J. Field Ornithol. 87(1):96–103, 2016 DOI: 10.1111/jofo.12132

Overcoming challenges to morphological and molecular identification of *Empidonax* flycatchers: a case study with a Dusky Flycatcher

Erin L. Heller,[1] Kevin C. R. Kerr,[2] Nor F. Dahlan,[3] Carla J. Dove,[3] and Eric L. Walters[1,4]

[1] *Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529, USA*
[2] *Toronto Zoo, 361A Old Finch Avenue, Toronto, Ontario M1B 5K7, Canada*
[3] *Division of Birds, Smithsonian Institution, P. O. Box 37012, MRC 116, Washington, D.C. 20013, USA*

Received 10 July 2015; accepted 5 October 2015

ABSTRACT. Flycatchers in the genus *Empidonax* are among the most difficult avian taxonomic groups to identify to species. Observers often rely on calls or songs in the field or detailed morphometrics in the hand to identify species. In January and February 2013, we twice captured an *Empidonax* flycatcher at the Virginia Zoo in Norfolk, Virginia. After being unable to identify the flycatcher to species level using morphometrics and photographs, we extracted DNA from two tail feathers collected during the second encounter to identify the individual genetically. Comparison of cytochrome *c* oxidase I (COI) with reference sequences in the Barcode of Life Database (BOLD) suggested that the specimen had a >99.8% probability of placement as a Dusky Flycatcher (*Empidonax oberholseri*). Additional comparisons of NADH dehydrogenase subunit 2 (ND2) to reference sequences in GenBank, however, suggested that the specimen was a Pine Flycatcher (*Empidonax affinis*), a species not represented in BOLD and confined geographically to a small area in Mexico and Guatemala. After analyzing both COI and ND2 from additional vouchered specimens, the bird caught in Virginia was determined to be a Dusky Flycatcher. We also suspect that some of the sequences in GenBank might derive from incorrectly identified specimens or otherwise could represent overlooked pseudogenes. Because the putative identification, based on GenBank sequences, would have represented the first record of Pine Flycatcher from the United States, our results reinforce the need for carefully vetted and taxonomically comprehensive molecular databases to allow definitive conclusions about sample identity. Further molecular phylogeographic review of this genus is warranted to resolve haplotype ambiguities.

RESUMEN. **Retos sobre la identificación morfológica y molecular de papamoscas del genero *Empidonax*: un caso de estudio con *Empidonax oberholseri***

Los papamoscas Empidonax se encuentran entre el grupo de aves más difíciles de identificar taxonómicamente a nivel de especies. Para identificar estos pájaros, los observadores dependen, particularmente, del canto o de las llamadas o de detalles que se pueden observar cuando tienen el ave en la mano En enero y febrero del 2013, en dos ocasiones capturamos un *Empidonax* en el Zoológico de Virginia, Virginia. No pudimos identificar la especie utilizando fotografías o rasgos morfométricos y a tales efectos extrajimos ADN de dos de las plumas del rabo, para tratar de identificar el segundo individuo genéticamente. La comparación de la oxidasa c citocromica I (COI), con la referencia secuencial en la base de datos del "Código de Barras de la Vida" (BOLD) sugirió que el espécimen tenía >99.8% de probabilidad de ser un *Empidonax oberholseri*. No obstante, una comparación de NADH dehidrogenasa, subunidad 2 (ND2) del Banco Genético (GenBank) sugería que el ave era *Empidonax affinis*, una especie que no estaba representada en BOLD y geográficamente confinada a un área pequeña de México y Guatemala. Luego de analizar el COI y el ND2 de especímenes adicionales, se determinó que el ave capturada en Virginia era *E. oberholseri*. Sospechamos que la misma secuencia en el GenBank se había obtenido de un individuo mal identificado, que de otra manera hubiera representado un pseudogene, pasado por alto. Debido a que la identificación putativa, basado en la secuencia del GenBank, hubiera representado el primer registro de un *E. affinis* en los Estados Unidos, nuestros resultados apoyan la necesidad de tener mucho cuidado con las bases moleculares para permitir la identificación conclusiva de muestras. Se necesita una revisión molecular filogeográfica de los *Empidonax* para resolver ambigüedades haplotípicas.

