Molecular Ecology Resources (2009) **9**, 1092–1101 doi: 10.1111/j.1755-0998.2009.02559.x

DNA BARCODING

DNA barcoding of stylommatophoran land snails: a test of existing sequences

ANGUS DAVISON, ROSE L. E. BLACKIE and GARY P. SCOTHERN
Institute of Genetics, School of Biology, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK

Abstract

DNA barcoding has attracted attention because it is a potentially simple and universal method for taxonomic assignment. One anticipated problem in applying the method to stylommatophoran land snails is that they frequently exhibit extreme divergence of mitochondrial DNA sequences, sometimes reaching 30% within species. We therefore trialled the utility of barcodes in identifying land snails, by analysing the stylommatophoran cytochrome oxidase subunit I sequences from GenBank. Two alignments of 381 and 228 base pairs were used to determine potential error rates among a test data set of 97 or 127 species, respectively. Identification success rates using neighbour-joining phylogenies were 92% for the longer sequence and 82% for the shorter sequence, indicating that a high degree of mitochondrial variation may actually be an advantage when using phylogeny-based methods for barcoding. There was, however, a large overlap between intra- and interspecific variation, with assignment failure (per cent of samples not placed with correct species) particularly associated with a low degree of mitochondrial variation (Kimura 2-parameter distance < 0.05) and a small GenBank sample size (< 25 per species). Thus, while the optimum intra/interspecific threshold value was 4%, this was associated with an overall error of 32% for the longer sequences and 44% for the shorter sequences. The high error rate necessitates that barcoding of land snails is a potentially useful method to discriminate species of land snail, but only when a baseline has first been established using conventional taxonomy and sample DNA sequences. There is no evidence for a barcoding gap, ruling out species discovery based on a threshold value alone.

Keywords: DNA barcoding, mitochondrial DNA, Mollusca, Stylommatophora

Received 3 October 2008; revision accepted 4 December 2008

Large-scale attempts to estimate recent losses of biodiversity are stalled by a poor knowledge of global species richness (Pimm *et al*. 1995), as fewer than 2 million species have been identified and described, against an estimated total number of species of at least 10 million (Caldecott *et al*. 1994). A further challenge is that taxonomists need to recall the names and diagnostic morphological features of species, a complex task considering that few professional biologists can reliably diagnose more than 1000 species. These circumstances have led to the recommendation that DNA barcodes be used as a rapid species identifier, resulting in a lively debate between proponents and detractors of the method (Hebert *et al*. 2003b; DeSalle *et al*. 2005; Dasmahapatra & Mallet 2006; Waugh 2007).

Correspondence: Angus Davison, Fax: 0115 823 0338; E-mail: angus.davison@nottingham.ac.uk

Animal barcoding studies to date have largely (but not wholly, Floyd *et al*. 2002) focused on mitochondrial DNA (mtDNA), particularly the most promising 'universal' candidate gene, cytochrome oxidase subunit I. Some of the major successes have been to highlight previously unrecognized species diversity. In the Neotropical skipper butterfly, 10 species are now recognized compared with one before, with similar results in other taxonomically diverse groups (Hebert *et al*. 2004a; Smith *et al*. 2007). The newly discovered species are real, because the designation based on DNA sequences is supported by subsequently discovered morphological and/or ecological data.

Despite the successes, there remains much controversy as to how widely DNA barcoding can be applied across taxa, and the extent to which it can or should replace traditional taxonomy (Tautz *et al*. 2003; Ebach & Holdrege 2005; Savolainen *et al*. 2005; Schindel & Miller 2005; Dasmahapatra

© 2009 Blackwell Publishing Ltd

& Mallet 2006). Opposing views argue that molecular methods of taxonomy focus solely on species identification and detract from other important aspects of organisms such as their functional morphology, behaviour and ecology, and that no single pattern can identify or define all species (Will & Rubinoff 2004; Smith 2005). Many opponents would argue that practically all barcoding papers have been based on profound taxonomic expertise.

The critical factor in successful DNA barcoding, and the one that has raised most concern, is its accuracy. It can only be relied on as a form of species identification if there is confidence that DNA sequences from a single species form a discrete unit. In an ideal world, there is a 'barcoding gap' — intraspecific variation has a discrete frequency distribution compared with interspecific variation. The extent of variation discovered within a group of sequences can then be used as the first step towards delineating species, in the complete absence of any a priori knowledge of species status.

Even the advocates of DNA barcoding acknowledge that while interspecific variation is usually greater than intraspecific variation, there is rarely a barcoding gap. In addition, the threshold used to delineate species must vary between taxonomic groups, as the extent of genetic variation inevitably depends upon a variety of factors, some of which may be phylum-specific (e.g. rates of molecular evolution), or others that are local (individual population history, hybridization and introgression). For each taxonomic grouping, therefore, it is preferable to test the effectiveness of DNA barcoding in a set of well-characterized species, hence establishing a baseline, before going on to make inferences on species status in others.

Although Mollusca is the second most speciose animal group after the arthropods, there have been rather few large-scale barcoding studies on the phylum, with some notable exceptions (Kelly *et al*. 2007; Campbell *et al*. 2008), the most prominent of which is a landmark study on marine gastropods (Meyer & Paulay 2005). Land snails, in particular, are one group for which there is a complete lack of baseline data for barcoding, despite the fact that it has long been recognized that they have an extraordinary mitochondrial diversity (Thomaz *et al*. 1996), begging an obvious question as to whether the diversity will be problematic for the method.

We therefore tested the validity of the DNA barcoding method in land snails of the infraorder Stylommatophora, a group that makes up the vast majority land snails, with a view to providing recommendations for future exploratory studies of less well-characterized land snails. This was carried out by downloading the extant cytochrome oxidase sequences from GenBank, and using these well-characterized sequences as exemplars and/or testers. Taking a lead from Meyer & Paulay (2005), two main questions were asked: what is the success rate in assigning an unknown individual to a species, given an exemplar data set, and what is the most suitable threshold of mitochondrial variation to use to delineate species?

