MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2013) **13**, 820–831 doi: 10.1111/1755-0998.12130

Increased accuracy of species lists developed for alpine lakes using morphology and cytochrome oxidase I for identification of specimens

KRISTY DEINER,*† ROLAND A. KNAPP,‡ DANIEL M. BOIANO§ and BERNIE MAY*

**Department of Animal Science, University of California-Davis, One Shields Ave, Davis, CA 95616, USA, †Department of Aquatic Ecology, Eawag, Überlandstrasse 133, P.O.Box 611 8600, Dübendorf, Switzerland, ‡Sierra Nevada Aquatic Research Laboratory, University of California, 1016 Mount Morrison Road, Mammoth Lakes, CA 93546, USA, §Sequoia and Kings Canyon National Parks, 47050 Generals Highway, Three Rivers, CA 93271, USA*

Abstract

The first step in many community ecology studies is to produce a species list from a sample of individuals. Community ecologists now have two viable ways of producing a species list: morphological and barcode identification. In this study, we compared the taxonomic resolution gained by a combined use of both methods and tested whether a change in taxonomic resolution significantly impacted richness estimates for benthic macroinvertebrates sampled from ten lakes in Sequoia National Park, USA. Across all lakes, 77 unique taxa were identified and 42% (32) were reliably identified to species using both barcode and morphological identification. Of the 32 identified to species, 63% (20) were identified solely by comparing the barcode sequence from cytochrome oxidase I to the Barcode of Life reference library. The increased resolution using a combined identification approach compared to identifications based solely on morphology resulted in a significant increase in estimated richness within a lake at the order, family, genus and species levels of taxonomy ($P < 0.05$). Additionally, young or damaged individuals that could not be identified using morphology were identified using their COI sequences to the genus or species level on average 75% of the time. Our results demonstrate that a combined identification approach improves accuracy of benthic macroinvertebrate species lists in alpine lakes and subsequent estimates of richness. We encourage the use of barcodes for identification purposes and specifically when morphology is insufficient, as in the case of damaged and early life stage specimens of benthic macroinvertebrates.

Keywords: benthic macroinvertebrates, Sierra Nevada, taxonomic resolution, universal primers

Received 4 October 2012; revision received 3 May 2013; accepted 9 May 2013

Introduction

Understanding the biology and ecology of aquatic species and their management starts with proper identification (Merritt *et al.* 2008). Aquatic species' identification, however, can be complicated and there is a large and diverse literature describing many of these challenges (Resh & McElravy 1993; Lenat & Resh 2001; Jones 2008; Pfrender *et al.* 2010; Bevilacqua *et al.* 2012). For example, identification of aquatic invertebrates to the species level is generally only possible using adults, and often adult males. However, adults of many taxa are often present only during narrow time windows. As a consequence, samples tend to be dominated by immature life stages (i.e. larvae and pupae), and these life stages generally lack the required morphological characteristics needed for species-level identifications (Lenat & Resh 2001; Jones 2008; Pfrender *et al.* 2010). Additionally, damaged specimens may be impossible to identify because they lack the required morphological characteristics used in most keys. Recently, the genetic approach of specimen identification by comparing a sequence from an unknown specimen to that of a referenced barcode library was proposed as a way to circumvent these challenges of identification and compliment morphology based approaches (Valentini *et al.* 2009; Pfrender *et al.* 2010; Baird & Sweeney 2011; DeWalt 2011).

Using DNA to identify a specimen can overcome some challenges associated with morphology-based identification; however, DNA-based identification relies on having a curated and annotated reference barcode library. For many animals, the region that is used for a reference barcode is a 658 base pair region of the mitochondrial cytochrome oxidase I gene (COI). A

Correspondence: Kristy Deiner, E-mail: alpinedna@gmail.com

© 2013 John Wiley & Sons Ltd

reference barcode is usually generated by sequencing this gene region, annotating it with metadata (e.g. raw sequence data, specimen museum number and picture of specimen) and depositing it in the Barcode of Life Data System (BOLD, Ratnasingham & Hebert 2007). Once a barcode is catalogued in BOLD, sequences from the COI gene region generated from unknown specimens can be compared to the referenced barcode to aid in identification. This is a molecular equivalent to the practice of type specimens for morphology (Cook *et al.* 2010). Additionally, barcode identification can be less subjective compared to morphology, which potentially allows for nonspecialists to accurately identify specimens (Valentini *et al.* 2009). However, BOLD is far from complete, given that, it was only started in 2007, but it is rapidly expanding and has already surpassed one million barcodes and is potentially becoming a useful tool for identification (International Barcode of Life project, IBOL 2012).

Here, we assess the current utility of barcode identification via BOLD from a set of benthic macroinvertebrate fauna sampled from 10 alpine lakes in Sequoia National Park, located in the Sierra Nevada mountains of California, USA (Table 1). These lakes were chosen because park staff were considering removing non-native fish populations from some of these sites and desired to identify the benthic macroinvertebrate communities and using this baseline information to evaluate future recovery. Sequoia National Park contains approximately 1000 lentic water bodies >1 ha, and most are located in the subalpine and alpine zones at elevations >3000 m. These lakes were naturally fishless (Knapp & Matthews 2000), and over the last 150 years several species of trout [*Oncorhynchus* (Suckley, 1861), *Salmo* (Linnaeus, 1758), and *Salvelinus* (J. Richardson, 1836)] were introduced to create recreational fisheries (Pister 2001).

Alpine aquatic systems are in need of restoration because trout introductions cause negative impacts to populations of many native alpine aquatic species including amphibians (Knapp & Matthews 2000; Pilliod *et al.* 2010), benthic macroinvertebrates (Carlisle & Hawkins 1998; Knapp *et al.* 2001; Nystrom *et al.* 2001) and zooplankton (Bradford *et al.* 1998; Knapp *et al.* 2001; Parker *et al.* 2001). Surveying lakes and accurately developing a taxon list are essential to determine recovery after fish removal (Knapp *et al.* 2001; Vredenburg 2004; and Knapp 2005).

In this study, we used benthic macroinvertebrate communities sampled from 10 lakes in Sequoia National Park to assess what increase in resolution is possible when morphology and COI sequences are congruently used for identification of taxa, rather than either method alone. We did not generate novel barcodes in this study, rather we used COI sequences to identify specimens. We then tested if a change in resolution affects estimated richness at different taxonomic levels. Lastly, we report on the ability of barcodes to solve challenges related to identification of young and damaged specimens.