Key words: Barcoding, BOLD, *Empidonax affinis*, *Empidonax oberholseri*, GenBank

Empidonax flycatchers are non-descript and morphologically similar insectivorous passerines in the family Tyrannidae (Sedgwick 1993, Johnson and Cicero 2002). Because of their non-descript plumage, these flycatchers are difficult to identify visually (Phillips et al. 1966, Johnson and Cicero 2002, Novitch et al. 2015). Of the 15 *Empidonax* species found in North America, many have ranges that overlap. Range overlaps become more pronounced during fall and spring migration when individuals are typically less

[4] Corresponding author. Email: ewalters@odu.edu

Fig. 1. Dusky Flycatcher (*Empidonax oberholseri*) caught on 27 January 2013 (left) and 23 February 2013 (right) in a mist-net at the Virginia Zoo in Norfolk, Virginia (photo credit: Erin L. Heller).

vocal (Sedgwick 2001). This leads to further challenges in correctly identifying individuals to species. Here, we summarize the identification process for an unknown *Empidonax* flycatcher captured during the winter in Virginia, USA. We describe the process by which we confirmed its identity and highlight the difficulty associated with morphological ambiguities and errors in genetic repositories for the genus *Empidonax*.

METHODS

On 27 January 2013, we captured an *Empidonax* flycatcher (Fig. 1) in a mist-net at the Virginia Zoo (Norfolk, VA; 36.8786°N, 76.2774°W). The bird was not banded because identification was not certain. An *Empidonax* flycatcher was again caught at the same location on 23 February 2013, presumably the same individual we captured in January. While being removed from the net, the bird shed four tail rectrices that we collected for later examination. Again, we photographed the bird (Fig. 1) before releasing it.

Two of four rectrices collected were used for genetic analysis at the Feather Identification Lab (Smithsonian Institution, Washington, D.C.). Calami were excised from the rectrices using sterile scissors and DNA extraction methods followed Dove et al. (2013). No other *Empidonax* specimens were handled in the lab at this time. Using the extracted DNA, a fragment of the cytochrome *c* oxidase I (COI) gene known as the "DNA barcode" was amplified using polymerase chain reaction (PCR). The gene fragment was sequenced and compared against a library of known DNA barcodes in the Barcode of Life Database (BOLD) using the BOLD Identification System (Ratnasingham and Hebert 2007; http://www.boldsystems.org/), which included 11 of the 15 *Empidonax* species. We then sequenced NADH dehydrogenase subunit 2 (ND2), for which all 15 *Empidonax* taxa were represented in GenBank.

Because the Dusky Flycatcher was only represented in GenBank by a single specimen, we sequenced both COI and ND2 genes from five additional museum specimens of Dusky Flycatchers and also from five museum specimens of the closely related Pine Flycatcher (*Empidonax affinis*; Table 1). PCR and sequencing methods followed Dove et al. (2008) for COI and Dove et al. (2013) for ND2. We constructed

Table 1. Samples used for comparison with the Barcode of Life Database and GenBank.

Museum number	Identified species	Date	Locality	Putative species
UWBM 82521	*E. affinis*	11 August 2006	Los Mimbres, Durango, MX	*E. affinis*
UWBM 82897	*E. affinis*	13 May 2006	Xocomanatlan, Guerrero, MX	*E. oberholseri*
UWBM 99693	*E. affinis*	23 January 2009	Ixtlan, Oaxaca, MX	*E. affinis*
UWBM 99851	*E. affinis*	31 January 2008	Teopisca, Chiapas, MX	*E. affinis*
UWBM 107797	*E. affinis*	25 April 2003	Bolaños, Jalisco, MX	*E. hammondii*
UWBM 80199	*E. oberholseri*	9 June 2003	Cle Elum WA	*E. oberholseri*
UWBM 80291	*E. oberholseri*	10 August 2003	Blue Lake, WA	*E. oberholseri*
UWBM 85193	*E. oberholseri*	27 May 2004	Mendocino Natl Forest, CA	*E. oberholseri*
UWBM 90450	*E. oberholseri*	16 June 2010	Looking Glass, OR	*E. oberholseri*
UWBM 104552	*E. oberholseri*	16 July 2002	Cass House Peak, NV	*E. oberholseri*
N/A	Unknown	23 February 2013	Norfolk, VA	*E. oberholseri*

UWBM, University of Washington Burke Museum.

maximum likelihood trees with additional *Empidonax* sequences from GenBank using PhyML 3.0 (Guindon et al. 2010) and added Fuscous Flycatcher (*Cnemotriccus fuscus*) sequences from GenBank as an outgroup.

