

## BARCODING

# Comprehensive DNA barcode coverage of North American birds

KEVIN C. R. KERR,\* MARK Y. STOECKLE,† CARLA J. DOVE,‡ LEE A. WEIGT,‡  
CHARLES M. FRANCIS§ and PAUL D. N. HEBERT\*

\*Department of Integrative Biology, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1,

†Program for the Human Environment, The Rockefeller University, New York, NY 10021, USA ‡Smithsonian Institution, National

Museum of Natural History, Washington, DC 20013-7012, USA, §National Wildlife Research Centre, Canadian Wildlife Service, Environment Canada, Ottawa, Ontario, Canada K1A 0H3

**OnlineOpen:** This article is available free online at [www.blackwell-synergy.com](http://www.blackwell-synergy.com)

## Abstract

DNA barcoding seeks to assemble a standardized reference library for DNA-based identification of eukaryotic species. The utility and limitations of this approach need to be tested on well-characterized taxonomic assemblages. Here we provide a comprehensive DNA barcode analysis for North American birds including 643 species representing 93% of the breeding and pelagic avifauna of the USA and Canada. Most (94%) species possess distinct barcode clusters, with average neighbour-joining bootstrap support of 98%. In the remaining 6%, barcode clusters correspond to small sets of closely related species, most of which hybridize regularly. Fifteen (2%) currently recognized species are comprised of two distinct barcode clusters, many of which may represent cryptic species. Intraspecific variation is weakly related to census population size and species age. This study confirms that DNA barcoding can be effectively applied across the geographical and taxonomic expanse of North American birds. The consistent finding of constrained intraspecific mitochondrial variation in this large assemblage of species supports the emerging view that selective sweeps limit mitochondrial diversity.

*Keywords:* Aves, cryptic species, cytochrome *c* oxidase, DNA barcoding, intraspecific mitochondrial variation, selective sweeps

*Received 25 October 2006; revision accepted 20 November 2006*

Mitochondrial DNA (mtDNA) analysis has been employed in the evolutionary study of animal species for more than 30 years (Brown *et al.* 1979; Mindell *et al.* 1997; Avise & Walker 1999). Its higher mutation rate and lower effective population size than nuclear DNA make mtDNA a powerful tool to probe for evidence of reproductive isolation between lineages. This fact provoked a proposal to standardize DNA-based species identification by analysing a uniform segment of the mitochondrial genome. With this approach, a library of sequences from

taxonomically verified voucher specimens serve as DNA identifiers for species, in short, DNA barcodes (Hebert *et al.* 2003). For animals, research has focused on a 648-bp segment of the mitochondrial gene cytochrome *c* oxidase I (COI), which can be readily recovered from diverse species with a limited set of primers. DNA barcoding translates expert taxonomic knowledge of diagnostic morphologic characters into a widely accessible format, DNA sequences, enabling more people to identify specimens. In addition to assigning specimens to known species, DNA barcoding can speed the discovery of new species, as large sequence differences in animal mtDNA generally signal species status.

For this approach to be effective, it must be possible to distinguish between intraspecific and interspecific mtDNA variation. Pseudogenes, retention of ancestral polymorphisms, hybridization, and the idiosyncrasies of mtDNA

Re-use of this article is permitted in accordance with the Creative Commons Deed, Attribution 2.5, which does not permit commercial exploitation.

Correspondence: Kevin C. R. Kerr, Fax: (519) 767 1656; E-mail: [kkerr@uoguelph.ca](mailto:kkerr@uoguelph.ca)

inheritance pose potential difficulties (Benasson *et al.* 2001; Moritz & Cicero 2004; Thalman *et al.* 2004; Will *et al.* 2005). The simplest test is whether genetic distances within species are less than those between species. Surprisingly, 23% of 2319 animal species failed this test in one review (Funk & Omland 2003), implying that mitochondrial gene sequences do not reliably capture species boundaries. However, the published studies that formed the basis for this estimate may be biased towards exceptional situations and groups in need of taxonomic revision, as further investigations on several vertebrate and invertebrate groups have shown that COI barcodes distinguish more than 95% of species (Ward *et al.* 2005; Hajibabaei *et al.* 2006).

Because birds have been the subjects of particularly intensive taxonomic analysis, they provide an excellent opportunity to test the efficacy of barcode-based species delimitation. With most recent species splits stemming from genetic studies, avian taxonomy could, in turn, benefit from a broad-scale genetic survey. In a preliminary survey of 260 North American bird species, COI sequence variation between species was generally much greater than that within species, and no two species shared barcodes (Hebert *et al.* 2004). As a result, COI sequence information enabled assignment of specimens to known species. Four of 120 species (3%) studied in greater detail contained two distinct barcode clusters, which appeared to reflect cryptic species, a conclusion supported by observations of subtle differences in song and morphology for three of the four cases (Rohwer 1976; Kroodsma 1989; Sibley & Monroe 1990). To test these results more stringently, we increase taxon coverage and sample sizes in this study, applying DNA barcoding to examine the taxonomic status of 643 species, representing 93% of the breeding and pelagic bird species from the USA and Canada (Fig. 1).

