

OPINION

Are molecular data supplanting morphological data in modern phylogenetic studies?

SETH M. BYBEE^{1,2}, JENNIFER M. ZASPEL^{1,3,4}, KYLE A. BEUCKE¹, CLARE H. SCOTT¹, BRADLEY W. SMITH¹ and MARC A. BRANHAM¹

¹Department of Entomology and Nematology, University of Florida, Gainesville, FL, U.S.A., ²Department of Ecology and Evolutionary Biology, University of California-Irvine, Irvine, CA, U.S.A., ³Department of Entomology, University of Minnesota, St. Paul, MN, U.S.A. and ⁴Bell Museum of Natural History, University of Minnesota, St. Paul, MN, U.S.A.

Introduction

Morphological data have long served as major sources of information for inferring phylogenetic relationships among taxa. With the advent of polymerase chain reaction and modern molecular approaches to phylogenetics, DNA has become a major source for phylogenetic inference. Combined analyses of molecular and morphological (CMM) data are not unusual. Studies examining the relative utility of morphological and molecular data derived from plants and animal groups have been performed (e.g. Sanderson & Donoghue, 1989), but the results of this research were limited by the relative novelty of molecular data to phylogenetics at the time. Differences of opinions exist among systematic biologists concerning the utility of morphological data (Scotland *et al.*, 2003; Tautz *et al.*, 2003; Wheeler, 2008). For example, some authors are of the opinion that morphological data for plants are inherently troublesome because of the large degree of convergence between plant structures (Scotland *et al.*, 2003), whereas others argue that morphology has become sidelined and depreciated as a source of data and needs to be examined using modern tools (Wheeler, 2008).

Insect systematics has a long and rich history of using morphological data as the primary source for phylogenetic estimation (Ross, 1973). Perhaps as a result of the high level of morphological diversity within the class Insecta, morphological data have served as an excellent resource for phylogenetic studies. Although molecular phylogenetic analyses have become common in insect systematics, it is unclear if molecular datasets are supplanting those derived from morphology. Here we examine the frequency with which morphological, DNA and CMM datasets have been used in published phylogenetic studies over a 15-year period.

Correspondence: Seth M. Bybee, Department of Ecology and Evolutionary Biology, University of California-Irvine, Irvine, CA 92697, U.S.A. E-mail: seth.bybee@gmail.com

We seek general trends regarding the use of each type of data and compare these trends among three major groups of organisms (insects, non-arthropod animals, and plants). Specifically, we address three questions. (i) What are the general trends shown in insect phylogenetic studies based on morphological, molecular and CMM datasets over the last 15 years? (ii) How do these trends compare with those for other groups of organisms? (iii) What are the implications of these findings regarding insect systematics with respect to the training of future graduate students?

Materials and methods

Seven peer-reviewed journals of broad systematic focus (BSJ) were searched between the years 1992 and 2007: *Biological Journal of the Linnean Society*, *Cladistics*, *Evolution*, *Science*, *Nature*, *Proceedings of the Royal Society of London B* and *Systematic Biology*. Two additional peer-reviewed journals that focus largely on insect phylogenetics (IJ), *Invertebrate Systematics* and *Systematic Entomology*, were searched to determine the use of phylogenetic datasets within insects specifically. Papers were selected from these journals based on the following criteria: (i) the paper must have a phylogeny figured, (ii) the dataset must be novel or represent the compilation of several data matrices in a novel analysis, and (iii) taxon sampling must be above the 'species level' (i.e. not at the population level or within a species group). Papers meeting the above criteria were divided into three major groups: insects (mostly composed of insect studies, although other arthropod groups were represented), non-arthropod animals (essentially mostly vertebrates) and plants (i.e. multicellular organisms capable of carrying out photosynthesis). Viruses, bacteria, prokaryotes, yeasts, etc. were excluded as they represent unicellular organisms, and little to no morphological work has been performed on these groups over the past 15 years. Phylogenetic studies that examined relationships at the 'kingdom level' were also excluded so as to keep datasets (i.e. insects, non-arthropod

