DNA-identification of a Dutch Semicollared Flycatcher *Ficedula semitorquata*, May 2021.

Peter de Knijff
Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.

This document with public access URL https://doi.org/10.5281/zenodo.7547214 was prepared as supplementary information to the following article to appear in Dutch Birding: *Balkanvliegenvanger bij Putten in mei 2021. Written by Jan Stronks, Folkert Jan Hoogstra, Gijs Baller & Enno B Ebels*

Summary
In early July 2021 I was approached by FJ Hoogstra (FJH) and Gijs Baller (GB), with the request if it was possible to identify a putative Semicollared Flycatcher *Ficedula semitorquata* based on fecal samples. The main problem was to exclude a possible hybrid between European Pied Flycatcher *F. hypoleuca* and Collared Flycatcher *F. albicollis*. Such hybrids can be deceptively similar to *semitorquata*. By sequencing an mtDNA ND6 fragment of 324 bp. in the fecal samples, the bird was found to be indeed a Semicollared Flycatcher *Ficedula semitorquata.*

Introduction
In early July 2021 I was approached by FJ Hoogstra (FJH) and Gijs Baller (GB), with the request if it was possible to identify a putative Semicollared Flycatcher *Ficedula semitorquata* based on fecal samples. The main problem was to exclude a possible hybrid between European Pied Flycatcher *F. hypoleuca* and Collared Flycatcher *F. albicollis*. Such hybrids can be deceptively similar to *semitorquata*.
I proposed to use a two-step approach. First a mitochondrial DNA analysis, using a short fragment of the ND6 gene of which sufficient *Ficedula* reference sequences were available in GenBank, including a few of *semitorquata*. If this would indicate either *albicollis* or *hypoleuca*, a second (autosomal) DNA analysis would be needed. If this mtDNA analysis would indicate *semitorquata*, a second autosomal DNA analysis would not be necessary.

DNA-sequencing
DNA-isolation
DNA-isolation from fecal samples was performed using the QIAamp Fast DNA Stool Mini Kit according to the manufacturer's instructions.

PCR reaction
In order to sequence a fragment of 324 bp. of the mitochondrial ND6 gene, two overlapping monoplex PCR's were performed with an input of 2.5-5µL of DNA-extract (see Table 1 for PCR-primer sequences). The PCR-mix contained Geneamp™ 10x PCR-buffer1 (Applied biosystems), 0.2mM dNTP's (GE healthcare), 0.4pmol per primer (Table 1) and 2U AmpliTAq gold DNA polymerase (Applied biosystems) in a total volume of 50µL. PCR's were run on a GeneAmp® PCR System 9700 with the following program. 94°C for 10 min, 36 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 2 minutes ending with 72°C for 10 minutes.

PCR purification
The PCR-products were visualized using the QIAxcel. Afterwards a purification step using the QIAquick® PCR purification kit (QIAGEN) was performed according to the protocol from this kit. Elution was performed in 20-70µL Aquabraun depending on the amount of PCR-product visible on the Qiaxcel.

Sequencing PCR
Forward and reverse sequencing PCR's (see Table 2 for sequencing primers) were performed using an input of 1-4µL of purified PCR-product. PCR-mix contained: 5x sequencing buffer big dye terminator V1.1 v3.1 (Life Technologies), 0.6pmol sequencing primer (tabel2) and 2µL BigDye® Terminator v3.1 ready reaction mix (Life Technologies) in a total volume of 10µL. Sequencing PCR's were run on a pre-

heated (96°C) GeneAmp® PCR System 9700 with the following program. 96°C for 1 min, 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes.

Sequencing PCR purification
12µL water was added to the sequencing PCR-product. Then the product was purified using the DyeEx® 2.0 Spin kit (QIAGEN), protocol for Dye-Terminator Removal.

Sequencing
Purified sequencing PCR-product was run on an AB3100 Genetic Analyzer.

Table 1. Forward and reverse primer sequences for the ND6 sequencing of *Ficedula* samples. Lowercase sequences tgtaaaacgacggccagt and caggaaacagctatgacc are M13 tails.