## Materials and methods

Most analytic methods followed those described in the earlier study (Hebert *et al.* 2004). DNA sources for this study included frozen tissue samples (muscle, liver, or blood), most of which were obtained from specimens with vouchers housed in museum collections. In addition to tissue samples, feathers (breast feathers or retrices) freshly collected at bird banding stations at six locales (Ontario,

New Brunswick, Nova Scotia, Yukon, North Carolina, Tennessee) were analysed. Feather samples were stored in a dark, dry location at room temperature.

DNA extraction, polymerase chain reaction (PCR), and sequencing reactions were performed at either the University of Guelph or the Smithsonian Institution. DNA was isolated using DryRelease (see Hajibabaei *et al.* 2005), QIAGEN DNeasy tissue extraction kit (QIAGEN), or the NucleoSpin96 tissue kit (Machery-Nagel). Feather samples were processed using the former method exclusively. PCR predominantly utilized a single primer pair: BirdF1 (TTCTCCAACCAC-AAAGACATT GGCAC) and BirdR1 (ACGTGGGAGAT-AATTCCAAATCCTG). If amplification was unsuccessful, an alternate forward primer, FalcoFA (TCAACAAA-CCACAAAGACATCGGCAC), or reverse primers, BirdR2 (ACTACATGTGAGATGATTCCG AATCCAG) and VertebrateR1 (TAGACTTCTGGGTGGCCAAAGAATCA), were employed. All reactions were run under the following thermal cycle program: 1 min at 94 °C followed by six cycles of 1 min at 94 °C, 1.5 min at 45 °C, and 1.5 min at 72 °C, followed in turn by 35 cycles of 1 min at 94 °C, 1.5 min at 55 °C, and 1.5 min at 72 °C, and finally 5 min at 72 °C. Forty-five cycles were run in place of 35 for DNA extracted from feather samples to compensate for lower yields of DNA. PCR products were visualized on precast 2% agarose gels using the E-gel 96 system (Invitrogen). PCR products were bidirectionally sequenced on an ABI 3100, 3130, or 3730. Contigs were assembled from forward and reverse reads using SEQUENCHER, version 4.5 (Gene Codes).

Specimen and collection data, sequences, and trace files are provided in the container project 'Birds of North America Phase 2' at <http://www.barcodinglife.org>. A Kimura 2-parameter distance metric was employed for sequence comparisons (Kimura 1980), genetic distances were calculated using the BOLD Management & Analysis System ([www.barcodinglife.org](http://www.barcodinglife.org)), bootstrap analysis was performed with 1000 replicates using MEGA, version 3.1 (Kumar *et al.* 2004), and scatter and box plots were generated with SIGMAPLOT 8.02 (SPSS). All new sequences have been deposited in GenBank under Accession nos DQ432694 to DQ433261, DQ433274 to DQ433846, and DQ434243 to DQ434805, while sequences from the earlier study (Hebert *et al.* 2004) are deposited in GenBank under Accession nos AY666171 to AY666596.

		RECOGNIZED SPECIES	
		YES	NO
DISTINCT BARCODE CLUSTERS	YES	A. SINGLE CLUSTERS MATCH SPECIES-LEVEL TAXONOMY	B. MULTIPLE CLUSTERS SUGGEST CRYPTIC SPECIES
	NO	C. OVERLAPPING CLUSTERS SUGGEST YOUNG SPECIES, HYBRIDIZATION, OR SYNONYMY	D. (BLANK)

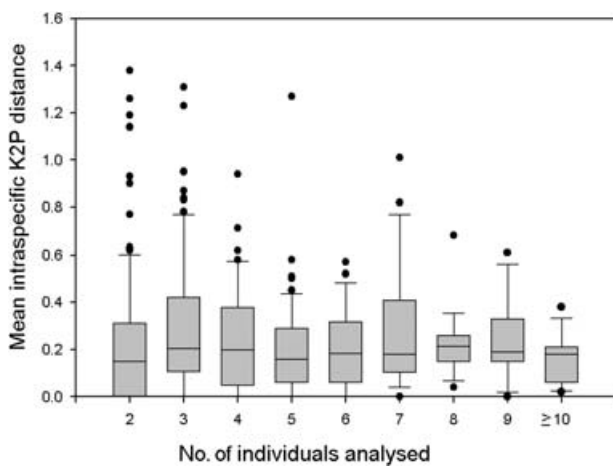
**Fig. 1** Comparing barcode sequence clusters with species-level taxonomy. Categories A–C are described in the figure; by definition, all potential splits recognized by barcoding have distinct barcodes, so D is blank.

