DNA-identification of the first Dutch field record of Lanceolated warbler *Locustella lanceolata* based on a single fecal sample.

Peter de Knijff, Rick de Leeuw, and Marije te Raa.
Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.

Summary
The identity of a "grasshopper warbler", discovered on October 4, 2016 at the Maasvlakte, Rotterdam, The Netherlands, was established by means of sequencing DNA that was isolated from a single small fecal sample collected in the field. Based on the sequence of short fragments of the mitochondrial cytochrome B (cytB) gene and three autosomal gene fragments, lactate dehydrogenase (LDH), myoglobine (MYO) and ornithine decarboxylase (ODC), the identification as a "pure" Lanceolatad warbler *Locustella lanceolata* was confirmed.

Lanceolated warbler *Locustella lanceolata*, Maasvlakte, Rotterdam, The Netherlands October 4, 2016. © Arnold Meijer. (source: https://www.dutchbirding.nl/gallery/detail/19166?page=0#navbar-collapse)

DNA-identification of the first Dutch field record of Lanceolated warbler *Locustella lanceolata* based on a single fecal sample.

Introduction

On October 4, 2016, a very confiding and skulking "grasshopper warbler" was discovered on the Maasvlakte, Rotterdam, The Netherlands. The bird was suspected to be a Lanceolatad warbler *Locustella lanceolata* (hereafter named *lanceolata*), an extremely rare bird in The Netherlands. If correct, it would constitute only the third Dutch record, after two birds captured at Bird-observatories and the first ever to be discovered in the field and " twitchable". Via the Dutch Birding alert-system, the Dutch twitching scene was informed, and the bird was seen by > 100 birdwatchers and extremely well documented (1,2) the same day. The next day, the bird was no longer present.

Soon after the initial discovery, some birdwatchers started to doubt the initial identification as a *lanceolata* and it was suggested that the bird was not a *lanceolata* but a close and sometimes similar looking relative, the Grasshopper warbler *Locustella naevia* (hereafter named *naevia*). At the Dutch Birding forum a long and heated debate about the identification in the end died out without a clear and generally accepted conclusion, although the subcurrent in the Dutch birders scene tended towards a defeat for those in the *lanceolata*-camp (3,4).

Thanks to a very sharp-eyed and alert birdwatcher a single fecal sample of the bird was collected for possible DNA-corroboration of the identification of this bird. I received this sample early December 2016. The initial plan was to use a small mitochondrial barcode to confirm its more-or-less commonly accepted identification as *naevia* and present these results at the forthcoming annual Dutch Birding meeting on February 4, 2017. The remainder of this document describes in some detail the laboratory approach we used in order to secure its DNA-based identification.

Samples used

In order to place the genetic results from the fecal sample of the 2016 Maasvlakte individual (to which we refer to as LL02) in a proper context, we made use of two different sets of reference sequences.
First, by searching GenBank and PubMed, we identified a suitable mitochondrial gene fragment and three autosomal gene fragments that could discriminate between *naevia* and *lanceolata* and of which reliable reference sequences were present in GenBank. Based on a set of key references (5 – 12) we decided to focus on the mitochondrial cytochrome B (cytB) gene and three autosomal gene fragments, lactate dehydrogenase (LDH), myoglobine (MYO) and ornithine decarboxylase (ODC). Of each of these four fragments we downloaded all available the sequences from GenBank (table 1).
Second, we assembled, with the help of others, a number of fresh feather samples or DNA-extracts which we used to produce new sequences of each of the four gene fragments. Similarly, we produced the sequences of each of these four gene fragments in the fecal sample of LL02 (table 2).

DNA-isolation

Details will be filled in later.

DNA-sequencing

Sanger sequencing

Details will be filled in later.

Guide-seq

Details will be filled in later.

Biostatistical analyses

All new sequences were initially aligned with the GenBank sequences using the Clustal W alignment tool (13) included in Bioedit (14) and fine-tuned by eye. Since all newly generated sequences were considerable shorter (in order to accommodate the very low DNA quality of the LL02 fecal sample) when compared with the sequences from GenBank, the resulting alignments were concatenated to the shorter new sequences. This resulted in the following four sequence fragment alignments: cytB 350 bp., LDH 249 bp., MYO 234 bp., and ODC 240 bp. Each of the final alignments (in fasta-format) were converted into Röhl Data Format (rdf) haplotype files by means of DnaSP-6 (15, 16). Each of the rdf-files was used to build median joining (MJ) networks using NETWORK v10 (17, 18). These MJ-networks were further processed by means of Network Publisher (19) and polished using Microsoft Powerpoint (Figures 1 – 4).

