

DNA Barcoding Korean Birds

Hye Sook Yoo, Jae-Yong Eah, Jong Soo Kim³, Young-Jun Kim², Mi-Sook Min², Woon Kee Paek¹, Hang Lee², and Chang-Bae Kim*

Biological Resource Center/Korean BioInformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-333, Korea;

¹ Natural History Team, National Science Museum, Daejeon 305-705, Korea;

² Conservation Genome Resource Bank for Korean Wildlife (CGRB), College of Veterinary Medicine, and BK21 Program for Veterinary Science, Seoul National University, Seoul 151-742, Korea;

³ Hankuk Academy of Foreign Studies, Yongin 449-854, Korea.

(Received August 31, 2006; Accepted November 13, 2006)

DNA barcoding, an inventory of DNA sequences from a standardized genomic region, provides a bio-barcode for identifying and discovering species. Several recent studies suggest that the sequence diversity in a 648 bp region of the mitochondrial gene for cytochrome c oxidase I (COI) might serve as a DNA barcode for identifying animal species such as North American birds, insects and fishes. The present study tested the effectiveness of a COI barcode in discriminating Korean bird species. We determined the 5' terminus of the COI barcode for 92 species of Korean birds and found that species identification was unambiguous; the genetic differences between closely related species were, on average, 25 times higher than the differences within species. We identified only one misidentified species out of 239 specimens in a genetic resource bank, so confirming the accuracy of species identification in the banking system. We also identified two potential composite species, calling for further investigation using more samples. The finding of large COI sequence differences between species confirms the effectiveness of COI barcodes for identifying Korean bird species. To bring greater reliability to the identification of species, increased intra- and interspecies sampling, as well as supplementation of the mitochondrial barcodes with nuclear ones, is needed.

Keywords: Bird; DNA Barcode; Korea; Mitochondrial Cytochrome c Oxidase I; Phylogeny.

* To whom correspondence should be addressed.

Tel: 82-42-860-4643 Fax: 82-42-879-8519

E-mail: changbae@kribb.re.kr

Introduction

A short section of DNA sequence is used to identify and discover species for DNA barcoding. Several studies suggest that a 648 bp region of the mitochondrial gene for cytochrome c oxidase I (COI), may serve as a DNA barcode for the identification of animal species (Hajibabaei *et al.*, 2006; Hebert *et al.*, 2004a; Ward *et al.*, 2005). This gene region is easily recovered, and provides good resolution, as evidenced by the fact that deep sequence divergences were the rule between 13,000 closely related pairs of animal species (Hebert *et al.*, 2003). As well as correctly identifying known species, a number of probably cryptic species have been discovered within what had previously been thought a single morphologically-based species tightly correlated with ecological data (Hebert *et al.*, 2004b; Smith *et al.*, 2006).

The initial success of DNA barcoding led rapidly to the formation of the Consortium for the Barcode of Life (CBOL, <http://barcoding.si.edu>), with the objective of obtaining DNA barcodes from all species on the planet. Advances in sequencing technology permit sequences to be obtained rapidly and cheaply, so that barcoding appears both plausible and worthwhile. The two essential components of an effective DNA barcode system are standardization on a uniform barcode sequence, such as COI, and a library of sequences linked to named voucher specimens.

The present study extends earlier investigations by testing the correspondence between species boundaries signaled by COI barcodes and those established by prior taxonomic work. For this analysis the barcode region of COI was sequenced in 92 of the 450 bird species that breed in Korea (Lee *et al.*, 2000). The sequences used were those

Abbreviation: COI, cytochrome c oxidase I.

available in a museum and a genetic resource bank.

Materials and Methods

We employed three steps of sampling designed to gain an overview of the patterns of COI sequence divergence among Korean birds. The first step examined a single individual from each of 92 species to assess COI divergence between species. These species were selected on the basis of accessibility without regard to known taxonomic issues. In the second step of sampling we examined one or two additional individuals from 45 of these species to provide a general sense of intraspecific sequence divergence as well as a preliminary indication of the variation within each species. When possible, the individuals sampled were obtained from widely separated localities in Korea. The third step of the analysis involved sequencing a maximum of six more individuals of the few species in which the second level detected more than 2.8% sequence divergence among individuals, or less than 2.8% sequence difference between species.