Primer	Reverse primer sequence
Fice_ND6_set1_21M13f	5'-tgtaaaacgacggccagtGCGGGAATTTATGATGCAGTTTG-3'
Fice_ND6_set1_M13rev	5'-caggaaacagctatgaccTCCCCTCAAGCCTCAGG-3'
Fice_ND6_set2_21M13f	5'-tgtaaaacgacggccagtGCCTCCAACCCTTCTCCCTA-3'
Fice_ND6_set2_M13rev	5'- caggaaacagctatgaccCARCCGAACTGAAGACAGCC-3'

Table 2. Overview of the forward and reverse primer sequences for the sequencing PCR

	Forward	Reverse
Sequencing PCR primers	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC

The resulting sequence was submitted to GenBank and released as record number MZ813091.

Network preparation and analyses
As a first step, we combined all (n=17) ND6 *Ficedula* sequences available from GenBank (see Table 3), with those from the present study into a single fasta file. Sequences were aligned and clipped to an overlapping size of 324 bp. As a second step, we used DnaSP v6 (from http://www.ub.edu/dnasp/), to extract and export a polymorphic sites only file in RDF-format from this fasta file. The resulting rdf file was used to prepare a median joining network using Network 10.2 (from https://www.fluxus-engineering.com/sharenet.htm). The network was adjusted using Network Publisher (purchased via https://www.fluxus-engineering.com/sharenet.htm), and subsequently exported as an .emf file. This .emf file was edited in PowerPoint which resulted in Figure 1 below.

References
Ekblom R, Smeds L, Ellegren H. Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. BMC Genomics 2014; 15: 467.
Sætre GP, Moum T, Bures S, Kra M, Adamjank M, Moreno J. Asexually selected character displacement in flycatchers reinforces premating isolation. Nature 1997; 387: 589 – 592.
Sætre GP, Borge T, Moum T. A new bird species? The taxonomic status of 'the Atlas Flycatcher' assessed from DNA sequence analysis. Ibis 2001; 143: 494 – 497.
Sætre GP, Borge T, Lindell J, Moum T, Primmer CR, Sheldon BC, Haavie J, Johnsen A, Ellegren H. Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. Molecular Ecology 2001; 10: 737 – 749.

Table 3. Summary of all *Ficedula* mtDNA ND6 sequences deposited or obtained from GenBank and used in this study.

GenBank nr	taxon	Location	Source
Y10215	*iberiae*	Spanje	Saetre et al, 1997
Y10308	*hypoleuca*	Tsechoslowakije	Saetre et al, 1997
Y10309	*albicollis*	Tsechoslowakije	Saetre et al, 1997
Y10310	*semitorquata*	Armenie	Saetre et al, 1997
AJ400988	*hypoleuca*	Noorwegen	Saetre et al, 2001a
AJ400989	*semitorquata*	Griekenland	Saetre et al, 2001a
AJ400990	*speculigeria*	Marocco	Saetre et al, 2001a
AJ400991	*speculigeria*	Marocco	Saetre et al, 2001a
AJ400992	*speculigeria*	Marocco	Saetre et al, 2001a
AJ400993	*speculigeria*	Marocco	Saetre et al, 2001a
AJ400994	*speculigeria*	Marocco	Saetre et al, 2001a
AJ299683	*hypoleuca*	Tsechoslowakije	Saetre et al, 2001b
AJ299684	*iberiae*	Spanje	Saetre et al, 2001b
AJ299685	*albicollis*	Tsechoslowakije	Saetre et al, 2001b
AJ299686	*semitorquata*	Armenie	Saetre et al, 2001b
AJ299687	*semitorquata*	Griekenland	Saetre et al, 2001b
KF293721	*albicollis*	Zweden	Ekblom et al, 2014
MZ813091	*semitorquata*	Nederland	Baller et al, 2021

MZ813091 is the GenBank number of the bird described here. All other sequences were deposited by others (see references).

Figure 1. Network of mtDNA-ND6 genetic variation among 18 *Ficedula* individuals.