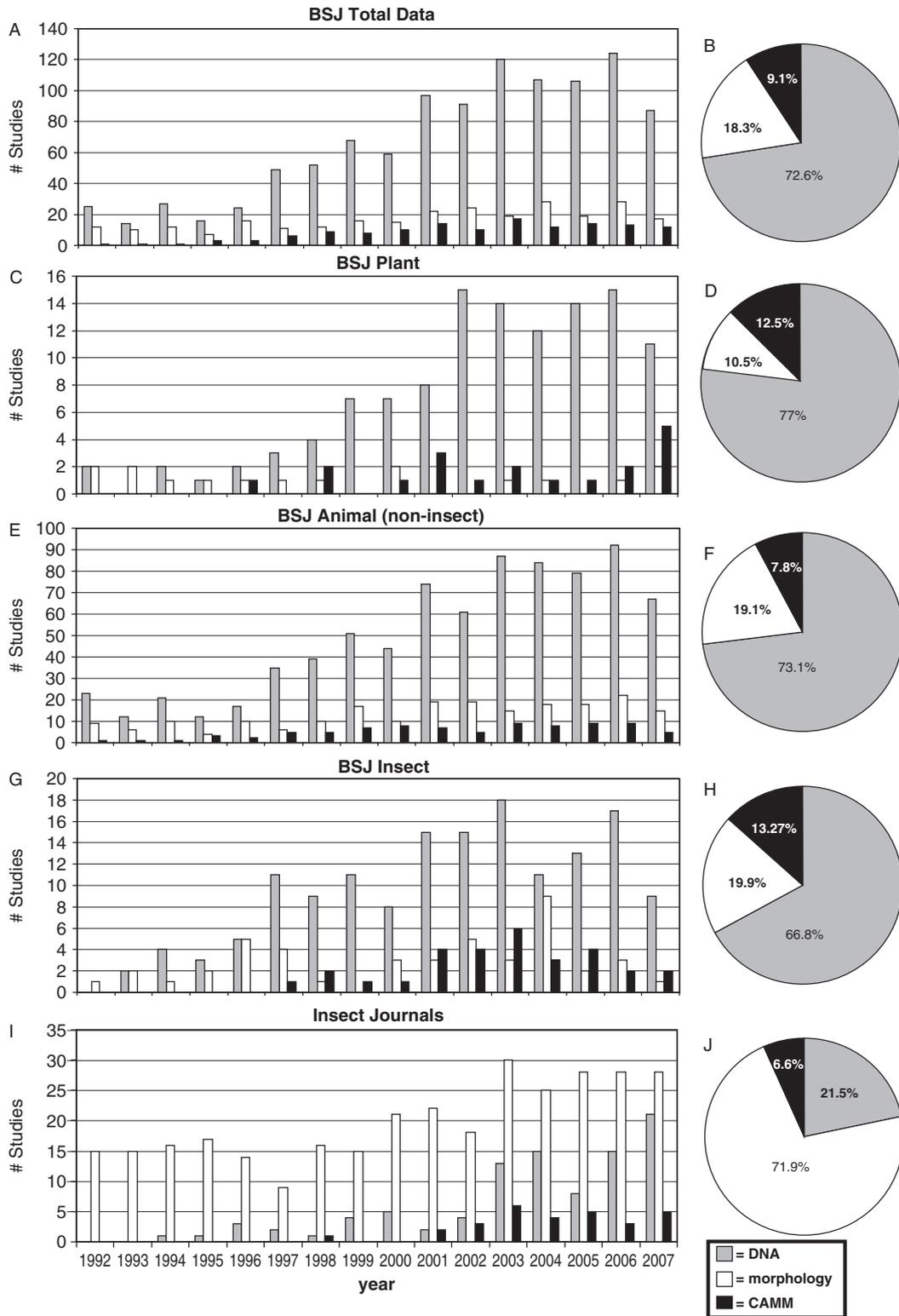


Fig. 1. Histograms of the number of morphological, molecular and CMM phylogenetic analyses published per year, and pie charts representing the breakdown for the total number of published phylogenetic analyses from 1992 to 2007. (A, B) total from journals of broad systematic focus (BSJ); (C, D) plants from BSJ; (E, F) non-arthropod animals from BSJ; (G, H) insects from BSJ; (I, J) insects from insect journals (IJ).

