

DNA barcoding at riverscape scales: assessing biodiversity among fishes of the genus *Cottus* (Teleostei) in northern Rocky Mountain streams

MICHAEL K. YOUNG, KEVIN S. MCKELVEY, KRISTINE L. PILGRIM and MICHAEL K. SCHWARTZ
U.S. Forest Service, Rocky Mountain Research Station, 800 East Beckwith Avenue, Missoula, MT 59801, USA

Abstract

There is growing interest in broad-scale biodiversity assessments that can serve as benchmarks for identifying ecological change. Genetic tools have been used for such assessments for decades, but spatial sampling considerations have largely been ignored. Here, we demonstrate how intensive sampling efforts across a large geographical scale can influence identification of taxonomic units. We used sequences of mtDNA cytochrome *c* oxidase subunit 1 and cytochrome *b*, analysed with maximum parsimony networks, maximum-likelihood trees and genetic distance thresholds, as indicators of biodiversity and species identity among the taxonomically challenging fishes of the genus *Cottus* in the northern Rocky Mountains, USA. Analyses of concatenated sequences from fish collected in all major watersheds of this area revealed eight groups with species-level differences that were also geographically circumscribed. Only two of these groups, however, were assigned to recognized species, and these two assignments resulted in intraspecific genetic variation (>2.0%) regarded as atypical for individual species. An incomplete inventory of individuals from throughout the geographical ranges of many species represented in public databases, as well as sample misidentification and a poorly developed taxonomy, may have hampered species assignment and discovery. We suspect that genetic assessments based on spatially robust sampling designs will reveal previously unrecognized biodiversity in many other taxa.

Keywords: Cottidae, cryptic species, sculpins, species discovery

Received 19 December 2012; revision received 24 January 2013; accepted 6 February 2013

Introduction

Projections of a rapidly changing climate and increasing human population in North America have led to calls for broad-scale biodiversity assessments that can serve as benchmarks for identifying ecological change. Assessing biological diversity requires identifying taxa of interest and describing their distributions. Species-level diversity has long been the primary metric by which biodiversity is measured, in part because organisms at the species level are often easily identified on the basis of their morphology, behaviour or acoustics. In many countries, diversity at levels below that of species is neglected with regard to conservation (Laike 2010), but to some degree that reflects the difficulty in cataloguing variation at lower taxonomic levels when using traditional methods. Increasingly, genetic tools are permitting more detailed and accurate assessments of biodiversity because of their ability to identify conservation units within species and

resolve cryptic species complexes (Bickford *et al.* 2007; Valentini *et al.* 2009).

Among the largest ongoing efforts to catalogue biodiversity are those predicated on DNA barcoding (Ratnasingham & Hebert 2007), which relies on the sequencing and comparison of a standardized portion of the genome—most often cytochrome *c* oxidase subunit 1 (COI) region of mtDNA (Hebert *et al.* 2003a)—for species delineation and identification. Although initially viewed as highly controversial (Rubinoff 2006), it has subsequently proven to be effective in many circumstances (Teletchea 2010). Although nuclear DNA sequencing has certain advantages and is becoming more feasible for biodiversity assessment (Taylor & Harris 2012), the lack of recombination, rarity of indels, ease and low cost of amplifying and sequencing, and high mutation rates of mtDNA have favoured its use (Zink & Barrowclough 2008).

Nevertheless, assigning samples collected in biodiversity surveys to recognized species is often problematic. The first issue arises when sequences of samples are

Correspondence: Michael K. Young, Fax: +1 406 543 2663;
E-mail: mkyoung@fs.fed.us

compared with those in reference collections. Public databases such as those maintained by the National Center for Biotechnology Information or Barcode of Life Data system (BOLD) contain tens of millions of reference sequences for hundreds of thousands of species. Although these databases represent an enormous, publically available catalogue of life, they suffer several shortcomings: incomplete taxonomic and geographical coverage (Nielsen & Matz 2006; Elias *et al.* 2007), inclusion of poor-quality sequences (Harris 2003) and misidentification of voucher specimens (Kvist *et al.* 2010). These can cause problems during the final step in the genetic assessment of biodiversity, which is to determine whether a sampled individual is a representative of an existing species or an undescribed one. Most examples of species identification from sequence data use distance-based methods and are founded on the observation that intraspecific genetic distances (generally <1%) tend to be much smaller than interspecific distances among congeners (Hebert *et al.* 2003b). Because minimum genetic distances of 1–3% are typical of species-level differences for many groups of vertebrates (Avice & Walker 1999; Ratnasingham & Hebert 2007; Ward 2009), this approach is generally accepted (Goldstein & DeSalle 2010) especially when corroborated by other methods (Ross *et al.* 2008; Zou *et al.* 2011). More controversial has been the use of distance-based methods for the discovery of new species (DeSalle *et al.* 2005). It has been argued that individuals with sequences that differ by some interspecific threshold (e.g. 2% among freshwater fishes; Ward 2009) from sequences of all other recognized forms may constitute new species (Hebert *et al.* 2004). Much of the critique of species discovery via this approach has focused on improved methods of calculation and selection of distance thresholds (Meyer & Paulay 2005; Meier *et al.* 2008; Collins *et al.* 2012) or use of alternatives such as character-based or model-based methods (Zou *et al.* 2011; Powell 2012). Some of the controversy is more fundamental, because somewhat arbitrary genetic distances are acceptable currency for species delineation only under certain hypotheses about what constitutes a species, for example, some variants of the phylogenetic species concept (Hausdorf 2011).

Surprisingly, the consequences of the geographical distribution of sampling for genetically based biodiversity assessments have been largely overlooked (Bergsten *et al.* 2012). Many studies rely on one or a few individuals of a species collected from a fraction of its distribution (Funk & Omland 2003; Meyer & Paulay 2005; Wiemers & Fiedler 2007). These individuals often reflect locations that were convenient to sample or were obtained as an adjunct to other activities, rather than resulting from a statistically robust sampling design. The field of phylogeography, however, is

founded on the notion that ancient and contemporary geological and climatic events, combined with the vagility of an organism, constrain and direct landscape-scale—or for stream-dwelling organisms, riverscape-scale (Fausch *et al.* 2002)—genetic structure in most species (Avice 2000). Thus, sampling schemes and reference databases must account for this structure to permit reliable delineation of species or major lineages. This is especially important when relying on contrasts in genetic distances within and among taxa because broad-scale sampling often results in increased intra-specific variation and less certainty about interspecific distance thresholds, that is the barcode gap (Fregin *et al.* 2012).

Fine-grained yet broad-scale sampling is critical for taxa that may exhibit limited dispersal and localized divergence, particularly among poorly studied groups. One such group comprises fishes in the genus *Cottus*, commonly known as sculpins, which are primarily freshwater, benthic species found in lakes, rivers and streams throughout the Northern Hemisphere. Sculpins are mid-trophic species that serve as prey for larger piscivores but also consume macroinvertebrates and the egg and larval stages of other fishes. Although they can represent the bulk of aquatic vertebrate biomass in small streams (Cheever & Simon 2009; Raggon 2010), their ecological significance is poorly understood. Contributing to this uncertainty is a lack of taxonomic clarity. Species in this genus are widely acknowledged as being among the most difficult freshwater fishes to identify (Jenkins & Burkhead 1994; Wydoski & Whitney 2003), in part because putatively diagnostic characteristics are geographically variable within species (McPhail 2007). Molecular analyses (Kinziger *et al.* 2005; Hubert *et al.* 2008; April *et al.* 2011) revealed that several species of North American *Cottus* contain deeply divergent lineages that might represent unrecognized species, thus the identity and distribution of many members of this genus remain unclear. And because the lifetime home ranges of individual fish can be relatively small, for example, <250 m (Petty & Grossman 2007; Hudy & Shiflet 2009), divergence at small spatial scales may be the norm.