Materials and methods

Although longer sequences are preferable for barcoding, we were limited by the fact that the sequences in GenBank were created for a variety of purposes, not necessarily for barcoding, and using different primer sets, so are a variety of lengths. We therefore created two data sets, aiming to maximize the length of sequence used in one ('long', 381 base pairs) and the number of species in the other ('short', 228 base pairs).

From both data sets, an exemplar alignment was created, by randomly removing all but one sequence per species (Table 1). This sequence was then used as the reference or *exemplar* barcode in subsequent analyses (Meyer & Paulay 2005; Meier *et al*. 2006). To test the effectiveness of barcodes in assigning land snail sequences to species, a *test* sequence for each species was randomly selected, and a phylogeny created using the neighbour-joining method and Kimura 2-parameter (K2P) distance measure in the PHYLIP 3.67 package. In the same manner as Meyer & Paulay (2005), identification was considered (i) correct, if the test sequence fell as the sister taxon of the corresponding exemplar species; (ii) incorrect, if it did not fall with the exemplar taxon; (iii) ambiguous, if the test sequence fell at a node below the exemplar taxon and another species from which it did not belong, making its assignment to one or the other equivocal (see Meyer and Paulay 2005 for further details).

The same data set was also used to estimate global intra/interspecific variation, with a view to establishing a threshold criterion for within or between species. This was done by compiling a file containing both the exemplar and test sequences. Distance matrices where then generated using the K2P model. Intraspecific distance was estimated as the K2P value between exemplar and test; interspecific distance as the K2P value between test and the nearest other species. For a given threshold value, the frequency of false positives (intraspecific variation misdiagnosed as 'novel' species) and false negatives (interspecific variation misdiagnosed as same species) were counted.

We also investigated the effect of sampling/haplotype frequency and cytochrome oxidase sequence variation on the ability of barcoding methods to assign sequences from the same species to a single monophyletic group. A phylogeny was constructed from the whole long data set. Then, for species with three or more haplotypes on GenBank, we counted the proportion of haplotypes in the largest monophyletic group for that species. If barcoding is effective, then all individuals of a species should in principle be monophyletic. Second, we estimated within-species cytochrome oxidase K2P distances and compared it to the degree of monophyly, to understand if sometimes large mitochondrial genetic diversity of land snails will be problematic for the method.

© 2009 Blackwell Publishing Ltd

Table 1 Samples used in this study and geographic origin

			Short data set (228 bp)		Long data set (381 bp)	
Family	Species	Location	Exemplar	Test	Exemplar	Test
Achatinellidae	*Achatinella apexfulva*	Hawaii	AF398185	AY148540	AF398185	AY148540
	Achatinella decipiens	Hawaii	AY148541	AY148543	AY148541	AY148543
	Achatinella fulgens	Hawaii	AY148550	AY148551	AY148550	AY148551
	Achatinella fuscobasis	Hawaii	AY148544	AY148549	AY148544	AY148548
	Achatinella lila	Hawaii	AY148552	AY148555	AY148553	AY148555
	Achatinella livida	Hawaii	AY044407	AY044409	AY044407	AY044408
	Achatinella mustelina	Hawaii	AY044338	AY044394	AY044338	AY044359
	Achatinella sowerbyana	Hawaii	AY044416	AY044423	AY044416	AY044422
	Newcombia cumingi	Hawaii	AY148505	AY148509	AY148505	AY148507
	Partulina crocea	Hawaii	AY148538	AY148539	AY148538	AY148539
	Partulina mighelsiana	Hawaii	AY148526	AY148527	AY148526	AY148527
	Partulina perdix	Hawaii	AY148514	AY148515	AY148514	AY148515
	Partulina physa	Hawaii	AY148510	AY148511	AY148510	AY148511
	Partulina proxima	Hawaii	AY148524	AY148525	AY148524	AY148525
	Partulina redfieldi	Hawaii	AY148512	AY148513	AY148512	AY148513
	Partulina semicarinata	Hawaii	AY148528	AY148529	AY148528	AY148529
	Partulina splendida	Hawaii	AY148516	AY148520	AY148516	AY148520
	Partulina tappaniana	Hawaii	AY148532	AY148536	AY148532	AY148534
	Partulina variabilis	Hawaii	AY148530	AY148531	AY148530	AY148531
	Perdicella helena	Hawaii	AY148522	AY148523	AY148522	AY148523
Agriolimacidae	*Deroceras laeve*	Western Europe	AF239733	EF128217	AF239733	EF128217
	Deroceras reticulatum	Western Europe	AM259702	AM259703	AF239734	AM259703
Arionidae	*Arion alpinus*	Western Europe	AY987867	DQ904249	AY987867	AY987868
	Arion circumscriptus	Western Europe	AY987872	AY987873	AY987872	AY987873
	Arion distinctus	Western Europe	AY423692	AY423701	AY423692	AY423698
	Arion fasciatus	Western Europe	AY987877	AY987878	AY987877	AY987879
	Arion flagellus	Western Europe	AY987880	AY987882	AY987880	AY987881
	Arion franciscoloi	Western Europe	AY987883	AY987884	AY987883	AY987884
	Arion fuscus	Western Europe	AJ809408	AJ809446	AJ809408	AJ809430
	Arion hortensis	Western Europe	AY423670	AY987889	AY423670	AY423691
	Arion intermedius	Western Europe	AY987890	AY987891	AY987890	AY987891
	Arion lusitanicus	Western Europe	AY987894	AY987895	AY987894	AY987895
	Arion owenii	Western Europe	AY423702	AY987898	AY423702	AY987898
	Arion rufus	Western Europe	AY987900	AY987903	AY987900	AY987902
	Arion silvaticus	Western Europe	AY987917	AY987918	AY987917	AY987918
	Arion subfuscus	Western Europe	AY987904	AY987916	AY987904	AY987915
	Arion transsylvanus	Western Europe	AY943858	AY943860	AY943858	AY943859
	Arion wiktori	Western Europe	AY987920	AY987921	AY987920	AY987921
	Kootenaia burkei	USA	AY382633	AY382636	AY382633	AY382635
Bradybaenidae	*Euhadra idzumonis*	Japan	AY251868	AY251869	AY251868	AY251869
	Euhadra peliomphala	Japan	AY251859	AY251862	AY251859	AY251862
	Satsuma adelinae	Ryukyu Islands	AB242532	AB242535	AB242532	AB242534
	Satsuma amanoi	Ryukyu Islands	AB242530	AB254964	AB242530	AB254964
	Satsuma eucosmia	Ryukyu Islands	AB242440	AB297744	AB242440	AB242490
	Satsuma hemihelvus	Ryukyu Islands	AB242553	AB254965	AB242553	AB242556
	Satsuma largillierti	Ryukyu Islands	AB242494	AB254962	AB242494	AB242521
	Satsuma omoro	Ryukyu Islands	AB242557	AB242558	AB242557	AB242558
	Satsuma shigetai	Ryukyu Islands	AB242546	AB297748	AB242546	AB297747
	Satsuma sororcula	Ryukyu Islands	AB242536	AB297746	AB242536	AB242545
	Satsuma tokunoshimana	Ryukyu Islands	AB242550	AB297749	AB242550	AB242552
Bulimulidae	*Bulimulus akamatus*	Galapagos Islands	DQ903721	DQ903722		
	Bulimulus amastroides	Galapagos Islands	DQ903728	DQ903729		
	Bulimulus blombergi	Galapagos Islands	DQ903738	DQ903739		
	Bulimulus darwini	Galapagos Islands	DQ903715	DQ903716		
	Bulimulus eschariferus	Galapagos Islands	DQ903730	DQ903731		

