Genetic calibration of species diversity among North America's freshwater fishes

Julien April[a,1], Richard L. Mayden[b], Robert H. Hanner[c], and Louis Bernatchez[a]

[a]Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec, QC, Canada G1V 0A6; [b]Department of Biology, Saint Louis University, Saint Louis, MO 63103; and [c]Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada N1G 2W1

Freshwater ecosystems are being heavily exploited and degraded by human activities all over the world, including in North America, where fishes and fisheries are strongly affected. Despite centuries of taxonomic inquiry, problems inherent to species identification continue to hamper the conservation of North American freshwater fishes. Indeed, nearly 10% of species diversity is thought to remain undescribed. To provide an independent calibration of taxonomic uncertainty and to establish a more accessible molecular identification key for its application, we generated a standard reference library of mtDNA sequences (DNA barcodes) derived from expert-identified museum specimens for 752 North American freshwater fish species. This study demonstrates that 90% of known species can be delineated using barcodes. Moreover, it reveals numerous genetic discontinuities indicative of independently evolving lineages within described species, which points to the presence of morphologically cryptic diversity. From the 752 species analyzed, our survey flagged 138 named species that represent as many as 347 candidate species, which suggests a 28% increase in species diversity. In contrast, several species of parasitic and nonparasitic lampreys lack such discontinuity and may represent alternative life history strategies within single species. Therefore, it appears that the current North American freshwater fish taxonomy at the species level significantly conceals diversity in some groups, although artificially creating diversity in others. In addition to providing an easily accessible digital identification system, this study identifies 151 fish species for which taxonomic revision is required.

DNA barcoding | cytochrome *c* oxidase I | biodiversity | evolutionarily significant units | aquatic ecosystem

Assessing the state of life in a world that faces a sixth mass extinction (1, 2) represents one of the largest challenges of modern science. This assessment is particularly needed for North American freshwater fishes. Indeed, 40% of the fauna is in peril (3), and it continues to support a multibillion dollar commercial and recreational fishery (4). Despite more than 2 centuries of descriptive taxonomic inquiry, problems inherent to species identification have hampered the study, conservation, and management of the richest diversity of temperate freshwater fish (5). Indeed, the exact amount of diversity is still unknown, and it is thought that 10% of North American freshwater fish are still formally undescribed (6, 7). Part of this problem may stem from the fact that there is no universally accepted operational species concept (8, 9). Most biologists agree that species are independently evolving lineages of populations or metapopulations (8, 9), but a debate surrounds the choice of an exact cutoff in the divergence continuum. Morphological differentiation has long been the criterion of choice for taxonomists because of the relative ease with which those characters are assessed. Unfortunately, such differences are not always caused by independent evolutionary history, and reproductively isolated taxa can sometimes be morphologically indistinguishable.

Generating rapid and accurate molecular identifications using standardized tools can help to resolve distorted views of biodiversity. Indeed, DNA barcoding (10) surveys using partial cytochrome *c* oxidase subunit I (COI) sequences have revealed cryptic diversity across the animal kingdom. For instance, previous DNA barcoding studies in other taxonomic groups have found as many as nine undescribed species embedded within a single known species of skipper butterfly (11). These numbers are even larger in less studied groups, such as parasitic hymenopterans (12). Whereas many species may need to be split into distinct evolutionary lineages, others may need to be combined, given that not all morphological differences are the result of cladogenesis. In an outstanding case documented in fishes, individuals that belonged to two different species were found to be the female and male of a single species (13).

In this study, we established a barcode reference library for more than 80% of the named freshwater fish species of North America. We used this survey of standing genetic diversity as an independent calibration of current taxonomic resolution within the North American fish fauna to reveal key areas of uncertainty where discrepancies between genetic data and morphologically based taxonomy arise. In contrast, where DNA sequences and traditional taxonomy exhibit congruence, our data serve as an accessible key for the molecular identification of North American freshwater fishes.

Results

We obtained mitochondrial barcodes for 5,674 fish specimens belonging to 50 families, 178 genera, and 752 species (Table 1). This coverage includes more than 80% of the 902 Canadian and American species listed by Nelson et al. (14). In accordance with the Fish Barcode of Life Campaign (15), we deposited all sequences and collateral specimen information within the Barcode of Life Data System (BOLD) (16), where this information can be queried by users, annotated, and curated in light of new information. This list of 752 species also includes 22 exotic invasive species (non-native species that adversely affect ecosystems), 4 species closely related to North American species, and 30 worldwide species of lampreys (17) because those species are notoriously difficult to identify and can have important ecological and economic impacts. For most species, multiple specimens (mean = 7.5 specimens per species) from distant localities (mean = 3.2 localities per species) were analyzed to document intraspecific variability. Only 62 species were represented by a single specimen, and 1 species (*Etheostoma radiosum*) was represented by 108 specimens. We observed a hierarchical increase in mean di-

Table 1. Summary of the fish taxa analyzed

Family	Species Barcoded	Indistinguishable using barcodes	With UCS (no. of UCS)
Cyprinidae	221	10	52 (121)
Percidae	197	18	45 (120)
Catostomidae	52	10	5 (10)
Centrarchidae	32	6	6 (18)
Cottidae	31	5	6 (16)
Ictaluridae	30	0	11 (29)
Salmonidae	30	7	0
Petromyzontidae	27	13	3 (8)
Fundulidae	26	2	3 (6)
Cichlidae	11	0	0
Poeciliidae	9	0	0
Acipenseridae	6	0	0
Esocidae	5	2	1 (2)
Gasterosteidae	5	0	1 (2)
Lepisosteidae	5	0	0
Clupeidae	4	2	0
Elassomatidae	4	0	2 (6)
Embiotocidae	4	0	0
Percichthyidae	4	0	0
Umbridae	4	0	0
Amblyopsidae	3	0	1 (3)
Cyprinodontidae	3	0	0
Geotriidae	3	0	0
Gobiidae	3	0	0
Osmeridae	3	0	0
Atherinopsidae	2	0	1 (2)
Characidae	2	0	0
Hiodontidae	2	0	0
Loricariidae	2	0	0
Sciaenidae	2	0	0
Achiridae	1	0	1 (2)
Amiidae	1	0	0
Anguillidae	1	0	0
Aphredoderidae	1	0	1 (2)
Ariidae	1	0	0
Belonidae	1	0	0
Channidae	1	0	0
Cobitidae	1	0	0
Dasyatidae	1	0	0
Doradidae	1	0	0
Elopidae	1	0	0
Fistulariidae	1	0	0
Gadidae	1	0	0
Lotidae	1	0	0
Myxinidae	1	0	0
Percopsidae	1	0	0
Pleuronectidae	1	0	0
Polyodontidae	1	0	0
Profundulidae	1	0	0
Sparidae	1	0	0
Total	752	75	138 (347)

