

BARCODING

DNA barcoding using chitons (genus *Mopalia*)

RYAN P. KELLY,*†INDRA NEIL SARKAR,† DOUGLAS J. EERNISSE‡ and ROB DESALLE*†

*Columbia University, Department of Ecology, Evolution, and Environmental Biology, 10th Floor Schermerhorn Ext., 1200 Amsterdam Avenue, New York, NY 10027, USA, †Division of Invertebrates, American Museum of Natural History, 79th Street at Central Park West, New York New York 10024, USA, ‡Department of Biological Science (MH-282), California State University, Fullerton, 800 North State College Blvd., Fullerton, CA 92831-3599, USA

Abstract

Incorporating substantial intraspecific genetic variation for 19 species from 131 individual chitons, genus *Mopalia* (Mollusca: Polyplacophora), we present rigorous DNA barcodes for this genus as per the currently accepted approaches to DNA barcoding. We also have performed a second kind of analysis that does not rely on BLAST or the distance-based neighbour-joining approach as currently resides on the Barcode of Life Data Systems website. Our character-based approach, called characteristic attribute organization system, returns fast, accurate, character-based diagnostics and can unambiguously distinguish between even closely related species based on these diagnostics. Using statistical subsampling approaches with our original data matrix, we show that the method outperforms BLAST and is equally effective as the neighbour-joining approach. Our approach differs from the neighbour-joining approach in that the end-product is a list of diagnostic nucleotide positions that can be used in descriptions of species. In addition, the diagnostics obtained from this character-based approach can be used to design oligonucleotides for detection arrays, polymerase chain reaction drop off diagnostics, TaqMan assays, and design of primers for generating short fragments that encompass regions containing diagnostics in the cytochrome oxidase I gene.

Keywords: character based, chiton, diagnostics, DNA barcoding, DNA taxonomy, *Mopalia*

Received 18 September 2006; revision accepted 30 October 2006

Until recently, DNA barcoding has focused on describing collections of species within geographical areas (Hebert *et al.* 2003a) rather than monophyletic groups [but see Meyer & Paulay 2005; the All Birds Barcoding Initiative (ABBI; <http://barcoding.si.edu/AllBirds.htm>) and the Fish Barcode of Life (FishBoL; <http://www.fishbol.org/>)]

Competing Interests. The authors have declared that no competing interests exist.

Author Contributions. R.P.K. collected and analysed the molecular data, collected a number of the samples from the field, and wrote the paper. I.N.S. conceived of and wrote the algorithm and the version implemented here. D.J.E. collected the majority of the specimens, preliminarily identified them, and helped to place the data set in a larger biological and intellectual context. R.D. conceived of using the data set for the purpose of testing the method presented herein and sponsored the work, which was carried out in his laboratory.

Correspondence: R. DeSalle, Fax: 212-769-5277;

E-mail: desalle@amnh.org

leaving open the question of whether sympatric, closely related species might be distinguished by the technique (Moritz & Cicero 2004). Using the genus *Mopalia* (Mollusca: Polyplacophora), we provide a challenging test case for the DNA barcoding approach, by comparing the efficacy of the approach using sympatric and recently radiated species within a monophyletic genus and covering a broad geographical range. We have generated DNA barcodes for a total of 19 species from the genus *Mopalia* and six closely related outgroup species using 131 individuals from these 25 total species. These individuals represent a significant amount of intraspecific variation in the genus.

Currently, the efficacy of DNA barcoding is assessed using tools established by the Consortium for Barcode of Life (CBoL) as outlined on the Barcode of Life Data Systems (BOLD) website (www.barcodinglife.org/). This approach uses a BLAST search (Altschul *et al.* 1997) to collect the top 100 BLAST hits and then constructs a neighbour-joining tree to allow for the attachment of a query sequence to this