© 2009 Blackwell Publishing Ltd

Table 1 *Continued*

Family	Species	Location	Short data set (228 bp)		Long data set (381 bp)	
			Exemplar	Test	Exemplar	Test
	Bulimulus indefatigabilis	Galapagos Islands	DQ903708	DQ903709		
	Bulimulus jacobi	Galapagos Islands	DQ903717	DQ903718		
	Bulimulus nesioticus	Galapagos Islands	DQ903710	DQ903711		
	Bulimulus nux	Galapagos Islands	DQ903698	DQ903700		
	Bulimulus ochsneri	Galapagos Islands	DQ903726	DQ903727		
	Bulimulus olla	Galapagos Islands	DQ903706	DQ903707		
	Bulimulus perrus	Galapagos Islands	DQ903748	DQ903749		
	Bulimulus pinzonensis	Galapagos Islands	DQ903777	DQ903779		
	Bulimulus quitensis	Galapagos Islands	DQ903696	DQ903697		
	Bulimulus reibischi	Galapagos Islands	DQ903724	DQ903725		
	Bulimulus sculpturatus	Galapagos Islands	DQ903712	DQ903713		
	Bulimulus snodgrassi	Galapagos Islands	DQ903732	DQ903736		
	Bulimulus tanneri	Galapagos Islands	DQ903755	DQ903756		
	Bulimulus tortuganus	Galapagos Islands	DQ903740	DQ903743		
	Bulimulus unifasciatus	Galapagos Islands	DQ903701	DQ903702		
	Bulimulus wolfi	Galapagos Islands	DQ903719	DQ903720		
Chondrinidae	*Solatopupa guidoni*	Western Europe	DQ305076	DQ305080		
	Solatopupa juliana	Western Europe	DQ305074	DQ305082		
	Solatopupa similis	Western Europe	DQ305073	DQ305085		
Helicidae	*Helix aspersa*	Western Europe	AY345052	AY546283	AY345052	AY546283
	Iberus gualtieranus	Iberia	DQ822155	DQ822157	DQ822155	DQ822158
Hygromiidae	*Candidula olisippensis*	Western Europe	AY238623	AY546266	AY238623	AY546266
	Candidula rugosiuscula	Western Europe	AY238620	AY546267	AY238620	AY546267
	Candidula spadae	Western Europe	AY238621	AY546268	AY238621	AY546268
	Candidula unifasciata	Western Europe	AY238617	AY546269	AY238617	AY546269
	Trochulus caelatus	Western Europe	DQ251522	DQ251526	DQ251522	DQ251526
	Trochulus clandestinus	Western Europe	DQ251503	DQ251520	DQ251503	DQ251505
	Trochulus lubomirskii	Western Europe	DQ251552	DQ251553	DQ251552	DQ251553
	Trochulus montanus	Western Europe	DQ251493	DQ251498	DQ251493	DQ251497
	Trochulus piccardi	Western Europe	DQ251492	DQ251547	DQ251492	DQ251547
	Trochulus villosulus	Western Europe	DQ251554	DQ251557	DQ251499	DQ251500
	Trochulus villosus	Western Europe	DQ251499	DQ251500	DQ251554	DQ251556
Limacidae	*Lehmannia marginata*	Western Europe	AM259707	AM259709	AM259707	AM259709
	Lehmannia valentiana	Western Europe	AM259710	AM259711	AM259710	AM259711
	Limax flavus	Western Europe	AM259712	AM259714	AM259712	AM259713
	Limax maximus	W. Europe	EF015443	EF015444	EF015443	EF015444
Milacidae	*Tandonia budapestensis*	Western Europe	AM259705	AM259706	AM259705	AM259706
Oreohelicidae	*Oreohelix haydeni*	USA	DQ858138	DQ858140	DQ858138	DQ858140
	Oreohelix cooperi	USA	DQ858073	DQ858075	DQ858073	DQ858091
Orthalicidae	*Placostylus bivaricosus*	Lord Howe Island	AY165836	AY165843	AY165836	AY165843
Partulidae	*Partula hyalina*	Polynesia	EU026179	EU026169	EU026179	EU026175
Polygyridae	*Ashmunella animasensis*	USA	AY823827	AY823829	AY823827	AY823828
	Ashmunella ashmuni	USA	AY823830	AY823831	AY823830	AY823831
	Ashmunella hebardi	USA	AY823795	AY823798	AY823795	AY823800
	Ashmunella mearnsii	USA	AY823787	AY823788	AY823787	AY823825
	Praticolella berlandieriana	USA + Central America	DQ086030	DQ086035	DQ086030	DQ086039
	Praticolella candida	USA + Central America	DQ086042	DQ086078	DQ086042	DQ086078
	Praticolella flavescens	USA + Central America	DQ086034	DQ086087	DQ086034	DQ086065
	Praticolella griseola	USA + Central America	DQ086025	DQ086095	DQ086025	DQ086062
	Praticolella pachylomae	USA + Central America	DQ086028	DQ086089	DQ086028	DQ086089
	Praticolella taeniata	USA + Central America	DQ086056	DQ086092	DQ086056	DQ086092
	Praticolella trimatrise	USA + Central America	DQ086074	DQ086075	DQ086074	DQ086075
Rhytididae	*Amborhytida dunniae*	New Zealand	DQ298453	DQ298465	DQ298453	DQ298465
	Amborhytida duplicata	New Zealand	DQ298466	DQ298470	DQ298466	DQ298467
	Amborhytida forsythi	New Zealand	DQ298471	DQ298483	DQ298471	DQ298477
	Amborhytida pycrofti	New Zealand	DQ298486	DQ298487	DQ298486	DQ298487

