Supporting Information

April et al. 10.1073/pnas.1016437108

SI Text

List of Species Pairs Exhibiting Shared Barcodes. References are given when other studies have found similar results for the same species.

Percidae. *Ammocrypta beanie/Ammocrypta bifascia*, *Etheostoma aquali/Etheostoma cinereum*, *Etheostoma caeruleum/Etheostoma uniporum* (1, 2), *Etheostoma wapiti/Etheostoma camurum*, *Etheostoma cervus/Etheostoma pyrrhogaster*, *Etheostoma percnurum/Etheostoma sitikuense*, *Etheostoma simoterum/Etheostoma tennesseense*, *Percina caprodes/Percina fulvitaenia*, *Percina tanasi/Percina uranidea*

Centrarchidae. *Lepomis cyanellus/Lepomis symmetricus/Lepomis marginatus/Lepomis megalotis* (3), *Micropterus punctulatus/Micropterus treculii*

Cyprinidae. *Campostoma anomalum/Campostoma pullum* (4), *Cyprinella callisema/Cyprinella garmani*, *Lythrurus fumeus/Lythrurus roseipinnis* (5, 6), *Notropis amabilis/Notropis jemezanus*, *Notropis buchanani/Notropis volucellus*

Catostomidae. *Carpiodes carpio/Carpiodes cyprinus/Carpiodes velifer*, *Catostomus microps/Catostomus occidentalis*, *Ictiobus bubalus/Ictiobus cyprinellus/Ictiobus niger*, *Moxostoma macrolepidotum/Moxostoma pisolabrum*

Fundulidae. *Fundulus grandis/Fundulus jenkinsi*

Clupeidae. *Alosa pseudoharengus/Alosa aestivalis* (7)

Esocidae. *Esox americanus/Esox niger* (7, 8)

Cottidae. *Cottus bairdii/Cottus chattahoochee/Cottus cognatus/Cottus hubbsi/Cottus tallapoosae* (7, 9, 10)

Salmonidae. *Coregonus artedi/Coregonus autumnalis/Coregonus hoyi/Coregonus kiyi*, *Coregonus laurettae/Coregonus nigripinnis/Coregonus zenithicus* (11)

Petromyzontidae. *Ichthyomyzon bdellium/Ichthyomyzon greeleyi* (12), *Ichthyomyzon castaneus/Ichthyomyzon gagei* (12), *Ichthyomyzon fossor/Ichthyomyzon unicuspis* (7, 12), *Entosphenus tridentatus/Entosphenus lethophagus/Entosphenus similis* (12), *Lampetra appendix/Lampetra alaskense/Lampetra camtschaticum/Lampetra kessleri* (12)

List of Species with Dissimilar Lineages (>2%) and Numbers of Lineages Within Species. ***Percidae.*** *Ammocrypta clara* (2), *Ammocrypta vivax* (2), *Crystallaria asprella* (2), *Etheostoma artesiae* (5), *Etheostoma blennioides* (3), *Etheostoma brevirostrum* (2), *E. caeruleum* (2), *E. camurum* (2), *Etheostoma chlorosomum* (4), *Etheostoma douglasi* (2), *Etheostoma edwini* (3), *Etheostoma flabellare* (7), *Etheostoma flavum* (3), *Etheostoma fricksium* (2), *Etheostoma fusiforme* (3), *Etheostoma gracile* (3), *Etheostoma jessiae* (2), *Etheostoma kennicotti* (3), *Etheostoma lawrencei* (2), *Etheostoma lepidum* (2), *Etheostoma lynceum* (4), *Etheostoma microperca* (2), *Etheostoma nigrum* (3), *Etheostoma olmstedi* (2), *Etheostoma parvipinne* (2), *Etheostoma proeliare* (3), *Etheostoma punctulatum* (3), *Etheostoma radiosum* (5), *Etheostoma smithi* (2), *Etheostoma spectabile* (4), *Etheostoma stigmaeum* (3), *Etheostoma swaini* (4), *E. uniporum* (2), *Etheostoma whipplei* (2), *Etheostoma zonale* (4), *Gymnocephalus cernuus* (2-exotic), *P. caprodes* (2), *Percina carbonaria* (2), *Percina evides* (2), *Percina nigrofasciata* (2), *Percina palmaris* (2), *Percina phoxocephala* (2), *Percina sciera* (2), *Percina shumardi* (2), *Percina vigil* (2)