100 best BLAST hit backbone tree (Hebert *et al.* 2003a; Hebert *et al.* 2003b; Steinke *et al.* 2005). While rapid and popular, approaches such as BLAST have some primary drawbacks: the occurrence of high rates of 'false-positives', and sampling-dependent accuracy is prevalent (Koski & Golding 2001). A false-positive result, where a query sequence is matched to a known sequence despite significant divergence, is a frequent consequence of using the BLAST algorithm by itself. This approach matches the query sequence with the least-distant sequence in the database, despite even significant divergence between the two. The loss of character information is also inherent to a distance-based approach. With respect to both BLAST and neighbour joining, by calculating pairwise distances, query sequences are reduced to vector distances expressing divergence between sequences; all character-state information is lost (DeSalle 2006). Finally, the accuracy of both BLAST and neighbour joining depends on the degree of disparity between intra- and interspecific variation (the 'barcoding gap'; Meyer & Paulay 2005). Insufficient taxon sampling will artificially inflate this disparity, increasing the apparent accuracy of the barcoding method.

Here, we present an alternative method of DNA barcoding that complements distance-based techniques. This character-based assessment called characteristic attribute organization system (CAOS; Sarkar *et al.* 2002a; 2002b) is rapid, avoids false-positives, retains evolutionary information contained in character-state data (Desalle *et al.* 2005) and is theoretically accurate independent of the degree of 'barcoding gap'. Further, because the character-based approach we suggest here adds a new perspective to the methods of DNA barcoding, a different kind of biological and taxonomic information is made available that is more in line with traditional taxonomy (DeSalle *et al.* 2005).

Materials and methods

Genus *Mopalia*

Mopalia is the most speciose chiton genus in the nearshore environment of the eastern Pacific Ocean (Eernisse *et al.* in press). Many species of the genus are large and common in the intertidal, presumably playing important roles in the ecology of the West Coast's rocky shorelines. However, the group is poorly studied, with *Mopalia* species being the focus of few recent publications (Clark 1991; Saito & Okutani 1991). Some ecological and life history details of *Mopalia* species have been described (Himmelman 1980; Piercy 1987), but much of the biology of these organisms remains undiscovered. Perhaps some of this lack of data can be attributed to the subtle morphological differences among the species; *Mopalia* encompasses 23 species (including a newly described taxon; D.J.E. *et al.* in preparation), many of which are difficult to distinguish

from one another by morphology alone. This difficulty in discerning species may be due to the recent origin and rapid radiation of the group, a scenario that is supported by both molecular and fossil data. The group's probable recent origin and the tremendous morphological similarity, therefore make *Mopalia* an attractive test case for examining DNA barcoding methodology. One of the promises of barcoding is that it may help researchers easily and accurately identify the organism that they have sampled, especially useful in difficult-to-identify and broadly sympatric species such as these.

Collection permits were obtained from the relevant agencies in Mexico, California, Oregon, Washington, British Columbia and Alaska. Voucher specimens corresponding to each DNA sequence were deposited in the mollusc collection of the Santa Barbara Museum of Natural History and tissue samples of each individual are maintained in Nitrogen vapor at $-110\text{ }^{\circ}\text{C}$ in the Monell Cryo Collection at the American Museum of Natural History.

Molecular data collection

Genomic DNA was extracted from specimens primarily collected from the field between 2002 and 2005 and kept in 95–100% EtOH at room temperature. In some cases, older tissues were used to supplement recent collections. Most of the older specimens whose genomic DNA was successfully extracted had been kept at $-80\text{ }^{\circ}\text{C}$ (D.J.E. collections, 1986–1989) or at room temperature in 80% EtOH (single *Mopalia cirrata* specimen, Los Angeles County Museum of Natural History, collected 1963). DNA was extracted from tissues using QIAGEN DNeasy kits, eluted in water and kept at $-20\text{ }^{\circ}\text{C}$ for short-term use. See Table 1 for a list of included species and numbers of individuals.

Polymerase chain reaction (PCR) was carried out on the DNA extracts, using standard reagents and the universal COI primers HCO and LCO (Folmer *et al.* 1994), annealing at $51\text{ }^{\circ}\text{C}$. In cases of degraded or low-concentration DNA extract, ready-made PCR beads (Amersham Biotech) were used in place of batch-mixed reagents to amplify product. PCR products were cleaned using either 96-well filter plates (Millipore Corp.) or AMPure beads (Agencourt Corp., protocols available from manufacturer).

