Phylogenetic Uncertainty and Taxonomic Re-revisions: An Example from the Australian Short-necked Turtles (Testudines: Chelidae)

Phillip Q. Spinks[1], Arthur Georges[2], and H. Bradley Shaffer[1]

Molecular data have greatly influenced our concepts of species and their relationships in the last few decades, and as a consequence the taxonomy of most vertebrate clades has been repeatedly revised to reflect phylogeny. However, as larger and more complete molecular data sets become available, the sometimes striking disparities between taxonomic revisions based on individual gene trees (particularly those based on mitochondrial DNA) and species trees has become increasingly apparent. Here, we present data from 13 nuclear and one mitochondrial gene. Our results demonstrate that the recent taxonomic proposal erecting the new Australian chelid genus *Flaviemys* (Testudines: Chelidae) was an unnecessary action, and that recognition of *Flaviemys* confuses, rather than clarifies, a phylogeny-based taxonomy of the group. Taxonomic actions have many broad repercussions, and we recommend that taxonomic changes should be proposed cautiously and only when they are based on the strongest possible data and analyses.

DURING the last decade, turtles (Testudines), perhaps more than any other vertebrate group, have been the focus of extensive taxonomic revisions. Essentially every major clade of turtle has been revised to some degree based on the results of one or more molecular phylogenetic analysis. Over the same time period, molecular systematics has grown as a field, both in terms of the amount of molecular data that are often available and the sophistication of downstream analyses. Consequently, one should expect that some of these recent taxonomic decisions may end up being partly or fully revised, reflecting progress in the phylogenetic knowledge on which they were based. Most recently, a number of studies (including some of our own) have analyzed one or more mitochondrial genes, perhaps augmented with a few nuclear loci, recovered trees that are mostly or fully resolved and well supported by the data in hand, discovered apparently paraphyletic taxa, and revised taxonomy under the guiding principle that named lineages should be monophyletic. However, as we now recognize, while the recovered trees might be well supported, the phylogenies may not accurately reflect the actual species phylogeny, particularly if one or a few gene trees dominate the analysis. This has led to a great deal of taxonomic confusion as names are changed, changed again, and sometimes changed back to the original configuration based on trees from different analyses that are each well supported but incongruent with one another.

Most of these previous analyses of testudines are based on analyses of relatively few independent loci (i.e., 1–6 mitochondrial DNA genes plus 1–6 nuclear loci). For some groups, this level of gene sampling appears adequate, while for others, phylogenies based on sparse gene sampling are inadequate. For example, Naro-Maciel et al. (2008) generated phylogenies for the sea turtles (Chelonioidea) from mtDNA (two genes) and nuDNA (five loci). Analyses of both data sets independently recovered the same tree topology, suggesting that the phylogeny for the sea turtles appears to be well resolved, and as a consequence the resulting taxonomy is probably stable. Revisions of other groups now appear to have been premature as subsequent phylogenetic analyses, based on expanded taxon or data sampling, produced phylogenies that are incongruent with those upon which the earlier taxonomic revisions were based. As one recent example, Iverson et al. (2013) generated phylogenies for the mud and musk turtles (family Kinosternidae) based on three mitochondrial DNA (mtDNA) and three nuclear DNA (nuDNA) loci, recovered the long-recognized turtle genus *Kinosternon* as paraphyletic with respect to *Sternotherus*, and reassigned six species of *Kinosternon* to the new genus *Cryptochelys*. However, in a follow up analysis, Spinks et al. (2014) generated phylogenies for the Kinosternidae based on 14 nuclear loci and recovered *Kinosternon* as monophyletic with respect to *Sternotherus* with strong support, but *Cryptochelys* as non-monophyletic with respect to the more restricted *Kinosternon*. In this case, the tree topology of Iverson et al. (2013) was driven by mitochondrial sequence variation, and not subsequently well supported by more extensive nuclear data. Based on the non-monophyly of *Cryptochelys,* Spinks et al. (2014) suggested that the recognition of *Cryptochelys* was premature because the phylogeny of the Kinosternidae is not stable (see also the Guidelines for Taxonomic Changes in Turtle Taxonomy Working Group [TTWG], 2014). Although taxonomy should always be based on the best available data, and data and analyses are always subject to change and revision, we also recognize that it is now much easier to bring larger, more robust data sets to critically important taxonomic revisions. Such data sets should help settle the sometimes unsettled taxonomy surrounding many groups.