Cyprinidae. *C. anomalum* (5), *Campostoma ornatum* (2), *C. pullum* (2), *Clinostomus funduloides* (2), *Cyprinella analostana* (2), *Cyprinella leedsi* (2), *Cyprinella venusta* (3), *Erimystax x-punctatus* (2), *Hybognathus placitus* (2), *Hybopsis amblops* (3), *Hybopsis rubrifrons* (2), *Hybopsis winchelli* (3), *Hybopsis zanema* (2), *Luxilus albeolus* (2), *Luxilus chrysocephalus* (2), *Luxilus coccogenis* (2), *Luxilus cornutus* (2), *L. fumeus* (2), *L. roseipinnis* (4), *Macrhybopsis aestivalis* (2), *Macrhybopsis storeriana* (2), *Margariscus margarita* (2), *Nocomis leptocephalus* (3), *Nocomis micropogon* (2), *Notemigonus crysoleucas* (2), *Notropis asperifrons* (2), *Notropis atherinoides* (2), *Notropis baileyi* (2), *Notropis boops* (3), *Notropis buccatus* (3), *Notropis chalybaeus* (2), *Notropis cummingsae* (3), *Notropis harperi* (2), *Notropis hudsonius* (3), *Notropis leuciodus* (2), *Notropis longirostris* (3), *Notropis lutipinnis* (3), *Notropis maculatus* (2), *Notropis nubilus* (2), *Notropis procne* (2), *Notropis rubellus* (2), *Notropis rubricroceus* (2), *Notropis skepticus* (2), *Notropis topeka* (2), *N. volucellus* (2), *Opsopoeodus emiliae* (2), *Pimephales notatus* (3), *Pimephales vigilax* (2), *Pteronotropis metallicus* (2), *Rhinichthys cataractae* (3), *Rhinichthys obtusus* (2), *Rhinichthys osculus* (2)

Catostomidae. *C. cyprinus* (2), *Erimyzon oblongus* (2), *Erimyzon sucetta* (2), *Hypentelium nigricans* (2), *Minytrema melanops* (2)

Ictaluridae. *Ameiurus platycephalus* (2), *Ameiurus brunneus* (2), *Noturus albater* (2), *Notorus hildebrandi* (2), *Noturus miurus* (2), *Noturus exilis* (4), *Noturus insignis* (2), *Noturus elegans* (2), *Noturus leptacanthus* (2), *Noturus flavus* (6), *Noturus phaeus* (3)

Cottidae. *Cottus rhotheus* (2), *C. cognatus* (2), *C. bairdii* (5), *Cottus carolinae* (2), *Cottus beldingii* (3), *Cottus hypselurus* (2)

Petromyzontidae. Eudontomyzon mariae (2), *Lampetra aepyptera* (4), *Lampetra planeri* (2)

Centrarchidae. *L. cyanellus* (2), *Lepomis gulosus* (2), *Lepomis macrochirus* (2), *L. marginatus* (5), *L. megalotis* (5), *Lepomis microlophus* (2)

Fundulidae. *Fundulus diaphanus* (2), *Fundulus catenatus* (2), *Leptolucania ommata* (2)

Elassomatidae. *Elassoma evergladei* (2), *Elassoma zonatum* (4)

Esocidae. *E. americanus* (2)

Gasterosteidae. *Culaea inconstans* (2)

Amblyopsidae. *Chologaster cornuta* (3)

Aphredoderidae. *Aphredoderus sayanus* (2)

Atherinopsidae. *Labidesthes sicculus* (2)

Achiridae. *Trinectes maculatus* (2)

SI Materials and Methods

Data Acquisition. Samples were gathered from different sources, mainly museums, universities, and government agencies [data accessible through the BOLD (13)]. Data from 190 Canadian species (1,360 individuals), published by Hubert et al. (7), are included in these analyses. In total, specimens from 752 species were obtained, which represents 83% of all Canadian and American freshwater fishes. The species-level identification was based on morphological characteristics according to the current literature and was conducted by taxonomists specialized in this fauna. Vouchers from 80% of the analyzed species are preserved in a permanent collection (see below).

Genetic material was obtained from either frozen or ethanol-preserved tissues (either muscle or fin). DNA extraction was performed using an automated glass fiber protocol (14). The DNA barcode region for animals, 652 bp of the COI mitochondrial gene, was amplified using the primer mixtures C_FishF1t1 and C_FishR1t1 from Ivanova et al. (15). The PCR thermal conditions were the following: 2 min at 95 °C; 35 cycles of 0.5 min at 94 °C, 0.5 min at 52 °C, and 1 min at 72 °C; 10 min at 72 °C; and held at 4 °C. The 12.5-μL PCR reactions included 6.25 μL of 10% (vol/vol) trehalose, 2.00 μL of water, and 1.25 μL of 10× PCR buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl, 0.625 μL of $MgCl_2$ [50 mM], 0.125 μL of each primer mixture [0.01 mM], 0.062 μL of each dNTP [10 mM], 0.060 μL of Platinum Taq

Polymerase (Invitrogen), and 2.0 μL of DNA template]. The amplified product was visualized on a 1.2% (wt/vol) agarose E-Gel (Invitrogen). Bidirectional sequencing was performed using the M13F and M13R primers and the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) on an ABI 3730 capillary sequencer (Applied Biosystems, Inc.) following the manufacturer's instructions.