DNA sequencing was carried out using ABI BigDye Terminator reactions (Applied Biosystems Inc.), with an annealing temperature of $50\text{ }^{\circ}\text{C}$. Cycle sequence products were cleaned with 70% isopropanol, 70% ethanol, resuspended in formamide, and read on an ABI 3730 automated sequencer (Applied Biosystems, Inc.). The resulting DNA sequences were verified by aligning reads from both 5' and 3' directions, using SEQUENCHER software (GeneCodes Corp.). Alignment was trivial, as there are no insertions or deletions present in the fragment sequenced for *Mopalia* or its outgroups. Sequences were finally edited in MACCLADE

Table 1 Number of individuals and number of localities collected for each species used for analysis. In addition, the number of pure (Pu) and private diagnostics in at least 80% Pr (> 80%) of the individuals of a species are also given

Species	<i>n</i> (Total)	Total localities (<i>n</i>)	Pu	Pr (> 80%)
<i>Mopalia ciliata</i>	4	2	22	0
<i>Mopalia cirrata</i>	1	1	35	0
<i>Mopalia hindsi</i>	7	3	19	9
<i>Mopalia imporata</i>	6	2	26	17
<i>Mopalia kennerleyi</i>	9	4	12	8
<i>Mopalia lignosa</i>	7	2	24	2
<i>Mopalia lionota</i>	7	2	65	71
<i>Mopalia lowei</i>	5	1	5	3
<i>Mopalia muscosa</i>	4	1	64	0
<i>Mopalia plumosa</i>	15	6	33	15
<i>Mopalia porifera</i>	4	1	54	0
<i>Mopalia retifera</i>	3	1	53	0
<i>Mopalia seta</i>	1	1	60	0
<i>Mopalia sinuata</i>	9	3	21	3
<i>Mopalia</i> sp. nov.	4	1	2	0
<i>Mopalia spectabilis/ferreirai</i>	11	2	8	6
<i>Mopalia swanii</i>	11	2	12	2
<i>Mopalia vespertina</i>	4	2	26	0
<i>Mopalia acuta</i>	4	1	2	0
Outgroups				
<i>Cryptochiton stelleri</i>	1	1	90	0
<i>Dendrochiton flectens</i>	4	1	74	0
<i>Dendrochiton thamnopus</i>	3	1	56	0
<i>Katharina tunicata</i>	3	3	91	0
<i>Placiphorella vellata</i>	1	1	36	0
<i>Tonicella lineata</i>	2	1	47	0
Total	131	N/A	1229	136

(Maddison & Maddison 2003), and MANIA (Swofford and Eernisse, unpublished) software aided in sequence editing and management. All DNA sequences were managed using the BOLD website which allowed us to deposit trace files with the DNA sequences into GenBank.

GenBank and BOLD data submission

We submitted the voucher information, sequence traces and sequence information to the BOLD using the spreadsheet templates provided by the BOLD website (www.barcodinglife.org). These spreadsheets allow for the inclusion of museum voucher information and establishment of barcode study specimen numbers for both archival studies and for manipulating specimens in the laboratory. Our barcoding project numbers for these specimens are MOPAL001–06 to MOPAL125–06. These barcode project numbers can be linked with three other collection numbers as listed in the project spreadsheets for this barcoding project (see BOLD website and Table S1,

Table 2 Overall accuracy values for each method of barcoding. Column 1 shows the accuracy of data partition B using partition A as either a guide tree (CAOS) or a database (BLAST). Column 2 shows the converse, and 3 shows the results of the methods when the same data were used both as guides/databases and query sequences

Method	Overall accuracy of partition B given A	Overall accuracy of partition A given B	Overall accuracy using entire data set
CAOS	100%	96%	100%
BLAST	69%	59%	100%
NJ	74%	79%	100%

Supplementary material) – AMNH AMCC frozen tissue archival numbers for the DNA samples, the various museum archive numbers for the voucher tissues and field collection numbers for the specimens. All of this archival information can be obtained by consulting the BOLD website using the barcoding project numbers listed above (MOPAL001–06 to MOPAL125–06). The sequences and traces generated for all specimens in this study are deposited under GenBank Accession nos EF159577–EF159701.