This list includes the number of indistinguishable species and the number of species with UCS (represented by lineages that diverge by over 2%), along with the total number of UCS (*SI Text*).

vergence from within species (mean = 0.73%, SE = 0.053) to within congeners (mean = 13.67%, SE = 0.004), within families (mean = 15.91%, SE = 0.002), and within orders (mean = 21.24%, SE = 0.002) (Table S1). Within a genus, the mean distance to the most closely related species (nearest neighbor) was 5.71% (SE = 0.012); therefore, the mean distance to the near-

est neighbor was 7.82 times higher than the mean intraspecific divergence.

The resolution of named species using barcodes approached 90% (676 of 752) (Table 1, Table S1, and *SI Text*). Those delineated species could be identified using a diagnostic nucleotide approach (18) and were represented by a unique haplotype, a single tight cluster of haplotypes, or distinct clusters of haplotypes. Using the more traditional distance-based identification approach, we obtained a success rate of 81%. However, the inflated proportion of indistinguishable species identified using the distance-based identification approach was mainly caused by the presence of cryptic diversity within "known" species. Thus, approximately half (72 of 141) of the 19% of indistinguishable species identified using the distance-based approach were those also considered as problematic using the diagnostic nucleotide approach. Over half of the remaining problematic cases (37 of 69) involved species diverging by over 2% from any other species but were considered as indistinguishable using the distance-based approach because of an even deeper intraspecific divergence (e.g., *Nocomis leptocephalus*; Fig. 1 and Fig. S1). For *N. leptocephalus*, some of the taxa exhibiting deep intraspecific divergence values were recovered as poly/paraphyletic in phylogenetic trees; nevertheless, all lineages remained clearly distinct from any other lineages using the diagnostic nucleotide approach and always differed by over 10 characters (e.g., *Nocomis* spp.; Fig. S1). The lineages of the remaining 32 species were less differentiated (range: 0.15–1.99%) but could still be identified using diagnostic nucleotides (e.g., Fig. S2). Identification at the genus level was completely accurate (100%) using both the diagnostic nucleotide and distance-based approaches.

DNA barcode species identification success rates varied from nearly 50% in Clupeidae and Petromyzontidae to 100% in 40 families (Table 1). A literature review of data from all 75 species that cannot be delineated using a DNA barcode showed that hybridization, ancestral polymorphism sharing, and inadequate taxonomy can explain those cases of haplotype sharing. The lack of divergence observed in lampreys is particularly intriguing because it involves 13 species included in only five clusters (Fig. 2 and Table 1). Each of those clusters included at least 1 parasitic and 1 nonparasitic species of lamprey that are morphologically quite distinct.
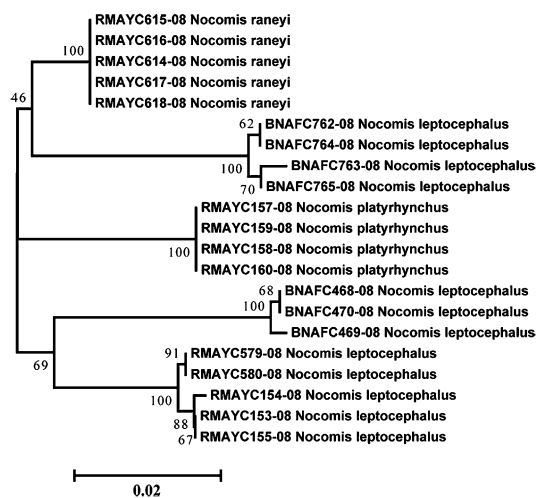

Fig. 1. Neighbor-joining tree shows the relationships among 615-bp COI sequences from three species of the genus *Nocomis*, producing five divergent clusters. The numbers at the nodes are bootstrap values based on 1,000 replications.
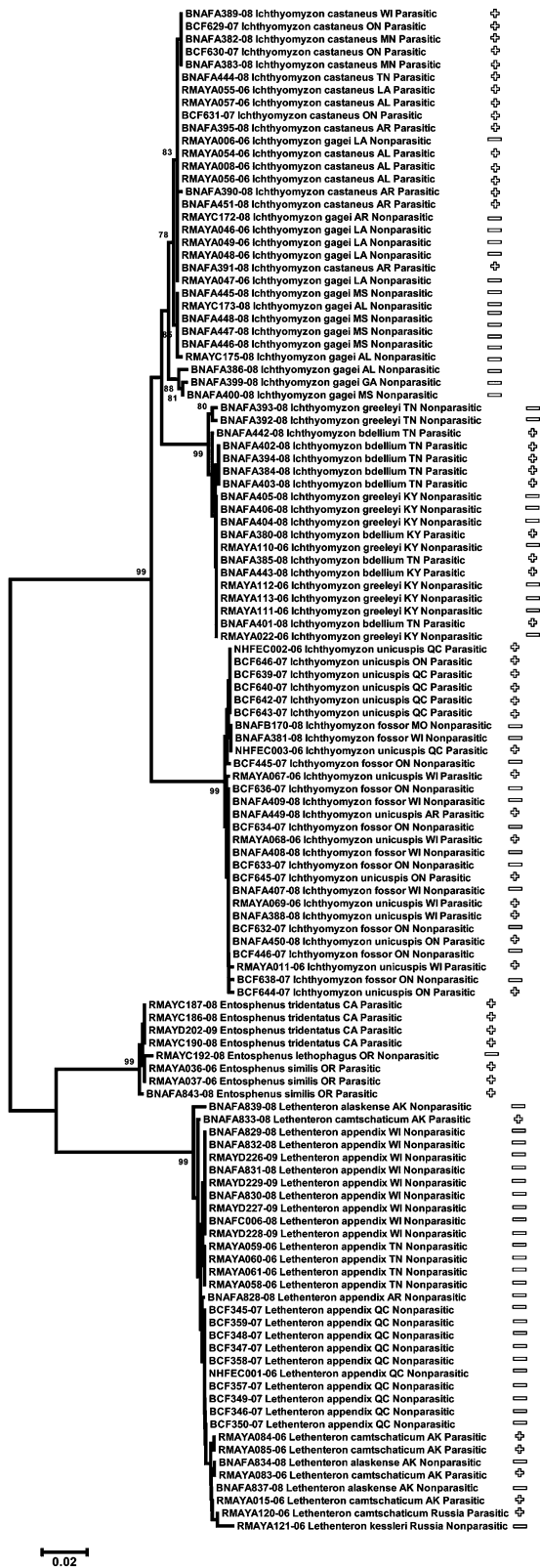
ENVIRONMENTAL SCIENCES