All recovered sequences were over 500 bp long (mean = 648 bp), and less than 1% of the reads were ambiguous. Because they were long for typical mtDNA inserts in the nuclear genome (nuclear mitochondrial DNA) and showed no insertions, deletions, or stop codons, all sequences appeared to be the expected functional mitochondrial loci COI as translated and examined for common contaminants using the BOLD (13).

COI sequences were submitted to two different open access servers: the BOLD (www.barcodinglife.org, 1) and GenBank (accession nos. EU522398–EU522464, EU523870–EU525162, HQ556931, HQ556937–HQ556979, HQ556989–HQ556990, HQ557037–HQ557038, HQ557067–HQ557069, HQ557071–HQ557076, HQ557086–HQ557089, HQ557095–HQ557097, HQ557114, HQ557121–HQ557132, HQ557136–HQ557222, HQ557262–HQ557272, HQ557285–HQ557286, HQ557301–HQ557365, HQ557375–HQ557395, HQ557397–HQ557464, HQ557467–HQ557471, HQ557475, HQ557489, HQ557495–HQ557497, HQ557524–HQ557555, HQ557720–HQ557733, HQ579002–HQ579067, HQ579071–HQ579136, HQ937011–HQ937054, HQ971430–HQ971434, and JN024710–JN028456). The data associated with each specimen (taxonomy, collection sites, and voucher accession numbers) and trace files associated with the sequenced specimens are available in the BOLD container "Freshwater fishes of North America."

Analysis. Sequence alignment was performed using the BOLD Management and Analysis System (13) and ClustalX software (16). Genetic distances [Kimura two-parameter (17)] and the neighbor-joining tree (17) were calculated with the BOLD Management and Analysis System (13). Further phylogenetic analysis was performed on some species of *Nocomis* and lampreys because they represent, respectively, examples of cryptic diversity and distinct recognized species that potentially represent single evolutionary lineages with important morphological and ecological polymorphisms. For those two groups, we performed neighbor-joining (17) analyses based on the Kimura two-parameter model (18) using Mega 4 (19). Branch support was estimated by bootstrapping with 1,000 replicates. Two cases in which identification failed with the distance-based approach but succeeded using a nucleotide diagnostic approach are represented with a haplotype network (*Nocomis* spp. in Fig. S1 and *Lythrurus* spp. in Fig. S2). The minimum spanning trees were obtained using Arlequin software (20) with pairwise distances.

1. Ray JM, Lang NJ, Wood RM, Mayden RL (2008) History repeated: Recent and historical mitochondrial introgression between the current darter Etheostoma uniporum and rainbow darter Etheostoma caeruleum (Teleostei: Percidae). *J Fish Biol* 72:418–434.
2. Bossu CM, Near TJ (2009) Gene trees reveal repeated instances of mitochondrial DNA introgression in orangethroat darters (percidae: etheostoma). *Syst Biol* 58:114–129.
3. Harris PM, Roe KJ, Mayden RL (2005) A mitochondrial DNA perspective on the molecular systematics of the sunfish genus Lepomis (Actinopterygii: Centrarchidae). *Copeia* 340–346.
4. Blum MJ, Neely DA, Harris PM, Mayden RL (2008) Molecular systematics of the cyprinid genus Campostoma (Actinopterygii: Cypriniformes): Disassociation between morphological and mitrochondrial differentiation. *Copeia* 2:360–369.
5. Snelson FFJ (1972) Systematics of the subgenus Lythrurus genus Nortopis Pisces Cyprinidae. *Bull Fla State Mus Biol Sci* 17:1–92.
6. Pramuk JB, et al. (2007) Phylogeny of finescale shiners of the genus Lythrurus (Cypriniformes: Cyprinidae) inferred from four mitochondrial genes. *Mol Phylogenet Evol* 42:287–297.
7. Hubert N, et al. (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE* 3:e2490.
8. Grande T, Laten H, Lopez JA (2004) Phylogenetic relationships of extant esocid species (Teleostei: Salmoniformes) based on morphological and molecular characters. *Copeia* 743–757.
9. Kinziger AP, Raesly RL, Neely DA (2000) New species of Cottus (Teleostei: Cottidae) from the middle Atlantic eastern United States. *Copeia* 1007–1018.
10. Neely DA, Williams JD, Mayden RL (2007) Two new sculpins of the genus Cottus (Teleostei: Cottidae) from rivers of eastern North America. *Copeia* 3:641–655.
11. 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.
12. Docker MF (2009) A review of the evolution of nonparasitism in lampreys and an update of the paired species concept. *Biology, Management, and Conservation of Lampreys in North America* 72:71–114.
13. Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol Ecol Notes* 7:355–364.
14. Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Mol Ecol Notes* 6:998–1002.
15. Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Mol Ecol Notes* 7:544–548.
16. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882.
17. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120.
18. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
19. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
20. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.