A region in which there has been little work on the biodiversity of sculpins includes the upper Columbia and Missouri River basins in northern Idaho and western Montana. Up to five species of sculpins—*Cottus bairdii*, *Cottus beldingii*, *Cottus cognatus*, *Cottus confusus* and *Cottus rhotheus*—have been thought to occur in small streams in this region, although there has been little consensus on their individual distributions (Lee *et al.* 1980; Wydoski & Whitney 2003; McPhail 2007). Confusion about the identity and distribution of sculpins in this region has also led to ambiguity in conservation

priorities. In Montana and part of British Columbia, *C. confusus* has at different times been regarded as either a species of special concern or as having never been present (Hendricks 1997; Committee on the Status of Endangered Wildlife in Canada (COSEWIC) 2010).

Our goal was to resolve the distribution and identity of members of this taxonomically difficult group from small streams throughout the U.S. Northern Rocky Mountains. First, we collected sculpins from a spatially comprehensive and randomly selected set of streams that represented all major river basins. Second, we sequenced two regions of the mitochondrial genome to permit comparisons with nearly all described species of North American freshwater sculpins in public databases. Third, we used consensus results from an array of methods to delineate and identify species.

Materials and methods

Sampling

Our data set consisted of sequences from samples collected in the field and sequences obtained from public sequence repositories. Field collections were made (using electrofishing) from 398 streams sampled from 2008 to 2011 on state and federal lands in the upper Columbia and Missouri River basins in northern Idaho and western Montana (Fig. 1). These streams formed part of the PACFISH/INFISH Biological Opinion Effectiveness Monitoring Program network (Kershner *et al.* 2004). This network comprises a random sample (Stevens & Olsen 1999) of about one-third of all 6th-code subbasins (area, 40–160 km²; Wang *et al.* 2011) with substantial federal ownership. Nearly, all sample sites consisted of the first low-gradient stream reach on public land (Kershner *et al.* 2004). Captured fish were held briefly in buckets containing stream water. Before releasing them, we retained upper caudal fin clips (on chromatography paper; LaHood *et al.* 2008) of up to 10 specimens captured at each site, but made no attempt to identify sculpins in the field because most species cannot be easily distinguished, and we sometimes handled hundreds of individuals at each site. We captured sculpins in 187 streams and analysed 1–5 fish (generally 2) from 119 streams (and at two sites on five of these streams). Our intent was to include 2–4 streams representing each 4th-code subwatershed in this area (<http://water.usgs.gov/GIS/regions.html>; Table S1, Supporting information). All collections were made under scientific collection permits issued (to MKY) by Montana Fish, Wildlife and Parks and the Idaho Department of Fish and Game. All tissue specimens and extracted DNA were voucherized at the Wildlife Genetics Laboratory, Missoula, MT.

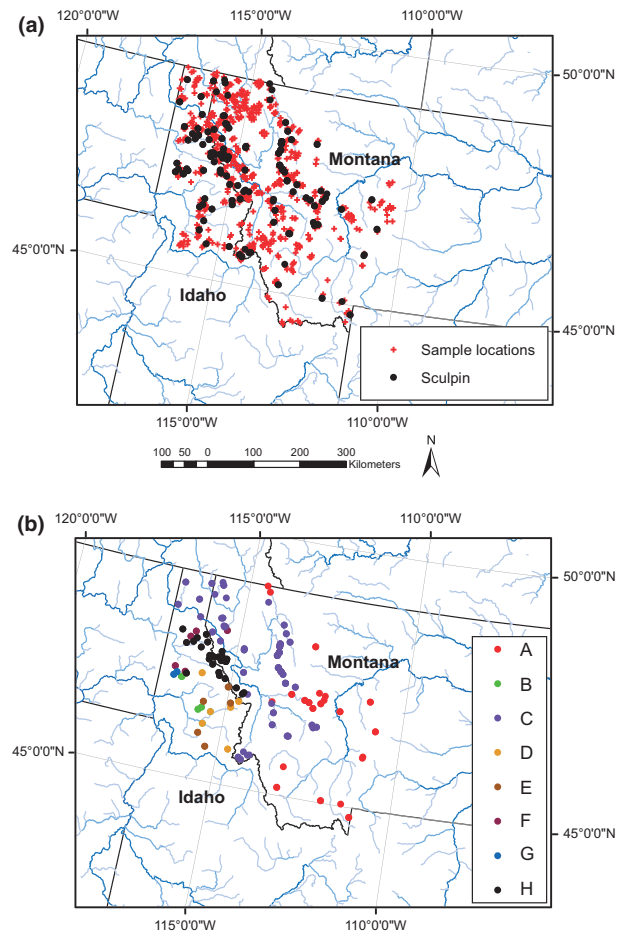


Fig. 1 Observations of *Cottus* in the upper Columbia River and Missouri River basins, Idaho and Montana, USA. These include (a) all sampled sites and those locations for which sculpins were sequenced and (b) the haplotype groups observed at each site.

DNA regions and phylogenetic analyses

Although sequences from many mtDNA regions have been used to identify fish species, we used COI and cytochrome *b* (cyt *b*) because their popularity (Ratnasingham & Hebert 2007; Page & Hughes 2010) provided the broadest coverage of sculpins in public databases. We sequenced the COI region for all sculpins ($n = 236$) in the sample (Table S1, Supporting information). A subsample of these ($n = 120$) that included most novel COI haplotypes was also sequenced at cyt *b*. GenBank accession numbers for these sequences are JX282572–282599 (for COI) and JX282526–282571 (for cyt *b*).

We used the QIAGEN DNeasy Blood and Tissue kit (QIAGEN Inc.) to extract genomic DNA from tissues, following the manufacturer's instructions for tissue. The COI region was amplified with primers *FF2d* and *FR1d*

(Ivanova *et al.* 2007). The *cyt b* region was amplified with primers *L14724* and *H15915* (Schmidt & Gold 1993). Reaction volumes of 50 μ L contained 50–100 ng DNA, 1 $\times$ reaction buffer (Life Technologies), 2.5 mM MgCl₂, 200 μ M each dNTP, 1 μ M each primer, 1 U Taq polymerase (Life Technologies). The PCR programme was 94 °C/5 min, [94 °C/1 min, 55 °C/1 min, 72 °C/1 min 30 s] $\times$ 34 cycles, 72 °C/5 min. The quality and quantity of template DNA were determined by 1.6% agarose gel electrophoresis. PCR products were purified with Exo-Sap-IT (Affymetrix-USB Corporation) according to manufacturer's instructions.

We used the Big Dye kit and the 3700 DNA Analyzer (ABI; High Throughput Genomics Unit) to obtain DNA sequences. We used the given primers to generate DNA sequence data for COI, whereas we used internal primers *LC1*, *LC2* and *LC3* (Kinziger & Wood 2003) to generate *cyt b* sequences in three fragments. Sequences were viewed and aligned with Sequencher (Gene Codes Corp.).