## Results

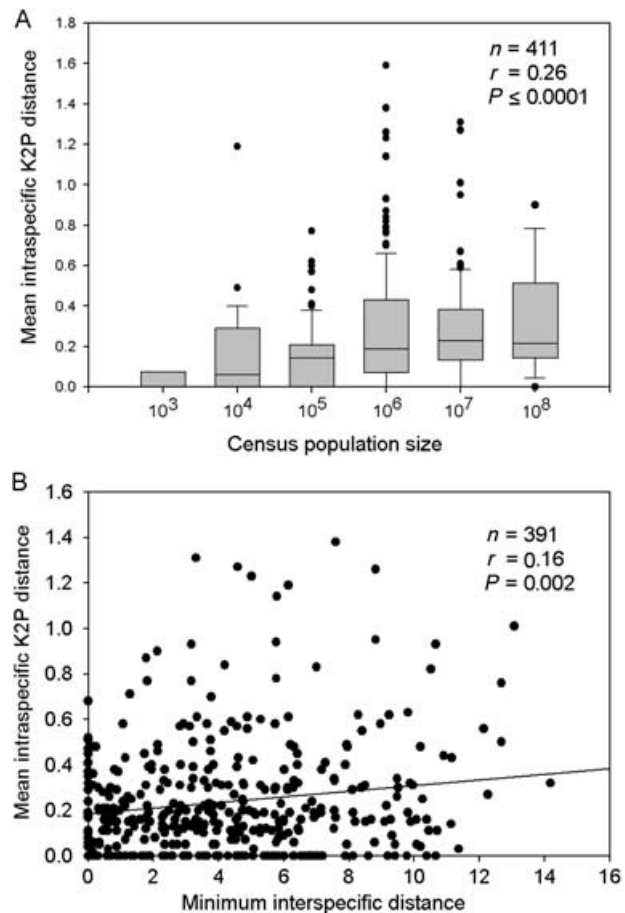
A standard set of primers amplified the target region of COI from all but one of 643 species. These taxa included representatives from 19 (70%) of the 27 extant orders of birds, distributed among 71 families and 286 genera (see Table S1, Supplementary material). Together with the 438 specimens analysed in the earlier study, we obtained COI sequences from 2590 individuals, 70% from vouchered specimens held in museum collections. The mean length of the products sequenced was 658 bp. We analysed multiple individuals (average = 4.1, range = 2–125) from 546 (85%) of the 642 species, including five or more individuals from 211 species (33%). In most cases, conspecific specimens derived from widely separated sites (Birds of North America Phase 2 project at [www.barcodinglife.org](http://www.barcodinglife.org)).

We detected presumptive pseudogenes in approximately 5% of the specimens. Because these were generally short, approximately 100–200 nucleotides, complete barcode sequences could be recovered with bidirectional sequencing. One presumptive pseudogene corresponding to the full-length barcode sequence was detected in three tyrannid flycatcher specimens (0.1%). Overall, pseudogenes were not an important limit to recovery of COI sequences.

Average intraspecific variation was unrelated to the number of individuals analysed, suggesting there was representative sampling (Fig. 2). Within the low and narrow band of intraspecific variation, there was a weak relationship to census population size, which ranges from a few thousand to over 300 million individuals (Fig. 3) (Wetlands International 2002; Rich *et al.* 2004). Intraspecific mitochondrial variation was only weakly associated with apparent species age (Fig. 3). The earlier North American bird study measured *mean congeneric distance*, the average distance



**Fig. 2** Mean intraspecific variation according to number of individuals analysed. Boxes indicate mean, 25th and 75th percentile; bars, 10th and 90th percentile; and dots, values above or below 90th or 10th percentile, respectively.



**Fig. 3** Intraspecific distance, population size, and apparent species age. (A) Linear regression of mean intraspecific distance and  $\log_{10}$  census population size. For illustration purposes, a box plot was generated as described in legend to Fig. 2. (B) Linear regression of mean intraspecific COI distance compared with apparent species age, as indicated by minimum interspecific Kimura 2-parameter (K2P) distance to nearest congeneric relative.

among all congeneric relatives. To more stringently test the discriminatory power of COI barcodes, the present study examined *nearest-neighbour distance*, the minimum genetic distance between a species and its closest congeneric relative. Nearest-neighbour distance averaged 4.3%, 19-fold higher than the mean within species and 11-fold higher than the average maximum intraspecific distance (Fig. 4). Including all species may give a more representative picture, as generic assignments may be incorrect, and 10% of birds are the sole members of their genus; in this case, average nearest-neighbour distance was higher at 5.9% (Fig. 4).

Levels of sequence difference varied across families: 35% of ducks, geese, and swans (Anatidae) showed nearest-neighbour differences of 1% or less, whereas all sandpipers (Scolopacidae), plovers (Charadriidae), and owls (Strigidae) had nearest-neighbour distances greater than 1%. COI barcodes separated 20 of the 23 taxonomic splits recognized