© 2009 Blackwell Publishing Ltd

Table 1 *Continued*

			Short data set (228 bp)		Long data set (381 bp)	
Family	Species	Location	Exemplar	Test	Exemplar	Test
	Paryphanta busbyi	New Zealand	DQ298489	DQ298500	DQ298489	DQ298500
	Paryphanta watti	New Zealand	DQ298505	DQ298506	DQ298505	DQ298506
	Rhytida greenwoodi	New Zealand	AF397671	DQ298508	DQ298508	DQ298507
	Rhytidarex johnsoni	New Zealand	DQ298510	DQ298512	DQ298510	DQ298512
	Schizoglossa novoseelandica	New Zealand	DQ298515	DQ298516	DQ298515	DQ298516
	Schizoglossa worthyae	New Zealand	DQ298513	DQ298514		
	Wainuia clarki	New Zealand	AF397657	AF397661		
	Wainuia edwardi	New Zealand	AF397654	AF397667		
	Wainuia fallai	New Zealand	AF397652	AF397653		
	Wainuia nasuta	New Zealand	AF397651	AF397665		
	Wainuia urnula	New Zealand	AF397650	AF397669		
Succineidae	*Catinella baldwini*	Hawaii	AY148571	AY150076	AY148571	AY150076
	Succinea caduca	Hawaii	DQ658422	DQ658436	DQ658422	DQ658433
	Succinea lumbalis	Hawaii	AY150073	AY150074	AY150073	AY150074
	Succinea lutulenta	Hawaii	AY150077	AY150079	AY150077	AY150079
	Succinea manuana	Hawaii	AY150088	AY150089	AY150088	AY150089

Results

On 12 November 2007, all stylommatophoran cytochrome oxidase subunit I sequences were downloaded from GenBank and aligned in BioEdit version 7.0.1. There were no insertion-deletion sites. Sequences of 'unknown' species were removed, as was any species that did not have at least one other representative. This produced a master file containing 824 sequences representing 129 species. A long core sequence alignment was then created, spanning 381 base pairs and included 97 species (75% of the original species list). Similarly, a short core sequence spanned 228 base pairs and included 127 species (98% of the original species list). A full list of all species used is in Table 1. These species are dominated by several large surveys that have taken place within relatively narrow species groups, including snails on Hawaii (Cowie & Holland 2008), the Galapagos (Parent & Crespi 2006), New Zealand (Spencer *et al*. 2006), America (K. E. Perez, unpublished; Weaver *et al*. 2006), Western Europe (Pfenninger *et al*. 2003, 2005) and the Ryukyu Islands (Kameda *et al*. 2007).

For the long data set, the mean K2P distance was 2.6% (± 5.5% SD) within species and 11.8% (± 7.6%) between species (Fig. 1). For the short data set, the equivalent figures were 2.5% (± 5.7%) and 10.0% (± 7.9%). When each exemplar sequence was placed in a phylogeny with a single test sequence, for the long data set identification was correct in 89/97 cases (91.8%); 6 species were incorrectly identified (6.2%) and 2 were ambiguous (2.1%). For the short data set, 104/127 species were correctly identified (81.9%), with 19 incorrect identifications (15.0%) and 4 were ambiguous (3.1%).

To identify the most appropriate threshold value for defining intra/interspecific variation, false positive and false negative error rates were estimated for a series of putative threshold values (Fig. 2). For both the long and short data sets, the threshold that produced the least number of false positives and negatives was 4%. At this threshold, in the long data set 14 interspecific comparisons (14.4%) would be misclassified as 'within species' (false negative),