RESULTS

The COI sequence (516 bp) from the unknown bird matched a reference sequence in BOLD for Dusky Flycatcher with the exception of one nucleotide, and had a 99.8% probability of placement based on the BOLD Identification System (Ratnasingham and Hebert 2007). Dusky Flycatchers are closely related to Pine Flycatchers (Johnson and Cicero 2002), a species not included in the BOLD library. Sequences from Johnson and Cicero (2002) demonstrated that these two taxa were virtually identical at the 3′ end of COI, but were more easily distinguished by ND2. We subsequently amplified the ND2 sequence for the unknown bird and compared it with samples in GenBank, which resulted in a closer match to the single Pine Flycatcher sequence in that database. Thus, based on BOLD matches, the unknown specimen was a Dusky Flycatcher, whereas, based on GenBank and ND2, it was a Pine Flycatcher. Both the COI and ND2 sequences from the unknown bird are provided in Supplementary Table S1.

We produced COI sequences for all but one Dusky Flycatcher (UWBM 104552) and one Pine Flycatcher (UWBM 99851), and ND2 was produced for all but one Pine Flycatcher (UWBM 99693) specimen. All new sequences from vouchered museum specimens were submitted to GenBank (accession numbers: KT806379-KT806395). Specimen UWBM 107797, originally identified as a Pine Flycatcher when collected, was identified as a Hammond's Flycatcher (*E. hammondii*) based on COI (Fig. 2). However, the ND2 sequence amplified for this individual was uniquely divergent and could represent a nuclear-mitochondrial pseudogene (see Discussion).

None of the five putative Dusky Flycatcher specimens matched the two ND2 sequences from GenBank for Dusky Flycatcher (Fig. 3). The ND2 sequence from GenBank for the Pine Flycatcher belonged to a clade containing the unknown specimen, four of the new Dusky Flycatcher sequences, and one of the new sequences for the Pine Flycatcher (UWBM 82897). The remaining two new Pine Flycatcher sequences formed their own unique clade, sister to the aforementioned clade. The COI results largely mirrored those of ND2 (Fig. 2), with the addition that UWBM 99693 also joined the new Pine Flycatcher clade, although with COI it appears to be the sister clade to Hammond's Flycatcher. Ultimately, the COI and ND2 sequences from the unknown Virginia bird matched clades that primarily include Dusky Flycatcher.