The COI data set consisted of 620-bp sequences, which included 182 variable and 124 parsimony-informative sites. Mean nucleotide frequencies were A, 0.2274; T, 0.2979; C, 0.2958; and G, 0.1789. The 1034-bp *cyt b* data set consisted of 370 variable and 256 parsimony-informative sites, for which mean nucleotide frequencies were A, 0.2337; T, 0.2861; C, 0.3247; and G, 0.1555. We observed no indels or gaps, amino acid translations did not reveal any stop codons, and all sequences exceeded lengths typical of nuclear DNA of mitochondrial origin (Zhang & Hewitt 1996).

We used a two-step approach to delineate and identify species of sculpin from the field collections. First, we assigned field-collected samples to particular haplotype groups using three methods to analyse concatenated COI and *cyt b* sequences ($n = 120$; Table S1, Supporting information): 95% maximum parsimony networks, maximum-likelihood phylogenetic trees and pairwise genetic distances. We used TCS 1.21 (Clement *et al.* 2000) to construct 95% maximum parsimony networks. Independent networks were regarded as candidate species (Hart & Sunday 2007). Ninety-five per cent maximum parsimony networks tend to be conservative estimators of species diversity because the probability of pooling distinct taxa into one network can be relatively high, whereas the likelihood of splitting a single taxon into separate networks is low (Hart & Sunday 2007; Chen *et al.* 2010). Next, we constructed mid-point-rooted maximum-likelihood phylogenetic trees under the strict tree-based method (Ross *et al.* 2008) to delineate groups. The evolutionary model GTR + G + I yielded the lowest AIC_c and highest log-likelihood values for these data. Using this model, we constructed maximum-likelihood trees with 1000 bootstrap replicates incorporating subtree-prune-and-graft

branch-swapping in MEGA 5.1 (Tamura *et al.* 2011). We inspected terminal clades in the majority consensus tree for reciprocal monophyly and for concurrence with the independent maximum parsimony networks. Because of incomplete lineage sorting, species can develop without exhibiting reciprocal monophyly (Funk & Omland 2003); thus, this method also tends to underestimate species richness. Finally, after assignments to putative individual species, we then compared the maximum intragroup distance to the minimum intergroup distance to assess whether a barcode gap was evident and consistent (Meier *et al.* 2006). We used the absolute maximum intragroup distance as the threshold for species delineation. We based calculations on uncorrected *p*-distances because these have been shown to perform as well or better for detection of barcode gaps than the more broadly used Kimura-2-parameter model (Collins *et al.* 2012; Srivathsan & Meier 2012).

The second step involved repeating these analyses with the inclusion of public sequences to facilitate species identification. These consisted of all publically available sequences of *Cottus bairdii*, *Cottus beldingii*, *Cottus cognatus*, *Cottus confusus* and *Cottus rhotheus*—those species believed to occupy the region we sampled—with unique haplotypes as well as single representatives of all other species of *Cottus* from North America present in public sequence repositories (Table S2, Supporting information). For these analyses, COI ($n = 28$ haplotypes from field sampling and 46 from public sources) and *cyt b* ($n = 46$ haplotypes from field sampling and 39 from public sources) sequences were examined separately because no individual in public databases was represented by sequences of both loci. After alignment, all sequences were trimmed to the same number of nucleotides. Sequences with ambiguous nucleotides were excluded from maximum parsimony networks because they can introduce errors in network structure (Joly *et al.* 2007) but were retained for other analyses. Following Kinziger *et al.* (2005), we used *Leptocottus armatus* as an outgroup in maximum-likelihood phylogenetic trees. Because the focus at this step was to assign each group to a described species, we adopted the liberal tree-based method with a distance threshold (Ross *et al.* 2008), that is, sequences that were sister to or within a monophyletic clade with a recognized species, and were less than the maximum intraspecific distance from that species, were identified as that species. We based distance thresholds on the absolute maximum intragroup distance for each marker-haplotype group combination for our field-collected samples, which led to a variable distance threshold. When haplotype groups were represented by a single sequence, we adopted the largest intragroup distance among all haplotype groups for that marker as the threshold distance.

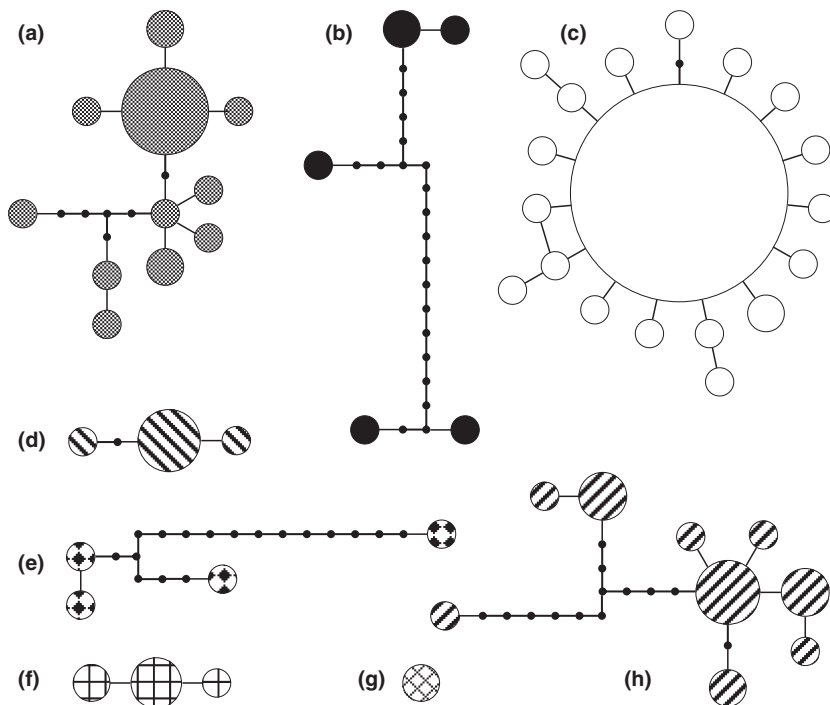


Fig. 2 Maximum parsimony networks of concatenated cytochrome *c* oxidase subunit 1 and cytochrome *b* oxidase sequences for sculpins collected in this study. Each circle represents a haplotype, and sizes are proportional to the number of individuals with that haplotype. Each line segment represents a single mutation, and small black dots represent unobserved haplotypes.

We interpreted consensus among the three methods and both markers—as well as geographical concordance—as evidence for (i) the assignment of a haplotype group to a described species or (ii) that group to represent a lineage or species not represented in public databases.

Results

Species delineation

On the basis of the concatenated sequences, there was substantial concordance among methods with respect to the classification of field-sampled sculpins. Calculation of 95% maximum parsimony networks generated eight distinct networks of haplotypes (Fig. 2) that were also resolved as reciprocally monophyletic clades in maximum-likelihood trees (Fig. 3). Comparisons of all possible pairwise genetic distances revealed that minimum intergroup distances (mean, 4.18%) tended to exceed maximum intragroup distances (mean, 0.58%; Table 1). Nevertheless, overlap between the minimum intergroup distance (haplotype groups F–G, 1.09%) and maximum intragroup difference (haplotype group B, 1.21%) precluded designating a fixed threshold (or barcode gap) that would delineate all groups. The eight haplotype groups (A–H) delimited by the network- and tree-building approaches, however, were also geographically distinct; most were confined to particular river basins within the study area. Consequently, we

regarded these eight groups as potential species in further analyses.