Here we revisit the recent phylogenetic analysis and taxonomic revision of the Australian short-necked turtles (Testudines: Chelidae). The phylogenetic and taxonomic history of this group is complicated by the fact that our understanding of species diversity and the geographic distribution of the contained species has changed dramatically in recent years. Legler (1981) foreshadowed splitting the genus *Elseya* into two major clades, one containing *Elseya dentata* and related species, the other containing *Elseya latisternum* and its close relatives, many of which were undescribed at that time. A subsequent study based on 54 allozyme loci (Georges and Adams, 1996) established the

[1] Department of Ecology and Evolutionary Biology, and La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, California 90095; E-mail: (PQS) pqspinks@ucla.edu; and (HBS) brad.shaffer@ucla.edu. Send reprint requests to PQS.
[2] Institute for Applied Ecology, University of Canberra, Canberra, ACT 2601, Australia; E-mail: georges@aerg.canberra.edu.au.
Submitted: 16 September 2014. Accepted: 29 January 2015. Associate Editor: D. S. Siegel.
 DOI: 10.1643/CH-14-147 Published online: July 31, 2015

''*latisternum* group'' as monophyletic, which was the foundation for the description of the new genus *Myuchelys* (Thomson and Georges, 2009) with four contained species (*M. purvisi, M. georgesi, M. bellii,* and *M. latisternum* as the type species). Two subsequent analyses based on mtDNA and/or a limited set of nuDNA sequence data were equivocal on the monophyly of *Myuchelys*. Georges et al. (1999) recovered *Myuchelys* as paraphyletic with respect to *Elseya* based on analyses of two mitochondrial genes (12S rRNA and 16S rRNA), but without statistical support. More recently, analyses of a single nuclear locus (*c-mos*) provided moderate support (83% bootstrap support values) for grouping *Emydura macquarii*, *M. latisternum*, and *M. georgesi* as a clade to the exclusion of *M. purvisi*, a result confirmed by analysis of mtDNA (Fielder et al., 2012). However, Georges and Adams (1996), Georges et al. (1999), and Fielder et al. (2012) all recognized that the uncertainty surrounding incongruence among these analyses should preclude taxonomic revisions and therefore did not propose revisions to correct the potential paraphyly of *Myuchelys* with respect to *Emydura*.

Most recently, Le et al. (2013) generated phylogenies for the chelid genera *Elseya, Emydura, Myuchelys,* and the monotypic genera *Elusor macrurus* and *Rheodytes leukops* using two mtDNA and a single nuDNA marker. The phylogeny recovered by Le et al. (2013) recovered *Myuchelys* as paraphyletic, again owing to the position of *M. purvisi*. Le et al. (2013) subsequently assigned *purvisi* to a new genus, *Flaviemys,* to maintain monophyly of *Myuchelys*.

In this study, we conducted an analysis of this problematic group of chelids using 13 additional independent nuDNA markers. Our analysis indicates that *purvisi* falls within a well-supported, monophyletic *Myuchelys* based on the weight of available evidence, and therefore that the taxonomic revision of Le et al. (2013) based on their more limited sequence information was premature.

MATERIALS AND METHODS

Taxon and data sampling.—Our taxon sampling consisted of 37 individuals including two *Chelodina* outgroups plus one or two individuals for all currently recognized species of *Elseya* and at least two individuals per species for all *Emydura*, *Elusor*, *Myuchelys*, and *Rheodytes*. In addition, we included samples from *Elseya* sp. aff. South Alligator (Georges and Adams, 1996) that may represent a currently undescribed species (Spinks et al., 2015: Appendix S1). We generated mtDNA data from the cytochrome oxidase subunit I (*COI*) gene, and 13 nuclear nuclear loci (Spinks et al., 2015: Appendix S1). DNA was extracted from blood or soft tissue samples using a salt extraction protocol, and partial sequences of all loci were generated using 20 μl volume PCR reactions. PCR conditions included an initial denaturation of 60 s at 95°C, followed by 40 cycles of denaturation (94°C for 30 s), annealing (45 s at 57–62°C), and template extension of (72°C for 60 s) with a final extension period (72°C for 10 min). Locus-specific annealing temperatures, extension times and primers for most markers can be found in Spinks et al. (2014). We redesigned primers for the *COI*, *HNFL,* and *TB73* loci because they sometimes failed to amplify or sequence well using the original primers. The redesigned primer sequences are provided in Appendix S1. All PCR products were sequenced bidirectionally by Beckman Coulter Genomics (http://www.beckmangenomics.com/).