CAOS algorithm

In addition to using the BOLD submission and identification system (<http://www.barcodinglife.org>), we applied the characteristic attribute organization system (CAOS) to extract character-based diagnostics from the data set. In this method, a guide tree is first produced from an existing data set of DNA sequences. The guide tree can be generated using either maximum-likelihood or parsimony methods. Phylogenetic analyses of sequence data were accomplished using PAUP* version 4.0b10 (Swofford 2003) and PAUPrat (Sikes & Lewis 2001). Descriptive rule sets were generated from the diagnostic character states at each node on the parsimony tree (Sarkar *et al.* 2002a, 2002b). Once a rule set is obtained, a novel (query) gene sequence can be aligned to the existing data set and evaluated using this rule set. If the query sequence contains sufficient information, it will be placed in the most exclusive diagnosable cluster of individuals and will be unambiguously indicated by its diagnostic characters. If the query sequence contains insufficient information or cannot be placed in a group given its characters, CAOS will stop the analysis; in this way, false-positives are avoided and a sequence will not be identified to a level unsupported by its character data.

Although our technique makes use of a gene tree in order to identify diagnostic characters that may distinguish species, we wish to stress that this tree does not represent a phylogenetic hypothesis of species relationships. A gene tree represents data from a single locus and is oftentimes

not sufficient to estimate phylogeny; however, it may be used to assess character-state changes and identify synapomorphies that describe clades within that single-locus tree. We also note that this character-based method is consistent with the phylogenetic species concept (Cracraft 1983; Nixon & Wheeler 1990; Davis & Nixon 1992; Goldstein & Rob DeSalle 2000), in contrast to a distance-based assessment of diversity, which may identify clusters of similar entities but lacks a philosophical and practical link by which species are identified and named via the science of taxonomy (Lipscomb *et al.* 2003; Seberg *et al.* 2003).

Results

The Mopalia data set

Sequences 569-bp long from COI were sequenced from 116 individuals in 19 ingroup species and 14 individuals from outgroup species. This sampling constitutes nearly the entire genus; the four species missing are rare and subtidal or else in a remote part of the northwestern Pacific Ocean. An average of 6.1 individuals per species was sampled; 10 of 19 ingroup species are represented by ≥ 5 individuals, and most species (12/19) were sampled at multiple locations to fully incorporate intraspecific variation (Table 1). These data were managed in spreadsheets obtained from the BOLD website and deposited on that website. Table 1 also shows the frequency of occurrence of diagnostic sites. The table lists the number of sites that are pure diagnostics on their own (Pu) and the number of private sites (Sarkar *et al.* 2002a) that occur in at least 80% of the individuals for that species (Pr; $> 80\%$). We also provide two novel kinds of information. The first is a list of diagnostics or DNA barcodes; sites are numbered with the first base of the Fulmer HCO primer being base number 1, along with the character state that is diagnostic for that site (Table S2, Supplementary material; 'State' column). Second, we present what we hope will be helpful oligonucleotide sequences relevant to DNA barcoding diagnosis (Table S2; 'Oligo' column). These oligonucleotides could be used for the following diagnostic purposes: (i) primers for PCR drop off or *TaqMan* assays for the various species, (ii) primer sets that optimize the occurrence of diagnostic sites in a PCR fragment and produce PCR DNA fragments of 100 bp or less. This kind of information could be useful to any researcher working with degraded DNA or who wants to avoid amplifying an entire large region for identifying query sequences, (iii) table for DNA DNA microarray approaches.