DISCUSSION

The COI sequence from our unknown specimen unambiguously matched a reference

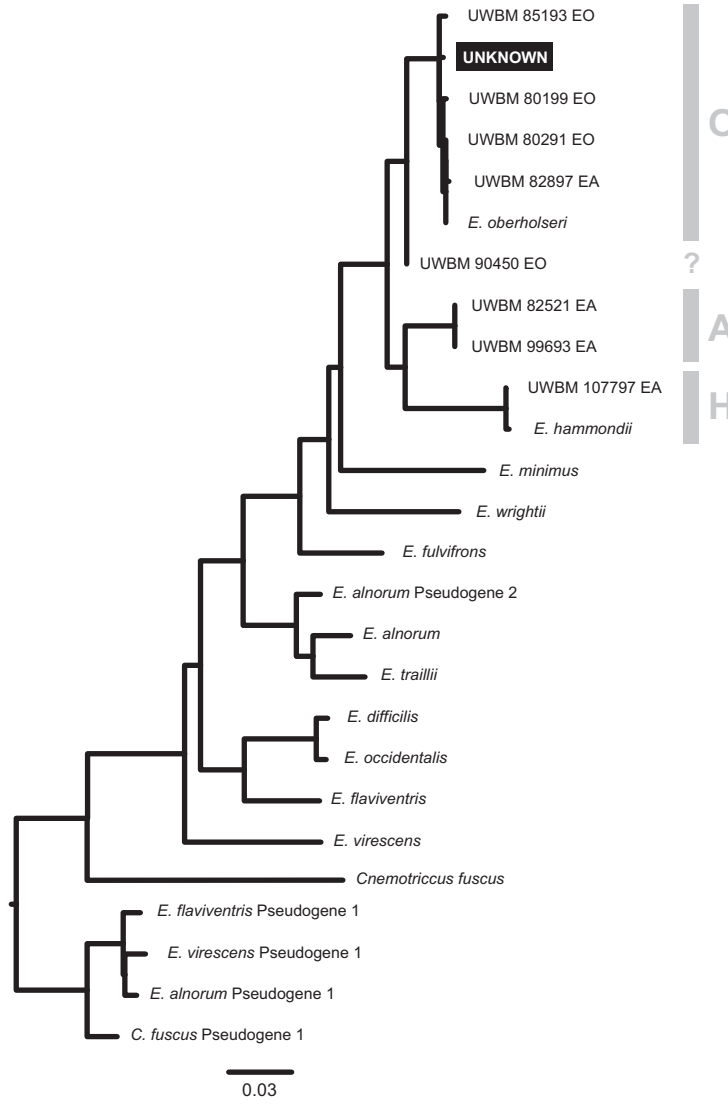

Fig. 2. Maximum likelihood tree based on cytochrome *c* oxidase I (COI) for 11 of the 15 North American *Empidonax* flycatchers, including Hammond's Flycatcher (*E. hammondii*), Dusky Flycatcher (*E. oberholseri*), Least Flycatcher (*E. minimus*), Gray Flycatcher (*E. wrightii*), Buff-breasted Flycatcher (*E. fulvifrons*), Willow Flycatcher (*E. traillii*), Alder Flycatcher (*E. alnorum*), Yellow-bellied Flycatcher (*E. flaviventris*), Pacific-slope Flycatcher (*E. difficilis*), Cordilleran Flycatcher (*E. occidentalis*), Acadian Flycatcher (*E. virescens*), and a sequence from the Fuscous Flycatcher (*Cnemotriccus fuscus*) as an outgroup. Five sequences from two clades of pseudogenes are also included for reference. The tree was run in PhyML 3.0 using a general time-reversible evolutionary model (Guindon et al. 2010). The suffix "EA" indicates specimens identified upon collection as the Pine Flycatcher (*E. affinis*); the suffix "EO" indicates those identified as *E. oberholseri*. The gray bars identify clades believed to belong to *E. affinis* (A), *E. oberholseri* (O), or *E. hammondii* (H). The scale represents substitutions per site. GenBank accession numbers for sequences not generated in this study include AY666171, DQ432910, DQ432913, DQ432914, DQ433602, DQ433606, DQ433613, DQ433617, DQ433622, DQ433624, GU116555, GU116559, HM033459, and HM396218. BOLD Process ID numbers for additional pseudogene sequences include PTYRN010-11, PTYRN020-12, and PTYRN023-12.