Species identification

Assignment of haplotype groups to known species depended on the region used and on the pool of species to which each haplotype group could be compared. Network analyses of COI sequences collapsed the haplotype groups into three networks, only two of which were unambiguous. Group B was connected to two haplotypes identified as *Cottus beldingii*. Haplotype groups D and E were joined in a single network unconnected to any other. The remaining haplotype groups were joined in a single complex network that included public sequences of *Cottus bairdii*, *Cottus caeruleomentum*, *Cottus cognatus* and *Cottus rhotheus*. In contrast, analyses of *cyt b* sequences returned seven unambiguous networks. Again, haplotype group B was connected to a public sequence identified as *Cottus beldingii*. The remaining networks consisted of one (A, C, F, G, H) or two (D–E) haplotype groups and were unconnected to sequences of any recognized species.

Similarly, maximum-likelihood trees based on COI sequences were less well resolved and supported than those based on *cyt b* sequences, but the topologies produced by both regions were comparable with regard to their sister taxa and to monophyly of the haplotype groups (Fig. 4). In the COI-based tree, bootstrap support was high ($\geq 86\%$) for all terminal clades containing

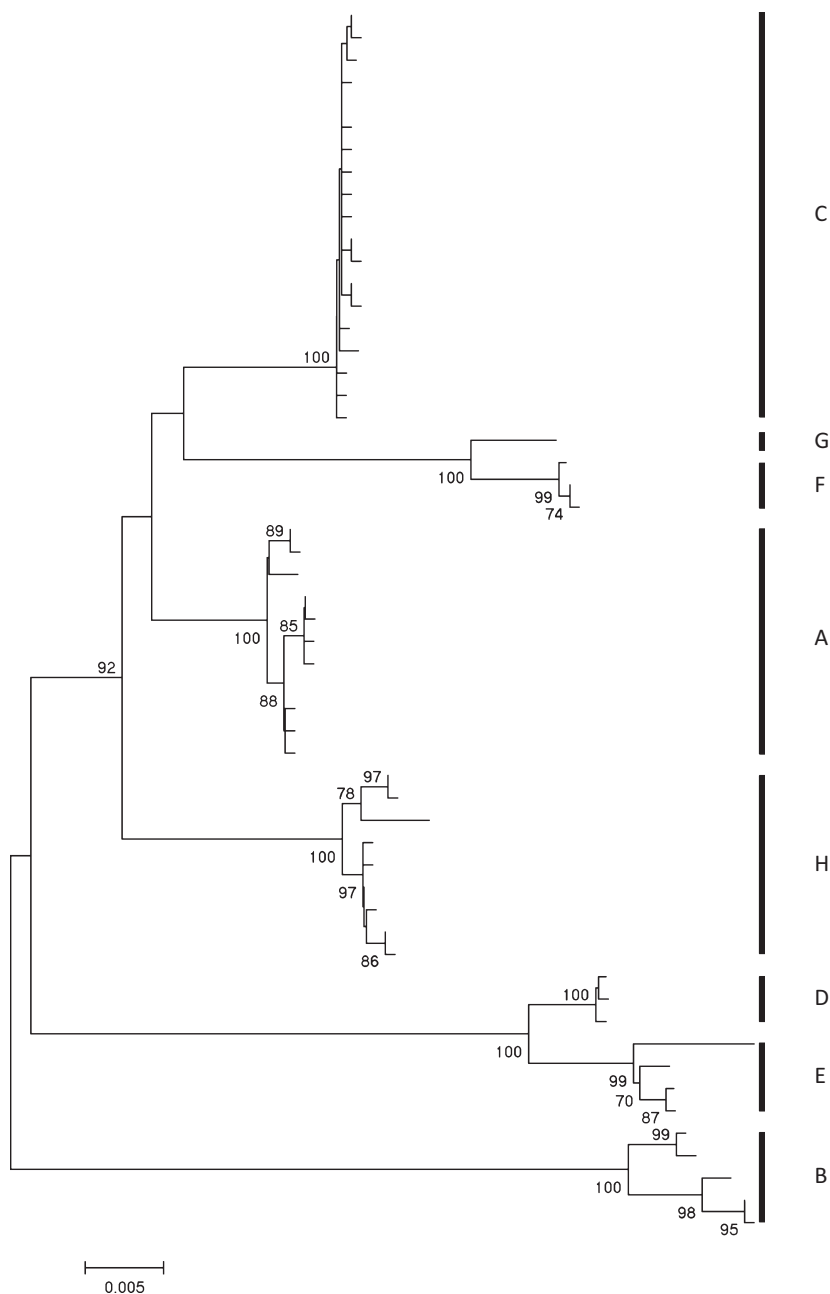


Fig. 3 Maximum-likelihood phylogeny of concatenated cytochrome *c* oxidase subunit 1 and cytochrome *b* oxidase sequences for sculpins collected in this study. Numbers on the branches are support with respect to 1000 bootstrap replications. Only values ≥ 70 are shown.

haplotype groups, and half of the groups were reciprocally monophyletic from all other sculpins. Applying the distance threshold (Table 2) resulted in identifying group B as *C. beldingii*, combining groups D and E as a single unidentified taxon and combining groups F and G and identifying them as *C. rhotheus*. The remaining groups were not assigned to any recognized species. In the *cyt b*-based tree, all haplotype groups were in terminal clades with very high bootstrap support ($\geq 97\%$). All were reciprocally monophyletic other than group B, which was again identified as *C. beldingii*. The other seven haplotype groups were too distant to be combined

with recognized species or one another and were regarded as unidentified species.

Comparisons of pairwise genetic distances indicated differentiation among haplotype groups and most recognized species (Table 2). The nearest neighbour of each haplotype group was inconsistent; for only three of the eight groups was the same nearest neighbour identified in analyses with both markers. Intragroup genetic distances tended to be larger for *cyt b* than for COI sequences (means, 0.59% vs. 0.44%). In all comparisons, genetic distances to the nearest neighbour were consistently greater with *cyt b* than with COI.

Table 1 Pairwise distance matrix (uncorrected *p*-distances, %) for concatenated sequences of field-sampled sculpins

	A	B	C	D	E	F	G	H
A*	0.48†							
B	4.95	1.21						
C	1.87	4.77	0.24					
D	4.04	5.97	4.53	0.18				
E	4.28	6.16	4.71	1.27	0.97			
F	2.78	6.64	2.96	5.43	5.49	0.12		
G	2.90	6.22	2.90	5.85	5.91	1.09	NC‡	
H	2.35	5.19	2.60	4.47	4.71	3.44	3.44	0.84

*We based group labels on independent 95% maximum parsimony networks and reciprocally monophyletic clades from maximum-likelihood trees.

†Values on the diagonal are maximum intragroup distances and off-diagonal values are minimum intergroup distances.

‡NC indicates no comparison because only one haplotype was observed.

Discussion

Delineation and identification of haplotype groups

Our spatially comprehensive sample likely provided a thorough assessment of haplotype diversity in stream-dwelling sculpins in the upper Columbia and Missouri River basins. In general, there was consensus among the three approaches and two markers for the existence of eight lineages that differed at levels typical of separate freshwater fish species (Ward 2009). The complex geological and climatic history in this region—involving the Cascadian orogeny, active volcanism, redirection of major river basins and Pleistocene glaciation—has led to diversification in an array of taxa (Shafer *et al.* 2010), so this diversity may be unsurprising. Quite surprising, however, was that the majority of these groups did not assign to recognized species, despite North American fishes having been generally well inventoried in genetic surveys (Hubert *et al.* 2008; April *et al.* 2011). Below, we briefly describe the likely taxonomic position of each group and its distribution.

