordre et la diversité du vivant,* pp. 161–203. Foundation Diderot/Librairie Artheme Fayard.

LaDuke, J. C. 1987. The existence of hypotheses in plant systematics or biting the hand that feeds you. *Taxon* 36: 60–64.

Maddison, W. P., M. J. Donoghue and D. R. Maddison. 1984. Outgroup analysis and parsimony. *Syst. Zool.* 33: 83–103.

Oersted, A. S. 1861. Til belysning af slaegten *Viburnum.* Vid. Medd. Naturh. For. Kjöbenhavn 13: 267–305.

Popper, K. R. 1968. *The logic of scientific discovery.* Harper & Row, New York.

Rehder, A. 1908. The viburnums of eastern Asia. *In:* C. S. Sargent (ed.), *Trees and shrubs,* vol. II, part II, pp. 105–116. Houghton Mifflin, Boston.

———. 1940. *Manual of cultivated trees and shrubs,* 2nd ed. Macmillan Co., New York.

Wanntorp, H. E. 1983. Cladistics misunderstood—Again. *Taxon* 32: 97–109.

Michael J. Donoghue, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, U.S.A.