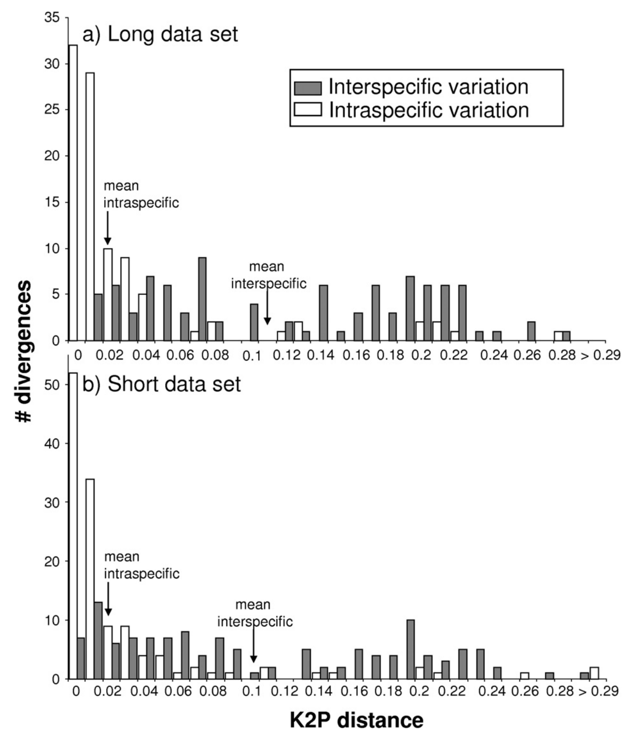

Fig. 1 Histograms illustrating intra- vs. interspecific variation in stylommatophoran land snails. Although the means are quite separate, the distributions of both intra- and interspecific variation are highly skewed.

© 2009 Blackwell Publishing Ltd

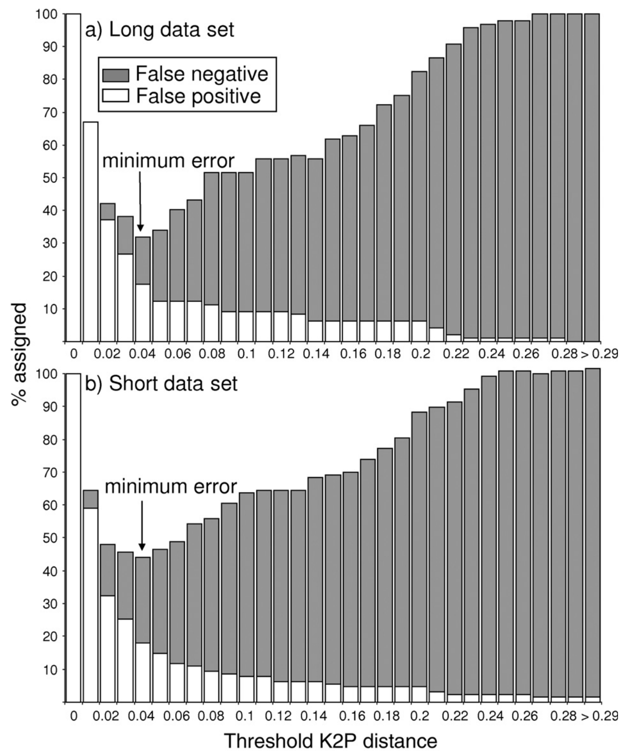

Fig. 2 Cumulative error based on false positives plus false negatives for each threshold value. The optimum threshold value is 4%, with error minimized at 34% (a) and 46.5% (b).

with 17 intraspecific comparisons (17.5%) classified as 'between species' (false positive). The equivalent values for the short data set were 18 (14.1%) and 26 (20.5%). Thus, the overall error was 31/97 (32.0%) and 44/127 (44.1%) for the long and short data sets, respectively.

In the analysis of sampling/haplotype frequency and cytochrome oxidase sequence variation, it was found that the species that were least likely to be assigned to monophyleytic groups were those with the lowest within-species K2P distances (K2P < 0.05; Table 3; Fig. 3a). It was also found that species with relatively few sequences (< 25) on GenBank are much less likely to be assigned to a single monophyletic group (Table 3; Fig. 3b).

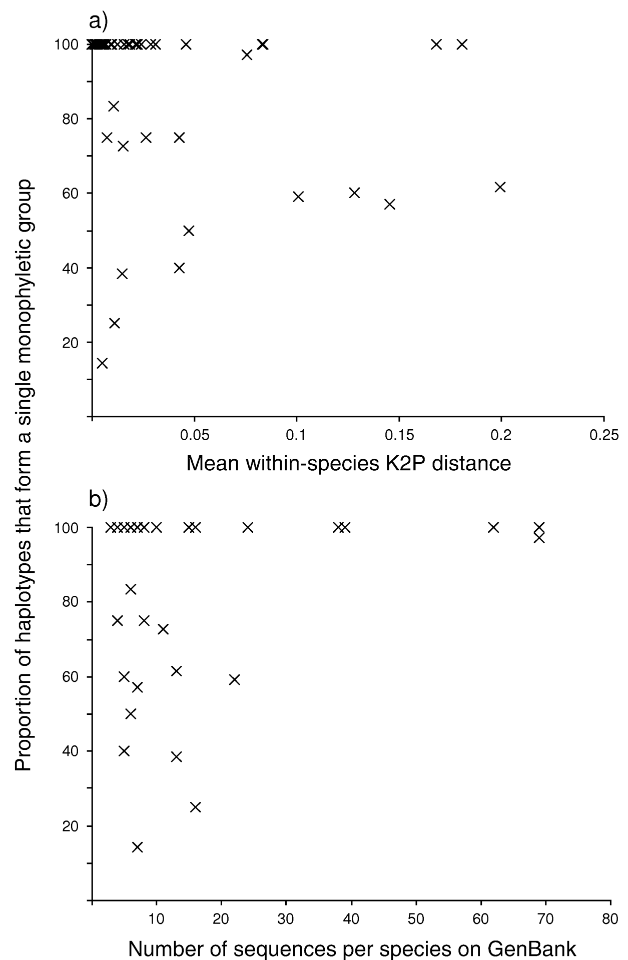

Fig. 3 Proportion of haplotypes from each species that fall in a single monophyletic group compared to within-species Kimura 2-parameter (K2P) distance (a) and number of samples on GenBank (b), the latter a proxy for sample size.