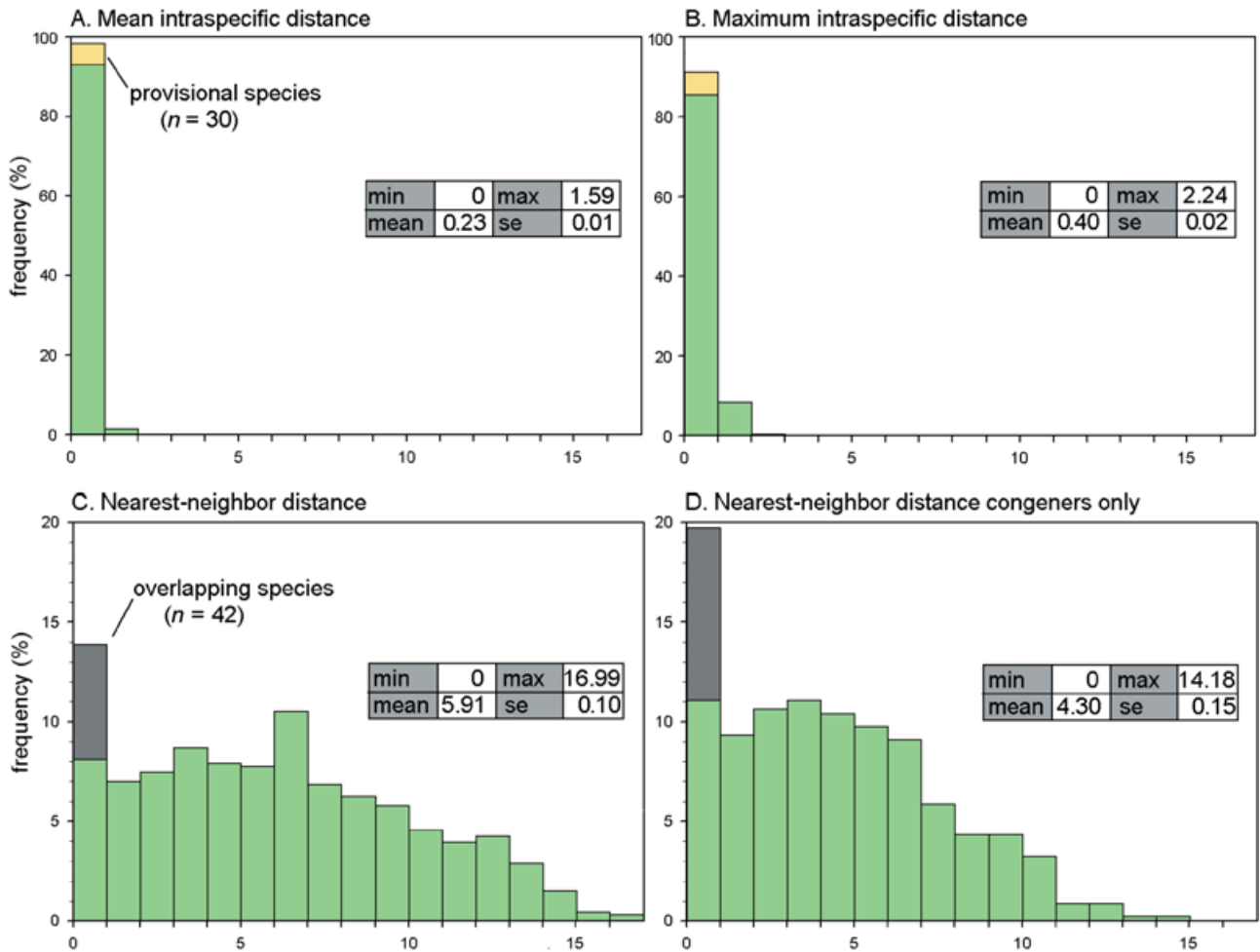


Fig. 4 Intraspecific and nearest-neighbour distances in North American birds. Applying these measures to the data set in the preliminary study (2) gave mean values of 0.24, 0.27, 8.02, and 5.86 for A–D, respectively.

in North American birds over the past 25 years with nearest-neighbour distances ranging from 0.3 to 6.0% (see Table S2, Supplementary material). Average bootstrap support for species nodes with multiple individuals was 97.8%. As expected, bootstrap values were lower among the most closely related species, averaging 79.8% for species with nearest-neighbour distances less than 1%, but 99.5% for species with distances above 1%.

Forty-two species (6.4%) shared sequences or had clusters of sequences overlapping with those of another species, including 14 pairs, two triplets, and one set of eight species (Table 1). The pattern of COI variation within these sets of overlapping species was indistinguishable from variation within single species, with the exception of mallards and black ducks, which are known to harbour two distinct mitochondrial lineages (Avise *et al.* 1990). By contrast, we detected 15 other species with intraspecific distances greater than 2.5% (Table 2); each contained two distinct sequence clusters typically comprised of individuals from different geographical areas. These clusters may

represent cryptic species. Treating these provisional species as distinct, average within-species variation for the COI barcode region was 0.23%.

## Discussion

The present study has reaffirmed that most North American bird species correspond to a single, tightly cohesive array of barcode sequences that are distinct from those of any other species. However, 15 species include two distinct barcode clusters, while 42 other species possess barcode sequences that are shared or overlap with those of other species. What explains these exceptional cases?