The specimens examined in this study were obtained from the bird collection in the National Science Museum, Korea and tissue samples in the Conservation Genome Resource Bank for Korean Wildlife, Seoul National University. Collection localities and other information about the specimens are available in Supplementary Data 1. Taxonomic assignments follow Lee *et al.* (2000) and Howard and Moore (1998).

DNA extracts were prepared from frozen samples or dry skin using an AccuPrep Genomic DNA Extraction Kit (Bioneer, Korea). 749-bp and 751-bp products were amplified using PCR primers, BirdF1, and BirdR1, and BirdR2 (Hebert *et al.*, 2004a), respectively. The 20 μ l PCR reaction mixes contained 8 μ l of water, 10 μ l of 2 \times DyeMix (Enzynomics, France) with 1.0 U of Taq polymerase, 0.5 μ l of each primer (0.1 mM), and 1 μ l of template DNA. The amplification protocol consisted of 5 min at 95°C followed by 40 cycles of 30 s at 95°C, 30s at 52°C, and 30 s at 72°C, and a final 7 min at 72°C. Direct sequencing reactions were carried out using Big Dye v3.1 and bidirectional PCR primers, and the products were analyzed on an ABI 377 sequencer. COI sequences were recovered from all the bird species and did not contain any insertions, deletions, nonsense, or stop codons.

Sequence divergence was calculated using the Kimura-2-parameter (K2P) model (Kimura, 1980). Phylogenetic analyses were performed by the Neighbor-Joining method (Saitou and Nei, 1987) with K2P distances, using the MEGA3 program (Kumar *et al.*, 2004) to provide a graphic representation of the pattern of divergence between species.

Results

All 92 bird species had different COI sequences. The COI sequences in the 45 species in which two or more individuals were sampled were either identical or most simi-

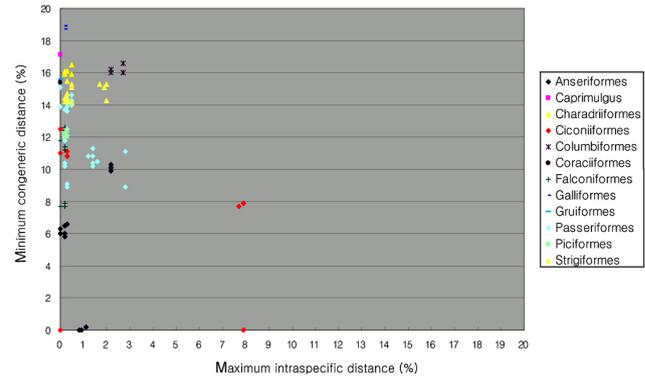


Fig. 1. Intraspecific and interspecific genetic distances for individual species based on COI sequences. For each species in which this comparison was possible ($n = 45$), maximum intraspecific variation is compared to minimum interspecific congeneric difference. For illustration purposes 2.8% was chosen as the cutoff between the usual values for intra- and interspecific variation.

lar to the other sequences of the same species. The COI sequence differences between closely related species were 25 fold higher than the differences within species, with average differences between and within species of 7.9% and 0.3%, respectively (Fig. 1).

In most cases the neighbor-joining (NJ) tree showed shallow intraspecific and deep interspecific divergence. However, in one exceptional case there was a deep divergence within a species (*Egretta intermedia*, Intermediate Egret). COI sequences in the polytypic species separated into pairs of divergent clusters in the NJ tree. The intraspecific distance in the exceptional species was 8.2%, 27 fold higher than the average distance (Fig. 2 and Supplementary Data 2).

Setting aside these polytypic species, the average intraspecific distance was very low, 0.26%, and the maximum average intraspecific difference only 2.8%. Most congeneric species pairs showed divergences well above this value, but the two species in the same genera had interspecific distances below 0.4%. They included *Anas platyrhynchos*, Mallard and *A. poecilorhyncha*, Spot-Billed Duck (Fig. 3 and Supplementary Data 3).