Materials and methods

Sampling

Benthic macroinvertebrates were sampled from 10 lakes in July and August 2008 using a D-net outfitted with a 0.5 mm mesh bag and a standard sweep method (Knapp *et al.* 2001) (Table 1). A total of 15 sweeps were taken from the littoral zone (<1.5 m deep) of each lake. Habitat types (bedrock, boulder, cobble, gravel, sand, silt and submerged vegetation) within each lake were sampled in rough proportion to their occurrence based on qualitative survey of lake habitats before sampling. Benthic macroinvertebrates were handpicked from other debris on site and were preserved in 95% ethanol. Ethanol was changed to fresh 95% ethanol immediately after samples were brought back to the laboratory to maximize DNA preservation. Samples were stored at room temperature

Table 1 Description of study lakes in Sequoia National Park

Basin	Lake ID	Fish status	Area (m^2)	Elevation (m)	Maximum depth (m)	Latitude	Longitude
Upper Kern	20224	Present	14024	3530	2.2	36.683391	−118.422364
	21018	Present	4465	3530	2	36.682204	−118.420540
	20225	Absent	11040	3590	2.5	36.687091	−118.421012
	20227	Present	16026	3310	12.25	36.681619	−118.414146
	21004	Absent	4329	3670	3	36.686953	−118.411485
Crytes	20279	Present	64627	3315	12	36.358442	−118.471205
	20280	Present	18421	3347	8.9	36.361932	−118.461978
	20278	Absent	6934	3343	3.5	36.352843	−118.482964
East Wright	20118	Present	10547	3490	4.5	36.618524	−118.350012
	20117	Absent	15529	3490	4	36.615355	−118.349712

© 2013 John Wiley & Sons Ltd

until morphological and barcode identifications were carried out.

Specimen identification

We use the term 'specimen' to refer to the individual being identified because the two identification methods potentially identify a specimen to a different taxonomic level. The term 'taxon' ('taxa' when plural), is used to refer to specimens identified to a taxonomic level such as order, family, genus or species. Specimen identification was conducted in two ways. First, they were identified based on morphology using taxonomic keys appropriate for each benthic macroinvertebrate taxon (Table 2). Each specimen was identified to the lowest possible taxonomic level based on the condition of the specimen, its developmental stage (i.e. instar), and the resolution of the key.

Second, a barcode identification was obtained using a COI sequence from a specimen by comparing the sequence generated from our specimens to referenced COI barcodes on BOLD v. 3.0 (Ratnasingham & Hebert 2007). Assignments made at the species level using the species-level barcode reference database and were based on the probability of placement calculated by the search engine. Probability of placement is based on the queried sequence match with the global alignment of all data stored in BOLD utilizing a hidden markov model, followed by a linear search of the database (Ratnasingham & Hebert 2007). To make assignments at the genus level or higher, we compared the COI sequence to the database of all barcode records on BOLD and inspected the neighbour-joining tree generated by the search engine which contained the 100 nearest neighbours (Ratnasingham & Hebert 2007). We determined assignment of a sequence to a taxon following the strict tree-based method of Wilson *et al.* (2011). The strict tree-based method assigns a sequence as belonging to a taxon if it was nested within a clade comprised of members from a single taxon (Wilson *et al.* 2011).

Lastly, an unrooted neighbour-joining tree was constructed from Tamura-Nei pairwise genetic distances and visualized using Geneious v 6.0 (Biomatters Ltd.). This tree was then used to detect similarities of sequences across study lakes and assess if the same or similar taxa were present when a species-level assignment was not assigned by BOLD (e.g. larvae of Taxon 21, Supplementary material).

Molecular protocols

DNA extraction was performed on 1–5 specimens per lake for each unique taxon first identified to the lowest level possible by morphology (Table 2). Many morphological taxonomic groups had <5 individuals sampled from each lake; therefore, in many cases (81%), the number of specimens extracted represented all the specimens sampled for each morphological group in each lake (Table 2).

DNA was extracted from a total of 429 individuals using the DNeasy Qiagen kit (Qiagen Inc.) and following the manufacturer's protocol. For specimens that were expected to have low DNA yields due to their small sizes in the families of Chironomidae and Arachnidae, we used the HotShot extraction method. HotShot extraction is carried out by boiling the specimen in a high-pH (~12) NaOH solution and then neutralizing the pH with Tris-HCL to 7.4 after boiling (Montero-Pau *et al.* 2008).

Following DNA extraction, universal COI primers (LCO1490 and HCO2198) were used to amplify DNA samples (Folmer *et al.* 1994). Each polymerase chain reaction (PCR) consisted of final concentrations of supplied TAQ buffer (1×) (Roche Inc.), BSA (1×), magnesium chloride (2mmol), dNTPs (0.20μmol), 0.05 units Faststart TAQ (Roche, Inc., Basel CH), 1 μL of DNA template (10–70 ng/μL), light and heavy primers (0.50μmol) and molecular grade water in a total reaction volume of 10 μL. PCR involved a 4-min denaturing at 94 °C before thermocycling for 35 cycles. The thermocycling profile was 94 °C denaturing for 30 s, 48 °C annealing for 30 s and 72 °C extension for 1 min, followed by a 7-min extension at 72 °C. PCR success was assessed on a 1.2% agarose gel stained with Gelstar (GE Healthcare). PCR products were purified using the ExoSap-it kit (GE Healthcare) and then sequenced bidirectionally on an ABI 3130/3730 sequencer using di-deoxy chain termination chemistry with BIG Dye v3.1 (Applied Biosystems) following recommended Applied Biosystems protocols. Sequences from light and heavy strands were aligned and edited with SEQUENCHER v4.8 (GeneCodes). A global alignment was generated using Geneious v 6.0 (Biomatters Ltd.). Sequences generated for this study have been submitted to GenBank (Accession numbers KF000103 – KF000348, Supplementary material).