Comparing CAOS to BLAST and BOLD NJ

Because the BOLD website uses both BLAST and neighbour-joining as part of their identification engine, we wanted

to test the efficacy of these two methods with respect to the CAOS approach with the chiton data set. In order to test all three of these barcoding methods rigorously, the data set of *Mopalia* COI sequences was divided into two sets of equal size. Since our original matrix was arranged in taxonomic order we decided that the best way to generate a trimmed matrix to construct a guide tree was to take every other individual and assign it to one data set, thus randomizing the inclusion of individuals into one of two data sets (called data set A and B), but ensuring most species were represented in each data set. Reciprocal tests were performed using one subdataset to create a 'guide' tree, and the other as 'new' query sequences in order to test the accuracy of the character-based barcoding. For testing BLAST, reciprocal tests were performed with the same data partitions: each subset was treated as a database and each individual (query) in the alternate set classified using BLAST (Altschul *et al.* 1997) where classification was based only on the top hit. For all the comparisons we did, the top hit was examined to ensure that it was truly the best top hit. For testing the BOLD NJ approach, reciprocal tests were performed with the same data partitions: again, each subset was treated as a database and each individual (query) in the alternate set classified using neighbour-joining analysis where classification was based on the position of the query individual in the resulting NJ tree.

Figure 1 shows the gene trees used to perform the test of CAOS, BLAST and BOLD NJ described above. By chance, one of the 50% subdatasets ('Set A') produced a well-resolved gene tree, whereas the other ('Set B') resulted in a highly polytomous tree. This had the fortuitous effect of allowing the CAOS method to be tested under both accommodating and adverse conditions. The classification's accuracy using CAOS was assessed by whether a taxon was correctly classified as belonging to a clade (relative to the complete data set tree, Fig. 1c) or as sister to an ancestral grouping (but not classified to a specific clade), but not over-classified to a more specific grouping than was correct according to the complete tree. Values for accuracy were compared following Sarkar *et al.* (2002b), which describes results as belonging to one of four categories: true-positive (TP), false-positive (FP; type I error), true-negative (TN), or false-negative (FN; Type II Error). Measures of accuracy were calculated as follows: recall, the fraction of time a sequence is placed into the clade in which it belongs, is $TP/(TP + FN)$; precision, the fraction of sequences placed in a clade that belong there, is $TP/(TP + FP)$; and overall accuracy, the proportion of sequences placed without any error, is $(TP + TN)/(TP + TN + FP + FN)$.

CAOS performed well even under the demanding conditions of a 50% reduction in data in the barcoding matrix and a poorly resolved guide tree. It correctly placed every query taxon in both reciprocal tests (100% recall), over-

classified few taxa (100% and 96% respective precision for guide trees A and B), and had no false-negative placements (100% recall for both guide trees). Thus, even given a basal polytomy in the guide tree, CAOS consistently classifies taxa into more specific groups based on greater character support.

BLAST had a higher rate of both FP and FN classifications (which for BLAST equates to FP classifications because BLAST always attempts to identify the most 'similar' result from a database), resulting in precision (and recall) values of 69% and 59% for guide trees A and B, respectively. In general, BLAST classified taxa into more specific groupings than were correct based on the complete tree (Fig. 1c); instead of stopping at a branch point, it over-classified a given taxon, going down an incorrect path. As a result, whereas CAOS will always classify correctly (but at possibly a more basal node), BLAST may erroneously misclassify a taxon because of the nature of the algorithm and how its results are represented and interpreted. BOLD NJ had lower precision (and recall) values than CAOS – 75% and 79%, respectively, for guide trees for data sets A and B. In this case, the higher rates of FP and FN classifications cannot automatically be attributed to FP classifications, because BOLD NJ will also attach query individuals at interior nodes in an NJ tree.

All three methods CAOS, BOLD NJ and BLAST had an overall accuracy of 100% when provided with the entire data set. Given that it performed with greater than 95% accuracy after a 50% reduction in data, CAOS appears to be more robust to missing data or small data sets than either BLAST or BOLD NJ. This further exemplifies the advantages of using a CAOS-based classification even at a lower sampling rate, which essentially increases the apparent 'gap'. Interestingly, in the few instances where all methods failed to correctly classify a query individual, the problem queries were the same individuals. However, with relatively complete databases, all three methods perform at 100% accuracy.

