Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 28 (2003) 610–613
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Short Communication

The major opsin gene is useful for inferring higher level phylogenetic relationships of the corbiculate bees

Sydney A. Cameron[a,*] and Patrick Mardulyn[b]

[a] *Department of Entomology, 505 South Goodwin Avenue, 320 Morrill Hall, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[b] *Unit of Evolutionary Genetics (CP 300), Free University of Brussels (ULB), 6041 Gosselies, Belgium*

Received 6 June 2002; revised 18 December 2002

The phylogeny of the corbiculate bees, which include the major groups of highly eusocial bees, is controversial. In an attempt to resolve the controversy, Ascher et al. (2001) argued that they have expanded a prior phylogenetic analysis of the corbiculate bees based on opsin (LW *Rh*) sequences reported by Mardulyn and Cameron (1999) "to more critically assess opsin's phylogenetic utility." They contended that all previous molecular studies of corbiculate bee phylogeny suffer from including too few outgroup taxa. Several additional conclusions stand out in their paper. One is that the monophyly of the ingroup (the corbiculate bees) cannot be unequivocally substantiated, just as Mardulyn and Cameron (1999) and, more recently, Cameron and Mardulyn (2001) found in their analyses. Second, in contrast to Mardulyn and Cameron, Ascher et al. stated that their expanded opsin data set shows little support for relationships among the ingroup corbiculate tribes. Third, they found that when they combined their data set with morphological and behavioral data for the corbiculate bees, the results supported the classical view of relationships among the tribes (((Apini + Meliponini) + Bombini) + Euglossini).

We take issue with each of these claims and stand by our original conclusions that LW *Rh* is useful to infer phylogenetic relationships of corbiculate bees. The opsin gene clearly contains phylogenetic information to resolve relationships among the tribes of corbiculate bees and, more importantly, this phylogenetic information is consistent with all other genes thus far studied in this group (28S, cytochrome *b*, and 16S; see Cameron and Mardulyn, 2001; Ef-1α, Sipes, Danforth, and Cameron, unpublished data). Moreover, Rokas et al. (2002) recently found this gene to be one of the most useful markers for within-tribe divergences of gallwasps (Hymenoptera: Cynipidae). It therefore seems a promising gene for higher level phylogenetics of insects.

While it is desirable to increase the number of available outgroups to the corbiculate clade, adding as many as 52 new outgroup taxa from 21 new tribes as done by Ascher et al. (2001), is unnecessary for the aim of resolving relationships among the four corbiculate tribes (although these data would be useful to further test the phylogenetic utility of the opsin fragment at higher taxonomic levels for bees in general). Moreover, opinion concerning the closest outgroups to the corbiculate bees has varied. Until recently (Roig-Alsina and Michener, 1993), the Xylocopinae were thought to be among the closest relatives (Sakagami and Michener, 1987), and based on sequences from different genes, this may still be the case. We contend that much work still needs to be done to resolve the corbiculate bee outgroup question.

With respect to the issue of monophyly of the corbiculate clade, Ascher et al. argued (p. 78) that their use of multiple outgroup exemplars from 21 additional apine tribes would enable them to *test* ingroup monophyly. A few pages later (p. 81) they stated that "monophyly of the corbiculate bees is uncontroversial." Why the need to test if it is uncontroversial? In fact, Ascher et al.'s inclusion of more outgroup taxa than in previous molecular studies has resulted in the paraphyly of the corbiculate bees, as one (parsimony analysis) or two (maximum likelihood) outgroup taxa fell within the ingroup. This is an interesting result, already noted by Mardulyn and Cameron (1999) and again in Cameron and Mardulyn (2001), using multiple outgroups and multiple genes. We have not been able in our own research to pinpoint a specific bias responsible for this

* Corresponding author. Fax: 1-217-244-3499.
E-mail address: scameron@life.uiuc.edu (S.A. Cameron).

doi:10.1016/S1055-7903(03)00055-1

surprising outcome (Cameron and Mardulyn, 2001). Although the specific outgroup falling within the ingroup varies from one gene to another (i.e., we cannot determine a consistent pattern of non-monophyly), and the combined data support monophyly (Cameron and Mardulyn, 2001), the paraphyly of the corbiculate bees should at least be considered as a plausible alternative hypothesis. Instead, Ascher et al. (2001) treated their result of non-monophyly as evidence that opsin is problematic.