Data analysis

After the taxon list was produced using a combined morphological and barcode method of identification for each lake (Table 2), we then calculated the morphology-based taxonomic resolution as the number of taxa identified to a particular level of taxonomy (order, family, genus or species) divided by the total taxa per lake identified to that level of taxonomy. Barcode-based taxonomic resolution was calculated in the same manner. The combined taxonomic resolution for each lake was calculated as the sum of all taxa identified to each level of taxonomy divided by the total taxa for each taxonomic level within a lake. Differences in resolution among identification

© 2013 John Wiley & Sons Ltd

Table 2 Taxonomic resolution of specimens identified using morphology and barcode identification of benthic macroinvertebrates from 10 lakes sampled in Sequoia National Park in 2008

Taxon ID	Class	Order	Family	Genus	Specific epithet	Morphological key	Morphological resolution	Barcode resolution	No. of specimens	No. of lakes occupied	No. of COI sequences	Probability of placement*
1	Arachnida	Sarcoptiformes				Smith *et al*. 2010	Class	Order	1	1	1	89
2	Arachnida	Trombidiformes	Lebertiidae	*Lebertia*	sp0942A	Smith *et al*. 2010	Genus	Species	1	1	1	98
3	Arachnida	Trombidiformes	Lebertiidae	*Lebertia*		Smith *et al*. 2010	Genus	Genus	15	5	13	99
4	Arachnida	Trombidiformes	Limnesiidae	*Hydryphantes*		Smith *et al*. 2010	Genus	Genus	1	1	1	93
5	Arachnida	Trombidiformes	Limnesiidae	*Limnesia*	sp0936BC	Smith *et al*. 2010	Genus	Species	94	7	15	99
6	Arachnida	Trombidiformes	Limnesiidae	*Limnesia*		Smith *et al*. 2010	Genus	Genus	1	1	1	94
7	Arachnida	Trombidiformes	Limnesiidae	*Oxus*		Smith *et al*. 2010	Genus	Family	1	1	1	85
8	Arachnida	Trombidiformes	Pionidae	*Piona*	sp0934B	Smith *et al*. 2010	Genus	Species	3	2	2	98
9	Arachnida	Trombidiformes	Sperchontidae	*Sperchon*		Smith *et al*. 2010	Genus	–	1	1	0	–
10	Bivalvia	Veneroida	Sphaeriidae	*Pisidium*		Smith 2001	Genus	Order	72	4	3	92
11	Clitellata	Haplotaxida	Naididae	*Nais*		Smith 2001	Class	Genus	5	1	2	92
12	Clitellata	Haplotaxida	Naididae			Smith 2001	Class	Family	5	1	5	83
13	Clitellata	Haplotaxida				Smith 2001	Class	Order	5	1	1	84
14	Clitellata	Rhynchobdellida	Glossiphoniidae	*Helobdella*		Smith 2001	Genus	Genus	3	3	2	89
15	Clitellata					Smith 2001	Class	–	13	1	0	–
16	Insecta	Coleoptera	Dytiscidae	*Acilius*	*abbreviatus*	Merritt *et al*. 2008 (Genus); Larson *et al*. 2000 (Sp.)	Species	–	4	1	0	–
17	Insecta	Coleoptera	Dytiscidae	*Agabus*	*tristis*	Merritt *et al*. 2008 (Genus); Larson *et al*. 2000 (Sp.)	Species	Species	8	4	8	99
18	Insecta	Coleoptera	Dytiscidae	*Hydroporus*	*erythrocephalus*	Merritt *et al*. 2008	Family	Species	1	1	1	98
19	Insecta	Coleoptera	Dytiscidae	*Hygrotus*	*patruelis*	Merritt *et al*. 2008	Family	Species	2	1	2	99
20	Insecta	Coleoptera	Dytiscidae	*Oreodytes*		Merritt *et al*. 2008	Genus	Family	4	1	3	89
21	Insecta	Coleoptera	Dytiscidae	*Stictotarsus*	*striatellus*	Merritt *et al*. 2008 (Genus); Larson *et al*. 2000 (Sp.)	Species	Family	129	9	26	99
22	Insecta	Coleoptera	Hydrophilidae	*Hydrochus*		Merritt *et al*. 2008	Genus	–	1	1	0	–
23	Insecta	Diptera	Ceratopogonidae			Merritt *et al*. 2008	Order	Family	1	1	1	87
24	Insecta	Diptera	Chironomidae	*Ablabesmyia*	*americana*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	1	1	1	97
25	Insecta	Diptera	Chironomidae	*Ablabesmyia*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	1	1	1	97
26	Insecta	Diptera	Chironomidae	*Chironomus*	*storai*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	1	1	1	99
27	Insecta	Diptera	Chironomidae	*Chironomus*	*whitseli*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	2	1	2	100
28	Insecta	Diptera	Chironomidae	*Cladotanytarsus*	*amandus*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	6	2	5	100
29	Insecta	Diptera	Chironomidae	*Corynoneura*	*arctica*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	5	3	5	100

© 2013 John Wiley & Sons Ltd

Table 2 (Continued)