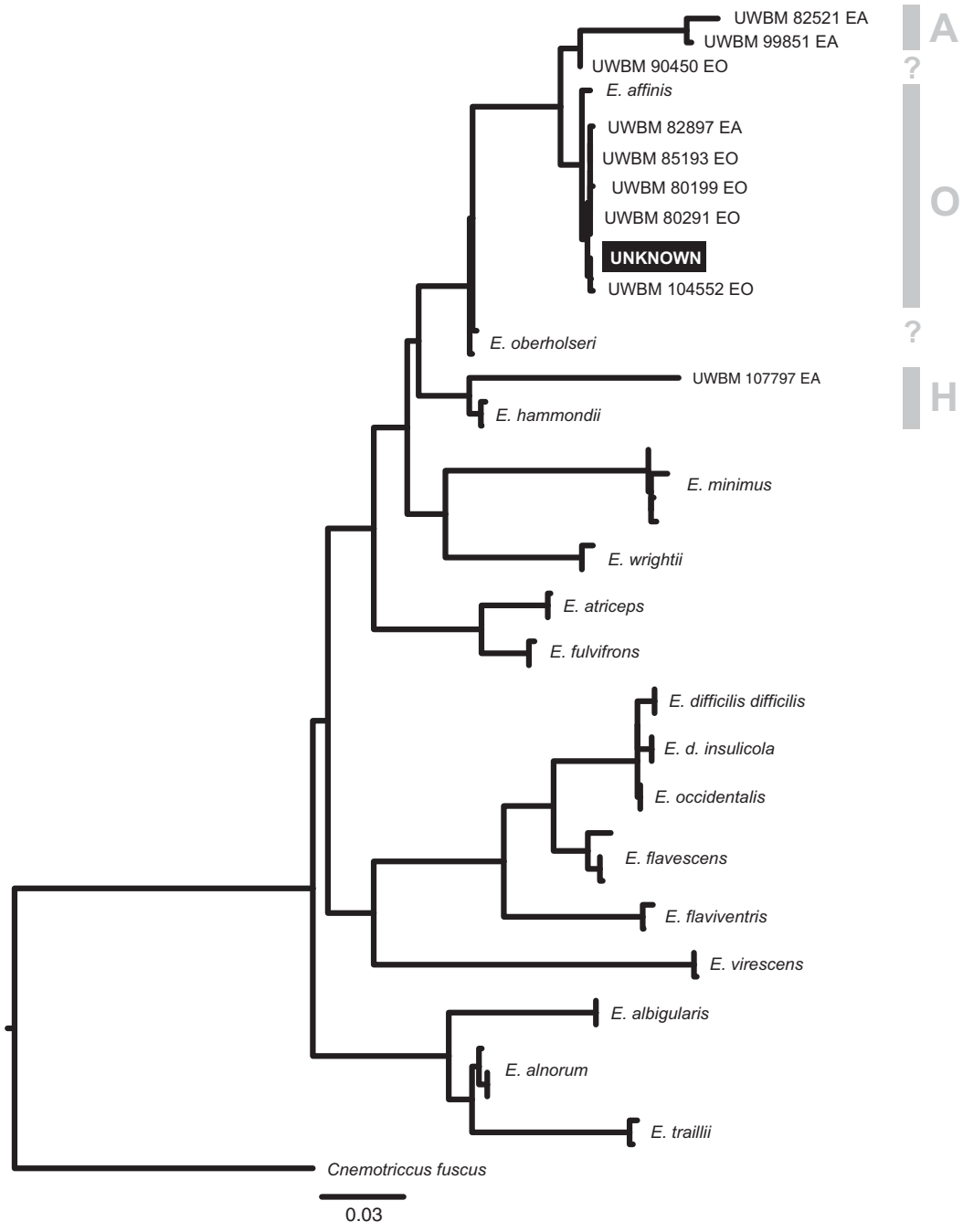

Fig. 3. Maximum likelihood tree based on NADH dehydrogenase subunit 2 (ND2), including all 15 North American *Empidonax* flycatchers plus a sequence from the Fuscous Flycatcher (*Cnemotriccus fuscus*) as an outgroup. The tree was run in PhyML 3.0 using a general time-reversible evolutionary model (Guindon et al. 2010). The suffix "EA" indicates specimens identified upon collection as *E. affinis*; the suffix "EO" indicates those identified as *E. oberholseri*. The gray bars identify clades believed to belong to *E. affinis* (A), *E. oberholseri* (O), or *E. hammondii* (H). The scale represents substitutions per site. Species names are provided in Fig. 2. GenBank accession numbers for sequences not generated in this study include AF447624-AF447629, AF447649, AY030124, AY030125, AY143209-AY143234, and DQ294544.

sequence for Dusky Flycatcher in BOLD, but only 11 of 15 *Empidonax* species were represented in the database. Comparing an ND2 sequence from the unknown specimen revealed a match to a purported Pine Flycatcher sequence in GenBank, but, based on analysis of additional specimens, the single voucher of the putative Pine Flycatcher (FMNH 393990) was probably originally misidentified.

For the two species we focused on, we would have expected to recover two distinctive clades, one representing each species, but, instead, we found five. Possible explanations for this include misidentified specimens and nuclear-mitochondrial pseudogenes. The former is highly probable given that many of the voucher specimens were collected in parts of Mexico where the winter ranges of North American migrant species overlap the ranges of local endemics (Johnson 1963, Morton and Pereyra 1985). For one specimen, both factors might have been at play. The putative Pine Flycatcher (UWBM 107797) produced a COI sequence that matched those of Hammond's Flycatcher, but its ND2 sequence, though closely allied to Hammond's Flycatcher, was unique. Given the expected linkage between COI and ND2, this suggests that one of the two ND2 haplotypes ("H" in Fig. 3) could represent a nuclear-mitochondrial pseudogene. This is further supported by the variation in branch length because nuclear-mitochondrial pseudogenes tend to exhibit slower substitution rates than mitochondrial genes (Bensasson et al. 2001).