Phylogenetic analyses.—*COI* and the nuclear exons were translated using Geneious v5.1 (available from http://www.geneious.com) to determine if pseudogenes may have been sequenced unintentionally. The presence of unexpected stop codons or frame shifts would signify possible pseudogene contamination. We used the MAFFT software (Katoh et al., 2002, implemented in Geneious) to align the sequence data and the PartitionFinder software (Lanfear et al., 2012) to choose both a partitioning strategy and models of molecular evolution. We performed partitioned-model Bayesian phylogenetic analyses using MrBayes v3.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) for the mtDNA and nuDNA data sets separately; the mtDNA and nuDNA were not concatenated. The mtDNA and nuDNA Bayesian analyses consisted of two independent runs each comprising four incrementally heated chains that ran for 10,000,000 generations, and we sampled the posterior distribution every 1000 generations. Stationarity was assumed when the potential scale reduction factor equaled 1, and the average standard deviation of split frequencies between independent runs approached 0. We examined the MCMC samples in Tracer and AWTY (Nylander et al., 2008; Rambaut et al., 2014) to confirm that all chains were sampling from the same target distribution. We discarded the first 25% of samples as burnin, provided the chains had reached stationarity prior to this point. Most MrBayes analyses were carried out through the CIPRES Web portal (Miller et al., 2010).

We also reconstructed a species tree from nuDNA loci using the *BEAST software (Heled and Drummond, 2010) from the BEAST v1.8.1 package (Drummond and Rambaut, 2007). For these analyses, trees, clock, and substitution models were unlinked, and we used a Yule species tree prior and piecewise linear and constant root for the population size model. The exponential uncorrelated relaxed clock model and the HKY model of nucleotide sequence evolution was used for each partition. We ran the analyses for 400 million generations and sampling every 10000 generations using the Cipres web portal (Miller et al., 2010). Convergence was assessed by examining MCMC samples in Tracer (Rambaut et al., 2014).

RESULTS

Mitochondrial phylogeny.—Our mtDNA data set was composed of up to 727 base pairs (bp) for 33 individuals (31 ingroup short-neck chelids, two outgroup long-neck *Chelodina*). The matrix was almost complete with ~1% missing data (Spinks et al., 2015), and all sequences generated here were submitted to GenBank (Spinks et al., 2015: Appendix S1). The mtDNA sequences were partitioned by codon for analysis. The majority-rule consensus of the posterior distribution of trees from the Bayesian analysis was not well supported, but was very similar to the tree of Le et al. (2013). For example, we recovered *Emydura* and *Elseya* each as monophyletic, but without strong support (Fig. 1). In addition, like Le et al. (2013) we recovered *Myuchelys* as paraphyletic due to the position of *M. purvisi*. However, we also recovered *Elseya irwini* from the Burdekin and North Johnstone rivers and *Emydura macquarii* as paraphyletic (Fig. 1).

Concatenated nuDNA phylogeny.—Our 13-locus nuDNA data set was composed of up to 9833 bp for 37 individuals. Excluding gaps imposed by the alignment, the minimum