animals, and plants) as discrete as possible. Those studies that were molecular in nature but did not use DNA data in a traditional phylogenetic format (e.g. aligned sequences) were omitted (i.e. allozyme, amplified fragment length polymorphism, microsatellite markers, isozyme, electrophoresis, restriction fragment length polymorphism), as these data are used more often to address population-level questions in which morphological data generally are not gathered for inclusion in these types of analyses. Each paper was placed in one of three categories, molecular, morphological or CAMM, to track the use of each type of data over the last 15 years. All data were examined using histograms and pie charts generated in Microsoft Excel© to visualize general trends.

Results

Journals of broad systematic focus

From the BSJ representing all groups of organisms, a total of 1469 studies were recorded from the 15-year period (Fig. 1A). Almost 73% of these studies were from the 'animal' group (excluding insects). Insects and plants (15 and 10%, respectively) made up the remaining studies (results not shown). Molecular studies made up nearly 73% of all studies performed, followed by 18% morphological and 9% CAMM studies (Fig. 1B). Over the 15-year period examined, plants were the group with the most molecular research (77% of all plant studies), followed by animals (73.1%) and insects (66.8%; Fig. 1D–F). Morphological data were used most commonly with insects (19.9%) and animals (19.1%); each of these groups had nearly twice the percentage of morphological studies as plants (10.5%; Fig. 1D–F). Insects had the highest number of CAMM studies (13.3%), followed by plants (12.5%) and animals (7.8%; Fig. 1D–F).

The general trend for the use of phylogenetic data over the last 15 years shows a steady increase in the number of molecular studies from 1997 to 2007 as well as a slight increase in morphological studies (Fig. 1A). CAMM studies become more common in the late 1990s, but there is no clear trend regarding their use. Plant molecular studies steadily increase from 1997 to 2001, with a dramatic increase starting in 2002 (Fig. 1C). Plant morphological and CAMM studies do not show any clear trends, although CAMM analyses become more common between 2000 and 2007. Animal molecular

studies also show a steady increase (Fig. 1E). Morphological and CAMM animal studies increase slightly in 1999 and remain relatively constant to 2007. Molecular studies of insects show an overall increase since 1992. Insect morphological studies appear to peak in 2004, whereas CAMM studies are relatively constant from 2001 onwards (Fig. 1G).

The average dataset and taxon sampling from the BSJ studies are listed in Table 1. Interesting results include the relative congruence between animal and insect datasets size and the difference in the number of morphological characters and size of taxon sampling between plants and animals (both insects and non-arthropod animals).

Insect journals

A total of 441 studies (i.e. data points) were recorded from the IJ from 1992 to 2007 (Fig. 1I, J). Nearly 72% of these used morphological data alone. Molecular datasets were estimated at 21.5%, and CAMM analyses at 6.6% (Fig. 1J). General trends from IJ include an increase in the number of morphological studies from 1992 onwards, an increase in molecular studies from 2003 onwards and an increase in CAMM studies in the years 2001–2003, which then remained relatively constant until 2007, the point at which we stopped collecting data (Fig. 1I).

Discussion and conclusions

Despite the emphasis on generating large molecular datasets that is currently seen in phylogenetics, morphological data remain both relevant and readily employed, particularly within insect phylogenetics. Although the data here are limited in providing an explanation for the common and continued use of morphological data in modern phylogenetics, possibilities could include extensive collections of pinned specimens, excellent preservation of specimens (both extant and fossil), rich variation of specialized structures (e.g. mouthparts, wings and terminalia), the sheer diversity of insects, and a historical emphasis on morphological study in insect systematics.

It is important to ask which skills students are currently being trained *in*. We see the current trends shown here, at least for studies of insects, as encouraging. Expertise in coding morphological variation for phylogenetic studies within entomology appears to be 'preserved' for the moment.

Table 1. Average number of characters and taxa within phylogenetic studies (1992–2007).