Taxon ID	Class	Order	Family	Genus	Specific epithet	Morphological key	Morphological resolution	Barcode resolution	No. of specimens	No. of lakes occupied	No. of COI sequences	Probability of placement*
30	Insecta	Diptera	Chironomidae	*Cricotopus*	*tibialis*	Merritt *et al*. 2008; Wiederholm 1983	Family	Species	2	1	2	98
31	Insecta	Diptera	Chironomidae	*Cricotopus*		Merritt *et al*. 2008; Wiederholm 1983	Family	Genus	2	2	2	96
32	Insecta	Diptera	Chironomidae	*Cricotopus*	*obnixus*	Merritt *et al*. 2008; Wiederholm 1983	Family	Species	2	1	1	98
33	Insecta	Diptera	Chironomidae	*Dicrotendipes*		Merritt *et al*. 2008; Wiederholm 1983	Genus	–	4	1	0	–
34	Insecta	Diptera	Chironomidae	*Diplocladius*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	1	1	1	92
35	Insecta	Diptera	Chironomidae	*Heterotrissocladius*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	3	3	4	93
36	Insecta	Diptera	Chironomidae	*Micropsectra*	*polita*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	1	1	1	100
37	Insecta	Diptera	Chironomidae	*Micropsectra*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	10	4	8	91
38	Insecta	Diptera	Chironomidae	*Microtendipes*	*pedellus*	Merritt *et al*. 2008; Wiederholm 1983	Species	–	2	2	0	–
39	Insecta	Diptera	Chironomidae	*Orthocladius*		Merritt *et al*. 2008; Wiederholm 1983	Family	Genus	3	2	3	93
40	Insecta	Diptera	Chironomidae	*Parakiefferiella*		Merritt *et al*. 2008; Wiederholm 1983	Genus	–	1	1	0	–
41	Insecta	Diptera	Chironomidae	*Parametriocnemus*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	1	1	1	91
42	Insecta	Diptera	Chironomidae	*Paratanytarsus*	*laccophilus*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	3	2	2	99
43	Insecta	Diptera	Chironomidae	*Paratanytarsus*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	8	2	1	86
44	Insecta	Diptera	Chironomidae	*Paratrichocladius*		Merritt *et al*. 2008; Wiederholm 1983	Family	Genus	3	1	3	94
45	Insecta	Diptera	Chironomidae	*Phaenopsectra*		Merritt *et al*. 2008; Wiederholm 1983	Genus	–	13	6	0	–
46	Insecta	Diptera	Chironomidae	*Procladius*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	8	4	4	96
47	Insecta	Diptera	Chironomidae	*Psectrocladius*	*psilopterus*	Merritt *et al*. 2008; Wiederholm 1983	Species	Family	1	1	1	89
48	Insecta	Diptera	Chironomidae	*Psectrocladius*	*sordidellus*	Merritt *et al*. 2008; Wiederholm 1983	Species	Genus	10	3	5	96
49	Insecta	Diptera	Chironomidae	*Psectrocladius*	*limbatellus*	Merritt *et al*. 2008; Wiederholm 1983	Species	Family	5	2	5	87
50	Insecta	Diptera	Chironomidae	*Psectrocladius*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Family	24	2	7	88
51	Insecta	Diptera	Chironomidae	*Psectrotanypus*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	1	1	1	98

© 2013 John Wiley & Sons Ltd

Table 2 (Continued)

Taxon ID	Class	Order	Family	Genus	Specific epithet	Morphological key	Morphological resolution	Barcode resolution	No. of specimens	No. of lakes occupied	No. of COI sequences	Probability of placement*
52	Insecta	Diptera	Chironomidae	*Pseudodiamesa*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	8	2	7	89
53	Insecta	Diptera	Chironomidae	*Tanytarsus*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Family	85	2	11	88
54	Insecta	Diptera	Chironomidae	*Tanytarsus*	*bathophilus*	Merritt *et al*. 2008; Wiederholm 1983	Genus	Species	14	4	8	100
55	Insecta	Diptera	Chironomidae	*Thienemannimyia*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Genus	63	3	17	96
56	Insecta	Diptera	Chironomidae	*Zavrelimyia*		Merritt *et al*. 2008; Wiederholm 1983	Genus	Family	2	2	2	90
57	Insecta	Diptera	Chironomidae			Merritt *et al*. 2008; Wiederholm 1983	Family	Family	17	7	5	90
58	Insecta	Diptera	Culicidae	*Culiseta*	*impatiens*	Merritt *et al*. 2008	Class	Species	3	1	3	100
59	Insecta	Diptera				Merritt *et al*. 2008	Class	Order	2	1	2	92
60	Insecta	Ephemeroptera	Ameletidae	*Ameletus*		Merritt *et al*. 2008	Genus	Genus	4	2	2	91
61	Insecta	Ephemeroptera	Baetidae	*Callibaetis*	*ferrugineus*	Merritt *et al*. 2008	Genus	Species	15	4	9	98
62	Insecta	Ephemeroptera	Siphlonuridae	*Siphlonurus*		Merritt *et al*. 2008	Genus	Genus	1	1	1	94
63	Insecta	Hemiptera	Corixidae	*Arctocorisa*	*sutilis*	Merritt *et al*. 2008 (Genus); Menke 1979 (Sp.)	Species	–	8	3	0	–
64	Insecta	Hemiptera	Corixidae	*Cenocorixa*	*kuiterti*	Merritt *et al*. 2008 (Genus); Menke 1979 (Sp.)	Species	–	4	1	0	–
65	Insecta	Hemiptera	Corixidae	*Cenocorixa*	*bifida*	Merritt *et al*. 2008 (Genus); Menke 1979 (Sp.)	Family	Species	3	1	1	99
66	Insecta	Hemiptera	Corixidae	*Dasycorixa*	*rawsoni*	Merritt *et al*. 2008 (Genus); Menke 1979 (Sp.)	Species	Species	1	1	1	99
67	Insecta	Hemiptera	Corixidae			Merritt *et al*. 2008;	Family	–	392	4	0	–
68	Insecta	Hemiptera	Gerridae	*Aquarius*	*incognitus*	Merritt *et al*. 2008 (Genus); Menke 1979 (Sp.)	Species	–	5	1	0	–
69	Insecta	Hemiptera	Notonectidae	*Notonecta*	*kirbyi*	Merritt *et al*. 2008 (Genus); Menke 1979 (Sp.)	Species	Species	4	2	1	100
70	Insecta	Lepidoptera	Noctuidae	*Dargida*	*procinctus*	Merritt *et al*. 2008	Class	Species	2	2	2	100
71	Insecta	Megaloptera	Sialidae	*Sialis*		Merritt *et al*. 2008	Genus	Genus	1	1	1	97
72	Insecta	Odonata	Aeshnidae	*Aeshna*		Merritt *et al*. 2008	Genus	Genus	3	1	1	100
73	Insecta	Trichoptera	Limnephilidae	*Desmona*		Merritt *et al*. 2008	Genus	Family	23	3	5	88
74	Insecta	Trichoptera	Limnephilidae	*Dicosmoecus*		Merritt *et al*. 2008	Genus	Family	15	4	7	91

© 2013 John Wiley & Sons Ltd

Table 2 (Continued)

Taxon ID	Class	Order	Family	Genus	Specific epithet	Morphological key	Morphological resolution	Barcode resolution	No. of specimens	No. of lakes occupied	No. of COI sequences	Probability of placement*
75	Insecta	Trichoptera	Limnephilidae	*Hesperophylax*	*designatus*	Merritt *et al.* 2008	Family	Species	2	1	2	99
76	Insecta	Trichoptera	Limnephilidae	*Psychoglypha*		Merritt *et al.* 2008	Genus	Genus	2	1	2	96
77	Turbellaria					Smith 2001	Class	–	87	6	0	–

*Probability of placement is calculated by BOLD and is a function of the sequence similarity and the number of nearest neighbours in the database. The number here represents the average of all sequences obtained for the taxon in this study. Taxonomic names listed represent the combined list generated from both morphology and barcode identification methods used in this study. A dash in the probability of placement column indicates that a barcode was not produced with universal primers. Specific epithet names with numbers and letters are interim identification codes on BOLD.

methods (morphology, barcode and combined) were tested using nonparametric Kruskal–Wallis one-way ANOVA and were followed by a Tukey's honestly significant difference (HSD) test. A nonparametric test was used because count data could not be normalized. Increased resolution of the combined approach and barcode approach was calculated as the difference between taxonomic resolution produced from the combined approach to that of the morphological approach alone. Increased resolution was calculated for each lake and across four levels of taxonomy (order, family, genus and species).