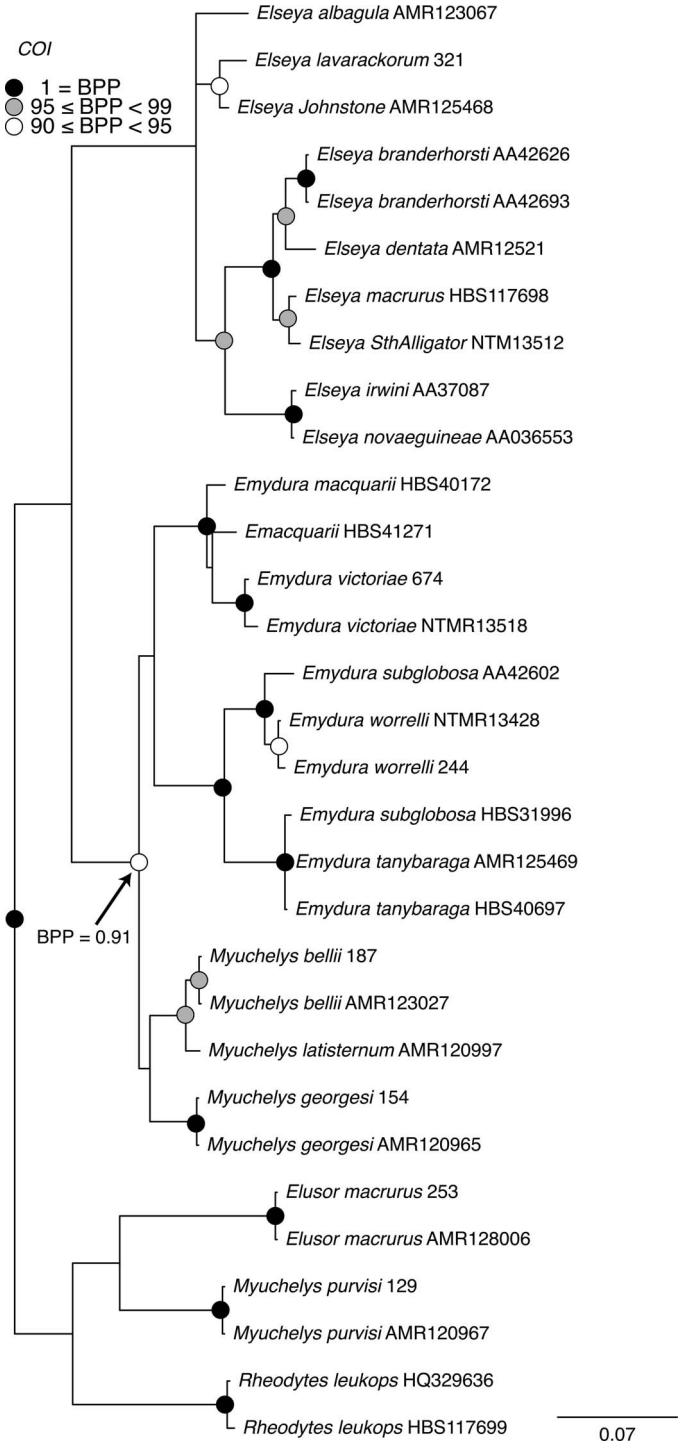

Fig. 1. Majority-rule consensus of the posterior distribution of trees from the Bayesian analysis of the *COI* data set (33 individuals, 727 bp). Bayesian posterior probabilities (BPP) greater than or equal to 90 are shown at appropriate nodes. *Myuchelys* is reconstructed as paraphyletic due to the position of *M. purvisi*, although support for the key node supporting this is relatively low (BPP = 0.91). Outgroups were removed for clarity of presentation.