Cases of deep barcode divergence within what are thought to be single species generally indicate cryptic taxa (Moritz 1994; Meyer & Paulay 2005). Our screen for provisional splits in species, employing a threshold that was 10× higher than the mean intraspecific variation, revealed 15 cases. Results from a thresholding approach must be interpreted with caution and are best used to flag species

**Table 1** Species with overlapping barcode clusters. The per cent similarity between related species (calculated using a Kimura 2-parameter distance metric) is provided

	Order	Common name	Scientific name	<i>n</i>	% similarity
1	Anseriformes	Snow goose	<i>Chen caerulescens</i>	5	99.8
		Ross's goose	<i>Chen rossii</i>	2	
2		American Black duck	<i>Anas rubripes</i>	8	99.4
		Mallard	<i>Anas platyrhynchos</i>	8	
		Mottled duck	<i>Anas fulvigula</i>	1	
3		Blue-winged teal	<i>Anas discors</i>	8	100.0
		Cinnamon teal	<i>Anas cyanoptera</i>	2	
4		King eider	<i>Somateria spectabilis</i>	5	99.7
		Common eider	<i>Scomateria mollissima</i>	1	
5	Galliformes	Sharp-tailed grouse	<i>Tympanuchus phasianellus</i>	3	99.7
		Greater prairie-chicken	<i>Tympanuchus cupido</i>	1	
		Lesser prairie-chicken	<i>Tympanuchus pallidicinctus</i>	5	
6	Podicipediformes	Western grebe	<i>Aechmophorus occidentalis</i>	2	99.7
		Clark's grebe	<i>Aechmophorus clarkii</i>	2	
7	Charadriiformes	Laughing gull	<i>Larus atricilla</i>	8	99.3
		Franklin's gull	<i>Larus pipixcan</i>	4	
8		California gull	<i>Larus californicus</i>	5	99.8
		Herring gull	<i>Larus argentatus</i>	7	
		Thayer's gull	<i>Larus thayeri</i>	4	
		Iceland gull	<i>Larus glaucooides</i>	1	
		Lesser black-backed gull	<i>Larus fuscus</i>	5	
		Western gull	<i>Larus occidentalis</i>	4	
		Glaucous-winged gull	<i>Larus glaucescens</i>	4	
		Glaucous gull	<i>Larus hyperboreus</i>	4	
		Red-naped sapsucker	<i>Sphyrapicus nuchalis</i>	5	
9	Piciformes	Red-breasted sapsucker	<i>Sphyrapicus ruber</i>	6	99.4
10	Passeriformes	Black-billed magpie	<i>Pica hudsonia</i>	3	99.6
		Yellow-billed magpie	<i>Pica nuttalli</i>	3	
11		American crow	<i>Corvus brachyrhynchos</i>	3	99.5
		Northwestern crow	<i>Corvus caurinus</i>	4	
12		Townsend's warbler	<i>Dendroica townsendi</i>	6	99.5
		Hermit warbler	<i>Dendroica occidentalis</i>	5	
13		Golden-crowned sparrow	<i>Zonotrichia leucophrys</i>	8	99.7
		White-crowned sparrow	<i>Zonotrichia atricapilla</i>	3	
14		Dark-eyed junco	<i>Junco hyemalis</i>	24	99.7
		Yellow-eyed junco	<i>Junco phaeonotus</i>	3	
15		Snow bunting	<i>Plectrophenax nivalis</i>	2	99.9
		McKay's bunting	<i>Plectrophenax hyperboreus</i>	1	
16		Great-tailed grackle	<i>Quiscalis mexicanus</i>	11	99.2
		Boat-tailed grackle	<i>Quiscalis major</i>	6	
17		Common redpoll	<i>Carduelis flammea</i>	2	99.7
		Hoary redpoll	<i>Carduelis hornemanni</i>	5	

in need of further research. Significantly, most of these hypothesized splits are supported by prior taxonomic work (Table 2). In total, nine of our 15 cases have been previously cited; eight have been proposed to represent species pairs, the exceptional case being the northern raven (Omland *et al.* 2000). Some of the species yield additional lineages when non-North American populations are included; for example, six lineages in total are suggested for the winter wren (Drovetski *et al.* 2004).

Regarding the 17 sets of species with overlapping barcodes, three processes may account for these findings.

First, some may be recently diverged sister taxa where COI has not yet accumulated sequence differences. In such cases, more extensive sequence information might allow resolution. Second, these taxa may share mtDNA because of hybridization. Most of our species sets with overlapping barcodes hybridize at least occasionally; many show extensive hybridization and produce fertile F1 hybrid offspring. Examples include snow goose and Ross's goose (Cook *et al.* 1995); blue-winged and cinnamon teal (Bolen 1979); mallard, mottled, and black ducks (McCracken *et al.* 2001); sharp-tailed grouse and greater prairie-chicken (Sparling

**Table 2** Provisional species. Provisional splits of recognized species with intraspecific distances above 2.5% threshold (\*) identified in earlier study (Hebert *et al.* 2004); (†) prior research supports split (see Table S3, Supplementary material); (‡) prior research cites genetic division but does not support species split (citations provided in this table). Bootstrap support for provisional species clusters are shown