Although species were the focus of this study, we noted that the NJ tree of COI sequences generally matched higher levels of avian classification, with most genera, families, and orders appearing as nested monophyletic lineages concordant with current taxonomy, with some exceptional splitting of orders.

To establish a threshold for assigning genetically divergent specimens to different provisional avian species we made a distance matrix including the DNA barcode information on Korean and North American birds (Hebert *et al.*, 2004a). The average distance of individuals within species was 0.3% compared with 8.2% for species within genera (Table 1). Overall, there was 27 times more variation

Table 1. Genetic divergence within various taxonomic levels of birds.

Comparisons within	Number of comparisons	Minimum	Mean Distance	Maximum	s.e.
Species	638	0	0.3	8.2 ^b	0.029
Genus	1148	0 ^a	8.2	16.9	0.102
Family	3921	4.8	13.8	23.0	0.043
Order	15218	8.0	16.0	22.0	0.016

Data are from 459 sequences of 158 species and 106 genera.

^a Including a case of misidentification of *Anas platyrhynchos* and *A. poecilorhyncha*.

^b Including *Egretta intermedia*.

among congeneric species than among conspecific individuals. Mean divergence among species within families increased to 13.8%, and among species within orders to 16.0% (Table 1, Fig. 4). The rate of increase declined in the higher taxonomic categories due to substitutional saturation.

Discussion

In this study we found that COI differences between most of the 92 Korean bird species far exceeded those within species. In order to conservatively test the effectiveness of COI barcodes as an identification tool, our sample should not have underestimated variability within species or overestimated it between species. Our measures of intraspecific variation could be underestimated if members of a species showing sequence divergence across their distribution failed to be tested. Forty-five species including more than two representatives were used to examine this issue by collecting from sites with the maximum distance being that between Jeju Island and Cheorwon, Gangwon Province. The data suggested that there was adequate representation of genetic diversity across their ranges. We have compared the sequence differences within species to the geographic distances between specimen collection points and found that these parameters were unrelated (data not shown). The lower level of intraspecific variation shown above has also been reported in previous studies (Hebert *et al.*, 2004a; Moore, 1995; Weibel and Moore, 2002). Based on these results, the levels of intraspecific divergence in COI of the 92 different Korean birds belonging to a variety of orders appear to be reasonable.

Our discovery of a polytypic species (*Egretta intermedia*) within one sample of 239 makes it likely there are other Korean birds with divergent populations that may represent hidden species. On the other hand it is possible that the one example is a result of misidentification, although this is not very likely. The differences of distances between the specimen (CIARin215) of *E. intermedia* and *E. alba*, the most closely relative species, are below the minimum interspecific variation. The NJ tree supports the close relationship between the specimen and *E. alba* (Fig. 2).

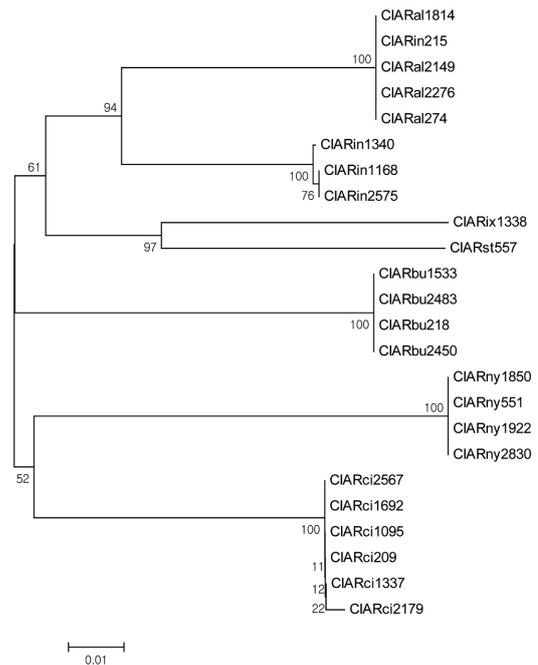


Fig. 2. Neighbor-joining tree of 24 COI sequences from 7 species of Ardeidae.