Fig. 2 (Neighbor-joining tree, taxon labels):

BNAFA389-08 Ichthyomyzon castaneus WI Parasitic
BCF629-07 Ichthyomyzon castaneus ON Parasitic
BNAFA382-08 Ichthyomyzon castaneus MN Parasitic
BCF630-07 Ichthyomyzon castaneus ON Parasitic
BNAFA383-08 Ichthyomyzon castaneus MN Parasitic
BNAFA444-08 Ichthyomyzon castaneus TN Parasitic
RMAYA055-06 Ichthyomyzon castaneus LA Parasitic
RMAYA057-06 Ichthyomyzon castaneus AL Parasitic
BCF631-07 Ichthyomyzon castaneus ON Parasitic
BNAFA395-08 Ichthyomyzon castaneus AR Parasitic
RMAYA006-06 Ichthyomyzon gagei LA Nonparasitic
RMAYA054-06 Ichthyomyzon castaneus AL Parasitic
RMAYA008-06 Ichthyomyzon castaneus AL Parasitic
RMAYA056-06 Ichthyomyzon castaneus AL Parasitic
BNAFA390-08 Ichthyomyzon castaneus AR Parasitic
BNAFA451-08 Ichthyomyzon castaneus AR Parasitic
RMAYC172-08 Ichthyomyzon gagei AR Nonparasitic
RMAYA046-06 Ichthyomyzon gagei LA Nonparasitic
RMAYA049-06 Ichthyomyzon gagei LA Nonparasitic
RMAYA048-06 Ichthyomyzon gagei LA Nonparasitic
BNAFA391-08 Ichthyomyzon gagei AR Nonparasitic
RMAYA047-06 Ichthyomyzon gagei LA Nonparasitic
BNAFA445-08 Ichthyomyzon gagei MS Nonparasitic
RMAYC173-08 Ichthyomyzon gagei AL Nonparasitic
BNAFA448-08 Ichthyomyzon gagei MS Nonparasitic
BNAFA447-08 Ichthyomyzon gagei MS Nonparasitic
BNAFA446-08 Ichthyomyzon gagei MS Nonparasitic
RMAYC175-08 Ichthyomyzon gagei AL Nonparasitic
BNAFA386-08 Ichthyomyzon gagei AL Nonparasitic
BNAFA399-08 Ichthyomyzon gagei GA Nonparasitic
BNAFA400-08 Ichthyomyzon gagei MS Nonparasitic
BNAFA393-08 Ichthyomyzon greeleyi TN Nonparasitic
BNAFA392-08 Ichthyomyzon greeleyi TN Nonparasitic
BNAFA442-08 Ichthyomyzon bdellium TN Parasitic
BNAFA402-08 Ichthyomyzon bdellium TN Parasitic
BNAFA394-08 Ichthyomyzon bdellium TN Parasitic
BNAFA384-08 Ichthyomyzon bdellium TN Parasitic
BNAFA403-08 Ichthyomyzon bdellium TN Parasitic
BNAFA405-08 Ichthyomyzon greeleyi KY Nonparasitic
BNAFA406-08 Ichthyomyzon greeleyi KY Nonparasitic
BNAFA404-08 Ichthyomyzon greeleyi KY Nonparasitic
BNAFA380-08 Ichthyomyzon bdellium KY Parasitic
RMAYA110-06 Ichthyomyzon greeleyi KY Nonparasitic
BNAFA385-08 Ichthyomyzon bdellium TN Parasitic
BNAFA443-08 Ichthyomyzon bdellium TN Nonparasitic
RMAYA112-06 Ichthyomyzon greeleyi KY Nonparasitic
RMAYA113-06 Ichthyomyzon greeleyi KY Nonparasitic
RMAYA111-06 Ichthyomyzon greeleyi KY Nonparasitic
BNAFA401-08 Ichthyomyzon bdellium TN Nonparasitic
RMAYA022-06 Ichthyomyzon greeleyi KY Nonparasitic
NHFEC002-06 Ichthyomyzon unicuspis QC Parasitic
BCF646-07 Ichthyomyzon unicuspis ON Parasitic
BCF639-07 Ichthyomyzon unicuspis QC Parasitic
BCF640-07 Ichthyomyzon unicuspis QC Parasitic
BCF642-07 Ichthyomyzon unicuspis QC Parasitic
BCF643-07 Ichthyomyzon unicuspis QC Parasitic
BNAFB170-08 Ichthyomyzon fossor MO Nonparasitic
BNAFA381-08 Ichthyomyzon fossor WI Nonparasitic
NHFEC003-06 Ichthyomyzon unicuspis QC Parasitic
BCF445-07 Ichthyomyzon fossor ON Nonparasitic
RMAYA067-06 Ichthyomyzon unicuspis WI Parasitic
BCF636-07 Ichthyomyzon fossor ON Nonparasitic
BNAFA409-08 Ichthyomyzon fossor WI Nonparasitic
BNAFA449-08 Ichthyomyzon unicuspis AR Parasitic
BCF634-07 Ichthyomyzon fossor ON Nonparasitic
RMAYA068-06 Ichthyomyzon unicuspis WI Parasitic
BNAFA408-08 Ichthyomyzon fossor WI Nonparasitic
BCF633-07 Ichthyomyzon fossor ON Nonparasitic
BCF645-07 Ichthyomyzon unicuspis ON Parasitic
BNAFA407-08 Ichthyomyzon fossor WI Nonparasitic
RMAYA069-06 Ichthyomyzon unicuspis WI Parasitic
BNAFA388-08 Ichthyomyzon unicuspis WI Parasitic
BCF632-07 Ichthyomyzon fossor ON Nonparasitic
BNAFA450-08 Ichthyomyzon unicuspis ON Parasitic
BCF446-07 Ichthyomyzon fossor ON Nonparasitic
RMAYA011-06 Ichthyomyzon unicuspis WI Parasitic
BCF638-07 Ichthyomyzon fossor ON Nonparasitic
BCF644-07 Ichthyomyzon unicuspis ON Parasitic
RMAYC187-08 Entosphenus tridentatus CA Parasitic
RMAYC186-08 Entosphenus tridentatus CA Parasitic
RMAYD202-09 Entosphenus tridentatus CA Parasitic
RMAYC190-08 Entosphenus tridentatus CA Parasitic
RMAYC192-08 Entosphenus lethophagus OR Nonparasitic
RMAYA036-06 Entosphenus similis OR Parasitic
RMAYA037-06 Entosphenus similis OR Parasitic
BNAFA843-08 Entosphenus similis OR Parasitic
BNAFA839-08 Lethenteron alaskense AK Nonparasitic
BNAFA833-08 Lethenteron camtschaticum AK Parasitic
BNAFA829-08 Lethenteron appendix WI Nonparasitic
BNAFA832-08 Lethenteron appendix WI Nonparasitic
RMAYD226-09 Lethenteron appendix WI Nonparasitic
BNAFA831-08 Lethenteron appendix WI Nonparasitic
RMAYD229-09 Lethenteron appendix WI Nonparasitic
BNAFA830-08 Lethenteron appendix WI Nonparasitic
RMAYD227-09 Lethenteron appendix WI Nonparasitic
BNAFC006-08 Lethenteron appendix WI Nonparasitic
RMAYD228-09 Lethenteron appendix WI Nonparasitic
RMAYA059-06 Lethenteron appendix TN Nonparasitic
RMAYA060-06 Lethenteron appendix TN Nonparasitic
RMAYA061-06 Lethenteron appendix TN Nonparasitic
RMAYA058-06 Lethenteron appendix TN Nonparasitic
BNAFA828-08 Lethenteron appendix AR Nonparasitic
BCF345-07 Lethenteron appendix QC Nonparasitic
BCF359-07 Lethenteron appendix QC Nonparasitic
BCF348-07 Lethenteron appendix QC Nonparasitic
BCF347-07 Lethenteron appendix QC Nonparasitic
BCF358-07 Lethenteron appendix QC Nonparasitic
NHFEC001-06 Lethenteron appendix QC Nonparasitic
BCF357-07 Lethenteron appendix QC Nonparasitic
BCF349-07 Lethenteron appendix QC Nonparasitic
BCF346-07 Lethenteron appendix QC Nonparasitic
BCF350-07 Lethenteron appendix QC Nonparasitic
RMAYA084-06 Lethenteron camtschaticum AK Parasitic
RMAYA085-06 Lethenteron camtschaticum AK Parasitic
BNAFA834-08 Lethenteron alaskense AK Nonparasitic
RMAYA083-06 Lethenteron camtschaticum AK Parasitic
BNAFA837-08 Lethenteron alaskense AK Nonparasitic
RMAYA015-06 Lethenteron camtschaticum AK Parasitic
RMAYA120-06 Lethenteron camtschaticum Russia Parasitic
RMAYA121-06 Lethenteron kessleri Russia Nonparasitic