The other extra haplotypes could also be attributed to nuclear-mitochondrial pseudogenes; in the absence of geographic isolation, mutation alone is an unlikely explanation given the degree of divergence. A scan of the sequences, however, did not reveal the typical hallmarks of pseudogenes (e.g., indels and frameshift mutations), but similar "cryptic" nuclear-mitochondrial pseudogenes have been identified previously from this avian family (Kerr 2010, Stoeckle and Kerr 2012). Based on its basal position on the tree (Fig. 3), we suggest that the ND2 haplotype reported for *E. hammondii* by Johnson and Cicero (2002) is more likely a nuclear-mitochondrial pseudogene. Of the three remaining haplotypes, one (UWBM 90450) is unique and discussed further. Of the other two, one consists of three Pine Flycatcher specimens, and the other includes all of the Dusky Flycatcher specimens as well as two Pine Flycatcher specimens and our unknown specimen. The Pine Flycatcher ND2 sequence in GenBank (accession number AY143209) could be from a misidentified specimen. Putative Pine Flycatcher UWBM 82897 was collected in southern Mexico, and re-extraction confirmed the bird's placement in a cluster with Dusky Flycatcher, suggesting that this specimen too might have been misidentified. There are records of Dusky Flycatchers from that part of Mexico at that time of the year, supporting the plausibility of a misidentification (Sedgwick 1993). Hybridization and introgression are unlikely explanations because the breeding ranges of these two species do not overlap (Nelson 1901, AOU 1983, Godfrey 1986).

Putative Dusky Flycatcher UWBM 90450 had unique sequences for both COI and ND2, but both were still most similar to the expected species identity (Dusky Flycatcher). Although this could represent a mutation, multiple copies of long and convincing pseudogenes have been recovered in other congeners, such as Alder Flycatchers (*E. alnorum*; Stoeckle and Kerr 2012), and so the presence of pseudogenes remains a possible explanation. Nonetheless, because we confirmed that sequences from our unknown specimen unequivocally matched the sequences of specimens collected within the breeding range of Dusky Flycatchers in North America, we believe that our unknown specimen is a Dusky Flycatcher. Misidentification is the most parsimonious explanation for the non-complying Pine Flycatcher specimens from Mexico.

This is the first record of a Dusky Flycatcher from Virginia (Rottenborn and Brinkley 2007, eBird 2015). However, vagrants of this species have been recorded at several other locations in eastern North America, including Pennsylvania (1969, Uhrich 2000), Ontario (1993, Ridout 1994), Delaware (2002, Burgiel et al. 2002), three records in Alabama (2009, Alabama Ornithological Society 2011), and Georgia (2012, Georgia Ornithological Committee 2013).

Our results reinforce the need for carefully vetted and taxonomically comprehensive molecular databases to allow definitive conclusions about sample identity. Deeper and more comprehensive genetic review of this and other Tyrannid genera is still warranted. Given their propensity for harboring nuclear-mitochondrial pseudogenes, extreme care should be exercised

when employing mitochondrial markers with Tyrant flycatchers. Measures can be taken to avoid inadvertently sequencing pseudogenes (Sorenson and Quinn 1998), but due to their cryptic nature in this group, whole genomic sequencing might be necessary to reveal the true extent of mitochondrial translocation that has occurred in this family. Ultimately, further study is needed to document geographical and genomic variation in the *Empidonax* genus and refine morphological distinctions between the species.

ACKNOWLEDGMENTS

We thank R. Sweeney and J. Lotz of the Virginia Zoo for granting us access to mist-net and band birds, the volunteer field assistants (A. Johnson, L. August, E. Cali, and S. Haskell) who assisted on the days when the Dusky Flycatcher was caught, and E. Brinkley who assisted in attempting to visually identify the bird and provided valuable input that aided in writing this paper. We also thank S. Birks, Genetic Resources Manager, at the Burke Museum in Seattle, WA, for sending specimens, D. Barshis and three anonymous reviewers for comments on a previous version of this manuscript, and G. Ritchison for helpful suggestions to improve the manuscript. The field component of this study was supported by awards to ELH by the Virginia Academy of Science Small Project Research Funds and the Virginia Society of Ornithology JJ Murray Research Award, and through start-up funds to ELW provided by Old Dominion University. The Smithsonian Feather Lab is supported by Interagency Agreements with the U.S. Air Force, Navy, and Federal Aviation Administration. The research reported here was conducted under permits from Old Dominion University, U.S. Department of the Interior, U.S. Fish and Wildlife Service, and the Virginia Department of Game & Inland Fisheries.