Results and Discussion

In each of the four DNA fragments the sequences haplotypes did not overlap between *lanceolata* and *naevia*. This enables a rather straightforward taxon-identification with each of these four fragments. Already early on, in January 2017, it became clear that, based on the few available cytB reference sequences, LL02, could be a *lanceolata*. This rather surprising results was communicated by me on the Dutch Birding meeting on February 4, 2017, and came as a shock to those who were present. I also discussed that, although the brids mtDNA suggested it to be a *lanceolata*, there were still some hypotheses to consider. These were (1) the bird is a genuine " pure" *lanceolata*, (2) the bird is a *naevia* with introgressed *lanceolata* mtDNA, (3) the bird is a *naevia/lanceolata* hybrid, and (4) we made a laboratory error.

It was clear, at that time, that we needed autosomal DNA-information in addition to mtDNA-information in order to be able to exclude hypotheses 2 – 4 and confirm hypothesis 1.
We could safely exclude hypothesis 4 (contamination / error) as we processed LL02 completely independent from other similar samples as at the time we only had a feather sample of the second Dutch bird, LL01 (see table 2) which was processed well before we received sample LL02. Obtaining reliable and reproducible autosomal DNA information from LL02 turned out to be extremely difficult because (1) the DNA in the fecal sample of LL02 was very degraded and (2) the autosomal DNA-concentration in this sample was very low. In the end, by designing a new and very complex (see DNA-sequencing) new sequence method that is based on combining two previously published methods (20, 21), GUIDE-seq and massively-parallel-sequencing (MPS) by means of targeted single primer enrichment sequencing. This new method was first tested on the available fresh reference samples of both *naevia* and *lanceolata*. Once suitably reliable, we used this method to obtain three different autosomal sequence fragments of LL02. This enabled us to place the genetic origin of LL02 in a reliable context.

As is clear from each of the four DNA-fragment MJ-networks (figures 1 – 4), LL02 is placed safely into a group of sequences from unquestionable *lanceolata* individuals, either obtained via GenBank, or by means of newly generated sequences. There is no evidence that, at least for the four genetic fragments analyzed here, LL02 carries any *naevia*-specific genetic information. From this we can conclude that according to all presently available genetic information, LL02 is a "pure" Lanceolatad warbler *Locustella lanceolata*.

Acknowledgements

First of all, we want thank the persons who discovered this bird xxx, and Christian Brinkman who was alert enough to collect the crucial fecal sample. We further wish to thank all contributors of fresh reference samples including Prof. Dr. Martin Collinson, the Ringers at the Ooijse Graaf, Kennmerduinen and Amsterdams Waterleidingsduinen, and Kees Schreven. Finally, we wish to thank the FLDO lab-staff for their moral support and patience throughout this ambitious project.

References.