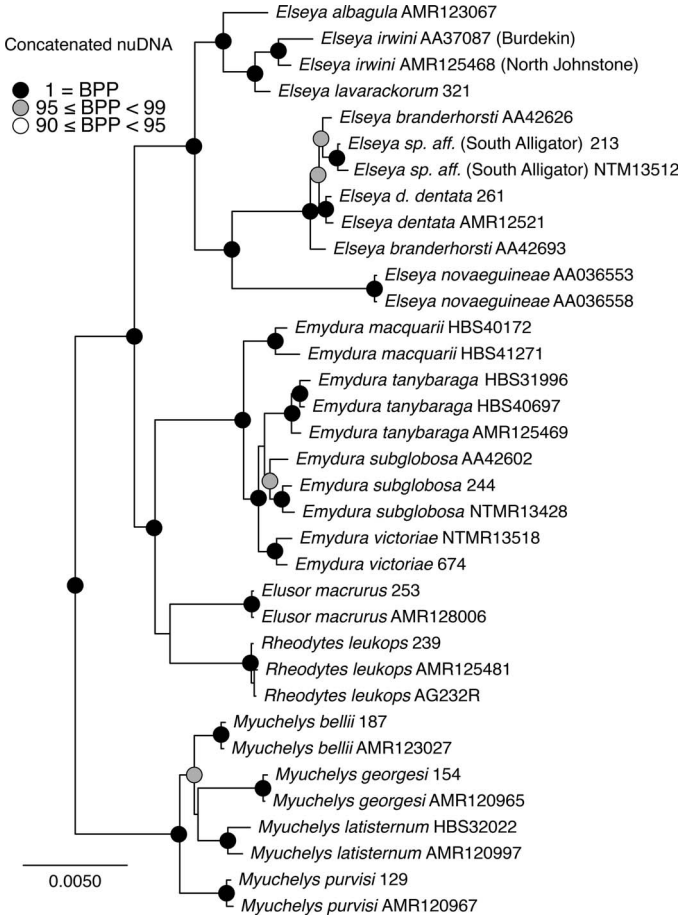

Fig. 2. Majority-rule consensus tree of the posterior distribution of trees from the Bayesian analyses of the concatenated nuclear loci data set (37 individuals, 13 loci, 9833 bp). Bayesian support values (BPP) are indicated, and the outgroups were removed for clarity of presentation.

sequence length = 7444 bp, and the maximum sequence length = 9754 bp. The matrix was also nearly complete with ~2.1% missing data. All sequences generated here were submitted to GenBank (Spinks et al., 2015: Appendix S1), and our nuDNA matrix was submitted to the Dryad Digital Repository (Spinks et al., 2015). The optimal partitioning strategy selected via PartitionFinder was a two-partition model with seven loci (*AHR, HMGB2, HNFL, P26s4, PAX, R35,* and *TB73*) assigned to partition 1 and the remaining six loci (*AIING, BDNF, BMP2, RAG, TB01,* and *ZFHX1B*) assigned to partition 2. We found no indication of pseudogenic sequences in our data.

The majority-rule consensus of the posterior distribution of trees from the Bayesian analysis of the partitioned data set was fully resolved and very well supported, with 26/34 ingroup nodes supported with Bayesian posterior probabilities (PP) of 1.0, four nodes with support >0.95, and only four nodes with <0.90 support. Importantly, we recovered strong support (PP = 1.0) for the reciprocal monophyly of the genera *Elseya*, *Emydura*, *Elusor*, *Rheodytes,* and *Myuchelys* inclusive of *M. purvisi* (Fig. 2). Although our taxon sampling is very limited, we recovered all other species for which we had more than one sample as monophyletic, with the exception of *Elseya branderhorsti* (Fig. 2).

Results from our species tree analysis were less well supported compared to the Bayesian phylogenetic analyses but were topologically very similar except for relationships within *Myuchelys*. Under the coalescent model implemented in *BEAST, *M. belli* is the sister taxon to the remaining *Myuchelys*, whereas the MrBayes analyses recovered *M. purvisi* as sister to the remaining *Myuchelys* (Figs. 2, 3). In addition, it is important to note that *BEAST is designed to estimate a species tree, and therefore requires that taxa be assigned to species *a priori*. Thus, the apparent monophyly