The lower genetic distance between *Anas platyrhynchos* and *A. poecilorhyncha* than the maximum intraspecific difference suggest that these species are recently diverged, hybridized, and synonymized. They are nested between these species in the NJ tree (Fig. 3).

Any critical test of the effectiveness of barcodes must also consider the possibility that our study has overestimated the variability between species. We therefore looked at species individually, comparing their minimum distance to a congener with the maximum divergence within each species. There were sufficient data to perform this analysis on the 45 of the 92 species examined that had more than two members (Fig. 1). The average maximum distances within these 45 species was 2.8%, while the average minimum distance to a congener was 5.8%. Previous studies of diverse animal groups suggest that the minimum distance for sister species is over 2% (Hebert *et*

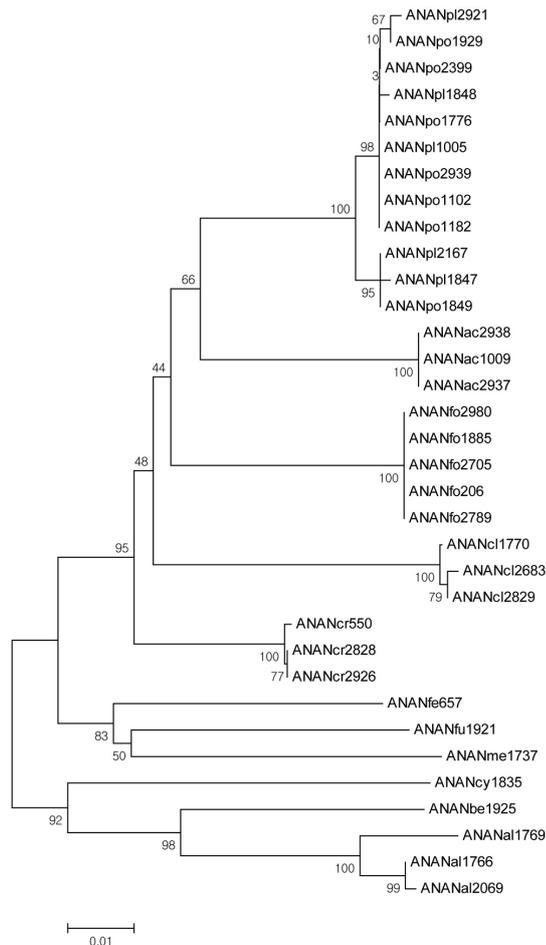


Fig. 3. Neighbor-joining tree of 34 COI sequences from 12 species of Anatidae.

al., 2004a; Johns and Avise 1998; Klicka and Zink, 1997; Ward *et al.*, 2005). Thus it appears that a COI barcode will permit the separation of most sister bird species.

The threshold for assigning genetically divergent specimens as provisional species should certainly be high enough to separate only specimens that very likely belong to different species. Hebert *et al.* (2004a) proposed a standard sequence threshold of 10x the mean intraspecific variation for the group under study. This threshold fits the outcomes of previous studies (Ward *et al.*, 2005). If applied to the Korean birds examined in this study (0.26% average intraspecific variation; 2.6% threshold), a 10x threshold would recognize over 95% of the 92 species, as well as one misidentified species.

For 92 of the 450 bird species breeding in Korea our evidence shows that COI barcodes separate individuals into species. The database of avian DNA barcodes linked to voucher specimens will be an enduring resource for ornithologists, public health officials and the interested public. DNA barcoding can be applied equally to intact specimens, single feathers collected from live birds in

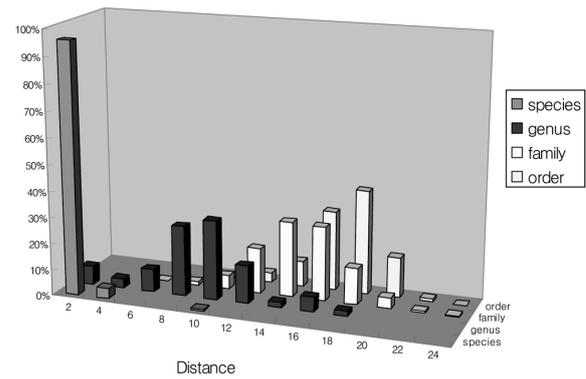


Fig. 4. Distribution of genetic distances based on the Kimura-2-parameter for COI within different taxonomic categories of birds.