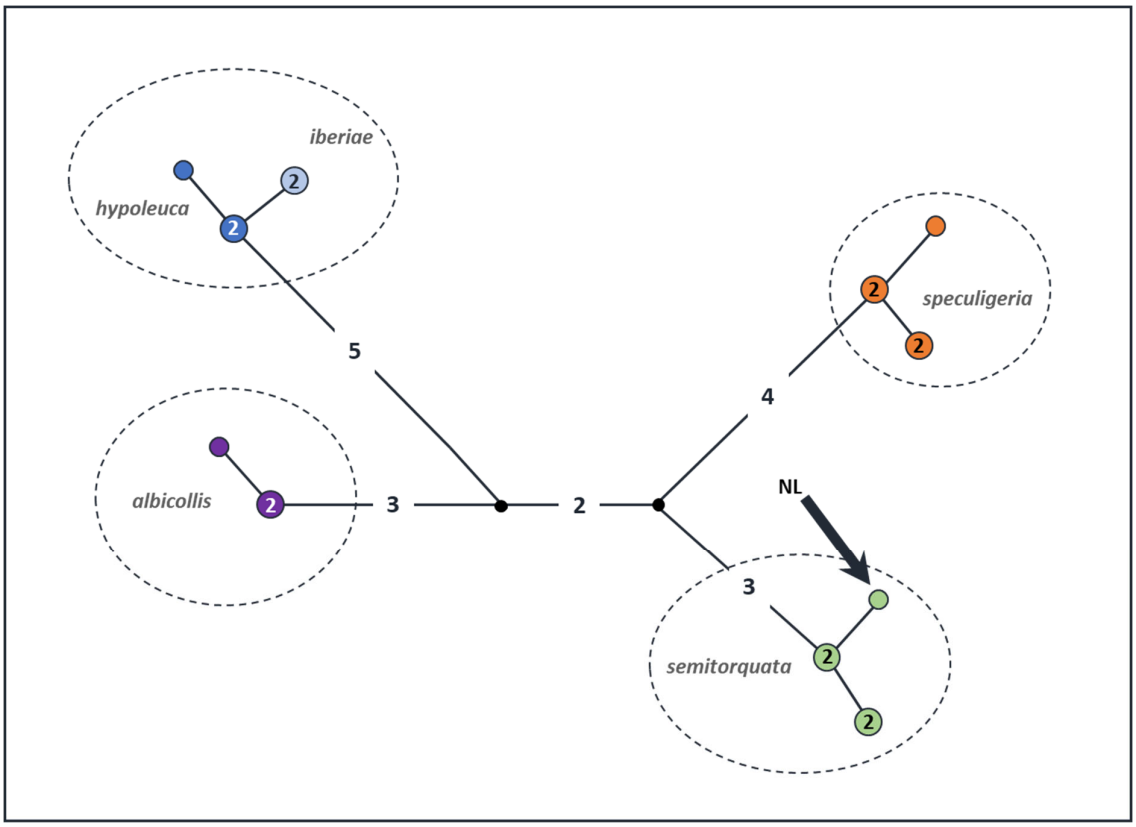

Figure 1. Network of variation in mtDNA ND6 gene fragment of *Ficedula* taxa. Each circle represents a unique ND6 sequence of 324 base pairs. Relative diameter of each circle is an indication of its frequency in the total dataset (n=18). Numbers inside circles: number of times that the sequence was observed. Circles without a number: sequences found in one individual only. Small black dots: inferred (not observed) sequences necessary to calculate and construct the network. Short lines between circles without number mark differences between two haplotypes on one position (out of total of 324 positions), other longer lines mark two or more (as indicated by numbers along lines) different positions. In this short ND6 sequence *hypoleuca (incl iberiae)*, *albicollis*, *speculigeria* and *semitorquata* can be reliably differentiated. *F. h. hypoleuca* en *F. h. iberiae* only differ at one (out of 324) positions. The ND6 sequence of the Dutch *F. semitorquata* is indicated with an black arrow