Taxonomic richness was calculated at the order, family, genus and species level using three methods of identification: morphology alone, barcode alone and the combined data set. Richness at each taxonomic level in a lake was calculated as the sum of unique taxa identified (e.g. Table 2). This means that if the specimen was not identified to the genus or species level, it was not included in the estimate of richness for that level. The reason for not including an unknown is because they are unidentified and so we did not want to assume that all unknowns of a genus belong to the same species or that all unknowns of a family belong to the same genus. To test whether or not there was a significant difference in estimated richness at each level of taxonomy (which was normally distributed), a paired *t*-test was performed between two methods of identification (morphology vs. barcode and morphology vs. combined barcode and morphology). For all statistical tests, Type I error was minimized by a sequential Bonferroni correction that adjusted the significance level (alpha of 0.05) for the number of tests run ($N = 4$) for the two identification methods. To reduce the likelihood of Type II error, significance was judged at the standard alpha of 0.05. Statistical analysis was carried out in JMP version 8 (SAS Institute).

Lastly, we calculated the success of barcode identification within each lake. Success was calculated as the sum of sequenced and identified taxa at the genus or species level divided by the total number of taxa for which DNA from at least one specimen was extracted.

Results

Identification of specimens with morphology and barcodes

Using morphology ($N = 1266$) and COI sequences ($N = 246$), we identified a total of 77 unique taxa to various levels of taxonomic resolution (Table 2). When a combined identification approach of barcode and morphology was used, all specimens were identified at the phylum and class levels of taxonomy. There were 14

© 2013 John Wiley & Sons Ltd

orders identified, and only two classes had specimens which could not be identified to order by either method of identification. There were 100 unidentifiable specimens belonged to the classes of Clitellata (13 individuals) and Turbellaria (87 individuals) (Table 2). In both of these cases, sequences were not obtained from specimens of these taxa and they were too damaged to identify using morphology.

We identified a total of 22 families and the only additional specimen (besides specimens in the classes of Clitellata and Turbellaria mentioned above) that could not be further identified at the family level was in the order of Sarcoptiformes (Table 2). We identified 55 genera, leaving seven higher taxa could not be identified to the genus level (Table 2). We identified 32 species, leaving the other 45 taxa identified to genus (Table 2). Of the taxa identified to species, barcodes accounted for 20 (63%) and half of these (10) were from the family Chironomidae. There were only two inconsistencies between barcode identification and morphology (2 of 246; Supplementary material). We considered these specimens to belong to the taxon identified from the barcode (Table 2) because a positive species or genus match was made to a barcode reference sequence on BOLD. However, taxon names obtained using both methods are listed in the Supplementary material.

Of the 429 specimens for which DNA was extracted, 246 produced a sequence ranging in size from 576 to 658 base pairs (Fig. 1). The 246 sequences contributed to identification at some level of taxonomy for 83% of taxa identified from the ten lakes (64 of 77, Table 2). Of the 13 unique taxa for which a sequence was not obtained, morphology resolved three to family or a higher level of taxonomy (i.e. order or class), five to genus and five to species.

For early-instar specimens for which a sequence was produced, 64% were identified (65 out of 102) to the genus level or lower. At any given sampling locality, an average of 62% (44%–79%) of early-instar specimens were identified using their COI sequence (Table 3). Sequences also allowed us to identify damaged specimens. For those that produced a sequence, 75% (6 of 8) were identified to the genus level or lower (Table 3).

Taxonomic resolution and richness estimates

There was a significant increase in the number of specimens identified at the genus and species level when a combined identification approach was used (Fig. 2). The largest average increase in resolution (24%) per lake was at the species level (Fig. 2). There was a significant difference in the estimated richness when the combined approach for identification was used in each lake at all levels of taxonomic resolution tested (Table 4). After Bonferroni correction, there was also a significant difference in estimated species richness using barcode identification compared to morphological identification (Table 4).

Discussion

Combined use of barcodes and morphology for identification led to a similar resolution at the order and family level and a more highly resolved taxon list at the genus and species level when compared to morphology or barcodes alone. The more highly resolved taxon list was due to specimens being identified from genus to species with barcodes (e.g. specimens in the family of Chironomidae). Many specimens did not produce a usable barcode, however, and therefore morphologically identified specimens aided in an increased resolution at the genus level. Increased resolution was also due to identifying young specimens (e.g. Dytiscidae larvae) and damaged individuals (e.g. ephemeropteran specimens) using barcodes where morphology could not be used. Therefore, when the best-resolved species list is needed for an aquatic community, identification via barcodes in addition to morphology has proved successful even given the very low representation of all macroinvertebrate species barcodes available in BOLD.

The increased resolution was most notable at the species level of identification, with 63% of species identifications for 77 taxa coming only from sequence-based identifications. This resulted in an average of 24% more identifications of taxa to species per sample locality. Barcodes strengthened and complimented morphological identification in this study because of the 32 taxa identified to species, 12 (38%) where identified to the species level when morphology could resolve them only to genus. Sweeney *et al.* (2011) had a similar result and observed an increase of 34–38%. As such, barcodes significantly changed the estimate of richness at the species level and changed the estimated local and regional species diversity measured within and among lakes. For example, we found that COI sequences from specimens in the genus *Tanytarsus* (Family Chironomidae) revealed that there was more than one species found in the study lakes (Table 2). The increased ability to resolve specimens to species is needed based on the most recent review of the effect of taxonomic resolution on ecological questions. Specifically, Bevilacqua *et al.* (2012) have shown that lumping unidentified specimens into higher taxonomic groups can bias ecological inferences because these groups often do not share adequate ecological similarity.