Group A was the only sculpin in the upper Missouri River basin and was also present in some portions of the Columbia River basin immediately adjacent to the Continental Divide. Although historically assigned to *Cottus bairdii* (Holton & Johnson 2003), sculpins from this portion of the study area are now believed to represent an unrecognized species (Neely 2003; McPhail 2007). Our results support this interpretation, but with a caveat: it may represent a range extension for *Cottus hubbsi*. Both phylogenetic trees suggest that the closest relatives of group A comprise those sculpins also formerly thought to be *C. bairdii*—*C. bairdii punctulatus*, *C. b. semiscaber*,

Cottus bendirei, *Cottus extensus* and *C. hubbsi* (Neely 2003). Because one individual of the latter species had a very similar COI haplotype, comprehensive sampling within its range should inform taxonomic placement of fish in group A.

Group B was found exclusively within the Clearwater River basin and aligned with some public sequences of *Cottus beldingii* in all analyses. Although we identify group B as likely members of that species, this assignment is problematic. Late in the 20th century, sculpins morphologically identified as *C. beldingii* were collected at locations close to those we sampled (Maughan 1976). Fish from one these locations, however, were once described as a separate species, *Cottus tubulatus* (Hubbs & Schultz 1932). Furthermore, haplotypes in this group—from the northernmost distribution of *C. beldingii*—were separated by genetic distances of 2.70–4.52% from those of *C. beldingii* present in the upper Snake River, Great Basin and Willamette River. This high intraspecific divergence warrants taxonomic attention, particularly given that the type location for this species is Lake Tahoe, Nevada (Lee *et al.* 1980).

Group C was found throughout the Pend Oreille and lower Kootenai River basin. Although sculpins from this area have traditionally been recognized as *Cottus cognatus* (Wydoski & Whitney 2003; McPhail 2007), public sequences labelled as *C. cognatus* were neither the nearest neighbour to this group (based on distance) nor among the most closely related (based on placement in the phylogenetic trees). Because the range boundary of *C. cognatus* in eastern North America is 500–1500 km distant (Lee *et al.* 1980), it is unlikely that intermediate haplotypes exist that would cause this species and group C to join in phylogenetic networks or trees. Moreover, a *C. cognatus* haplotype from extreme eastern Siberia also differs greatly (2.80–2.99%) from group C. Consequently, we regard this group as a potentially undescribed species.

Groups D and E were found in overlapping portions of the upper Clearwater River basin, were most similar to one another in all analyses and were sometimes joined as a single group. Their nearest neighbour in any analysis is a single sequence of *C. confusus* collected from near the type location of this species in the Salmon River near Ketchum, Idaho (Lee *et al.* 1980), but this haplotype is relatively distant (2.22–2.70%) from either groups. Nevertheless, *C. confusus* has been described as the only sculpin present in the upper Clearwater River (Maughan 1976). Interpreting the identity of groups D and E as part of a highly variable species or as independent, unnamed lineages will require analyses of additional samples from throughout the historical range of *C. confusus*.

The distribution of groups F and G are consistent with that of *C. rhotheus* (Maughan *et al.* 1980; Wydoski

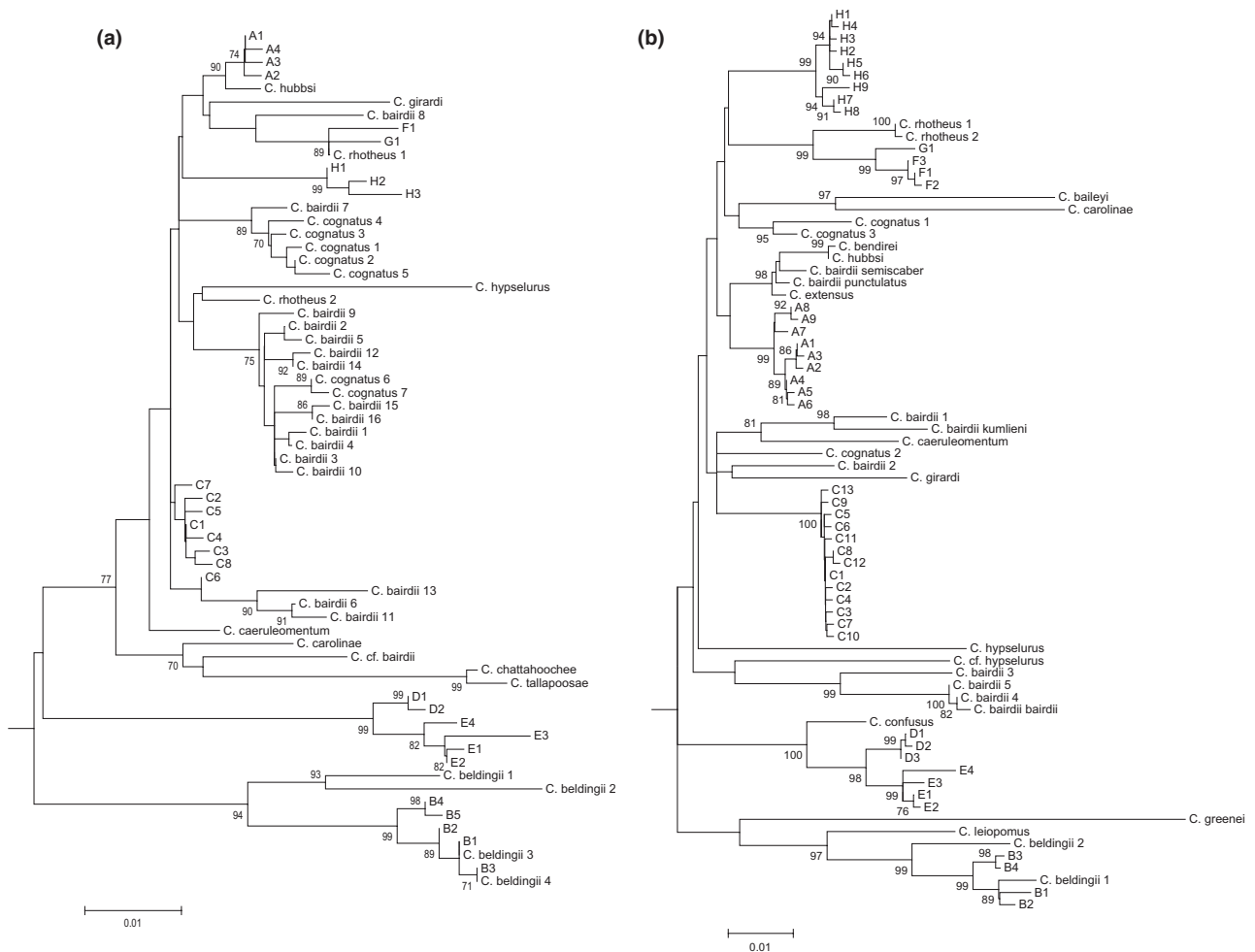


Fig. 4 Maximum-likelihood phylogeny for sculpins. These represent fish collected during this study and species of Cottidae from North America represented in GenBank for (a) cytochrome *c* oxidase subunit 1 (COI) and (b) cytochrome *b* oxidase (*cyt b*) sequences. Numbers on the branches are support with respect to 1000 bootstrap replications. Only values ≥ 70 are shown. The outgroup (*Leptocottus armatus*) and distantly related *Cottus* of the *C. asper* clade (*C. aleuticus*, *C. asper*, *C. asperimus*, *C. gulosus*, *C. klamathensis*, *C. marginatus*, *C. perplexus*, *C. pitensis*, *C. princeps*, and *C. tenuis*; Kinziger et al. 2005) and *C. ricei* are not shown.