The unweighted parsimony analyses reported by Ascher et al. (2001) did not provide strong support for most of the higher level bee relationships, and in particular for the relationships among the corbiculate tribes inferred by Mardulyn and Cameron (1999). However, Mardulyn and Cameron (1999) showed that transitions in third positions were saturated, and that weighted parsimony and maximum likelihood analyses, which are known to correct for saturation, resulted in much higher support for the clade grouping Bombini and Meliponini together (87 and 71% bootstrap support, respectively). When Ascher et al. accounted for saturation, support for a Bombini + Meliponini clade increased from <50 to 76%. (Although they first stated (p. 78) that they downweighted transversions (TVs) by 1/4, and then (p. 87) that they downweighted transitions (TNs) by 1/4, we assume that they used, as they mentioned, the weighting scheme of Mardulyn and Cameron (1999); i.e., that TNs in third positions were given four times less weight).

Ascher et al.'s simultaneous parsimony analysis of the morphology + behavior + opsin data sets does not seem appropriate because it excluded two-thirds of the available molecular data; sequences of both 16S and cytochrome *b* were available from GenBank at the time of their study, but were disregarded. Instead, the authors combined 20 morphological characters and two correlated behavioral characters with a fraction of the molecular characters available (i.e., those for opsin). If the idea was to analyze all available characters together, why did they not include the other sequence data? If 16S and cytochrome *b* sequences were unavailable for all of the 52 new outgroups, they could have selected a few outgroups that were available and performed a simultaneous analysis with all available data (at least for the purpose of resolving relationships within the corbiculate bees). This was remedied in the above-cited paper by Cameron and Mardulyn (2001), in which data from four genes plus morphological characters from Roig-Alsina and Michener (1993) were analyzed simultaneously, resulting in strong support for a Bombini + Meliponini clade.

Another concern with their simultaneous analysis is the manner in which the authors carried out their test of character congruence (pp. 88–89). The problem with their methods was the use of an incomplete morphology + behavior data set for the 52 outgroup taxa they included in their comparative analyses. They presented a character matrix for the corbiculate bees taken from the literature (Schultz et al., 1999). Parsimony analysis of those characters resulted in the M44 tree (their Fig. 1a) containing an Apini + Meliponini clade. However, rather than coding all the outgroup taxa for additional relevant characters, the authors merely assigned 0's and ?'s to the 52 outgroups for this limited data set. This would obviously lead to a complete lack of resolution among the outgroups, essentially constraining any possibility for resolution to only the corbiculate bees. Hence, it is not surprising that the tree length (TL) they reported for the morphology-based parsimony analysis is only 27 steps. Their "simultaneous" analysis resulted in the M44 tree, with a TL of 1717 (four steps longer than the sum of the two independent TL's, which represents 0.24% of the total tree length, not 0.0024% as written by Ascher et al.). From this result they concluded that there is only weak incongruence between the two data sets, with the additional conflict coming from a slight increase in homoplasy in the opsin partition. As a result of their incomplete character resolution for their 52 outgroups, they did not give the morphology much chance to have homoplasy across the taxa, whereas the DNA had to resolve itself across all 69 taxa. It would therefore seem more appropriate if the authors had done one of the following: (a) reduced the number of taxa to include only the corbiculate bees and the two outgroups used in Mardulyn and Cameron's (1999) study, then run an ILD analysis, or (b) included additional morphology and behavior characters in the data matrix for resolution of outgroup relationships, not just assign 0's and ?'s to outgroups for a limited data set. Such character data were available from Roig-Alsina and Michener (1993).