Table 2 The 10 most divergent intraspecific comparisons using the long data set

K2P distance	Species	Exemplar	Test
0.27	*Praticolella griseola*	DQ086025	DQ086062
0.21	*Euhadra peliomphala*	AY251859	AY251862
0.20	*Praticolella pachylomae*	DQ086028	DQ086089
0.20	*Arion subfuscus*	AY987904	AY987915
0.20	*Euhadra idzumonis*	AY251868	AY251869
0.19	*Praticolella flavescens*	DQ086034	DQ086065
0.12	*Arion rufus*	AY987900	AY987902
0.12	*Succinea lutulenta*	AY150077	AY150079
0.11	*Praticolella berlandieriana*	DQ086030	DQ086039
0.08	*Satsuma eucosmiandria*	AB242440	AB242490

K2P, Kimura 2-parameter.

Discussion

DNA barcoding methods must ideally have the ability to deliver accurate species identification to be an effective tool for molecular taxonomy. Using a trial data set, we estimated that the mean within-species divergence (~3%) of land snails is considerably less than the between-species divergence (~12%). However, the problem is that there is a skew in the degree of variation (Fig. 1), so that within-species variation is sometimes very wide (~30%), and between-species variation is frequently quite narrow (~1%). By constructing neighbour-joining phylogenies with a relatively short test and exemplar sequence, barcodes were nonetheless able to

© 2009 Blackwell Publishing Ltd

Table 3 The proportion of haplotypes that form a single monophyletic group for each species, and the mean, standard deviation and range of Kimura 2-parameter (K2P) distances within species

Family	Species	No. of haplotypes	Proportion monophyletic	K2P distances			
				Mean	SD	Min	Max
Achatinellidae	*Achatinella decipiens*	3	100	0.012	0.004	0.008	0.016
	Achatinella fuscobasis	6	83	0.011	0.014	0	0.032
	Achatinella lila	4	100	0.001	0.001	0	0.003
	Achatinella livida	3	100	0	0	0	0
	Achatinella mustelina	69	100	0.031	0.015	0	0.055
	Achatinella sowerbyana	8	75	0.007	0.009	0	0.019
	Newcombia cumingi	5	100	0	0	0	0
	Partulina splendida	6	100	0.005	0.005	0	0.016
	Partulina tappaniana	6	100	0.002	0.002	0	0.005
Agriolimacidae	*Deroceras reticulatum*	3	100	0	0	0	0
Arionidae	*Arion alpinus*	3	100	0.005	0.005	0	0.008
	Arion distinctus	10	100	0.006	0.005	0	0.019
	Arion fasciatus	3	100	0.021	0.009	0.011	0.027
	Arion flagellus	3	100	0.168	0.089	0	0.268
	Arion fuscus	39	100	0.022	0.014	0	0.052
	Arion hortensis	24	100	0.006	0.004	0	0.019
	Arion owenii	5	100	0.004	0.002	0	0.008
	Arion rufus	4	100	0.083	0.055	0.008	0.124
	Arion subfuscus	13	62	0.200	0.075	0	0.273
	Arion transsylvanus	3	100	0.016	0.009	0.005	0.021
	Kootenaia burkei	4	100	0.002	0.001	0	0.003
Bradybaenidae	*Satsuma adelinae*	4	75	0.027	0.024	0.003	0.052
	Satsuma amanoi	4	100	0.003	0.002	0	0.005
	Satsuma eucosmia	69	97	0.076	0.054	0	0.286
	Satsuma hemihelvus	5	100	0.031	0.034	0	0.075
	Satsuma largillierti	38	100	0.046	0.017	0	0.073
	Satsuma shigetai	6	50	0.048	0.024	0.005	0.071
	Satsuma sororcula	11	73	0.015	0.011	0	0.032
	Satsuma tokunoshimana	4	75	0.043	0.037	0.003	0.077
Helicidae	*Iberus gualtieranus*	4	100	0.029	0.006	0.021	0.035
Hygromiidae	*Trochulus caelatus*	4	100	0.021	0.012	0.003	0.041
	Trochulus clandestinus	5	100	0.010	0.009	0	0.022
	Trochulus montanus	6	100	0.002	0.003	0	0.008
	Trochulus piccardi	7	100	0.004	0.003	0	0.008
	Trochulus villosulus	4	100	0	0	0	0
Limacidae	*Lehmannia marginata*	3	100	0	0	0	0
	Limax flavus	3	100	0.022	0.017	0.003	0.033
Oreohelicidae	*Oreohelix haydeni*	62	100	0.001	0.003	0	0.016
	Oreohelix cooperi	3	100	0.005	0.005	0	0.008
Orthalicidae	*Placostylus bivaricosus*	8	100	0.016	0.017	0	0.050
Partulidae	*Partula hyalina*	16	100	0.024	0.019	0	0.054
Polygyridae	*Ashmunella animasensis*	3	100	0.007	0.006	0	0.011
	Ashmunella hebardi	7	14	0.005	0.003	0	0.011
	Ashmunella mearnsii	16	25	0.011	0.005	0	0.021
	Praticolella berlandieriana	22	59	0.101	0.075	0	0.222
	Praticolella candida	5	60	0.128	0.105	0	0.222
	Praticolella flavescens	5	100	0.181	0.057	0.019	0.209
	Praticolella griseola	7	57	0.145	0.076	0	0.274
	Praticolella taeniata	5	40	0.043	0.025	0.008	0.072
Rhytididae	*Amborhytida dunniae*	13	38	0.015	0.007	0	0.029
	Amborhytida duplicata	5	100	0.018	0.007	0.005	0.027
	Amborhytida forsythi	8	100	0.019	0.012	0.003	0.040
	Paryphanta busbyi	16	100	0.019	0.013	0	0.035
	Rhytidarex johnsoni	3	100	0.005	0.003	0.003	0.008
Succineidae	*Succinea caduca*	15	100	0.009	0.008	0	0.019
	Succinea lutulenta	3	100	0.084	0.061	0.013	0.123

© 2009 Blackwell Publishing Ltd

correctly identify stylommatophoran land snails to species in 92% of cases. This relatively high value illustrates the power of the method when applied to short data sets (Hajibabaei *et al*. 2006; Meusnier *et al*. 2008). Nonetheless, although the success rate is promising, it is not of the standard demonstrated in some other studies.