& Whitney 2003; McPhail 2007). In addition, COI haplotypes of *C. rhotheus* were more similar to those of F and G than they were to one another and the type location—Spokane River Falls, Washington—was relatively close to some of our sampling sites. Consequently, we regard these groups as possible members of this species, but note that this results in an intraspecific genetic distance among all available samples (at *cyt b*) exceeding 2.5%, a threshold generally associated with species-level differences.

Group H was distributed throughout the Coeur d'Alene and St. Joe River basins in Idaho, with a disjunct range in the central Clark Fork River basin in Montana. This distribution is not consistent with that of any described species. In some treatments, sculpins from near the Idaho sampling locations were identified as

C. confusus (Bailey & Bond 1963), but members of group H differ by more than 4.05% from that species (and groups D and E) and are unrelated on the basis of our analyses. Given the large differences between group H and all other species and haplotype groups of sculpins, we consider it likely that this group constitutes an undescribed species.

Did genetically based biodiversity assessment succeed?

We regard the characterization of sculpin biodiversity in the U.S. northern Rocky Mountains as a partial success. Genetic analyses of spatially comprehensive collections of *Cottus* revealed eight geographically and genetically delineated groups of sculpins that could be regarded as distinct taxa. Nevertheless, we were able to assign (and

Table 2 Pairwise genetic distances (%) within haplotypes groups (maxima, bolded) and to their three nearest neighbours (range)

Haplotype group	Nearest neighbour group	COI	Nearest neighbour group	cyt b
A		0.32*		0.77
	<i>Cottus hubbsi</i>	0.48–0.65	<i>Cottus extensus</i>	1.54–1.83
	C	0.97–1.29	<i>Cottus bairdii punctulatus</i>	1.64–1.93
	<i>Cottus cognatus</i>	1.13–2.10	<i>C. cognatus</i>	1.74–2.70
B		1.13		1.26
	<i>Cottus beldingii</i>	0.00–4.52	<i>C. beldingii</i>	0.77–3.19
	C	3.87–4.84	<i>Cottus leiopomus</i>	4.16–4.83
	<i>C. hubbsi</i>	4.52–4.84	<i>Cottus bairdii punctulatus</i>	4.83–5.02
C		0.32		0.29
	<i>Cottus caeruleomentumum</i>	0.81–1.13	<i>C. bairdii punctulatus</i>	2.41–2.61
	A	0.97–1.29	A	2.41–2.90
	<i>Cottus bairdii</i>	0.97–1.94	<i>C. extensus</i>	2.51–2.70
D		0.16		0.19
	E	0.97–1.94	E	1.26–1.83
	<i>C. hubbsi</i>	4.03–4.19	<i>Cottus confusus</i>	2.22–2.41
	A	4.19–4.52	A	3.86–4.34
E		0.97		1.06
	D	0.97–1.94	D	1.26–1.83
	<i>C. hubbsi</i>	4.19–4.84	<i>C. confusus</i>	2.32–2.70
	C	4.19–5.00	<i>C. cognatus</i>	3.67–5.14
F		NA†		0.19
	<i>Cottus rhotheus</i>	0.65–2.90	G	1.06–1.16
	G	1.13	<i>Cottus rhotheus</i>	2.51–2.70
	<i>C. bairdii</i>	1.77–3.55	A	3.19–3.67
G		NA		NA
	<i>C. rhotheus</i>	0.48–2.74	F	1.06–1.16
	F	1.13	<i>C. rhotheus</i>	2.41–2.51
	C	1.77–1.94	A	3.28–3.57
H		0.65		0.97
	C	1.61–2.26	<i>C. bairdii punctulatus</i>	2.22–2.61
	A	1.94–2.74	<i>C. extensus</i>	2.22–2.61
	<i>C. rhotheus</i>	1.94–3.23	<i>C. bairdii semiscaber</i>	2.32–2.70

*Values are from uncorrected *p*-distances.

†NA indicates that no intragroup distance could be calculated because a haplotype group was represented by a single haplotype.

tenuously at best) few of the haplotype groups to a recognized species on the basis of genetic sequences. Is it possible that we have discovered up to eight previously undescribed taxa (e.g. species or subspecies) or is the current approach to genetically based identification suspect?

A primary tenet of such identification is that all species that are potentially related to an unknown sample are available for comparison. We expected to meet this criterion because we examined sequences from nearly all named species of *Cottus* in North America and because barcode coverage of North American freshwater fishes is regarded as nearly complete (Hanner *et al.* 2011). Although it is plausible that the groups we observed constitute novel taxa, representatives of sculpins in public databases from the region we sampled were almost nil, which we regard as indicative of a more fundamental issue. Current protocols for global, genetically based species inventories aim for five individuals of most species, and up to 25 individuals of widely distributed taxa

(Steinke & Hanner 2011). If one presupposes that intra-specific genetic variation is uniformly low, a few samples of an organism—perhaps even one—would be adequate for species diagnosis or discovery. But characterizing biodiversity at and below the level of species appears to require several-fold larger samples (Bergsten *et al.* 2012). Moreover, we observed geographically complex and highly divergent genetic structure among many of our haplotype groups (Fig. 2), which may be common among organisms that exhibit limited vagility or occupy patchily distributed habitats (Hammer *et al.* 2010) in environments of varying stability over evolutionary time. In such instances, reference collections of convenience samples collected at coarse scales will misrepresent taxonomic diversity by underestimating intraspecific genetic diversity and failing to include novel lineages that constitute new species.

Improving this situation requires progress on several fronts. Foremost is the addition of data sets derived from

systematic, fine-scale sampling. Based both on sculpin biology and on the wide genetic divergence between reference samples putatively belonging to the same species, we believe that river-basin-level genetic structuring of sculpins (or any taxa with comparable life histories) is likely across North America, particularly in regions that have not been recently glaciated (Bernatchez & Wilson 1998). We echo the suggestion that a thorough revision of the taxonomy of sculpins is warranted (Kinziger *et al.* 2005), but emphasize that far more attention to the spatial scale and grain of sampling will be necessary to achieve a robust phylogeny, if for no other reason than such sampling is more likely to collect all extant species (Hillis *et al.* 2003).

Correct assignment to described species or discovery of new species also requires that reference specimens are correctly identified. The extent of divergence within some species of sculpin, for example, *C. bairdii*, *C. beldingii*, *C. confusus* and *C. rhotheus*, suggests that multiple taxa may be concealed under these species designations (April *et al.* 2011). Although this may reflect an imprecise taxonomy, misidentification of specimens also seems likely and plagues many samples in reference collections (Becker *et al.* 2011; Cerutti-Pereyra *et al.* 2012). A system of tissue or DNA vouchers with precise georeferencing of the locations of collection, as now required for submission to some sequence repositories (Ratnasingham & Hebert 2007; Puillandre *et al.* 2012), would permit resequencing of problematic individuals or recollection at those sites and begin to establish geographical boundaries for haplotype groups. And particularly for species that are genetically structured at small spatial scales, a critical adjunct to this process may be to sequence individuals from the type locations of many species with an aim towards assigning a reference sequence to an individual representative of the holotype (Kvist *et al.* 2010; Lowenstein *et al.* 2011).