The lack of support for relationships among any of the corbiculate tribes found by Ascher et al. in their equal weights parsimony analysis, the paraphyly of the corbiculate bees observed in all their analyses, and the results of their combined analysis, were used by the authors as arguments in their discussion to reject the phylogenetic relationships among corbiculate tribes inferred in the study of Mardulyn and Cameron (1999). Their conclusion is not supported by the data. Indeed, all published molecular studies to date, using four different genes (28S, 16S, cytochrome *b*, and opsin), support the same Bombini + Meliponini clade. Ascher et al.'s study does not contradict this finding, as this clade is also present in their parsimony and ML trees (Figs. 3, 6, 7, and 10). The poor jackknife and bootstrap support found for this clade in their equal weight parsimony analysis appears to be due to the fact that that analysis did not take into account saturation of transitions at third position sites. Therefore, although their inclusion of many more outgroups resulted in paraphyly of corbiculate bees, the authors did not show that the use of more outgroups decreased support for the

Bombini + Meliponini clade found in Mardulyn and Cameron's (1999) study.

Ascher et al. suggested that the placement of *Apis* outside the ((B + M) + E) clade was due to a long branch problem coupled with a skewed base composition problem. They did not, however, test specifically for these biases. We tested those potential biases (Cameron and Mardulyn, 2001), using parametric bootstrapping (Huelsenbeck, 1997) and spectral analyses on LogDet distances (Steel et al., 2000), and found no evidence for them in any of the four molecular data sets included in that study (of course, other biases may exist that we are not aware of). Ascher et al.'s conclusion that their simultaneous analysis provided strong support for a monophyletic highly eusocial Apini + Meliponini clade is contradicted by the far more comprehensive simultaneous analysis of Cameron and Mardulyn (2001), which soundly rejected an A + M clade and showed strong support for a B + M clade.

To further test this result, we have conducted likelihood ratio tests separately on three molecular data sets (opsin, 28S, and cytochrome *b*) published in Cameron and Mardulyn (2001). For this purpose, we performed ML analyses with PAUP* version 4.0b10 (Swofford, 1998) on each data set (HKY 85 model; Ti/Tv ratio, proportion of invariable sites and gamma shape parameter to be estimated; use of empirical frequencies), with and without the constraint of an A + M clade. A Kishino–Hasegawa (KH) test (Kishino and Hasegawa, 1989) was conducted using PAUP (RELL approximation, 1000 replicates) to compare the best overall ML tree that includes a B + M clade (hypothesis inferred by Cameron (1993) using 16S sequences), with the best ML tree containing an A + M clade (hypothesis suggested by morphological characters). All three KH tests resulted in significant p values (0.026 for opsin, 0.048 for 28S, and 0.007 for cytochrome *b*), rejecting the null hypothesis that there is no difference between the two topologies. It is not appropriate to perform a KH test on the 16S data set analyzed in Cameron and Mardulyn (2001) as most of those sequences were used in Cameron (1993), which first suggested the B + M hypothesis (a posteriori testing, see Goldman et al., 2000).

Ascher et al. also suggested that Mardulyn and Cameron (1999) ignored numerous instances of incongruence between their favored tree (see Ascher et al.'s Fig. 1d) and their trees based on different modeling approaches, such as maximum likelihood or weighted parsimony, or trees based on different genes. Ascher et al.'s Fig. 1 shows a tree (Fig. 1g) that is supposed to represent a putative "incongruent" opsin ML tree from Mardulyn and Cameron's paper. This tree is in fact not the ML tree from Mardulyn and Cameron (Fig. 5). Instead, they show a tree with a B + E clade in place of the correct B + M clade. This is an error, given that the actual opsin ML tree reported in Mardulyn and Cameron's Fig. 5 recovers a B + M clade with 71% bootstrap support.

Ascher et al.'s introductory claim that "not a single morphological character has ever been reported that would support Bombini + Meliponini" (technically inaccurate—see data set of Roig-Alsina and Michener, 1993) offers nothing to discount Mardulyn and Cameron's results. *All* molecular data sets thus far analyzed, going back to the earliest attempts to sequence the corbiculate bees in the late 1980s, to recent analyses (reviewed in Cameron and Mardulyn, 2001), are entirely consistent in their findings that Bombini + Meliponini form a clade. The number of genetic markers showing this relationship is now up to five, including a preliminary EF-1α data set (Sipes, Danforth and Cameron, unpublished data).