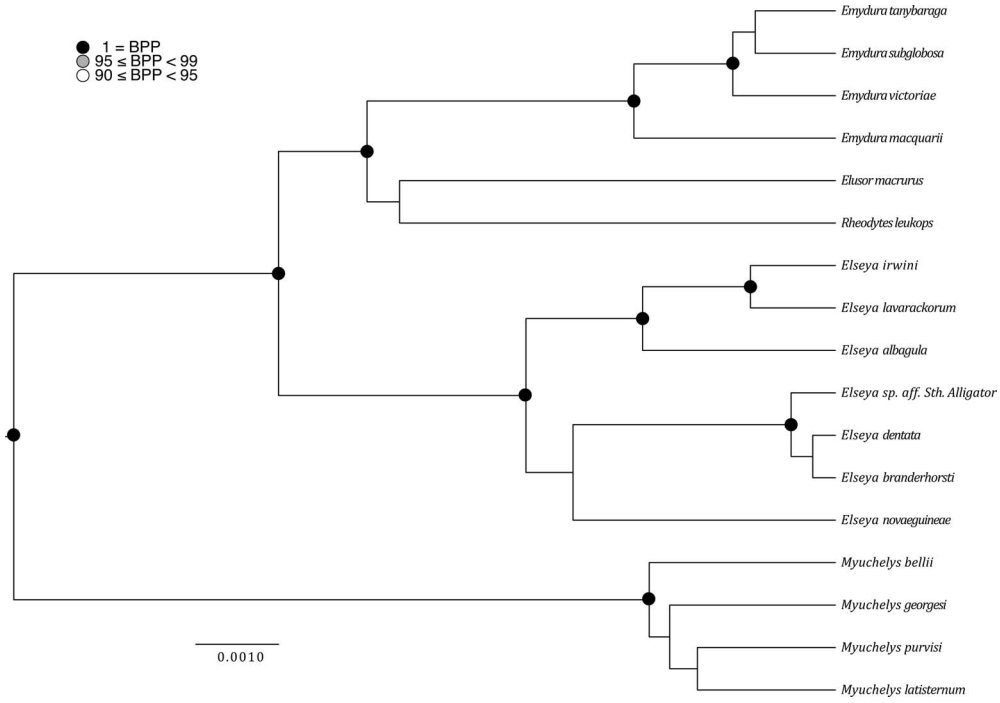

Fig. 3. Majority-rule consensus tree of the posterior distribution of trees from the species tree analyses of the concatenated nuclear loci data set (37 individuals, 13 loci, 9833 bp). Bayesian support values (BPP) are indicated, and the outgroups were excluded from this analysis.

of *E. branderhorsti* in Figure 3 is a necessary consequence of the *a priori* taxon assignments, rather than an empirical test of monophyly for these species.

Single gene analyses.—Phylogenies generated from single nuDNA loci varied somewhat in their overall resolution and support values, but were mostly well resolved at the generic level, and congruent with the tree from the concatenated analysis (Spinks et al., 2015: Supplementary Figures S4–S7). Importantly, *Myuchelys* inclusive of *M. purvisi* was monophyletic with strong support at 10/13 nuclear loci, and analyses of the remaining three loci either recovered *Myuchelys* as monophyletic but with low support, or were not well resolved. *Myuchelys* including *M. purvisi* was never recovered as paraphyletic with strong support (Spinks et al., 2015: Supplementary Figures S4–S7).

DISCUSSION

Phylogenetic analyses and resulting taxonomic revisions based on relatively sparse geographic or genetic coverage are often tenuous, and therefore prone to change as additional, more comprehensive analyses based on expanded taxon and nucleotide sequence data sets become available. As these more extensive data sets become available, molecular phylogenetic analyses will almost certainly continue to identify cases of putative paraphyly. The tension for the systematics community is determining when the data are sufficiently compelling that a taxonomic change is justified and appropriate. We favor a conservative approach and suggest that existing taxonomy should only be revised when the weight of evidence from multiple independent lines of evidence indicates that the existing taxonomy is in conflict with phylogeny. Caution is especially warranted when analyses are largely or exclusively driven by a single marker, as is often the case for fast-evolving, phylogenetically informative mtDNA gene trees. Because the potential for analyses of mtDNA to produce trees that do not reflect species trees is now well known and thoroughly demonstrated (Moore, 1995; Funk and Omland, 2003; Dupuis et al., 2012; Toews and Brelsford, 2012), we hope that such single-gene dominated analyses will be treated as hypotheses for additional testing rather than the basis for taxonomic change. Further, the addition of a few nuclear loci does not necessarily incorporate sufficient information to adequately characterize the nuclear genome. For example, the phylogenetic signal from one or a few nuclear loci rarely equals that of a mitochondrial locus; thus, when a few nuDNA loci are combined with mtDNA, the resulting tree often still reflects the stronger mitochondrial phylogenetic signal (Moore, 1995; Hudson and Coyne, 2002; Zhang and Hewitt, 2003; Brito and Edwards, 2009).

In the current case, our current data set suggests that splitting *Myuchelys* and the shift of *M. purvisi* into the monotypic *Flaviemys* was unwarranted, and we strongly recommend the return of *M. purvisi* to the genus *Myuchelys*. In retrospect, following a more conservative approach might have led Le et al. (2013) to call attention to a potential problem with the taxonomy of *Myuchelys,* but not to propose a taxonomic revision until additional data were available to confirm their results. Such a course of action would have avoided, for example, the most recent compilation of testudine taxonomy (TTWG, 2014) from incorporating *Flaviemys* into their yearly nomenclatural update, only to potentially reject it in a future revision. Although slower to correct non-monophyly issues in our current taxonomy, we view this approach as preferable to the change-reconsider-change again approach to taxonomy that we face in a number of groups at present. Given the twin goals of stability and phylogenetic accuracy in taxonomy, our strong recommendation is that taxonomic revisions should be approached cautiously and carried out only when data from multiple lines of evidence provide a well-supported phylogeny that is at odds with the existing taxonomy (Spinks et al., 2014; TTWG, 2014).