0.02

Fig. 2. Neighbor-joining tree shows the relationships among 620-bp COI sequences from six parasitic and seven nonparasitic species of lampreys, producing five clusters. The numbers at the nodes are bootstrap values based on 1,000 replications. The life history characteristic [parasitic (+) or nonparasitic (−)] and origin of the samples (states/provinces) are given for each individual.

We found deeply divergent intraspecific clusters (>2%) within 138 of the 752 analyzed species (Fig. 1, Table 1, Fig. S1, and SI Text). Those divergent intraspecific clusters, which correspond to divergent evolutionary lineages, were restricted to 15 of the 50 analyzed fish families (Table 1 and SI Text). The number of lineages by species varied from 2 to 7, for a total of 347 divergent lineages among 138 named species. Deeply divergent intraspecific lineages (>2%) were almost always (88%) found in different geographical locations.

Discussion

This continent-wide genetic survey of North American freshwater fishes offers a new perspective on their diversity. Indeed, results reveal that current species-level taxonomy significantly conceals diversity in some groups, although artificially creating diversity in others (Figs. 1 and 2 and Table 1). Furthermore, DNA barcodes provide a straightforward identification system when a perfect match exists between morphology-based taxonomy and genetic divergence.

Comparison Between Taxonomy and DNA Barcodes. The DNA barcodes library provides an identification system with many applications, including the identification of fish parts or remnants, such as fish filet, sushi, smoked fish, caviar, eggs, and larvae, that are not recognizable using morphological characters (19, 20). It may facilitate tracking exotic invasive species through water samples (21) and aid food web reconstruction through gut content or fecal sample analyses (e.g., ref. 22). Barcodes are therefore highly valuable for enhancing wildlife protection from poaching and illegal trade by easing the application of different laws (e.g., Convention on International Trade in Endangered Species) and also help to protect consumers from market fraud (19).