Discussion

The character-based CAOS method for DNA barcoding provides unambiguous, fast, and accurate identification of query sequences given even a modest 'guide' data set. Unambiguous classification is restricted to monophyletic clades: in the case of *Mopalia ferreirai*/*Mopalia spectabilis*, in which *M. spectabilis* is paraphyletic and *M. ferreirai* nests within *M. spectabilis*, individuals were simply classified as belonging to the *M. ferreirai*/*M. spectabilis* clade rather than being over-diagnosed as one or the other.

CAOS was sufficiently sensitive to distinguish each taxon in a monophyletic group of closely related species (including sister species pairs) using only a short stretch of mtDNA sequence data. Further, it provides an alternative to distance-based methods by preserving evolutionary information in the form of character data and by avoiding

false-positive identification of query sequences. The number of diagnostics supporting the placement of a query sequence in a clade can also function as a confidence value for the identification (see Table 1). For instance, *Mopalia porifera* with 54 diagnostics can be considered more precisely diagnosed than *Mopalia lowei* that has only five diagnostics. Finally, the 'barcoding gap' (the difference between within- and between-species distances (Meyer & Paulay 2005) is irrelevant using CAOS, as diagnostic character states distinguish monophyletic groups without reference to the relative degree of divergence within and among taxa.

In addition to the currently used BLAST and BOLD NJ approaches in DNA barcoding we suggest that character-based DNA barcodes can be used as a means of cataloguing diagnostic molecular characters useful for distinguishing species. These diagnostics can complement, but not replace, traditional morphological taxonomy (DeSalle *et al.* 2005). The CAOS approach is also a useful tool for identifying significant intraspecific divergences (particularly using distance methods), although species descriptions must depend on multiple lines of evidence. We further suggest that genetic information is one of several lines of evidence that might contribute to a new species description (Desalle *et al.* 2005). Additional evidence could include morphological, geographical, ecological, behavioural, or other quantifiable data that would be diagnostic for the new taxon. Hypotheses of new taxa should be corroborated so as to avoid circularity of reasoning (i.e. a new species description cannot be substantiated by the same evidence as that which is the basis for the description). By preserving character state information and reporting diagnostic differences for each node and hence for each species, our DNA barcoding method may be useful to taxonomists in identifying and describing divergent organismal groups (whether putative new species, ESUs, management units, etc.) that can be singled out for further scrutiny (Blaxter 2004). This method renders threshold values for 'species-level' divergences (Hebert *et al.* 2004; Meyer & Paulay 2005) unnecessary, addressing another critical issue for distance-based approaches.

The CAOS algorithm as applied to DNA barcoding avoids some of the 'pitfalls' of distance-based barcoding (Moritz & Cicero 2004) and puts the endeavour in a character-based framework. Including evolutionary and taxonomic information creates a context in which sequence data may be interpreted and makes real the promise of DNA barcoding, enabling researchers to rapidly identify and make sense of the diversity that confronts them.

Acknowledgements

Funding for this work was generously provided by Lerner-Gray Grants for Marine Research (2005, to R.P.K.), the National Science Foundation GK-12 program (DGE0231875 to Columbia University/R.P.K.), the Lounsberry Foundation (to R.D.) and the Lewis B and Dorothy Cullman Program in Molecular Systematics (to

R.D./I.N.S.). We gratefully acknowledge the cooperation of the following individuals and institutions, without whom this work would not have been possible: Collectors (A. Draeger, A. Rodriguez, G. Eckert, B. Sirenko, R. Clark, J. Sigwart and H. Keller), Field Assistants (E.C. Stone, M.-M. Villa, G. Snyder, J. Manke, R. Cole, C. Kelly, P. Kelly, C.A.S. Kelly, E. Atkinson), Curators (M. Coppolino, P. Mikkelsen, A. Corthals, J. Feinstein, L. Groves and P. Valentich-Scott), Institutions (University of Washington Friday Harbor Laboratories, Ambrose Monell Cryo Collection at the American Museum of Natural History), D. Stillman and A. Baker.

Supplementary material

The supplementary material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/MEN/MEN1641/MEN1641sm.htm>

Table S1 CBoL spread sheet showing the collection data, accession numbers AMNH sequence tags, AMCC accession numbers and Genbank accession numbers.