The observed combined morphology and barcode identification success rate of 84% across taxa identified to the genus or species level are likely an overestimate of

© 2013 John Wiley & Sons Ltd

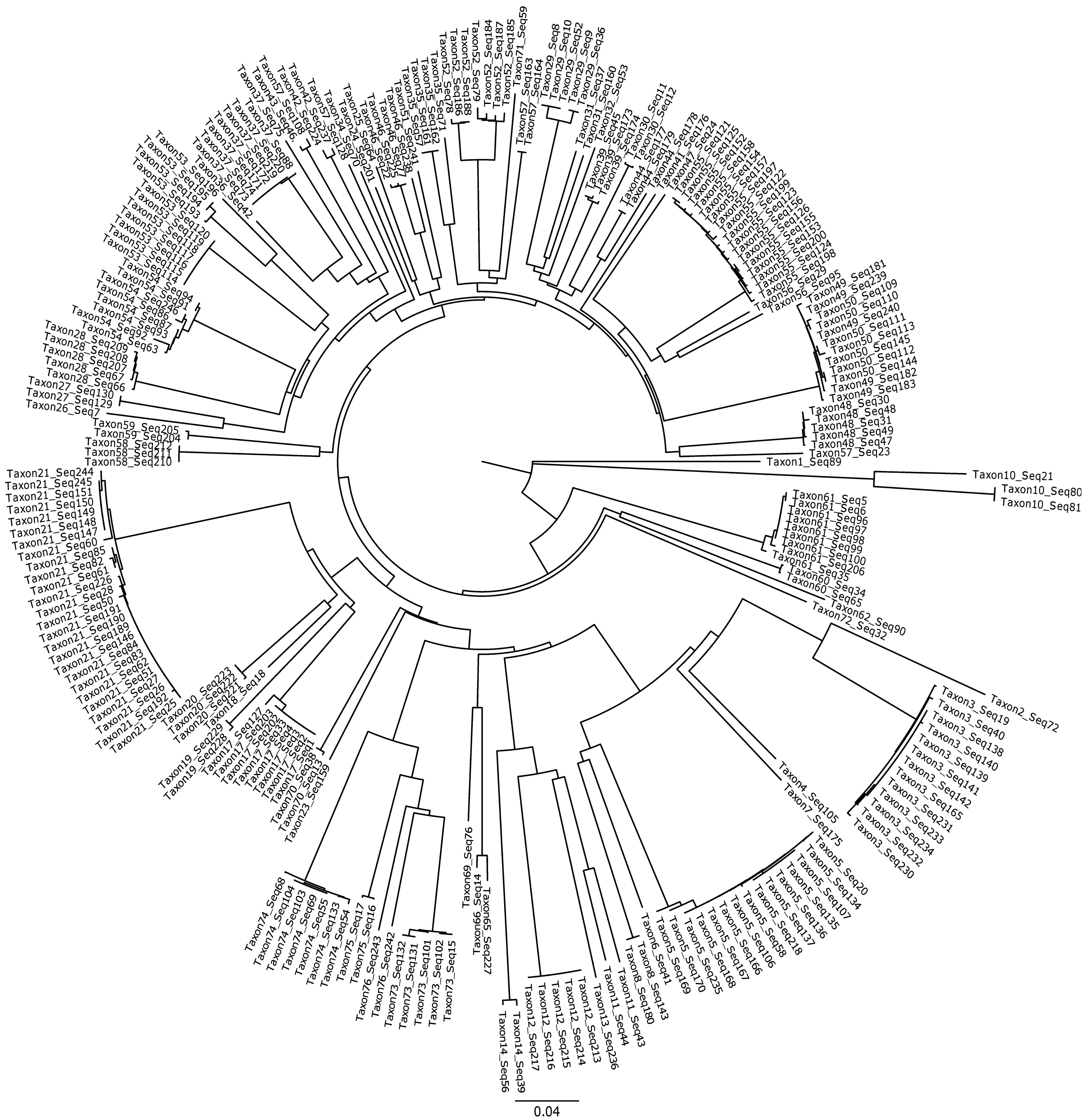

Fig. 1 An unrooted neighbour-joining tree of all COI sequences used for identification. Tree illustrates similarities of COI sequences assigned to taxonomic groups (e.g. 'Taxon 21') from specimens across all lakes. Branch lengths and scale bar are Tamura–Nei genetic distances.

success because taxa resolved to the family and genus level of taxonomy could contain more than one taxon if there is more than one specimen. However, the groups that would likely have the greatest impact on the total number of taxa are the Platyhelminthes (flatworms). We sampled a total of 87 specimens from six lakes and they were only identified to the level of class in our study. This group, as well as members in the order of Hemiptera (e.g. Corixidae and Gerridae) did not produce a sequence usable for identification with universal primers. Sweeney *et al.* (2011) excluded Platyhelminthes for similar reasons. Platyhelminthes are also difficult to identify morphologically because when they are collected in benthic samples and preserved in the field, it is

© 2013 John Wiley & Sons Ltd

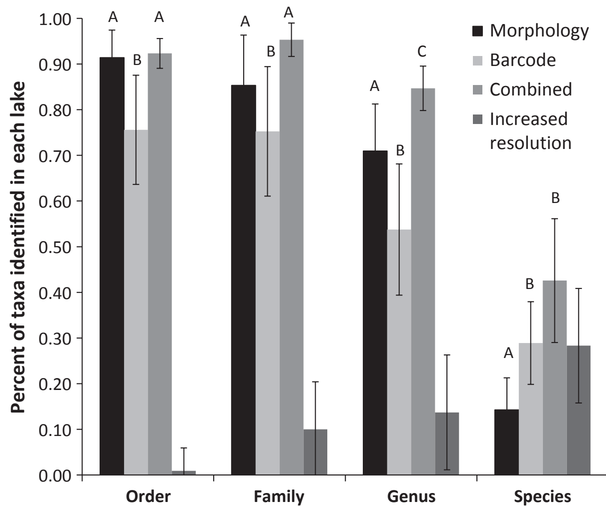

Fig. 2 Resolution obtained for different methods of identification. Comparison of the performance of different specimen identification methods. Percentage identified is based on the 170 possible taxa that were identifiable across all lakes (N = 10) sampled in this study. Morphological identification was carried out with dichotomous keys, and barcode identification was based on a match with the BOLD database from a 658 bp cytochrome oxidase I sequenced PCR product. 'Combined' includes the total identified by barcodes matches and morphology. 'Increased resolution' is the difference in number of specimens identified by a combined approach compared to only using morphology. Bars not identified by the same letter are significantly different at a P-value of <0.05 within each taxonomic level. Black lines are standard deviations based on estimates from 10 lakes.

typically performed in a manner that compromises their morphology (Smith 2001; Merritt *et al.* 2008).

