# Comparing two genetic markers used in the identification of diving ducks (Aythyinae) involved in birdstrikes

Damani Eubanks, Carla Dove Ph.D, Faridah Dahlan, Sergei Drovetski Ph.D

Smithsonian
*National Museum of Natural History*

REU Site, EAR-1062692

## Abstract

Knowing the species of birds involved in damaging collisions with U.S. military and civil aircraft (birdstrikes) is paramount to understanding and preventing human-wildlife conflicts in this field. The Feather Identification Lab, Smithsonian Institution, identifies over 9,000 birdstrike cases each year using feather morphology and DNA barcoding. While the DNA barcode marker (CO1) is successful at identifying many species of birds, it falls short in species that are very closely related or hybridize frequently. This project tested the effectiveness of two mitochondrial genetic markers, cytochrome oxidase 1 (CO1) and NADH dehydrogenase subunit 2 (ND2) used for identifying species of waterfowl within the genus *Aythya*. Because these diving ducks are commonly involved in damaging birdstrikes, the most reliable method of DNA identification is needed for species designation of birdstrike samples. To compare CO1 and ND2 markers, we sequenced 19 specimens from the 9 different species of *Aythya* that are commonly involved in birdstrikes. We obtained additional sequences from GenBank to increase sample size and geographic coverage. We then constructed gene-specific phylogenetic trees and compared genetic distances between closely related species and support for monophyly of species-specific clades. While neither genetic marker was 100% effective at DNA identifications of all taxa within this group, ND2 was significantly more effective than CO1 at genetic separation of *A. fuligula* (Tufted Duck), *A. nyroca* (Ferruginous Duck), *A. australis* (Hardhead Duck), *A. americana* (Redhead), *A. collaris* (Ring-neck Duck), *A. affinis* (Lesser Scaup) and *A. marila* (Greater Scaup). *Aythya valisineria* (Canvasback) and *A. ferina* (Common Pochard) formed a single clade and could not be distinguished in either tree. Understanding the reliability of the two genetic markers in this group of birds improves lab efficiency and DNA identification success of diving ducks involved in birdstrikes.

## Introduction

The Feather Identification Lab (Division of Birds - Smithsonian Institution, National Museum of Natural History) is one of the premier labs in the world for identifying birds involved in collisions with aircraft (birdstrikes). The lab uses multiple methods to identify the birds, including whole feather comparison with museum specimens, microscopic analysis, and DNA barcoding. The DNA 'barcode' genetic marker is also known as cytochrome oxidase 1 (CO1). This particular marker is usually very effective at identifying different species of birds, however it is less effective when taxa are closely related or when species hybridize frequently. One particular genus of diving ducks (*Aythya*) includes many closely related species that hybridize frequently making it impossible to obtain confident DNA identifications using the CO1 marker. In this project, we investigated another mitochondrial DNA marker known as NADH dehydrogenase subunit 2 (ND2) to determine if this marker could better resolve the relationships within the genus *Aythya*. Because ND2 (2.9% per Ma) is a faster evolving gene than CO1 (1.6% per Ma) (Lerner *et al.* 2011) and has a longer sequence length (1041 bp vs. 654 bp), we hypothesized that it would be more effective at identifying birds within this genus. The purpose of this project was to sequence the CO1 and ND2 gene for multiple specimens from nine different *Aythya* species that are commonly involved in birdstrikes to determine which marker is more effective at identifying these birds. This project has important real-world implications because it is essential for military and civil aviation to know the exact species involved in birdstrike events in order to implement proper mitigation measures to reduce birdstrike risks to aviation safety while at the same time helping to protect birds.

## Methods

For this project 19 tissue samples and/or DNA extracts were borrowed through the Division of Birds loan process from the Smithsonian Biorepository and combined with 34 CO1 (GenBank accessions AF090337, DQ432760, DQ433344, DQ433345, DQ434306 - DQ434308, DQ434313 - DQ434316, DQ434322 - DQ434325, DQ434331 - DQ434334, GQ481387, GQ481388, GU571273 - GU571275, JF499098, JF499099, JF499101, JF499102, KP252170, NC000877, and BOLD BROMB434, BROMB440, BROMB744, BROMB853) and 9 ND2 sequences (GenBank AF090337, EU585684 - EU585689, KJ710708, KJ722069) obtained from Genebank and BoLD to allow for increased sample size and better geographic coverage. DNA was extracted using the AutoGen© (Holliston, MA) Gene Prep DNA extraction system according to the manufacturer's protocols. DNA concentration in the extracts was measured using a Qubit 2.0 fluorometer. For samples containing less than 10 ng/µl of DNA the amount of sample used in PCR was increased, and samples containing over 20 ng/µl were diluted. PCRs for the CO1 and ND2 were performed for all the samples. For the CO1, four different primers (CO1F, CO1R, DGF, DGR) were used, representing the forward and reverse direction. For the ND2, three different primers were used for sequencing: in addition to amplification primers L5215 (Hackett 1996) and H1064 (Drovetski et al. 2004), we used internal primer L347 (Drovetski et al. 2004) . Following PCR, gel electrophoresis was performed on an 1.5% agarose gel to test for successful DNA amplification. Amplicons were prepared for Sanger sequencing and submitted to Smithsonian L.A.B. for sequencing on the ABI 3730 Genetic Analyzer (Applied Biosystems Inc., Foster City, CA). The sequences were aligned automatically and manually confirmed in Sequencher 5.0.1 (Gene Codes Corporation, Ann Arbor, MI).