Table S2 Table showing the diagnostics for the species barcoded in this paper. The columns indicate the position of the diagnostic base relative to the first base in the primers listed in the materials and methods. The base pair diagnostic states are listed as well as a typical primer sequence for a microarray or a PCR dropout primer.

References

- Altschul SF, Thomas LM, Alejandro AS, Zhang J, Zhang Z, Webb M & David JL (1997) GAPPED BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.
- Blaxter ML (2004) The promise of DNA taxonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 669–679.
- Clark RN (1991) A new species of *Mopalia* (Polyplacophora: Mopaliidae) from the Northeast Pacific. *The Veliger*, **34**, 309–313.
- Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, **1**, 159–187.
- Davis JJ, Kevin CN (1992) Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, **41**, 421–435.
- DeSalle R (2006) What's in a character? *Journal of Biomedical Informatics*, **39** (1), 6–17.
- Desalle R, Mary Egan G, Mark S (2005) The unholy trinity: taxonomy, species delimitation, and DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1905–1916.
- Eernisse DJ, Clark RN, Draeger A (in press) Class Polyplacophora (Chitons). In: *Light and Smith Manual: Intertidal Invertebrates from Central California to Oregon*, 4th Ed. (ed. Carlton JT). University of California Press, Berkeley, California.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA Primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**.
- Goldstein PZ, Rob D (2000) Phylogenetic species, nested hierarchies, and character fixation. *Cladistics*, **16**, 364–384.
- Hebert PDN, Alina C, Shelley LB, Jeremy Rd (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **270**, 313–321.
- Hebert PDN, Sujevan R, Jeremy Rd (2003b) Barcoding animal life: cytochrome c oxidase subunit I divergences among closely related species. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **270**, S96–S99.
- Hebert PDN, Mark YS, Tyler SZ, Charles MF (2004) Identification of birds through DNA barcodes. *PLoS Computational Biology*, **2**, 1657–1663.
- Himmelman JH (1980) Reproductive cycle pattern in the chiton genus *Mopalia* (Polyplacophora). *Nautilus*, **94**, 39–49.
- Koski LB, Golding GB (2001) The closest BLAST hit is often not the nearest neighbor. *Journal of Molecular Ecology*, **52**, 540–542.
- Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology & Evolution*, **18**, 65–66.
- Maddison DR, Maddison WP (2003) *MACCLADE: Analysis of Phylogeny and Character Evolution*, Version 4.06. Sinauer Associates, Sunderland, Massachusetts.
- Meyer CP, Gustav Paulay (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Computational Biology*, **3**, e422.
- Moritz C, Cicero C (2004) DNA barcoding: promises and pitfalls. *PLoS Computational Biology*, **2**, e354.
- Nixon KC, Wheeler QD (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211–224.
- Piercy RD (1987) Habitat and food preferences in six Eastern Pacific chiton species (Mollusca: Polyplacophora). *The Veliger*, **29**, 388–393.
- Saito H, Okutani T (1991) Taxonomy of Japanese species of the genera *Mopalia* and *Plaxiphora* (Polyplacophora: Mopaliidae). *The Veliger*, **34**, 172–194.
- Sarkar IN, Joseph WT, Paul JP, David HF, Bernd S, Rob D (2002a) An automated phylogenetic key for classifying homeoboxes. *Molecular Phylogenetics and Evolution*, **24**, 388–399.
- Sarkar IN, Paul JP, Bael TE, Stanley SE, Siddall M, DeSalle R, David HF (2002b) Characteristic attributes in cancer microarrays. *Journal of Biomedical Informatics*, **35**, 111–122.
- Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology & Evolution*, **18**.
- Sikes DS, Lewis PO (2001) PAUP*: PAUP* implementation of the parsimony ratchet. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut.
- Steinke D, Vences M, Salzburger W, Meyer A (2005) TAXI: a software tool for DNA barcoding using distance methods. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1975–1980.
- Swofford DL (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.0. Sinauer Associates, Sunderland, Massachusetts.