Finally, the methods of species discovery and the markers used for it merit further attention. Shortcomings of individual genetic methods and markers have been widely reported (Meier *et al.* 2006), and we observed similar issues. For example, one weakness of most network- or tree-based methods is that they fail to discriminate between closely related species (Hart & Sunday 2007). This was particularly evident among 95% maximum parsimony networks, and to a lesser extent maximum-likelihood trees, based on COI sequences. In contrast, both methods yielded comparable results—a high degree of haplotype group discrimination—when based on *cyt b* sequences. Because COI sequences are below the threshold at which adding more nucleotides greatly increases and stabilizes phylogenetic resolution (c. 1000–1200 nucleotides; Pollock *et al.* 2002; Roe & Sperling 2007), relying solely on COI as a biodiversity tag, at least for this

group of fishes, appears problematic. An additional critique has been that networks and trees are probabilistic methods and that diagnostic portions of a sequence, that is, one or more nucleotide-location combinations that are fixed and unique, would provide a more definitive assignment of species identity and be more consistent with historical taxonomic practice (DeSalle *et al.* 2005; Zou *et al.* 2011). Yet, until multiple individuals from throughout the geographical range of each species have been examined, character sets that are truly diagnostic will remain speculative (Elias *et al.* 2007). Finally, we regard mtDNA-based delineations of haplotype groups as evidence for particular hypotheses. Analyses with nuclear DNA are essential to confirm or refute these hypotheses and to examine the influence of ancient or modern introgression on these groups and on difficulties in morphological species identification (Nolte *et al.* 2009).

Is species identification necessary?

Strictly speaking, genetic analyses of diversity do not require taxonomic assignment of the groups that constitute components of that diversity. Yet, assignment to a taxonomic group—typically a species—has heuristic value for studies of systematics, phylogeography and ecology, as well as being a prerequisite for designation as warranting conservation attention (Turner 1999; Mace 2004). Consequently, the designation of provisional taxa is widely practised (Blaxter *et al.* 2005; April *et al.* 2011), although its legitimacy is sometimes questioned because of disagreements about defining meaningful conservation units (Valentini *et al.* 2009; Funk *et al.* 2012) or even what constitutes a species (de Queiroz 2007; Frankham *et al.* 2012). Also, these taxonomic placeholders function as an ad hoc form of DNA taxonomy, which in its fullest expression as a replacement for the Linnaean system has been roundly criticized by many in the taxonomic community (Seberg *et al.* 2003). Nevertheless, we agree with the widespread consensus that they fill a need by highlighting potentially valid taxa (Janzen *et al.* 2009; Goldstein & DeSalle 2010) and that a more formal system for contributing and recognizing this diversity is urgently needed (Maddison *et al.* 2012) to ensure that these quasitaxonomic entities have standing until integrated taxonomic practices provide a more thorough treatment of their position in the tree of life (Padiál *et al.* 2010).

Acknowledgements

We thank S. Adams and D. Schmetterling for piquing our interest in the investigation of sculpins in this region. P. Unmack, R. Collins and M. Knight provided helpful comments on an earlier version of this manuscript.

References

- April J, Mayden RL, Hanner RH, Bernatchez L (2011) Genetic calibration of species diversity among North America's freshwater fishes. *Proceedings of the National Academy of Sciences of the USA*, **108**, 10602–10607.
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, New Haven, Connecticut.
- Avise JC, Walker D (1999) Species realities and numbers in sexual vertebrates: perspective from an asexually transmitted genome. *Proceedings of the National Academy of Sciences of the USA*, **96**, 992–995.
- Bailey RM, Bond CE (1963) Four new species of freshwater sculpins, genus *Cottus*, from western North America. *Occasional Papers of the Museum of Zoology, University of Michigan*, **634**, 1–27.
- Becker S, Hanner R, Steinke D (2011) Five years of FISH-BOL: brief status report. *Mitochondrial DNA*, **22**(Suppl 1), 3–9.
- Bergsten J, Bilton DT, Fujisawa T *et al.* (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, **61**, 851–869.
- Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**, 431–452.
- Bickford D, Lohman DJ, Sodhi NS *et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**, 148–155.
- Blaxter M, Mann J, Chapman T *et al.* (2005) Defining operational taxonomic units using DNA barcode data. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **360**, 1935–1943.
- Cerutti-Pereyra F, Meekan MG, Wei NWV, O'Shea O, Bradshaw CJA (2012) Identification of rays through DNA barcoding: an application for ecologists. *PLoS ONE*, **7**, e36479.
- Cheever BM, Simon KS (2009) Seasonal influence of brook trout and mottled sculpin on lower trophic levels in an Appalachian stream. *Freshwater Biology*, **54**, 524–535.
- Chen H, Strand M, Norenburg JL *et al.* (2010) Statistical parsimony networks and species assemblages in cephalotrichid nemerteans (Nemertea). *PLoS ONE*, **5**, e12885.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.
- Collins RA, Boykin LM, Cruickshank RH, Armstrong KF (2012) Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods in Ecology and Evolution*, **3**, 457–465.
- Committee on the Status of Endangered Wildlife in Canada (COSEWIC) (2010) *COSEWIC Assessment and Status Report on the Shorthead Sculpin *Cottus confusus* in Canada*. Committee on the Status of Endangered Wildlife in Canada, Ottawa, Ontario.
- DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **360**, 1905–1916.
- 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 of London, Series B: Biological Sciences*, **274**, 2881–2889.
- Fausch KD, Torgersen CE, Baxter CV, Li HW (2002) Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience*, **52**, 483–498.
- Frankham R, Ballou JD, Dudash MR *et al.* (2012) Implications of different species concepts for conserving biodiversity. *Biological Conservation*, **153**, 25–31.
- Fregin S, Haase M, Olsson U, Alström P (2012) Pitfalls in comparisons of genetic distances: a case study of the avian family Acrocephalidae. *Molecular Phylogenetics and Evolution*, **62**, 319–328.
- Funk DJ, Omland DE (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.
- Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, **27**, 489–496.
- Goldstein PZ, DeSalle R (2010) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *BioEssays*, **33**, 135–147.
- Hammer MP, Unmack PJ, Adams M, Johnson JB, Walker KF (2010) Phylogeographic structure in the threatened Yarra pygmy perch *Nannoperca obscura* (Teleostei: Percichthyidae) has major implications for declining populations. *Conservation Genetics*, **11**, 213–223.
- Hanner R, DeSalle R, Ward RD, Kolokotronis S-O (2011) The fish Barcode of Life (FISH-BOL) special issue. *Mitochondrial DNA*, **22**(Suppl 1), 1–2.
- Harris DJ (2003) Can you bank on GenBank? *Trends in Ecology and Evolution*, **18**, 317–319.
- Hart MW, Sunday J (2007) Things fall apart: biological species form unconnected parsimony networks. *Biology Letters*, **3**, 509–512.
- Hausdorf B (2011) Progress toward a general species concept. *Evolution*, **65**, 923–931.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, 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 of London, Series B: Biological Sciences*, **270**(Suppl 1), 96–99.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptus fulgerator*. *Proceedings of the National Academy of Sciences*, **101**, 14812–14817.
- Hendricks P (1997) *Status, Distribution, and Biology of Sculpins (Cottidae) in Montana: A Review*. Montana Natural Heritage Program, Helena.
- Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? *Systematic Biology*, **52**, 124–126.
- Holton GD, Johnson HE (2003) *A Field Guide to Montana Fishes*, 3rd edn. Montana Fish, Wildlife and Parks, Helena.
- Hubbs CL, Schultz LP (1932) *Cottus tubulatus*, a new sculpin from Idaho. *Occasional Papers of the Museum of Zoology, University of Michigan*, **242**, 1–9.
- Hubert N, Hanner R, Holm E *et al.* (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE*, **3**, e2490.
- Hudy M, Shiflet J (2009) Movement and recolonization of Potomac sculpin in a Virginia stream. *North American Journal of Fisheries Management*, **29**, 196–204.
- Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**, 544–548.
- Janzen DH, Hallwachs W, Blandin P *et al.* (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources*, **8**(Suppl 1), 1–26.
- Jenkins RE, Burkhead NM (1994) *Freshwater Fishes of Virginia*. American Fisheries Society, Bethesda, Maryland.
- Joly S, Stevens MI, van Vuuren BJ (2007) Haplotype networks can be misleading in the presence of missing data. *Systematic Biology*, **56**, 857–862.
- Kershner JL, Archer EK, Coles-Ritchie M *et al.* (2004) Guide to effective monitoring of aquatic and riparian resources. U.S. Forest Service General Technical Report, RMRS-GTR-121.
- Kinziger AP, Wood RM (2003) Molecular systematics of the polytypic species *Cottus hyscelurus* (Teleostei: Cottidae). *Copeia*, **2003**, 624–627.
- Kinziger AP, Wood RM, Neely DA (2005) Molecular systematics of the genus *Cottus* (Scorpaeniformes: Cottidae). *Copeia*, **2005**, 303–311.
- Kvist S, Ocegüera-Figueroa A, Siddall ME, Eerséus C (2010) Barcoding, types and the *Hirudo* files: using information content to critically evaluate the identity of DNA barcodes. *Mitochondrial DNA*, **21**, 1998–2005.
- LaHood ES, Miller JJ, Aplan C, Ford MJ (2008) A rapid, ethanol-free fish tissue collection method for molecular genetic analyses. *Transactions of the American Fisheries Society*, **137**, 1104–1107.
- Lairke L (2010) Genetic diversity is overlooked in international conservation policy implementation. *Conservation Genetics*, **11**, 349–354.