More investment in molecular tools is required for taxa when the universal barcode primers do not work. The development of taxon specific primers can allow for sequence data to be produced. For example, Lázaro *et al.* (2009) reported on the success of primers designed for Platyhelminthes of the genus *Dugesia*. Primers such as these should be more broadly tested in aquatic studies. A concerted and organized effort by the molecular and aquatic taxonomic communities will be necessary to develop barcode primers for troublesome taxa. This would greatly aid aquatic ecologists in their pursuit to identify individuals in communities when morphology is not sufficient. This, however, requires reporting and publishing which taxa are consistently not amplifying such as what we have carried out here in Table 2.

The significant increase in estimates of species richness attained when a combined approach was used opens the discussion as to whether we should accept a standard phrase such as 'best available' taxonomy when based solely on one method of identification.

Table 3 Barcode success, percentage of early-instar and damaged individual identified to the genus or species level using barcodes

Lake ID	Barcode success	% Young identified	% Damaged identified
20224	0.52	0.83	N.A.
21018	0.48	1.00	0.00
20227	0.50	1.00	N.A.
20225	0.38	0.80	1.00
21004	0.65	1.00	0.50
20279	0.73	0.67	N.A.
20280	0.81	0.67	N.A.
20278	0.60	0.44	N.A.
20118	0.65	0.80	1.00
20117	0.45	0.67	1.00
Average	0.58	0.79	0.70

N.A. indicates not applicable because all specimens had traits that could allow for morphological identification. Barcode success was calculated as the ratio of number of reliable barcodes produced to the number extracted within each lake using universal primers.

Table 4 Difference in estimated taxon richness when identifications were made using morphology compared to (1) using barcodes and (2) using both barcodes and morphology

		Barcode		Combined	
Resolution	d.f.	M_d	P-value	M_d	P-value
Order	9	0.0	1.00	0.5	**0.05**
Family	9	1.3	0.74	1.4	**0.01**
Genus	9	−2.2	0.06	2.3	**0.01**
Species	9	2.7	**0.00**	4.2	**0.00**

Bolded values were significant after sequential Bonferroni correction for multiple tests. M_d is the mean difference between groups and d.f. stands for degrees of freedom.

This may especially be important for restoration efforts when uncertainty in the reference state should be estimated (Clewell *et al.* 2005). More notably, taxon richness is of high importance in conservation because it is a standard metric used to determine attributes of places such as biodiversity hotspots (Myers *et al.* 2000). For some well-known taxa, where morphology is able to resolve identifications accurately to the species level, this is less of an issue, but for estimates of taxa richness in groups such as aquatic insects, evidence in this study and others (e.g. Pilgrim *et al.* 2011; Sweeney *et al.* 2011) indicate that we are underestimating biodiversity.

Barcode identification also helps with juveniles, damaged specimens and a reduced reliance on expert taxonomic identification of every specimen, but it does

© 2013 John Wiley & Sons Ltd

not solve all of the challenges faced when creating a list of taxa for any sample. Identifications of taxa will remain troublesome, as is the case here, when universal tools cannot be employed or when libraries such as BOLD lack a reference barcode for the taxon of interest. Morphological and barcode identification methods suffer from the same lack of reference data, in that if a species is not included in a taxonomic key or the neighbour-joining tree used for higher level identification here, then a proper identification cannot be made (Wilson *et al.* 2011) and should be treated with caution. Further, we found that information on the precise methods used by BOLD to assign taxonomic names to be less transparent than what is ideally required to assess individual assignments and suggest a more detailed public explanation (e.g. explain methods on database website) of how assignments are made to increase utility of the database. Lastly, there is a burgeoning literature on additional ways to account for identification uncertainty. Specifically, Cayuela *et al.* (2011) developed a method to incorporate the effect of taxon uncertainty into hypothesis testing, and although we did not test the use of accounting for uncertainty statistically, we suggest future studies consider incorporating both barcode and statistical methods in addition to morphology for dealing with uncertainty in taxon identification, thus allowing for more robust ecological interpretations.

In conclusion, barcode identification combined with morphological identification can increase the resolution of a taxonomic list for macroinvertebrates in alpine lakes. The ability to accurately identify specimens will continue to increase with quality-controlled contributions to BOLD and we encourage this sustained collaboration between taxonomists and molecular ecologists. We found that COI sequences generated with the universal barcode primers can provide species-level identifications for many macroinvertebrate groups (e.g. Chironomidae) and from early instar and damaged specimens. We reiterate that the use of barcodes for identification purposes increases taxonomic resolution especially at the species level for aquatic benthic invertebrate communities, and we strongly encourage the use of barcodes for identification, especially when morphology is insufficient.

Acknowledgements

We would like to thank M. Stephens for laboratory advice during the project. We thank B. Hammock for morphological identification assistance with chironomids and mites. Funding for this work was provided by the Kern Community Foundation, University of California-Davis Jastro-Shields and Block Grants and Valentine Eastern Sierra Nevada Reserve Fund.

References

Baird DJ, Sweeney BW (2011) Applying DNA barcoding in benthology: the state of the science. *Journal of the North American Benthological Society*, **30**, 122–124.

Bevilacqua S, Terlizzi A, Claudet J, Fraschetti S, Boero F (2012) Taxonomic relatedness does not matter for species surrogacy in the assessment of community responses to environmental drivers. *Journal of Applied Ecology*, **49**, 357–366.

Bradford DF, Cooper SD, Jenkins TM, Kratz K, Sarnelle O, Brown AD (1998) Influences of natural acidity and introduced fish on faunal assemblages in California alpine lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 2478–2491.