	Common name	Scientific name	Maximum intraspecific distance	Bootstrap	Citation (if applicable)
1	Northern fulmar	<i>Fulmaris glacialis</i>	3.1	100/—	
2	Solitary sandpiper*	<i>Tringa solitaria</i>	5.4	100/100	
3	Western screech owl	<i>Megascops kennicottii</i>	3.1	100/100	
4	Warbling vireo*†	<i>Vireo gilvus</i>	4.0	100/99	Sibley & Monroe 1990
5	Mexican jay†	<i>Aphelocoma ultramarina</i>	5.3	100/—	Rice <i>et al.</i> 2003
6	Western scrub-jay†	<i>Aphelocoma californica</i>	3.2	77/—	Rice <i>et al.</i> 2003
7	Common raven‡	<i>Corvus corax</i>	4.3	100/92	Omland <i>et al.</i> 2000
8	Mountain chickadee†	<i>Poecile gambeli</i>	3.7	100/100	Gill <i>et al.</i> 1993
9	Bushtit	<i>Psaltriparus minimus</i>	3.6	100/100	
10	Winter wren†	<i>Troglodytes troglodytes</i>	6.4	100/100	Drovetski 2004
11	Marsh wren*†	<i>Cistothorus palustris</i>	7.9	100/100	Kroodsmma 1989
12	Bewick's wren	<i>Thyromanes bewickii</i>	4.8	100/100	
13	Hermit thrush	<i>Catharus guttatus</i>	3.2	100/100	
14	Curve-billed thrasher†	<i>Toxostoma curvirostre</i>	7.4	100/—	Zink & Blackwell-Rago 2000
15	Eastern meadowlark*†	<i>Sturnella magna</i>	4.6	100/100	Rohwer 1976

1980); red-naped and red-breasted sapsuckers (Johnson & Johnson 1985); Townsend's and hermit warblers (Morrison & Hardy 1983); and the eight species of large white-headed gulls (California, Glaucous, Glaucous-winged, Herring, Iceland, lesser black-backed, western, Thayer's; Olsen & Larsson 2004). These taxa may be in the indeterminate zone between differentiated populations and distinct species (de Queiroz 2005), or well-formed species that are losing genetic identity due to secondary contact and hybridization. Third, some of the pairs with overlapping barcodes may be a single species (Johnston 1961).

Although there is an abundance of *subspecific* assignments in North American birds — 5.5 per species according to one survey — many do not show any evidence of genetic divergence (Zink 2004). Barcode analyses can serve as a quick screening tool for those lineages with deep genetic divergence, aiding detection of overlooked species. In fact, all past barcode surveys have identified new taxonomic units, either as named species, provisional species, evolutionarily significant units (ESUs), or molecular operational taxonomic units (MOTUs) in 4–40% of the species examined (Meyer & Paulay 2005; Monaghan *et al.* 2005; Saunders 2005; Smith *et al.* 2005; Hajibabaei *et al.* 2006; Scheffer *et al.* 2006). These results suggest that 'an iterative process of DNA barcoding ... followed by taxonomic study' will be a productive path to cataloguing biodiversity (Barber & Boyce 2006). In the present study, most provisional species were small to medium-sized, plainly coloured birds, whereas most species with overlapping barcodes were large and/or brightly coloured, which might reflect a natural taxonomic tendency toward undersplitting inconspicuous birds and/or oversplitting more conspicuous species.

Over 30 years ago, Richard Lewontin concluded that intraspecific variation is tightly constrained and recognized that both genetic drift and natural selection offer possible explanations for this fact (Lewontin 1974). Under genetic drift, recent population bottlenecks could account for low intraspecific variation. It might be argued that the low levels of mitochondrial variation detected in our study reflect the unique history of North American birds, most of which have expanded into their present ranges from smaller populations following retreat of glaciers. However, restricted intraspecific mitochondrial variation also exists in many vertebrate and invertebrate species from tropical, temperate, marine, and terrestrial environments (Barrowclough & Shields 1984; Bucklin & Wiebe 1998; Meyer & Paulay 2005; Saunders 2005; Hajibabaei *et al.* 2006), implying a more general explanation. Effective population size for nuclear genes can reach an asymptotic limit due to linkage; this effect is strongest for organisms with large genomes, with the result that the effective population size of vertebrates might not exceed  $10^4$  (Gillespie 2000; Lynch *et al.* 2006). Although not directly applicable to mitochondria, this effect does reveal the complexities of estimating effective population sizes and predicting the role of drift in scouring variation.

Low mitochondrial variation might alternatively (or additionally) reflect recurrent selective sweeps; repeated diffusions of new, selectively favoured variants across the breeding range of a species could purge mitochondrial diversity. Although 98% of the nucleotide differences in COI barcode sequences in our study between nearest neighbours were synonymous, selection on any nucleotide position in the mitochondrial genome would result in the

loss of variation in the barcode region because mtDNA is inherited as a single linkage group, due to its asexual transmission. Mutations in nuclear or mitochondrial loci important in nuclear-mitochondrial co-adaptation might be particularly important (Catalano *et al.* 2006). A recent analysis of patterns of substitution in nuclear and mtDNA concluded that reduced mitochondrial diversity in animals is due to selective sweeps (Bazin *et al.* 2006). Although these authors found no correlation between census population size and intraspecific mitochondrial variation, the range of variation was less than expected given census population sizes. This latter finding, together with our results showing trends toward increased diversity in larger populations and older species, imply that genetic drift does influence mitochondrial variation, but only weakly.