We used the Bayesian information criterion (BIC) implemented in jModelTest (Posada 2008) to select substitution models for phylogenetic analyses of our loci. For both loci, jModelTest selected TrN+I sub-model of the generalized time reversible (GTR) model (Tavaré 1986) where transversions are weighted equally and the proportion of invariable sites (I) was included.

Phylogenetic analyses were carried out separately for each locus in BEAST 2.0.2 (Drummond et al. 2012). We incorporated a Yule process speciation prior for our BEAST analysis and the strict molecular clock prior. The MCMC analysis consisted of 107 generations with a 1000 generation burn-in. Parameters sampled every 1000 steps. Tracer 1.5 (http://beast.bio.ed.ac.uk/ Tracer) was used to determine the effective sample size of each parameter and calculate its mean and 95% highest posterior density (95% HPD) interval. Tree topologies were assessed using TreeAnnotator 2.0.2 (Drummond et al. 2012)] and visualized in FigTree 1.3.1 (http:// tree. bio.ed.ac.uk/software/figtree/).

## Results

For our phylogenetic reconstruction we used 25 (16 new and 9 from GenBank) ND2 sequences and 51 (17 new and 34 from GenBank/BoLD) CO1 sequences. Although we were not able to sequence 3 samples for ND2 and for 2 samples for CO1, this indicates similar efficiency of ND2 and CO1 primers.

Genetic distances between closely related species in the ND2 tree were 1.8 - 5.3 times greater than those in the CO1 tree (Fig. 1; Table 1). When number of segregating sites is considered, the differences are even greater due to the differences in sequence length of the two markers.

Both the differences in evolutionary rate and the sequence length between the two loci had a strong effect on posterior probability of monophyly of conspecific haplotypes. Although both loci failed to support reciprocal monophyly of *Aythya ferina* and *A. valisineria*, all other species-specific clades were strongly supported in the ND2 tree (Fig. 1). The monophyly of *A. affinis* had 99% posterior probability (PP), whereas the monophyly of all other species (*marila*, *fuligula*, *nyroca*, *americana* and *collaris*) had 100% PP (Table 2). In contrast, only monophyly of *A. fuligula* and *A. collaris* was statistically significant (both PP values = 1), whereas support for monophyly of other species varied between 23% and 91%. This comparison suggests that ND2 sequence has much greater probability of correct assignment to a species.

Table 1. Genetic distance between closely related species (% sequence divergence per site)

| Species | ND2 | CO1 |
|---|---|---|
| *A. affinis* | 99% | 91% |
| *A. marila* | 100% | 23% |
| *A. fuligula* | 100% | 100% |
| *A. nyroca* | 100% | 87% |
| *A. americana* | 100% | 44% |
| *A. collaris* | 100% | 100% |
| *A. ferina/valisineria* | 100% | 100% |

Table 2. Posterior Probability values for monophyly of species

| Clade | ND2 | CO1 |
|---|---|---|
| *A. affinis, A. marila* | 0.41% | 0.21% |
| *A. affinis, A. marila , A. fuligula* | 2.79% | 0.77% |
| *A. australis, A. nyroca* | 0.92% | 0.51% |
| *A. americana, A. collaris* | 2.06% | 0.38% |
| *A. americana, A. collaris, A. ferina* | 2.40% | 0.81% |

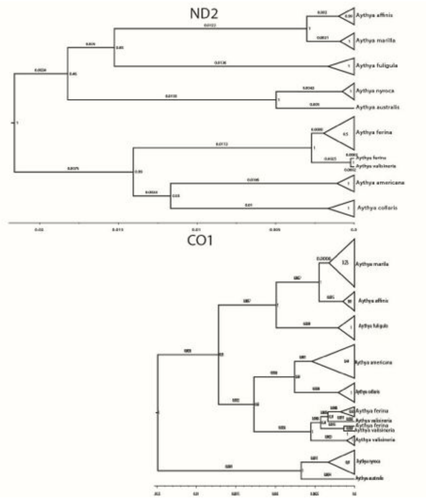
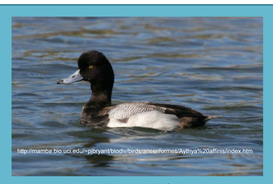
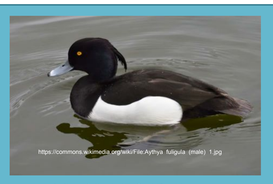



Figure 1. Comparison of phylogenetic trees using DNA makers ND2 (top) and CO1 (bottom).