1. Pictures of the bird on Dutch Birding Site. https://www.dutchbirding.nl/gallery/1?species=588&group=NL
2. Pictures of the bird at Waarneming.nl. https://waarneming.nl/species/1113/photos/?advanced=on
3. A discussion-thread on the Dutch Birding Forum. https://www.dutchbirding.nl/gallery/detail/20547?page=0#navbar-collapse.
4. A discussion-thread on the Dutch Birding Forum. https://www.dutchbirding.nl/gallery/detail/19795?page=0#navbar-collapse.
5. Per Alström, Alice Cibois, Martin Irestedt, Dario Zuccon, Magnus Gelang, Jon Fjeldså, Michael J. Andersen, Robert G. Moyle, Eric Pasquet, Urban Olsson. Comprehensive molecular phylogeny of the grassbirds and allies (Locustellidae) reveals extensive non-monophyly of traditional genera, and a proposal for a new classification. Mol Phyl Evol 2018; 127: 367 - 375.
6. Ma Y, Guo H, Niu L. Comparative study of the relationship among 16 passerines based on mitochondrial and nuclear intron sequencing (unpublished).
7. Xin Lei, Zuohua Yin, Zhenmin Lian, Cungen Chen, Chuanyin DAI, Anton Krištín, Fumin Lei. Phylogenetic relationships of some Sylviidae species based on complete mtDNA cyt b and partial COI sequence data. Chinese Birds 2010; 1: 175 – 187.
8. Per Alström, Silke Fregin, Janette A. Norman, Per G.P. Ericson, Les Christidis, Urban Olsson. Multilocus analysis of a taxonomically densely sampled dataset reveal extensive non-monophyly in the avian family Locustellidae. Mol Phyl Evol 2011; 58: 513 – 526.
9. Silke Fregin, Martin Haase, Urban Olsson, Per Alström. New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers. BMC Evol Biol 2012, 12:157.
10. Silke Fregin, Martin Haase, Urban Olsson, Per Alström. Multi-locus phylogeny of the family Acrocephalidae (Aves: Passeriformes) – The traditional taxonomy overthrown. Mol Phyl Evol 52; 2009: 866–878.
11. Jérôme Fuchs, Eric Pasquet, Arnaud Couloux, Jon Fjeldså, Rauri C.K. Bowie. A new Indo-Malayan member of the Stenostiridae (Aves: Passeriformes) revealed by multilocus sequence data: Biogeographical implications for a morphologically diverse clade of flycatchers. Mol Phyl Evol 53; 2009: 384–393.
12. Carl H. Oliveros, Sushma Reddy, Robert G. Moyle. The phylogenetic position of some Philippine ‘‘babblers’’ spans the muscicapoid and sylvioid bird radiations. Mol Phyl Evol 65; 2012: 799–804.
13. JD Thompson, DG Higgins, TJ Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673 - 4680.
14. Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 1999; 41:95-98.
15. Rozas J, Rozas R. DnaSP, DNA sequence polymorphism: an interactive program for estimating Population Genetics parameters from DNA sequence data. Comput Applic Biosci 1995; 11: 621-625.
16. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Mol. Biol. Evol. 2017; 34: 3299-3302. DOI: 10.1093/molbev/msx248
17. See https://www.fluxus-engineering.com/sharenet.htm.
18. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 1999; 16: 37-48.
19. See https://www.fluxus-engineering.com/nwpub.htm.
20. Shengdar Q Tsai, Zongli Zheng, Nhu T Nguyen, Matthew Liebers, Ved V Topkar, Vishal Thapar, Nicolas Wyvekens, Cyd Khayter, A John Iafrate, Long P Le, Martin J Aryee, J Keith Joung. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature biotechnology 2015; 33: 187 – 197.
21. Quan Peng, Chang Xu , Daniel Kim, Marcus Lewis, John DiCarlo & Yexun Wang. Targeted single primer enrichment sequencing with single end Duplex-UMI. Scientific Reports 2019; 9: 4810. doi.org/10.1038/s41598-019-41215-z 3.

Table 1. Sample information of GenBank reference sequences used in this study.

TAXON	GenBank nr	Code	Locus	SampleID	Country	Location	Reference
Locustella lanceolata	DQ119525	LL0B	CYTB	11A-2	CHINA		6
Locustella lanceolata	DQ119524	LL0C	CYTB	11A-1	CHINA		6
Locustella lanceolata	HQ608849	LL0E	CYTB	Y29-102576	CHINA	HEBEI	7
Locustella lanceolata	HQ706139	LL0A	CYTB	U1949	CHINA	HEBEI	8
Locustella lanceolata	JX398904	LL0D	CYTB	KU4248	CHINA		12
Locustella naevia	HQ706147	LU0A	CYTB	U1951	SWEDEN		8
Locustella lanceolata	MH089215	LL0F	LDH	NRM20006473	SWEDEN		5
Locustella lanceolata	JX236319	LL0G	LDH	B0761	CHINA	BEIDAIHE	9
Locustella lanceolata	FJ883088	LL0H	LDH		CHINA	HEBEI	10
Locustella lanceolata	HQ706195	LL0A	LDH	U1949	CHINA	HEBEI	8
Locustella naevia	JX236320	LU0B	LDH		GERMANY		9
Locustella naevia	HQ706202	LU0A	LDH	U1951	SWEDEN		8
Locustella lanceolata	HQ706235	LL0A	MYO	U1949	CHINA	HEBEI	8
Locustella lanceolata	FJ883091	LL0L	MYO		CHINA	HEBEI	10
Locustella naevia	HQ706242	LU0A	MYO	U1951	SWEDEN		8
Locustella naevia	GQ369648	LU0E	MYO	E42	FRANCE		11
Locustella lanceolata	HQ706313	LL0A	ODC	U1949	CHINA	HEBEI	8
Locustella lanceolata	FJ883160	LL0K	ODC		CHINA	HEBEI	10
Locustella naevia	MH089169	LU0C	ODC	DZUG-5184	SWEDEN		5
Locustella naevia	MH089168	LU0D	ODC	DZUG-5183	SWEDEN		5
Locustella naevia	HQ706320	LU0A	ODC	U1951	SWEDEN		8

Table 2. Sample information of individuals that were sequenced in this study.