banding or customs operations, or fragments of birds involved in aircraft collisions recently of interest to airport officials and the Air Force in Korea. The majority of avian genetic samples utilized in this study were part of a tissue collection in the Conservation Genome Resource Bank for Korean Wildlife (CGRB; www.cgrb.org) in Seoul National University. Only one of the 239 samples from the CGRB turned out to be a misidentified species (0.41%), confirming the reliability of species verification procedures in the banking system. Thus our current study demonstrates the feasibility of a DNA barcoding system as a quality control measure for national genetic resource collection systems in Korea.

The utility of barcoding relies on the assumption that genetic variation within a species is much smaller than variation between species. This assumption was verified in previous studies (Hajibabaei *et al.*, 2006; Hebert *et al.*, 2004a; 2006b; Ward *et al.*, 2005). To bring greater reliability to the identification of species using short DNA sequences, a move should be made to supplement the mtDNA-based barcode with nuclear barcodes. This would reduce the problem of reliance on a single character and help identify cases where mtDNA behaves differently from the nuclear genome.

The present study provides an initial set of COI barcodes for about 20% of Korean birds. More detailed sampling of COI sequences is needed for these species, and barcodes need to be gathered for the remaining Korean birds and for those in other Far Eastern Asian regions.

Acknowledgments This study was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. M10508040002-06N0804-00210), the Brain Korea 21 Program for Veterinary Science funded by the Ministry of Education, and the Korea National Research Resource Collection Program funded by the Korea Science and Engineering Foundation. We express our apprecia-

tion to Mr. Tae Wook Kang for assistance with data analysis and special thanks to Tae Young Choi, Dong Gi Choi, and Chun Gweon Choi for donating road-killed samples collected for the Eco-technopia 21 Project, the Korean Government's R&D Program on environmental technology development. The authors also thank the Korean Association for Bird Protection in Cheorwon, Korea and the Wildlife Rescue Center in Suncheon for samples.

References

- Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., and Hebert, P. D. N. (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA* **103**, 968–971.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 313–321.
- Hebert, P. D. N., Stoeckle, M. Y., Zemplak, T. S., and Francis, C. M. (2004a) Identification of birds through DNA barcodes. *PLOS Biology* **2**, e312.
- Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., and Hallwachs, W. (2004b) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. USA* **101**, 14812–14817.
- Howard, R. and Moore, A. (1998) *A Complete Checklist of the Birds of the World*. Academic Press, London.
- Johns, G. C. and Avise, J. C. (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Mol. Biol. Evol.* **15**, 1481–1490.
- Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.
- Klicka, J. and Zink, R. M. (1997) The importance of recent ice ages in speciation: A failed paradigm. *Science* **277**, 1666–1669.
- Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**, 150–163.
- Lee, W.-S., Koo, T.-H., and Park, J.-Y. (2000) *A Field Guide to the Birds of Korea*. LB Evergreen Foundation, Seoul.
- Moore, W. S. (1995) Inferring phylogenies from mtDNA variation: Mitochondrial gene trees versus nuclear-gene trees. *Evolution* **49**, 718–726.
- Saitou, N. and Nei, M. (1987) The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W., and Hebert, P. D. N. (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proc. Natl. Acad. Sci. USA* **103**, 3657–3662.
- Ward, R. D., Zemplak, T. S., Innes, B. H., Last, P. R., Hebert, P. D. N. (2005) Barcoding Australia's fish species. *Phil. Trans. R Soc. B.* **360**, 1847–1857.
- Weibel, A. C. and Moore, W. S. (2002) Molecular phylogeny of a cosmopolitan group of woodpeckers (genus *Picoides*) based on COI and cyt b mitochondrial gene sequences. *Mol. Phylogenet. Evol.* **22**, 65–75.