Most researchers agree that species are a key unit of biological systems, but quarrel about how best to define them. Hence, theoretical and operational species concepts proliferate, each emphasizing different aspects of present-day biology and evolutionary history (Wheeler & Meirer 2000). Some believe that a basic taxonomic unit does not exist, instead viewing species as a convenient taxonomic construct, 'an arbitrary cut-off somewhere along a branch in the tree of life' (Mishler & Shapley 2004). The tight clustering of mtDNA sequences within species observed in our study not only bolsters the view that species are fundamental biological units, but also reveals that their identification is usually uncomplicated.

In summary, most North American bird species appear to have a similar genetic structure, each being a single tight cluster of mtDNA variants distinct from the clusters of closely related species. High bootstrap support for species nodes in this study and in other animal groups suggests neighbour-joining analysis of COI barcode sequences will be widely effective (Ward *et al.* 2005; Hajibabaei *et al.* 2006). The few species with higher intraspecific diversity were comprised of two such clusters, many of which appear to represent cryptic species. It seems likely that further study will reveal additional lineages within some species, but leave unchanged the underlying pattern of segregation of mitochondrial diversity into distinct clusters (Zink 2004). Together these observations imply a general constraint on mitochondrial diversity in birds.

## Acknowledgements

This study was funded by grants from NSERC, the Canada Research Chairs program, Genome Canada through the Ontario Genomics Institute, Environment Canada, and the Gordon and Betty Moore Foundation to PDNH and by grants from the Federal Aviation Administration, DoD Legacy Resource Management Program, Air National Guard – Environmental Division and US Air Force to C.J.D. For providing tissue specimens and sharing collection data, we are indebted to Allan Baker, Sharon Birks, Mark Peck, Michel Gendron, Bob Montgomerie, and the staff at

the Royal Ontario Museum (599 specimens), Burke Museum (495 specimens), and Canadian Wildlife Service (400 specimens) for samples sent to Guelph. Also, Carla Cicero, Joel Cracraft, Scott Edwards, Shannon Hackett, A. Townsend Peterson, Nate Rice, Fred Sheldon, Jeremiah Trimble, and the staff at the Academy of Natural Sciences, Philadelphia (144 specimens), American Museum of Natural History (116 specimens), Field Museum (99 specimens), Museum of Comparative Zoology (35 specimens), Louisiana State University (58 specimens), Museum of Southwestern Biology (16 specimens), Museum of Vertebrate Zoology (41 specimens), National Museum of Natural History (407 specimens), University of Alaska-Fairbanks (2 specimens), and University of Kansas (9 specimens) for samples loaned to the Smithsonian. We thank the Toronto FLAP for additional specimens, banding stations in Canada and the USA for providing feathers, and Irby Lovette for sharing unpublished COI sequences of *Cardellina rubrifrons* and *Limnothlypis swainsonii*. We thank Charles Cabral Jr, Greg Concepcion, and Marcy Heacker at the SI and Alex Borisenko, Natalie Ivanova, and Angela Hollis at Guelph for aiding laboratory work. Sujeevan Ratnasingham and Rob Dooh provided much help with data analysis. Finally, we thank Jesse Ausubel, Allan Baker, David Thaler, and Paul Waggoner for helpful commentary.

## Supplementary material

The supplementary material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/MEN/MEN1670/MEN1670sm.htm>

**Table S1** Species and number of individuals analysed

**Table S2** COI barcode distances between recent taxonomic splits in North American birds

## References

- Avise JC, Ankney CD, Nelson WS (1990) Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. *Evolution*, **44**, 1109–1119.
- Avise JC, Walker D (1999) Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome. *Proceedings of the National Academy of Sciences, USA*, **96**, 992–995.
- Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatod larvae. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **273**, 2053–2061.
- Barrowclough GF, Shields GF (1984) Karyotypic evolution and long-term effective population sizes of birds. *Auk*, **101**, 99–102.
- Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–572.
- Benasson D, Zhang D, Hart DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, **16**, 314–321.
- Bolen EG (1979) Blue-winged × cinnamon teal hybrid from Oklahoma. *Wilson Bulletin*, **91**, 367–370.
- Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences, USA*, **76**, 1967–1971.