Within North American freshwater fishes, the combined use of the distance-based identification approach with, when necessary, the diagnostic nucleotide approach appears to deliver the most reliable identifications using COI. For some of the 10% of indistinguishable species, the additional use of nuclear DNA markers (e.g., 28S, ITS1, microsatellites, amplified fragment length polymorphisms) may facilitate species-level identification. For some other indistinguishable species, it is quite possible that no genetic marker would ever allow delineation because they might not represent an isolated gene pool, and thus might not represent separate species.

With a correct species identification rate of 90%, freshwater fishes are among the groups of animals harboring the most frequent cases of interspecific haplotype sharing (Table S1): 8% in Canadian freshwater fishes, 4% in Cuban freshwater fishes, 2% in Australian marine fishes, 6% in North American birds, and 1% in Lepidoptera (23–27). The relatively elevated proportion of species sharing haplotypes in this study (10%) has four possible nonexclusive explanations: hybridization, incomplete lineage sorting, inadequate taxonomy, and erroneous identification (15). Hybridization is problematic for DNA barcode identification because mtDNA is maternally inherited; therefore, a hybrid will inevitably be diagnosed as its maternal species. Because of incomplete lineage sorting, some haplotypes can remain identical in two isolated gene pools. This situation mainly occurs when divergence is recent. Inadequate taxonomy can give the impression that two different species share barcodes if, in fact, the two "named" species are part of the same gene pool, and are therefore from the same actual species. Finally, erroneous identifications within the reference library could produce an apparent case of haplotype sharing if, in fact, specimens from the same species are wrongly given two different names by the identifier.

In this study, most cases of haplotype sharing appear to result from hybridization and inadequate taxonomy. A literature review involving all species lacking barcode divergence shows that 47% of the cases reported in this survey are corroborated by other studies that also report a lack of genetic divergence or hybrid-

ization between the same species (*SI Text*). According to available literature, hybridization is the most common explanation, consistent with the status of freshwater fishes as the group of vertebrates with the highest occurrence of hybridization (28). Here, the proportion of species sharing haplotypes with another species (10%) was identical to the classic, and still widely accepted, estimated proportion of hybridizing freshwater fishes (10%) (28). An example of a suspected case of hybridization that was determined to be of unexpectedly large scale involves two pike species: *Esox americanus* and *Esox niger* (23, 29). Nuclear markers and morphology clearly distinguish these species and also the two subspecies of *E. americanus* (29). Grande et al. (29) found a lack of mtDNA divergence between one specimen of *E. americanus americanus* and one specimen of *E. niger* but a clear differentiation between *E. americanus vermiculatus* and *E. niger*/ *E. americanus americanus* Here, our analyses of 54 specimens from 10 different states/provinces showed that the DNA barcodes from the chain pickerel (*E. niger*) and red fin pickerel (*E. americanus americanus*) are all part of the same cluster, whereas all sequences from the grass pickerel (*E. americanus vermiculatus*) belong to a distinct cluster that diverges by over 2% from any other Esocidae. The most likely explanation for this pattern is that the mitochondrial genome of *E. niger* has been completely replaced over its entire distribution range (i.e., over thousands of square kilometers) by the mtDNA from red fin pickerel (*E. americanus americanus*) through introgressive hybridization. A literature review indicates that many other nondistinguishable species are part of a "species complex" for which taxonomic boundaries are still debated, and phylogenetic studies have also revealed a lack of monophyly among recognized species. These include North American ciscoes [*Coregonus* spp., Salmonidae (23, 30)], sunfishes [*Lepomis* spp., Centrarchidae (31)] and sculpins [*Cottus bairdi* species complex, Cottidae (32)] (*SI Text*). It remains impossible to completely rule out the possibility that misidentifications explain some of the uncorroborated 5% of cases of haplotype sharing. However, this possibility is unlikely, because morphological identifications were performed by specialized taxonomists and most specimens are from museum collections. Furthermore, if there are some cases of misidentification, it is indicative of a pressing need for complementary identification tools, such as DNA barcodes. Overall, this survey highlights the failure of many recognized North American "species" to meet all the criteria of biological (33), phylogenetic (34), and phenetic (35) species concepts.

This study is among the first to show that DNA barcodes can differentiate recently radiated species. Indeed, we found that DNA barcodes allowed the identification of 94% of the species within the two most diversified and recently radiated genera of the North American freshwater fauna, *Notropis* (71 species) and *Etheostoma* (139 species). This identification success rate is even higher than the overall mean for North American freshwater fishes. It has been suggested that such an evolutionary history of adaptive radiations would illustrate inherent pitfalls on DNA barcoding, as for some recently diverged African cichlids (15) and the *Coregonus* species complex (23, 30). On the contrary, our results clearly show that DNA barcodes can correctly distinguish these often morphologically similar species in many circumstances.

Deep Intraspecific Divergence and New Candidate Species. The mean level of intraspecific divergence of 0.73% observed in North American freshwater fishes was approximately two to three times higher than for any other animal groups thoroughly surveyed with DNA barcodes (Table S1), including the following: 0.39% in Australian marine fishes, 0.23% in North American birds, and 0.43% in Lepidoptera (25–27). Such a high level of intraspecific divergence may be explained by the effect of highly restricted gene flow attributable to the fragmented nature of freshwater ecosystems. The limited dispersal capabilities of freshwater fishes

relative to flying or marine organisms (36) can, in turn, promote lineage divergence and enhanced speciation rates (37, 38). Indeed, nearly 90% of the deep intraspecific lineages recovered through this survey are allopatric. This finding reinforces the fact that such lineages have independent evolutionary histories (39) and do not reflect patterns of genetic variation suggestive of a single large population. This intraspecific diversification among lineages from different geographical regions demonstrates that individuals from some taxa can be identified not only according to species but linked to a particular watershed.

The proportion of species that possess deep intraspecific lineages (>2%) among North American freshwater fishes is high relative to the other animal groups surveyed (19%) (Table S1). The proportion of species that exceed this divergence threshold is only 5.1% for North American Lepidoptera (27), 3.2% for North American birds (26), and 2.1% for Australian marine fishes (25). This proportion of cryptic diversity within North American freshwater fishes is surprising, considering that they represent one of the best taxonomically studied groups of organisms. Perhaps, unsurprisingly, most cryptic diversity was found within taxa of minimal direct economic value.