Unfortunately, there is no barcoding gap in land snails, ruling out species discovery based on a threshold value alone. Thus, while the optimum threshold is 4%, it masks a high error rate (32%), of which about equal numbers might be novel species lumped with another species (false negative), or within-species variation that is incorrectly identified as originating from a different species. As a similar trend was found for each of the two separate long and short data sets, this indicates that barcoding may be a useful tool for the prospecting of novel snail species, but that instead, strict identification must be backed up by phylogenetic methods and conventional taxonomy. Rates of success may improve as further sequences accumulate — if there is a previously identified exemplar or reference sequences for comparison, then eventually a high rate of identification may be attainable, but this will not be the case if many species are poly- or paraphyletic for mtDNA. Hebert *et al*. (2003a) suggest a threshold of 3% COI divergence as sufficient to characterize different species and a standard sequence threshold of 10 times the mean intraspecific variation in order to screen for novel taxa, while minimizing false positives (Hebert *et al*. 2004b). Clearly, the latter is impractical in land snails — a very high divergence would tend to miss the vast majority of species.

Compared to previous studies, the intraspecific variation revealed in this study of stylommatophoran land snail species is generally much higher than in other organisms (for discussion on this issue, see Davison 2006). This is one of the main reasons for the high error rate produced by using a threshold (Table 2). We are certain that this is not an artefact of sampling or reporting, because many studies have previously reported extreme within-species variation of land snail mtDNA, and there is no evidence to suggest these are cryptic species (Thomaz *et al*. 1996; Chiba 1999; Davison *et al*. 2005; Pinceel *et al*. 2005). Therefore, while other DNA barcoding studies have generally found intraspecific cytochrome oxidase variation to be less than 2–3% (with exceptions', Elias *et al*. 2007; Meier *et al*. 2006), land snails frequently have intraspecific variation of between 10% and 30%, depending upon the species and the gene used.

A justifiable criticism of the trial test we carried out is that the data from GenBank are biased: intraspecific variation will be overestimated, because identical sequences are not usually submitted to the database; interspecific variation is probably also overestimated, because insufficient taxa are represented by gene sequences. Thus, one argument is that both a greater within-species sampling effort and greater taxonomic depth will improve the overall identification success rate, therefore reducing the threshold. The main caveat, already mentioned, is that in-depth sampling may also reveal mitochondrial poly- or paraphyly.

The relatively few instances where test sequences could not be assigned with exemplars (8/97 in the long data set) may be due to insufficient DNA sequence for phylogeny reconstruction. Some studies have shown that shorter lengths are successful in identifying species (Hajibabaei *et al*. 2006), whereas others argue that shorter lengths of COI sequences are not useful and may lead to incorrect or ambiguous placement in a neighbour-joining tree (Min & Hickey 2007). However, we suspect that the more likely explanation is incomplete lineage sorting because of recent speciation or introgression, although specific biological investigations will be required to uncover the cause on a case-by-case basis. Another key issue in sampling land snails is the geographical range, as sampling over a restricted geographical area may result in underestimation of intraspecific variation (Dasmahapatra & Mallet 2006). As snails have an extreme genetic structure (e.g. Davison & Clarke 2000), then this issue is particularly pertinent to them.

We also attempted to test the effect of sample frequency on the effectiveness of barcoding methods by making the assumption that if there are more sequences on GenBank, then this is representative of a greater sampling effort (efforts to determine the absolute number of samples by reference to the original papers failed because of a lack of detail for many and/or unpublished studies). It was found that species with relatively few sequences on GenBank (< 25) are least likely to be assigned to a single monophyletic group using the K2P and neighbour-joining algorithms (Fig. 3b). Similarly, species with low within-species K2P distances (< 0.05) were also least likely to be assigned to monophyleytic groups. As the estimate of mean K2P distance was not correlated with the number of sequences on GenBank (not shown), then a reasonable and not unexpected conclusion is that recently diverged species (i.e. those that have a low K2P distance) are problematic for barcoding. The extreme divergence of cytochrome oxidase sequences in some land snails is not a problem for barcoding using phylogenetic methods (Fig. 3a), but actually an advantage because the most divergent sequences tend to fall in monophyletic groups. The other implication is that rather many samples (> 25) of each species are required if barcoding is to be effective. A cautionary note, however, is that relatively few species were sampled many times in our data set, and hence, further investigation is warranted on this issue. Certainly, for accuracy widespread geographical sampling is then another requirement.

It is evident that barcoding cannot be employed alone as a method of snail identification; yet used in conjunction with other methods of identification, it could potentially become a useful taxonomic tool. DeSalle *et al*. (2005) suggest using a combination of molecular approaches and traditional morphological taxonomy as the best method of species

© 2009 Blackwell Publishing Ltd

identification, a view already put into practice in some research establishments such as in assessment of taxonomic splits in North American birds (by the American Ornithologists Union) whose evidence is based not only genetic data but morphological and vocal differences. The use of multiple genetic markers such as both mtDNA and nuclear DNA, as well as morphological and ecological information is suggested to provide an accurate perspective on both the evolutionary history of an organism and its taxonomic relationships (Funk & Omland 2003; Rubinoff 2006). This may thus provide more accurate species identification because different genes may reflect different evolutionary histories. Kelly *et al*. (2007) present the character-based approach to DNA barcoding which proved to be highly successful and claims to reduce the occurrence of false-positive identifications that are frequent with programs such as BLAST. Such a method is also theoretically accurate, independent of the degree of 'barcoding gap', thus providing a more reliable method for species identification (Kelly *et al*. 2007) and one which may become widely accepted in the future.