- Bucklin A, Wiebe PH (1998) Low mitochondrial diversity and small effective population sizes of the copepods *Calanus finmarchicus* and *Nannocalanus minor*: possible impact of climatic variation during recent glaciation. *Journal of Heredity*, **89**, 383–392.
- Catalano D, Licciulli F, Turi A *et al.* (2006) MitoRes: a resource of nuclear-encoded mitochondrial genes and their products in Metazoa. *BMC Bioinformatics*, **7**, 36.
- Cooke F, Rockwell RF, Lank DB (1995) *The Snow Geese of la Pérouse Bay: Natural Selection in the Wild*. Oxford University Press, Oxford, UK.
- de Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences, USA*, **102**, 6600–6607.
- Drovetski SV, Zink RM, Rohwer S *et al.* (2004) Complex biogeographic history of a Holarctic passerine. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **271**, 1471–2954.
- Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.
- Gill FB, Mostrom A, Mack AL (1993) Speciation in North American chickadees: patterns of mtDNA genetic divergence. *Evolution*, **47**, 195–212.
- Gillespie JH (2000) The neutral theory in an infinite population. *Gene*, **261**, 11–18.
- Hajibabaei M, deWaard JR, Ivanova NV *et al.* (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1959–1967.
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcoding distinguishes species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the USA*, **103**, 968–971.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 313–321.
- Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *Public Library of Science Biology*, **2**, 1657–1663.
- Johnson NK, Johnson CB (1985) Speciation in sapsuckers (*Sphyrapicus*): II. Sympatry, hybridization, and mate preference in *S. ruber daggetti* and *S. nuchalis*. *Auk*, **102**, 1–15.
- Johnston DW (1961) *The Biosystematics of American Crows*. University of Washington Press, Seattle.
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **15**, 111–120.
- Kroodsma DE (1989) Two North American song populations of the marsh wren reach distributional limits in the central Great Plains. *Condor*, **91**, 332–340.
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinformatics*, **5**, 150–163.
- Lewontin RC (1974) *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
- Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genomic architecture. *Science*, **311**, 1727–1730.
- McCracken KG, Johnson WP, Sheldon FH (2001) Molecular population genetics, phylogeography, and conservation biology of the mottled duck (*Anas fulvigula*). *Conservation Genetics*, **2**, 87–102.
- Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *Public Library of Science Biology*, **3**, 2229–2237.
- Mindell DP, Sorenson MD, Huddleston CJ *et al.* (1997) Phylogenetic relationships among and within select avian orders based on mitochondrial DNA. In: *Avian Molecular Evolution and Systematics* (ed. Mindell DP), pp. 214–247. Academic Press, New York.
- Mishler BB, Shapley RL (2004) Presentation at assembling the tree of life – PI meeting, National Science Foundation, November 19–21, [http://ucjeps.berkeley.edu/bryolab/ATOL/pdf/ATOL\\_2004\\_Mishler\\_PEG.pdf](http://ucjeps.berkeley.edu/bryolab/ATOL/pdf/ATOL_2004_Mishler_PEG.pdf).
- Monaghan MT, Balke M, Gregory TR, Vogler AP (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1925–1933.
- Moritz C (1994) Defining ‘evolutionarily significant units’ for conservation. *Trends in Ecology & Evolution*, **9**, 373–375.
- Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *Public Library of Science Biology*, **2**, 1529–1531.
- Morrison ML, Hardy JW (1983) Hybridization between Hermit and Townsend’s Warblers. *Murrelet*, **64**, 65–72.
- Olsen KM, Larsson H (2004) *Gulls of Europe, Asia and North America*, 2nd edn. Princeton University Press, Princeton, New Jersey.
- Omland KE, Tarr CL, Boarman WI, Marzluff JM, Fleischer RC (2000) Cryptic genetic variation and paraphyly in ravens. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **267**, 2475–2482.
- Rice NH, Martinex-Meyer E, Peterson AT (2003) Ecological niche differentiation in the *Aphelocoma* jays: a phylogenetic perspective. *Biological Journal of the Linnean Society*, **80**, 369–383.
- Rich TD, Beardmore CJ, Berlanga H *et al.* (2004) *Partners in Flight: North American Landbird Conservation Plan*. Cornell Laboratory Ornithology, New York.
- Rohwer SA (1976) Specific distinctness and adaptive differences in southwestern meadowlarks. *Occasional Papers of the Museum of Natural History*, **44**, 1–14. The University of Kansas.
- Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1879–1888.
- Scheffer SJ, Lewis ML, Joshi RC (2006) DNA barcoding applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines. *Annals of the Entomological Society of America*, **99**, 204–210.
- Sibley CG, Monroe BL Jr (1990) *Distribution and Taxonomy of Birds of the World*. Yale University Press, New Haven, Connecticut.
- Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1825–1834.
- Sparling DW Jr (1980) Hybridization and taxonomic status of greater prairie-chickens and sharp-tailed grouse. *Prairie Naturalist*, **12**, 92–101.
- Thalman O, Hebler J, Poinar JN, Paabo S, Vigilant L (2004) Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Molecular Ecology*, **13**, 321–335.
- Ward RW, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia’s fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1847–1857.



- Wetlands International (2002) *Waterbird Population Estimates*, 3rd edn. *Wetlands International Global Series*, no. 12. Wageningen, The Netherlands.
- Wheeler QD, Meirer R (2000) *Species Concepts and Phylogenetic Theory*. Columbia University Press, New York.
- Will KW, Mishler BD, Wheeler Q (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.
- Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **271**, 561–564.
- Zink RM, Blackwell-Rago RC (2000) Species limits and recent population history in the curve-billed thrasher. *Condor*, **102**, 881–886.