ACKNOWLEDGMENTS

This work was supported in part by funding from NSF-DEB 0817042 and DEB 1239961.

LITERATURE CITED

Brito, P. H., and S. V. Edwards. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135:439–455.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Dupuis, J. R., A. D. Roe, and F. A. Sperling. 2012. Multilocus species delimitation in closely related animals and fungi: one marker is not enough. Molecular Ecology 21:4422–4436.

Fielder, D., K. Vernes, E. Alacs, and A. Georges. 2012. Mitochondrial variation among Australian freshwater turtles (genus *Myuchelys*), with special reference to the endangered *M. bellii*. Endangered Species Research 17:63–71.

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution and Systematics 34:397–423.

Georges, A., and M. Adams. 1996. Electrophoretic delineation of species boundaries within the short-necked chelid turtles of Australia. Zoological Journal of the Linnean Society 118:241–260.

Georges, A., J. Birrell, K. M. Saint, W. P. McCord, and S. C. Donnellan. 1999. A phylogeny for side-necked turtles (Chelonia: Pleurodira) based on mitochondrial and nuclear gene sequence variation. Biological Journal of the Linnean Society 67:213–246.

Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27:570–580.

Hudson, R. R., and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755.

Iverson, J. B., M. Le, and C. Ingram. 2013. Molecular phylogenetics of the mud and musk turtle family Kinosternidae. Molecular Phylogenetics and Evolution 69:929–939.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059–3066.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29:1695–1701.

Le, M., B. N. Reid, W. P. McCord, E. Naro-Maciel, C. J. Raxworthy, G. Amato, and A. Georges. 2013. Resolving the phylogenetic history of the short-necked turtles, genera *Elseya* and *Myuchelys* (Testudines: Chelidae) from Australia and New Guinea. Molecular Phylogenetics and Evolution 68:251–258.

Legler, J. M. 1981. The taxonomy, distribution, and ecology of Australian freshwater turtles (Testudines: Pleurodira: Chelidae). National Geographic Society Research Reports 13:391–404.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p. 1–8. *In:* Proceedings of the Gateway Computing Environments Workshop (GCE). 14 November 2010, New Orleans, Louisiana.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726.

Naro-Maciel, E., M. Le, N. N. FitzSimmons, and G. Amato. 2008. Evolutionary relationships of marine turtles: a molecular phylogeny based on nuclear and mitochondrial genes. Molecular Phylogenetics and Evolution 49:659–662.

Nylander, J. A., J. C. Wilgenbusch, D. L. Warren, and D. L. Swofford. 2008. AWTY (Are We There Yet): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Spinks, P. Q., A. Georges, and H. B. Shaffer. 2015. Data from: Phylogenetic uncertainty and taxonomic re-revisions: an example from the Australian Short-necked Turtles (Testudines: Chelidae). Dryad Digital Repository: http://doi.org/10.5061/dryad.6q6sq

Spinks, P. Q., R. C. Thomson, M. Gidiş, and H. B. Shaffer. 2014. Multilocus phylogeny of the New-World mud turtles (Kinosternidae) supports the traditional classification of the group. Molecular Phylogenetics and Evolution 76:254–260.

Thomson, S., and A. Georges. 2009. A new genus for *Elseya latisternum* and related forms of Australian freshwater turtle (Testudines: Pleurodira: Chelidae). Zootaxa 2053:32–42.

Toews, D. P., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology 21:3907–3930.

Turtle Taxonomy Working Group [P. P. van DiJk, J. B. Iverson, A. G. J. Rhodin, H. B. Shaffer, and R. Bour]. 2014. Turtles of the World. Seventh edition. Annotated checklist of taxonomy, synonymy, distribution with maps, and conservation status. Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group. *In*: A. G. J. Rhodin, P. C. H. Pritchard, P. P. van Dijk, R. A. Saumure, K. A. Buhlmann, J. B. Iverson, and R. A. Mittermeier. Chelonian Research Monographs 5(7):000.329–479. DOI: 10.3854/crm5.000.checklist.v7. 2014.

Zhang, D. X., and G. M. Hewitt. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology 12:563–584.