*Aythya marila* (Greater Scaup)



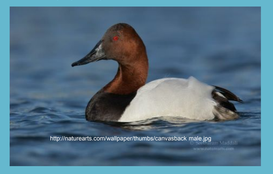

*Aythya affinis* (Lesser Scaup)




*Aythya fuligula* (Tufted Duck)


*Aythya nyroca* (Ferruginous duck)

*Aythya australis* (Hardhead)




*Aythya valisineria* (Cavasback)


*Aythya collaris* (Ring-necked duck)


*Aythya americana* (Redhead)


*Aythya ferina* (Common Pochard)

**Damani Eubanks (NHRE Intern) and Faridah Dahlan (Co-sponsor) in the L.A.B. facility.**

## Discussion

We compared performance of two mtDNA markers (ND2 and CO1) for the molecular identification of species in a taxonomically challenging genus *Aythya* (Aves Anatidae). The standard protocol developed by the Barcode of Life Database (BoLD) network recommends using a fragment of the mtDNA CO1 gene. However, in some cases, CO1 appears to not have enough resolution to distinguish closely related taxa (Dove et al. 2013) and many ornithologists working in avian phylogeography prefer ND2 as a mtDNA marker for their studies. Our comparison reveled that the values for divergence between closely related species in the ND2 dataset were 2-5 times greater than the CO1 divergence for the same sister species and clades. Except for a single pair of sibling species, which could not be separated by either marker, all 6 species we used in our comparison were grouped into clades by ND2 sequences with probability ≥ 0.99 CO1 failed to statistically support monophyly of 4 of the 6 species (0.23 ≤ PP ≤ 0.91). Our comparison suggests that ND2 is more effective marker for distinguishing different species within the *Aythya* genus than standard barcoding marker (CO1).

Accurate species identification is essential for birdstrike cases involving *Aythya* species because these birds are typically large, heavy and are known to cause damage to aircraft. We therefore recommend using ND2 in the cases involving *Aythya* ducks when species identifications are important to the investigation (i.e. damaging strikes).

Neither marker was able to distinguish all species within *Aythya*. *A. ferina* and *A. valisineria* lacked reciprocal monophyly in both locus-specific trees (Figure 1) which renders both markers incapable of confident molecular identification. Fortunately, these two species do not overlap geographically so identifications are still possible if the sequence analysis assigns a sample to the *A. ferina* + *A. valisineria* clade.

This project will enable the Feather Lab to determine more effective genetic markers to identify species within the genus *Aythya*. Further implications of the project could include testing other closely related taxa of birds to determine whether ND2 is more effective in distinguishing those species as well.

## Acknowledgements

I would like to thank Carla Dove, Faridah Dahlan, Sergei Drovetski and everyone in the Feather Identification lab and the Division of Birds for all the help and support on this project. All of the molecular laboratory work was conducted with the support of the L.A.B. facility of the National Museum of Natural History. I would also like to thank Gene Hunt, Elizabeth Cottrell and Virginia Power for making this experience possible with their countless hours of work and support. Finally I would like to thank NMNH, the Smithsonian Institution and NSF for providing funding for this project. In addition to finding an effective genetic marker useful for distinguishing *Aythya* I also learned valuable information about conducting research and working in a molecular lab that will help me in the future. While conducting the research for my project I learned about obstacles encountered while working in a lab. Tissue loan processing time, training on new equipment, lack of experience in a working lab, extraction failures, variable PCR results, sample contamination, labeling errors and human contamination were all impediments to this project.

Overall, I would say that my project was a success and that I thoroughly enjoyed my summer working at with the Smithsonian.

## References

Dove C, Dahlan N, Drovetski S (2013) MtDNA ND2 sequence identifies Streaked Horned Lark (*Eremophila alpestris strigata*) from birdstrike to US Air Force F-15 at Portland International Airport, Oregon. *Conservation Genetics Resources* **5**, 997-999.

Drovetski SV, Zink RM, Fadeev IV, *et al.* (2004) Mitochondrial phylogeny of *Locustella* and related genera. *Journal of Avian Biology* **35**, 105-110.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.

Hackett SJ (1996) Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). *Molecular Phylogenetics and Evolution* **5**, 368-382.

Lerner HRL, Meyer M, James HF, Hofreiter M, Fleischer RC (2011) Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biology* **21**, 1838-1844.

Posada D (2008) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In: *Some Mathematical Questions in Biology - DNA Sequence Analysis* (ed. Miura RM), pp. 57-86. American Mathematical Society, Providence.