The highest proportion of cryptic diversity was found among pygmy sunfish [Elassomatidae, an increase of 100% (4 of 4)], cave fish [Amblyopsidae, an increase of 67% (2 of 3)], and catfish [Ictaluridae, an increase of 60% (18 of 30)] (Table 1 and *SI Text*). Nonetheless, 70% of all cryptic diversity occurs in the two most diversified families, Percidae and Cyprinidae. Indeed, *Etheostoma* (family Percidae) and *Notropis* (family Cyprinidae), already the most species-rich genera (with species often distinguished by only a few subtle morphological characters), still harbor a large proportion of cryptic diversity (Table 1 and *SI Text*). For three species (*Etheostoma brevirostrum*, *Etheostoma artesiae,* and *Aphredoderus sayanus*), the maximum intraspecific divergence was over 15%, and thus closer to the level of divergence observed among genera (13.5%) and families (15.9%) than between sister species (5.7%). Those extreme cases of intraspecific divergence are unlikely to be the result of misidentification. Indeed, it is highly implausible that the highlighted diverging clades corresponded to misidentified specimens from another described species not included in this analysis, creating a false excess of intraspecific divergence. For the genus *Etheostoma,* for example, we have already included nearly all described species. Thus, we included specimens from 139 *Etheostoma* species, whereas for the same genus, Nelson et al. (14) listed a total of 131 described Canadian and American species. More recently, 142 species (98.6%) were listed in FishBase (17). For *A. sayanus*, the extreme level of intraspecific divergence is unlikely the result of misidentification because it is the sole member of the Aphredoderidae family, which is quite distinct from any other fish family. Overall, it appears that, just like for other components of biodiversity, the distribution of cryptic diversity is not uniform. Although little is known about factors that have an impact on the distribution of cryptic diversity both taxonomically and regionally, some answers might be gained from the ever-growing coverage of DNA sequence data. Because we detected a significantly higher intraspecific divergence for fishes found in the United States (mean = 0.73, SE = 0.05) compared with those from Canada (mean = 0.27, SE = 0.01), as already suggested (37), a promising avenue may be to investigate further a possible relationship between latitude and cryptic diversity.

This exhaustive genetic survey of North American freshwater fishes revealed a significant amount of previously unrecognized cryptic diversity. By comparing our results with studies that have previously reported cases of deep (>2%) intraspecies divergence, we conclude that our study highlights some 87 new taxa. This represents an increase of 42% relative to what has been previously reported over more than 20 y of phylogeographical research. This estimate is likely conservative, because for 36 species analyzed at two to three sites in the literature, it is likely that our study, which detected divergent lineages within those species,

ENVIRONMENTAL SCIENCES

flagged different, and therefore new, lineages. Finally, we also found divergent lineages that have not been detected before in 8 species for which phylogeographical studies have been conducted, reflecting the failure of many published phylogeographical studies to cover the entire species range.

The genetically dissimilar taxa flagged in this study may represent new species. For the time being, however, we view those lineages as unconfirmed candidate species (UCS) (9). Our calibration suggests that taxonomic revisionary work is warranted for those UCS, with an emphasis on watersheds where genetic variants were detected. We first suggest a careful reexamination of the morphological variation within those described species that possibly harbor cryptic species, because, at this point, it remains possible that some of the newly identified lineages possess some slight morphological differences that have simply been overlooked. The reproductive biology and ecology of such cryptic lineages also require investigation. However, because species are lost at an alarming rate and looking for reproductive isolation is time-consuming, the precautionary principle suggests that the lineages highlighted here should be considered evolutionarily significant units that need to be taken into account in conservation strategies (40, 41). Despite uncertainties surrounding mtDNA molecular clock calibration, deeply divergent mtDNA clusters (e.g., >2%) are likely indicative of a million years or more of unique evolutionary legacy (39, 42). This represents a relatively long evolutionary time period compared with the mean time between origination and extinction, which is estimated to be 2.5 million years for mammals (43), the vertebrate group with the best estimates of species life span. Conserving as many "genetic building blocks" (*sensu* ref. 44) as possible will hopefully enhance the capacity of biodiversity to evolve and adapt to an ever-changing environment and reduce these species' chances of extinction.

Lack of Interspecific Divergence and Distinct Species That May Be Single Species. At the other end of the spectrum were the cases where different species were found to form a single genetic cluster, and probably a single evolutionary lineage. This situation seems particularly recurrent in lampreys. Indeed, there are 13 recognized species included in only five clusters, and each one of those clusters includes at least 1 parasitic and 1 nonparasitic species of lamprey (Fig. 2 and Table 1). Interestingly, it has been known for decades that most lamprey genera harbor paired species with morphologically similar larvae (45, 46). Nevertheless, following metamorphosis, adults differ both morphologically and ecologically and are either parasitic to other fish or nonparasitic (i.e., do not eat anything at all). Adults of nonparasitic species do not migrate, are smaller, and have less developed teeth than parasitic lampreys (46). Although the current taxonomy considers all parasitic and nonparasitic lampreys as distinct species, our results, along with those of several previous studies (46–53), stress the need to revisit lamprey taxonomy.

The lack of significant genetic divergence between parasitic and nonparasitic lampreys appears to be a ubiquitous worldwide pattern. Our analysis, based on 174 specimens representing 30 lamprey species that were characterized by ~650 bp of mtDNA, documents new relationships among lampreys and further corroborates results of other studies. In the most exhaustive previous study, 46 specimens from 23 species had been characterized by 384 bp of mtDNA (48). Lack of divergence between two pairs of lamprey has also been observed at allozymes [*Lampetra fluviatilis/Lampetra planeri* (49) and *Lampetra ayresii /Lampetra richardsoni* (50)]. The relationship between *L. fluviatilis* and *L. planeri* was previously analyzed using mtDNA (47, 49). Okada et al. (51) and Yamazaki et al. (52) also investigated the relationships between the nonparasitic *Lethenteron kessleri* and two other nonparasitic species using mtDNA. Our analysis complements that study by showing that this clade also comprises two other nonparasitic species (*Lampetra alas-*

kense and *Lampetra appendix*) and the parasitic species *Lampetra camtschaticum* (Fig. 2).