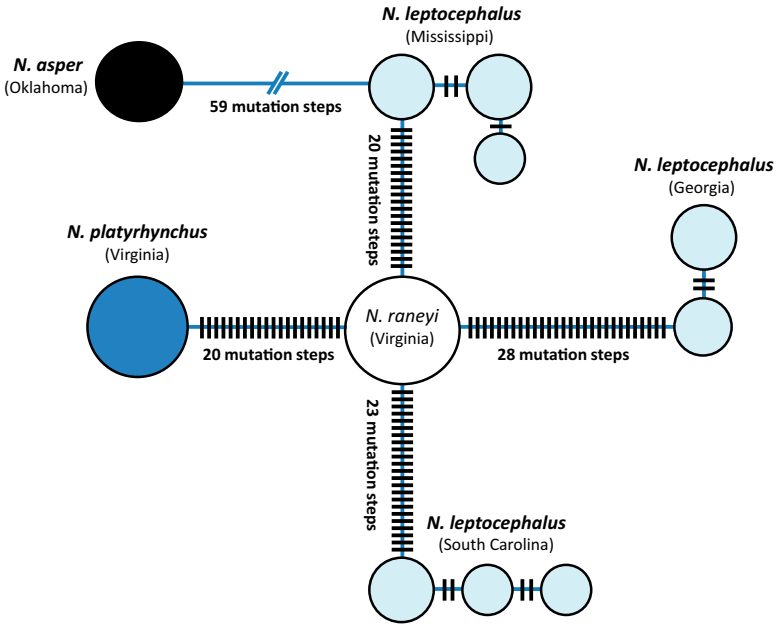

Fig. S1. Statistical parsimony network that connects the haplotypes documented in *Nocomis leptocephalus*, *Nocomis raneyi*, *Nocomis platyrhynchus*, and *Nocomis asper*. Haplotypes are represented by circles, which have a size proportional to the number of analyzed specimens with this haplotype. Black bars represent the number of mutation steps between haplotypes.

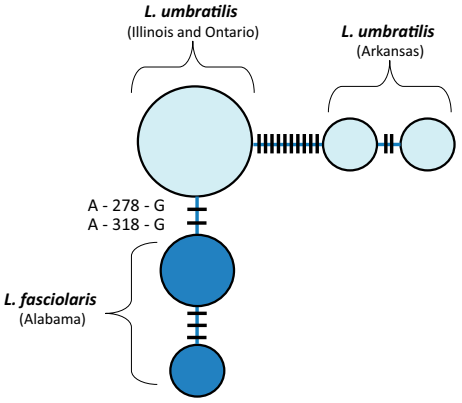

Fig. S2. Statistical parsimony network connecting the haplotypes documented in *Lythrurus umbralitis* and *Lythrurus fasciolaris*. Haplotypes are represented by circles, which have a size proportional to the number of analyzed specimens with this haplotype. Black bars represent the number of mutational steps between haplotypes. Here, two diagnostic nucleotides distinguish the two species: a mutational change of A in *L. umbralitis* to G in *L. fasciolaris* at positions 278 and 318.

Table S1. Summary statistics that describe the data and results from this study, along with information obtained from some of the most comprehensive DNA barcoding surveys

Group	Species, *n*	Individuals/ species, Mean	Species sharing barcodes, %	Intragenus divergence		Intraspecific divergence		
				Mean	SE	Mean	SE	% of species with UCS
North American freshwater fish (this study)	751	7.5	10	5.71	0.186	0.73	0.053	19
Australian marine fish (1)	207	3.7	0	9.93	0.096	0.39	0.031	2.1
North American birds (2)	643	4.1	6	5.54	0.021	0.23	0.01	2
North American Lepidoptera (3)	1,327	8.5	1	7.7	0.033	0.43	0.017	5.1

UCS refers to intraspecific lineages that diverge by 2% or more from any other taxa.

1. 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.
2. Kerr KCR, et al. (2007) Comprehensive DNA barcode coverage of North American birds. *Mol Ecol Notes* 7:535–543.
3. Hebert PDN, Dewaard JR, Landry JF (2010) DNA barcodes for 1/1000 of the animal kingdom. *Biol Lett* 6:359–362.