		Avg. no. of characters/matrix			Avg. no. of taxa sampled/study		
		DNA	Morph	CAMM	DNA	Morph	CAMM
BSJ	Total	1602.5	83.6	1618.2	45	32.3	38.4
	Plant	1760.7	24.7	1248.3	59	18	44.8
	Animal	1503.75	83.9	1625.3	39.7	32.4	37.2
	Insect	1501	84.5	1622.6	45.2	32.3	40.3
Insect journals		1400.6	55.4	1719.5	39.2	29.9	31.1

BSJ, journals of broad systematic focus; Morph, morphological studies; CAMM, combined analyses of molecular and morphological data.

What will be more interesting perhaps is to repeat the efforts of this paper in the future to examine if morphology as a whole is still being utilized as a viable phylogenetic dataset and if the numbers of CAMM phylogenetic studies have increased. Furthermore, it would be helpful to identify for which groups of insects morphological data have been coded for inclusion in phylogenetic analyses and at what level of classification (e.g. familial, ordinal, etc.) these data are being employed. Perhaps more valuable is determining trends in morphological expertise for generating phylogenetic data. For example, who is generating morphological matrices (e.g. students or long-time experts, first authors or collaborators), and are the morphological data included in matrices novel, gathered from taxonomic literature, or pulled from other morphological matrices? Further still, it would be helpful to identify how many morphological studies were supported with federal or other competitive funds, as many entomologists, both amateur and professional, perform morphological research without any such funding. Morphological research conducted by non-professionals often contributes important findings to the study of entomology and is critical to groups such as Coleoptera, Lepidoptera and Odonata.

In entomology, the use of morphological data for phylogenetic analysis remains prevalent. However, there is no easy way to quantitatively analyse the data presented here in order to address the above questions. From our brief inspection of the studies included in Fig. 1, it is our impression that many morphological matrices published in BSJ, especially those included within CAMM analyses, were not largely original and represented tests of existing morphological synapomorphies and/or classifications constructed from morphological data. By comparison, studies published in IJ represented mostly original morphological codings that were used to refine existing or constructing the first ever phylogenetic hypotheses for a particular group.

As in any other branch of science, the more data and breadth of those data will provide more interesting results and observations. Regardless of what aspect of evolutionary biology one enjoys, truly interesting observations that have broad scientific and general public appeal, almost without exception, involve some aspect of morphology. It is important not only to preserve our current knowledge but to extend it in the form of new hypotheses for classification and

evolutionary origins. We encourage all systematists to focus on the importance of morphology and its relevance to a stable classification scheme for all organisms, to taxonomy, and to palaeontology. If morphological expertise is lost or even diminished within entomology, we will lose the ability to properly identify, describe, and keep our science focused on the organism. Indeed, it would be a shame to lose such expertise for one of the most diverse groups of organisms on the planet.

Although methods for generating molecular data are becoming easier, faster and more cost-effective, it is encouraging to see trends that suggest that students are still being exposed to morphological research and that collections are being actively used by systematists. If these trends continue, the application of morphological expertise within insect phylogenetics should be preserved, and taxonomists, palaeontologists and collections will not be marginalized.

Acknowledgements

We would like to thank K.B. Miller, S.L. Cameron and G.J. Svenson for helpful discussions, and Q.D. Wheeler and the journal editors, P.S. Cranston and L.B. Vilhelmsen, for helpful comments.

References

- Ross, H.H. (1973) Evolution and phylogeny. *History of Entomology* (ed. by R.F. Smith, T.E. Mittler and C.N. Smith), pp. 171–184. Annual Reviews Inc., Palo Alto.
- Sanderson, M.J. & Donoghue, M.J. (1989) Patterns of variation in levels of homoplasy. *Evolution*, **43**, 1781–1795.
- Scotland, R.W., Olm, R.G. & Bennett, J.R. (2003) Phylogeny reconstruction: the role of morphology. *Systematic Biology*, **52**(4), 539–548.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**, 70–74.
- Wheeler, Q.D. (2008) Undisciplined thinking: morphology and Hennig's unfinished revolution. *Systematic Entomology*, **33**, 2–7.

Accepted 2 June 2009

First published online 20 August 2009