Recurrent lack of genetic differentiation suggests that many genera of lamprey possess an intrinsic capacity to develop two alternative morphotypes that correspond to parasitic and non-parasitic life history strategies. Therefore, it is plausible that when facing different environmental conditions, larvae from several taxa may adopt one of two trophic tactics, each with its own costs and benefits, in a similar fashion to the winged or nonwinged morphs in pea aphids (54). This hypothesis is also supported by rare observations of plasticity in feeding type for some populations of lampreys (53) and the distribution overlap between parasitic and nonparasitic species (46). If the large anadromous parasitic lampreys and the small toothless non-parasitic lampreys effectively represent alternative life history tactics within the same gene pool, this places lampreys among the most phenotypically plastic vertebrates known.

Conclusion

Overall, this study demonstrated the ability of DNA barcoding to help calibrate current taxonomic resolution and shed new light on the biodiversity of North American fishes. Indeed, although differently described species might, in fact, represent single evolutionary lineages, as in the case of lampreys, as much as 28% of the American and Canadian freshwater fish species could be waiting for a formal taxonomic description. This finding reinforces the status of North American fresh waters as harboring some of the world's greatest fish diversity and deserving more important conservation efforts (6). Further DNA barcoding surveys will reveal whether the extremely high proportion of cryptic evolutionary lineages detected in this study is a common characteristic of freshwater organisms. Finally, our results stress the need for more taxonomic research, because it appears that even for economically important vertebrates that have benefited from over a century of scientific inquiry, additional work is required to create a more accurate picture of species diversity.

Materials and Methods

For each of the 902 freshwater fish species occurring north of Mexico and recorded by Nelson et al. (14), our objective was to sample five individuals from each main watershed within the species range. More individuals were analyzed when many specimens were readily available from the museum collection. Despite our effort to get a representative sampling design, because of our somewhat opportunistic approach and the rarity of some species, it was not possible to achieve a randomly structured sampling design with respect to biogeographical regions and drainages. Samples and DNA sequence acquisition, as well as the analyses of genetic distance and phylogenetic trees, are described in *SI Text*. Specimen and sequence data are accessible in the BOLD projects "Freshwater fishes of North America and Barcoding of Canadian freshwater fishes," and sequences have also been submitted to GenBank.

The diagnostic nucleotide identification method assumes that a described species can be correctly identified using DNA barcodes if all reference specimens morphologically identified as this species by taxonomists possess one or more unique and nonhomoplasic (i.e., diagnostic) nucleotides relative to the other species (18). The distance-based approach assumes that a species can be correctly identified when the mean distance to the most closely related species (nearest neighbor) is higher than the maximum intraspecific distance. The latter approach offers more stringent criteria because all analyzed specimens of the species of interest not only need to have private haplotypes with a diagnostic nucleotide (as in the former approach) but need to appear monophyletic in a phylogenetic tree, because the level of divergence between species needs to be higher than the intraspecific variability. Even if neither of the two approaches is absolutely impervious to the possibility that an unsampled population presents a case of haplotype sharing, other DNA barcoding studies (26, 27) and some of the species analyzed here using a nearly complete geographical coverage have shown that this situation is certainly rare. Furthermore, both the number of samples and the number of localities by species remain relatively high compared with other large-scale DNA barcoding projects (Table S1).

A threshold of 2% of maximum divergence was applied to refer to cryptic diversity. This level of divergence is commonly observed between distinct

vertebrate species (55) and has also been used as a standard in recent DNA barcoding projects (27). It is noteworthy that a lower threshold would allow recovery of more cryptic diversity but could also erroneously inflate species diversity before further investigations are performed on the candidate species. In contrast, a higher threshold would give more conservative estimates of diversity but might also miss many lineages that should be described as species. The number of divergent lineages within recognized species was calculated as the number of haplotypes, or clusters of haplotypes, with a mean divergence of over 2% from any other haplotypes or clusters of haplotypes.

1. Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of Earth's ecosystems. *Science* 277:494–499.
2. Sala OE, et al. (2000) Global biodiversity scenarios for the year 2100. *Science* 287:1770–1774.
3. Jelks HL, et al. (2008) Conservation status of imperiled North American freshwater and diadromous fishes. *Fisheries (Bethesda, Md)* 33:372–407.
4. FAO (2009) *The State of World Fisheries and Aquaculture 2008* (Fisheries and Aquaculture Department, Food and Agriculture Organization of the United Nations, Rome).
5. Abell RA, et al. (2000) *Freshwater Ecoregions of North America: A Conservation Assessment* (Island Press, Washington, DC).
6. Warren ML, et al. (2000) Diversity, distribution, and conservation status of the native freshwater fishes of the southern United States. *Fisheries (Bethesda, Md)* 25:7–31.
7. Butler RS, Mayden RL (2003) Cryptic biodiversity. *Endangered Species Bulletin* 28:24–26.
8. Mayden RL (1997) *Species, the Units of Biodiversity*, eds Claridge MF, Dawah HA, Wilson MR (Chapman & Hall, London), pp 381–424.
9. Pedial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Front Zool* 7:16.
10. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proc Biol Sci* 270:313–321.
11. 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 Astraptes fulgerator. *Proc Natl Acad Sci USA* 101:14812–14817.
12. Smith MA, et al. (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proc Natl Acad Sci USA* 105:12359–12364.
13. Byrkjedal I, Rees DJ, Willassen E (2007) Lumping lumpsuckers: Molecular and morphological insights into the taxonomic status of Eumicrotremus spinosus (Fabricius, 1776) and E. eggvinii Koefoed, 1956 (Teleostei: Cyclopteridae). *J Fish Biol* 71:111–131.
14. Nelson JS, et al. (2004) *Common and Scientific Names of Fishes from the United States, Canada, and Mexico* (American Fisheries Society, Bethesda, MD)Special Publication 29, 6th Ed.
15. Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. *J Fish Biol* 74:329–356.
16. Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol Ecol Notes* 7:355–364.
17. Froese R, Pauly D (2010) FishBase. Available at www.fishbase.org, version (07/2010). Accessed August 15, 2010.
18. Wong EHK, Shivji MS, Hanner RH (2009) Identifying sharks with DNA barcodes: Assessing the utility of a nucleotide diagnostic approach. *Mol Ecol Resour* 9:243–256.
19. Wong EHK, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood. *Food Res Int* 41:828–837.
20. Victor BC, Hanner R, Shivji M, Hyde J, Caldow C (2009) Identification of the larval and juvenile stages of the Cubera Snapper, Lutjanus cyanopterus, using DNA barcoding. *Zootaxa* 2215:24–36.
21. Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biol Lett* 4:423–425.
22. Kaartinen R, Stone GN, Hearn J, Lohse K, Roslin T (2010) Revealing secret liaisons: DNA barcoding changes our understanding of food webs. *Ecol Entomol* 35:623–638.
23. Hubert N, et al. (2008) Identifying Canadian Freshwater Fishes through DNA Barcodes. *PLoS ONE* 3:e2490.
24. Lara A, et al. (2010) DNA barcoding of Cuban freshwater fishes: Evidence for cryptic species and taxonomic conflicts. *Mol Ecol Resour* 10:421–430.
25. Ward RD, Zemlak TS, Innes BH, Last PR, et al. (2005) DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci* 360:1847–1857.
26. Kerr KCR, et al. (2007) Comprehensive DNA barcode coverage of North American birds. *Mol Ecol Notes* 7:535–543.
27. Hebert PDN, Dewaard JR, Landry JF (2010) DNA barcodes for 1/1000 of the animal kingdom. *Biol Lett* 6:359–362.