DNA barcodes will only really be significantly useful in identifying species from previously well-described taxa, and the ability to use it to identify individuals can only be fully utilized following previous taxonomic description and construction of a complete database (Dayrat 2005). The rate of accumulation of cytochrome oxidase sequences is increasing — the earliest studies involving a few hundred analyses (Hebert *et al*. 2003a) took over a year to complete, whereas the more recent survey of 1000 fish specimens representing 1% of all known fish species by Ward *et al*. (2005) was completed in only 10 days. It is thus a method of future promise for the identification of snails, but it should probably be considered only an aid for the description of their biodiversity, and used alongside other branches of taxonomy in order to maximize its reliability.

Acknowledgements

Thanks to the University of Nottingham for financial support of this project. Space limitations prevented us citing the bibliographic source of each DNA sequence, but we are grateful to all these indirect contributors to the project. Thanks also to two anonymous referees and the editor for comments on the manuscript.

References

Caldecott JO, Jenkins MD, Johnson T, Groombridge B (1994) *The Priorities for Conserving Global Species Richness and Endemism*. World Conservation Monitoring Centre, Cambridge, UK.

Campbell DC, Johnson PD, Williams JD *et al*. (2008) Identification of 'extinct' freshwater mussel species using DNA barcoding. *Molecular Ecology Resources*, **8**, 711–724.

Chiba S (1999) Accelerated evolution of land snails *Mandarina* in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. *Evolution*, **53**, 460–471.

Cowie RH, Holland BS (2008) Molecular biogeography and diversification of the endemic terrestrial fauna of the Hawaiian Islands. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 3363–3376.

Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. *Heredity*, **97**, 254–255.

Davison A (2006) The ovotestis: an underdeveloped organ of evolution. *Bioessays*, **28**, 642–650.

Davison A, Clarke B (2000) History or current selection? A molecular analysis of 'area effects' in the land snail *Cepaea nemoralis*. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1399–1405.

Davison A, Chiba S, Barton NH, Clarke BC (2005) Speciation and gene flow between snails of opposite chirality. *Public Library of Science, Biology*, **3**, e282.

Dayrat B (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society*, **85**, 407–415.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1905.

Ebach MC, Holdrege C (2005) DNA barcoding is no substitute for taxonomy. *Nature*, **434**, 697–697.

Elias M, Hill RI, Willmott KR *et al*. (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2881–2889.

Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology*, **11**, 839.

Funk DJ, Omland KE (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.

Hajibabaei M, Smith MA, Janzen DH *et al*. (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society Series B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society Series B: Biological Sciences*, **270**, S96–S99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *Public Library of Science, Biology*, **2**, 1657.

Kameda YC, Kawakita A, Kato M (2007) Cryptic genetic divergence and associated morphological differentiation in the arboreal land snail Satsuma (Luchuhadra) largillierti (Camaenidae) endemic to the Ryukyu Archipelago, Japan. *Molecular Phylogenetics and Evolution*, **45**, 519–533.

Kelly RP, Sarkar IN, Eernisse DJ, Desalle R (2007) DNA barcoding using chitons (genus *Mopalia*). *Molecular Ecology Notes*, **7**, 177–183.

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Meusnier I, Singer GAC, Landry JF *et al*. (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *Public Library of Science, Biology*, **3**, 2229–2238.

© 2009 Blackwell Publishing Ltd

Min XJ, Hickey DA (2007) Assessing the effect of varying sequence length on DNA barcoding of fungi. *Molecular Ecology Notes*, **7**, 365–373.

Parent CE, Crespi BJ (2006) Sequential colonization and diversification of Galapagos endemic land snail genus Bulimulus (Gastropoda, Stylommatophora). *Evolution*, **60**, 2311–2328.

Pfenninger M, Eppenstein A, Magnin F (2003) Evidence for ecological speciation in the sister species *Candidula unifasciata* (Poiret, 1801) and *C. rugosiuscula* (Michaud, 1831) (Helicellinae, Gastropoda). *Biological Journal of the Linnean Society*, **79**, 611–628.

Pfenninger M, Hrabakova M, Steinke D, Depraz A (2005) Why do snails have hairs? A Bayesian inference of character evolution. *BMC Evolutionary Biology*, **5**, 59.

Pimm SL, Russell GJ, Gittleman JL, Brooks TM (1995) The Future of Biodiversity. *Science*, **269**, 347–350.

Pinceel J, Jordaens K, Backeljau T (2005) Extreme mtDNA divergences in a terrestrial slug (Gastropoda, Pulmonata, Arionidae): accelerated evolution, allopatric divergence and secondary contact. *Journal of Evolutionary Biology*, **18**, 1264–1280.

Rubinoff D (2006) Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology*, **20**, 1026–1033.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions* of. *Proceedings of the Royal Society B: Biological Sciences*, **360**, 1805.

Schindel DE, Miller SE (2005) DNA barcoding a useful tool for taxonomists. *Nature*, **435**, 17–17.

Smith VS (2005) DNA barcoding: perspectives from a 'Partnerships for Enhancing Expertise in Taxonomy' (PEET) debate. *Systematic Biology*, **54**, 841–844.

Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proceedings of the National Academy of Sciences, USA*, **104**, 4967–4972.

Spencer HG, Brook FJ, Kennedy M (2006) Phylogeography of Kauri snails and their allies from Northland, New Zealand (Mollusca: Gastropoda: Rhytididae: Paryphantinae). *Molecular Phylogenetics and Evolution*, **38**, 835–842.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution*, **18**, 70–74.

Thomaz D, Guiller A, Clarke B (1996) Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 363–368.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1847–1857.

Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. *Bioessays*, **29**, 188–197.

Weaver KF, Anderson T, Guralnick R (2006) Combining phylogenetic and ecological niche modeling approaches to determine distribution and historical biogeography of Black Hills mountain snails (Oreohelicidae). *Diversity and Distributions*, **12**, 756–766.

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, **20**, 47–55.

© 2009 Blackwell Publishing Ltd