- Lee DS, Gilbert CR, Hocutt CH, Jenkins RE, McAllister DE, Stauffer JR (1980) *Atlas of North American Freshwater Fishes*. North Carolina State Museum of Natural History, Raleigh.
- Lowenstein JH, Osmundson TW, Becker S, Hanner R, Stiassny MJ (2011) Incorporating DNA barcodes into a multi-year inventory of the fishes of the hyperdiverse Lower Congo River, with a multi-gene performance assessment of the genus *Labeo* as a case study. *Mitochondrial DNA*, **22**(Suppl 1), 52–70.
- Mace GM (2004) The role of taxonomy in species conservation. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **359**, 711–719.
- Maddison DR, Guralnick R, Hill A, Reysenback A-L, McDade LA (2012) Ramping up biodiversity discovery via online quantum contributions. *Trends in Ecology and Evolution*, **27**, 72–77.
- Maughan OE (1976) A survey of fishes of the Clearwater River. *Northwest Science*, **50**, 76–86.
- Maughan OE, Edmundson EE, Farris AE, Wallace RL (1980) A comparison of fish species above and below Palouse Falls, Palouse River, Washington-Idaho. *Northwest Science*, **54**, 5–8.
- McPhail JD (2007) *The Freshwater Fishes of British Columbia*. University of Alberta Press, Edmonton.
- Meier R, Kwong S, 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.
- Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest inter-specific distances exaggerates the size of the “barcoding gap” and leads to misidentification. *Systematic Biology*, **57**, 809–813.
- Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, 2229–2238.
- Neely DA (2003) *A Systematic and Taxonomic Reassessment of the Mottled Sculpin Species Complex, Cottus bairdii Girard (Teleostei: Cottidae)*. Doctoral dissertation, University of Alabama, Tuscaloosa.
- Nielsen R, Matz M (2006) Statistical approaches for DNA barcoding. *Systematic Biology*, **55**, 162–169.
- Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin hybrid zones suggest genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.
- Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 16.
- Page TJ, Hughes JM (2010) Comparing the performance of multiple mitochondrial genes in the analysis of Australian freshwater fishes. *Journal of Fish Biology*, **77**, 2093–2122.
- Petty JT, Grossman GD (2007) Size-dependent territoriality of mottled sculpin in a southern Appalachian stream. *Transactions of the American Fisheries Society*, **136**, 1750–1761.
- Pollock DD, Zwickl DJ, McGuire JA, Hillis DM (2002) Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology*, **51**, 664–671.
- Powell JR (2012) Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods in Ecology and Evolution*, **3**, 1–11.
- Puillandre N, Bouchet P, Boisselier-Dubayle M-C et al. (2012) New taxonomy and old collections: integrating DNA barcoding into the collection curation process. *Molecular Ecology Resources*, **12**, 396–402.
- de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.
- Raggon MF (2010) *Seasonal Variability in Diet and Consumption by Cottid and Salmonid Fishes in Headwater Streams in Western Oregon*. Master's thesis, Oregon State University, Corvallis, USA.
- Ratnasingham S, Hebert PDN (2007) BOLD: the barcoding of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.
- Roe AD, Sperling FAH (2007) Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*, **44**, 325–345.
- Ross HA, Murugan S, Li WLS (2008) Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology*, **57**, 216–230.
- Rubinoff D (2006) Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology*, **20**, 1026–1033.
- Schmidt TR, Gold JR (1993) Complete sequence of the mitochondrial cytochrome b gene in the cherryfin shiner, *Lythrurus roseipinnis* (Teleostei: Cyprinidae). *Copeia*, **1993**, 880–883.
- Seberg O, Humphries CJ, Knapp S et al. (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology and Evolution*, **18**, 63–65.
- Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589–4621.
- Srivathsan A, Meier R (2012) On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, **28**, 190–194.
- Steinke D, Hanner R (2011) The FISH-BOL collaborators' protocol. *Mitochondrial DNA*, **22**(Suppl 1), 10–14.
- Stevens DL Jr, Olsen AR (1999) Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics*, **4**, 415–428.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, **12**, 377–388.
- Teletchea F (2010) After 7 years and 1000 citations: comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA*, **21**, 206–226.
- Turner GF (1999) What is a fish species? *Reviews in Fish Biology and Fisheries*, **9**, 281–297.
- Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology and Evolution*, **24**, 110–117.
- Wang L, Infante D, Esselman P et al. (2011) A hierarchical spatial framework and database for the National River Fish Habitat Condition Assessment. *Fisheries*, **36**, 436–449.
- Ward RD (2009) DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources*, **9**, 1077–1085.
- Wiemers M, Fiedler K (2007) Does the barcoding gap exist?—A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, **4**, 8.
- Wydoski RS, Whitney RR (2003) *Inland Fishes of Washington*, 2nd edn. American Fisheries Society, Bethesda, Maryland.
- Zhang DX, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution*, **11**, 247–251.
- Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.
- Zou S, Li Q, Kong L, Yu H, Zheng X (2011) Comparing the usefulness of distance, monophyly and character-based DNA barcoding methods in species identification: a case study of Neogastropoda. *PLoS ONE*, **6**, e26619.

M.K. Young, K.S. McKelvey, and M.K. Schwartz designed the study and conducted the fieldwork. K.L. Pilgrim and M.K. Young extracted and sequenced the DNA. M.K. Young conducted the statistical analyses. M.K. Young wrote the paper with contributions from all co-authors.

Data Accessibility

DNA sequences: GenBank accessions JX282572–282599 (for COI) and JX282526–282571 (for cyt b).

Supporting Information

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

Table S1 Haplotypes and locations of *Cottus* collected in the upper Columbia and Missouri River basins

Table S2 Cytochrome *c* oxidase subunit 1 and cytochrome *b* oxidase haplotype identifiers, accession numbers, and sampling locations of sculpins in GenBank used in these analyses