In summary, the major goals of Ascher et al. were to analyze corbiculate bee relationships with multiple outgroups, and to examine the phylogenetic signal present in the opsin gene. However, the authors provide no new perspective. Their inability to recover monophyly of the corbiculate clade is mirrored by other genes, and the opsin gene in fact (even in their analyses) corroborates the Bombini + Meliponini relationship inferred for the corbiculate tribes using all other genes, independently or combined. Their yardstick for success appears to be whether molecular results are concordant with tree topologies based on prior morphological character sets (primarily Roig-Alsina and Michener, 1993). As that previous study contained many clades with relatively low support (as Roig-Alsina and Michener discuss), this is not a robust approach to measuring phylogenetic accuracy. Accuracy can only be determined with reference to a known, or at least strongly corroborated, phylogeny. At present, this is provided only by the molecular data.

In conclusion, none of the molecular data sets from different genes, all of which are entirely consistent among one another, are concordant with the most definitive morphological data set (Roig-Alsina and Michener, 1993) for the corbiculate bees. The more significant question is *why*? (Lockhart and Cameron, 2001). Determining the cause of this lack of concordance will only be achieved by collecting additional data from diverse sources, including morphology and DNA, and analyzing them rigorously. Of key interest is to understand what is going on with the non-monophyly of the corbiculate bees: Is this the result of a long branch attraction problem (Mardulyn and Cameron, 1999 tested for this), some other kind of bias, or are the corbiculate bees indeed not monophyletic?

Acknowledgments

We thank two anonymous reviewers for useful comments on a previous version of this manuscript. P.M. is

a Research Associate at the F.N.R.S. (National Fund for Scientific Research) in Belgium. Our research on the corbiculate bees was funded by the NSF (Grant No. GER-9450117).

References

Ascher, J.S., Danforth, B.N., Ji, S., 2001. Phylogenetic utility of the major opsin in bees (Hymenoptera: Apoidea): a reassessment. Mol. Phylogenet. Evol. 19, 76–93.

Cameron, S.A., 1993. Multiple origins of advanced eusociality in bees inferred from mitochondrial DNA sequences. Proc. Natl. Acad. Sci. USA 90, 8687–8691.

Cameron, S.A., Mardulyn, P., 2001. Multiple data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). Syst. Biol. 50, 194–214.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46, 69–74.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Lockhart, P.J., Cameron, S.A., 2001. Trees of bees. TREE 16, 84–88.

Mardulyn, P., Cameron, S.A., 1999. The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. Mol. Phylogenet. Evol. 12, 168–176.

Roig-Alsina, A., Michener, C.D., 1993. Studies of the phylogeny and classification of long-tongued bees (Hymenoptera: Apoidea). Univ. Kansas Sci. Bull. 55, 124–162.

Rokas, A., Nylander, J.A.A., Ronquist, F., Stone, G.N., 2002. A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. Mol. Phylogenet. Evol. 22, 206–219.

Sakagami, S.F., Michener, C.D., 1987. Tribes of Xylocopinae and origin of the Apidae. Ann. Entomol. Soc. Am. 80, 439–450.

Schultz, T., Engel, M.S., Prentice, M., 1999. Resolving conflict between morphological and molecular evidence for the origin of eusociality in the "corbiculate" bees (Hymenoptera: Apidae): a hypothesis-testing approach. Univ. Kansas Nat. Hist. Mus. Spec. Publ. 24, 125–138.

Steel, M., Huson, D.H., Lockhart, J.P., 2000. Invariable sites models and their use in phylogeny reconstruction. Syst. Biol. 49, 225–232.

Swofford, D.L., 1998. PAUP* 4.0: Phylogenetic Analysis Using Parsimony (And Other Methods). Sinauer, Sunderland, MA.