28. Hubbs CL (1955) Hybridization between fish species in nature. *Syst Zool* 4:1–20.
29. Grande T, Laten H, Lopez JA (2004) Phylogenetic relationships of extant esocid species (Teleostei: Salmoniformes) based on mophological and molecular characters. *Copeia* 4:743–757.
30. Turgeon J, Estoup A, Bernatchez L (1999) Species flock in the North American Great Lakes: Molecular ecology of Lake Nipigon Ciscoes (Teleostei: Coregonidae: Coregonus). *Evolution* 53:1857–1871.
31. Harris PM, Roe KJ, Mayden RL (2005) A mitochondrial DNA perspective on the molecular systematics of the sunfish genus Lepomis (Actinopterygii: Centrarchidae). *Copeia* 2:340–346.
32. Kinziger AP, Raesly RL, Neely DA (2000) New species of Cottus (Teleostei: Cottidae) from the middle Atlantic eastern United States. *Copeia* 4:1007–1018.
33. Mayr E (1942) *Systematics and the Origin of Species* (Columbia Univ Press, New York).
34. Cracraft J (1983) Species concepts and speciation analysis. *Curr Ornithol* 1:159–187.
35. Sokal RR, Crovello TJ (1970) The biological species concept: A critical evaluation. *Am Nat* 104:127–153.
36. Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater and anadromous fishes. *J Fish Biol* 44:213–232.
37. Bernatchez L, Wilson CC (1998) Comparative phylogeography of nearctic and palearctic fishes. *Mol Ecol* 7:431–452.
38. Coyne JA, Orr HA (2004) *Speciation* (Sinauer Associates, Sunderland, MA).
39. Avise JC (2000) *Phylogeography: The History and Formation of Species* (Harvard Univ Press, Cambridge, MA).
40. Waples RS (1991) Pacific salmon, Oncorhynchus spp., and the definition of "species" under the Endangered Species Act. *Mar Fish Rev* 53:11–22.
41. Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: Towards a unified concept for defining conservation units. *Mol Ecol* 10:2741–2752.
42. Brown WM, George M, Jr., Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 76:1967–1971.
43. Alroy J (2000) New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* 26:707–733.
44. Waples RS (1995) Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *Evolution and the Aquatic Ecosystem: Defining Unique Units in Population Conservation* 17:8–27.
45. Zanandrea G (1959) Speciation among lampreys. *Nature* 184:380.
46. Docker MF (2009) Biology, management, and conservation of lampreys in North America. *Am Fish Soc Symp* 72:71–114.
47. Espanhol R, Almeida PR, Alves MJ (2007) Evolutionary history of lamprey paired species Lampetra fluviatilis (L.) and Lampetra planeri (Bloch) as inferred from mitochondrial DNA variation. *Mol Ecol* 16:1909–1924.
48. Blank M, Jurss K, Bastrop R (2008) A mitochondrial multigene approach contributing to the systematics of the brook and river lampreys and the phylogenetic position of Eudontomyzon mariae. *Can J Fish Aquat Sci* 65:2780–2790.
49. Schreiber A, Engelhorn R (1998) Population genetics of a cyclostome species pair, river lamprey (*Lampetra fluviatilis* L.) and brook lamprey (*Lampetra planeri* Bloch). *J Zoological Syst Evol Res* 36:85–99.
50. Beamish RJ, Withler RE (1986) *Indo-Pacific Fish Biology: Proceedings of the Second International Conference on Indo-Pacific Fishes*, eds Uyeno T, Arai R, Taniuchi T, Matsuura K (Ichthyological Society of Japan, Tokyo, Japan), pp 31–49.
51. Okada K, Yamazaki Y, Yokobori S, Wada H (2010) Repetitive sequences in the lamprey mitochondrial DNA control region and speciation of Lethenteron. *Gene* 465:45–52.
52. Yamazaki Y, Yokoyama R, Nishida M, Goto A (2006) Taxonomy and molecular phylogeny of Lethenteron lampreys in eastern Eurasia. *J Fish Biol* 68:251–269.
53. Cochran PA (2008) Observations on giant American brook lampreys (Lampetra appendix). *J Freshwat Ecol* 23:161–164.
54. Podjasek JO, Bosnjak LM, Brooker DJ, Mondor EB (2005) Alarm pheromone induces a transgenerational wing polyphenism in the pea aphid, *Acyrthosiphon pisum. Can J Zool* 83:1138–1141.
55. Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proc Biol Sci* 265:1707–1712.

ENVIRONMENTAL SCIENCES