LITERATURE CITED

Alabama Ornithological Society [online]. 2011. Alabama Bird Records Committee review list. <http://www.aosbirds.org/ABRCreports/abrc_review%20list_march%202011_update.pdf> (Accessed 10 July 2015).

American Ornithologists' Union (AOU). 1983. Check-list of North American birds, 6th ed. American Ornithologists' Union, Washington, D.C.

Bensasson, D., D. X. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology and Evolution 16: 314–321.

Burgiel, J. C., R. O. Paxton, and D. A. Cutler. 2002. Hudson-Delaware. North American Birds 56: 158–161.

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

———, N. C. Rotzel, M. Heacker, and L. A. Weigt. 2008. Using DNA barcodes to identify bird species involved in birdstrikes. Journal of Wildlife Management 72: 1231–1236.

EBIRD [online]. 2015. eBird: an online database of bird distribution and abundance. eBird, Ithaca, NY. <http://www.ebird.org> (Accessed 10 July 2015).

Georgia Ornithological Society [online]. 2013. Georgia Ornithological Checklists & Records Committee Activity. <http://www.gos.org/gcrc-activity-2013> (Accessed 10 July 2015).

Godfrey, W. E. 1986. The birds of Canada, revised ed. National Museum of Natural Science, Ottawa, ON, Canada.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59: 307–321.

Johnson, N. K. 1963. Biosystematics of sibling species of flycatchers in the *Empidonax hammondii-oberholseri-wrightii* complex. University of California Press, Berkeley, CA.

———, and C. Cicero. 2002. The role of ecologic diversification in sibling speciation of *Empidonax* flycatchers (Tyrannidae): multigene evidence from mtDNA. Molecular Ecology 11: 2065–2081.

Kerr, K. C. R. 2010. A cryptic, intergeneric cytochrome *c* oxidase I pseudogene in tyrant flycatchers (family: Tyrannidae). Genome 53: 1103–1109.

Morton, M. L., and M. E. Pereyra. 1985. The regulation of egg temperatures and attentiveness patterns in the Dusky Flycatcher (*Empidonax oberholseri*). Auk 102: 25–37.

Nelson, E. W. 1901. Descriptions of five new birds from Mexico. Auk 18: 46–49.

Novitch, N. R., M. Westberg, and R. M. Zink. 2015. Migration of Alder Flycatchers (*Empidonax alnorum*) and Willow Flycatchers (*Empidonax traillii*) through the Tuxtla Mountains, Veracruz, Mexico, and the identification of migrant flycatchers in collections. Wilson Journal of Ornithology 127: 142–145.

Phillips, A. R., M. A. Howe, and W. E. Lanyon. 1966. Identification of the flycatchers of eastern North America, with special emphasis on the genus *Empidonax*. Bird-Banding 37: 153–171.

Ratnasingham, S., and P. D. N. Hebert. 2007. The barcode of life data system. Molecular Ecology Notes 7: 355–364.

Ridout, R. 1994. The autumn migration August 1-November 30, 1993: Ontario region. American Birds 48: 81–156.

Rottenborn, S. C., and E. S. Brinkley (eds.). 2007. Virginia's birdlife: an annotated checklist. Virginia Society of Ornithology, Ervington, VA.

Sedgwick, J. A. 1993. Dusky Flycatcher (*Empidonax oberholseri*). In: The Birds of North America, No. 78 (A. Poole, and F. Gill, eds.), The Academy of Natural Sciences, Philadelphia, PA, and The American Ornithologists' Union, Washington, D.C.

———. 2001. Geographic variation in the song of Willow Flycatchers: differentiation between *Empidonax*

traillii adastus and *E. t. extimus*. Auk 118: 366–379.

SORENSON, M. D., AND T. W. QUINN. 1998. Numts: a challenge for avian systematics and population biology. Auk 115: 214–221.

STOECKLE, M. Y., AND K. C. R. KERR. 2012. Frequency matrix approach demonstrates high sequence quality in avian BARCODEs and highlights cryptic pseudogenes PLoS ONE 7: e43992.

UHRICH, B. 2000. Reading eagle: lifestyle. Reading Eagle Company, Reading, PA.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Table S1. Sequence fragments in FASTA format of COI and ND2 obtained from the unknown *Empidonax* flycatcher.