Carlisle DM, Hawkins CP (1998) Relationships between invertebrate assemblage structure, 2 trout species, and habitat structure in Utah mountain lakes. *Journal of the North American Benthological Society*, **17**, 286–300.

Cayuela L, de la Cruz M, Ruokolainen K (2011) A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, **34**, 94–102.

Clewell A, Rieger J, Munro J (2005) *Society for Ecological Restoration International: Guidelines for Developing and Managing Ecological Restoration Projects*, 2nd edn, pp. 16. Society for Ecological Restoration International, Tuscon, AZ.

Cook L, Edwards R, Crisp M, Hardy N (2010) Need morphology always be required for new species descriptions? *Invertebrate Systematics*, **24**, 322–326.

DeWalt RE (2011) DNA barcoding: a taxonomic point of view. *Journal of the North American Benthological Society*, **30**, 174–181.

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**, 294–299.

IBOL (2012) International Barcode of life. Available from: http://ibol.org/resources/barcode-library/. Accessed 03 October 2012.

Jones FC (2008) Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. *Environmental Reviews*, **16**, 45–69.

Knapp RA (2005) Effects of nonnative fish and habitat characteristics on lentic herpetofauna in Yosemite National Park, USA. *Biological Conservation*, **121**, 265–279.

Knapp RA, Matthews KR (2000) Non-native fish introductions and the decline of the mountain yellow-legged frog from within protected areas. *Conservation Biology*, **14**, 428–438.

Knapp RA, Matthews KR, Sarnelle O (2001) Resistance and resilience of alpine lake fauna to fish introductions. *Ecological Monographs*, **71**, 401–421.

Larson DJ, Alarie Y, Roughley RE (2000) *Predaceous Diving Beetles (Coleoptera: Dytiscidae) of the Nearctic Region with Emphasis on the Fauna of Canada and Alaska, pp xiv + 982*. NRC Research Press, Ottawa.

Lázaro EM, Sluys R, Pala M, Stocchino GA, Baguñà J, Riutort M (2009) Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus Dugesia in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular phylogenetics and evolution*, **52**, 835–845.

Lenat DR, Resh VH (2001) Taxonomy and stream ecology: the benefits of genus-and species-level identifications. *Journal of the North American Benthological Society*, **20**, 287–298.

Menke AS (ed.) (1979) *The Semiaquatic and Aquatic Hemiptera of California (Heteroptera: Hemiptera)*, pp. 1–161. University of California Press, Berkeley, CA.

Merritt RW, Cummins KW, Berg B (2008) *An Introduction to the Aquatic Insects of North America*. Kendall/Hunt Publishing Company, Dubuque, Iowa.

Montero-Pau J, Gómez A, Muñoz J (2008) Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. *Limnology and Oceanography Methods*, **6**, 218–222.

© 2013 John Wiley & Sons Ltd

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nystrom P, Svensson O, Lardner B, Bronmark C, Graneli W (2001) The influence of multiple introduced predators on a littoral pond community. *Ecology*, **82**, 1023–1039.

Parker BR, Schindler DW, Donald DB, Anderson RS (2001) The effects of stocking and removal of a nonnative salmonid on the plankton of an alpine lake. *Ecosystems*, **4**, 334–345.

Pfrender ME, Hawkins CP, Bagley M *et al.* (2010) Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *The Quarterly Review of Biology*, **85**, 319–340.

Pilgrim EM, Jackson SA, Swenson S *et al.* (2011) Incorporation of DNA barcoding into a large-scale biomonitoring program: opportunities and pitfalls. *Journal of the North American Benthological Society*, **30**, 217–231.

Pilliod DS, Hossack BR, Bahls PF *et al.* (2010) Non-native salmonids affect amphibian occupancy at multiple spatial scales. *Diversity and Distributions*, **16**, 959–974.

Pister EP (2001) Wilderness fish stocking: history and perspective. *Ecosystems*, **4**, 279–286.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Resh VH, McElravy EP (1993) Contemporary Quantitative Approaches to Biomonitoring Using Benthic Macroinvertebrates. In: *Freshwater Biomonitoring and Benthic Macroinvertebrates* (eds Rosenberg M & Resh VH), pp. 159–194. Chapman and Hall, New York.

Smith DG (2001) *Pennak's Freshwater Invertebrates of the United States: Porifera to Crustacea*. Jolm Wiley and Sons Inc., New York.

Smith IM, Cook DR, Smith BP (2010) Water mites (Hydrachnidiae) and other arachnids. In: *Ecology and Classification of North American Freshwater Invertebrates*(eds Thorp JH & Covich AP), 3rd edn, pp. 485–586. Academic Press, San Diego, CA.

Sweeney BW, Battle JM, Jackson JK, Dapkey T (2011) Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? *Journal of the North American Benthological Society*, **30**, 195–216.

Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology and Evolution*, **24**, 110–117.

Vredenburg VT (2004) Reversing introduced species effects: experimental removal of introduced fish leads to rapid recovery of a declining frog. *Proceedings of the National Academy of Sciences*, **101**, 7646–7650.

Wiederholm T (1983) Chironomidae of the holarctic region. Keys and diagnoses. Part 1. Larvae. *Entomologica Scandinavica Supplement*, **19**, 1–457.

Wilson J, Rougerie R, Schonfeld J *et al.* (2011) When species matches are unavailable are DNA barcodes correctly assigned to higher taxa? An assessment using sphingid moths. *BMC Ecology*, **11**, 1–14.

K.D., R.A.K., D.M.B. and B.M. designed research; K.D. and D.M.B. performed research; B.M. contributed new reagents or analytical tools, K.D. and B.M. analysed data, K.D., R.A.K., D.M.B. and B.M. wrote the paper.

Data Accessibility

DNA sequences: GenBank Accessions KF000103 - KF000348. Final edited DNA sequences, alignment and tree file and all files used for statistical analysis available at DRYAD entry doi:10.5061/dryad.n5 h13. Online supplementary material includes Table S1 and Table S2 both in excel format.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Extended taxonomic table listing each specimen from each lake and identifying the sequence ID for each COI sequence from each specimen depicted in Fig. 1

Table S2 Metadata for each sequence used for identification in this study including the link between the SeqID used in Fig. 1 and the Genbank accession number

© 2013 John Wiley & Sons Ltd