DNA nr	FLDO code	Taxon	Location	Country	Source material	Ring nr	Date
X16-012-049	LL01	*Locustella lanceolata*	Ooipolder	Nederland	feather		05-10-2013
X16-012-236	LL02	*Locustella lanceolata*	Maasvlakte	Nederland	fecal		04-10-2016
X19-004-025	LL03	*Locustella lanceolata*	HornØya, Spitsbergen	Norway	feather (mummyfied)		08-08-2018
X20-004-003	LL04	*Locustella lanceolata*	Fair isle	Schotland	feather	JTA712	11-09-2018
X20-004-004	LL05	*Locustella lanceolata*	Fair isle	Schotland	feather	JTA703	05-09-2018
X20-004-005	LL06	*Locustella lanceolata*	Quendale	Schotland	feather	CTP920	11-09-2017
X17-013-016	LU01	*Locustella naevia*	VRS Awduinen	Nederland	feather	?	26-05-2017
X17-013-020	LU02	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BF92288	27-05-2017
X17-013-024	LU03	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BF93258	03-09-2017
X17-013-025	LU04	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BF93257	03-09-2017
X17-013-026	LU05	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BF93250	03-09-2017
X17-013-027	LU06	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BF93248	03-09-2017
X18-004-009	LU07	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BG91963	09-08-2018
X18-004-010	LU08	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BG91626	01-08-2018
X18-004-018	LU09	*Locustella naevia*	VRS Awduinen	Nederland	feather	BH50914	04-10-2018
X20-004-012	LU10	*Locustella naevia*	VRS Van Lennep	Nederland	feather	BF93536	24-09-2017

Table 3. GenBank numbers of new submitted sequences.

	Gene (fragment)						
	CYTB	LDH		MYO		ODC	
Code / Taxon / DNA number	GenBank nr	GenBank nr	GenBank nr	GenBank nr	GenBank nr	GenBank nr	GenBank nr
FLDO LU01 Locustella naevia X17-013-016	MW267954	MW267970	-	MW267992	-	MW268008	MW268009
FLDO LU02 Locustella naevia X17-013-020	MW267955	MW267971	-	MW267993	-	MW268010	MW268011
FLDO LU03 Locustella naevia X17-013-024	MW267956	MW267972	-	MW267994	-	MW268012	-
FLDO LU04 Locustella naevia X17-013-025	MW267957	MW267973	-	MW267995	-	MW268013	MW268014
FLDO LU05 Locustella naevia X17-013-026	MW267958	MW267974	-	MW267996	MW267997	MW268015	-
FLDO LU06 Locustella naevia X17-013-027	MW267959	MW267975	-	MW267998	-	-	-
FLDO LU07 Locustella naevia X18-004-009	MW267960	MW267976	MW267977	MW267999	-	MW268016	-
FLDO LU08 Locustella naevia X18-004-010	MW267961	MW267978	-	MW268000	-	MW268017	-
FLDO LU09 Locustella naevia X18-004-018	MW267962	MW267979	MW267980	MW268001	-	MW268018	MW268019
FLDO LU10 Locustella naevia X20-004-012	MW267963	MW267981	-	MW268002	-	MW268020	-
FLDO LL01 Locustella lanceolata X16-012-049	MW267964	MW267982	-	MW268003	-	MW268021	MW268022
FLDO LL02 Locustella lanceolata X16-012-236	MW267965	MW267983	-	MW268004	-	MW268023	MW268024
FLDO LL03 Locustella lanceolata X19-004-025	MW267966	-	-	-	-	-	-
FLDO LL04 Locustella lanceolata X20-004-003	MW267967	MW267984	-	MW268005	-	MW268025	-
FLDO LL05 Locustella lanceolata X20-004-004	MW267968	MW267985	-	MW268006	-	MW268026	-
FLDO LL06 Locustella lanceolata X20-004-005	MW267969	MW267986	-	MW268007	-	MW268027	MW268028

Please note that LDH, MYO, and ODC are autosomal gene fragments. As a consequence, individuals can – but not must - have two different sequences. With the sequencing method used by us, these two different sequences can easily be distinguished. If in the table above, for these three fragments there is only one GenBank number this means that in this individual two identical copies have been observed.

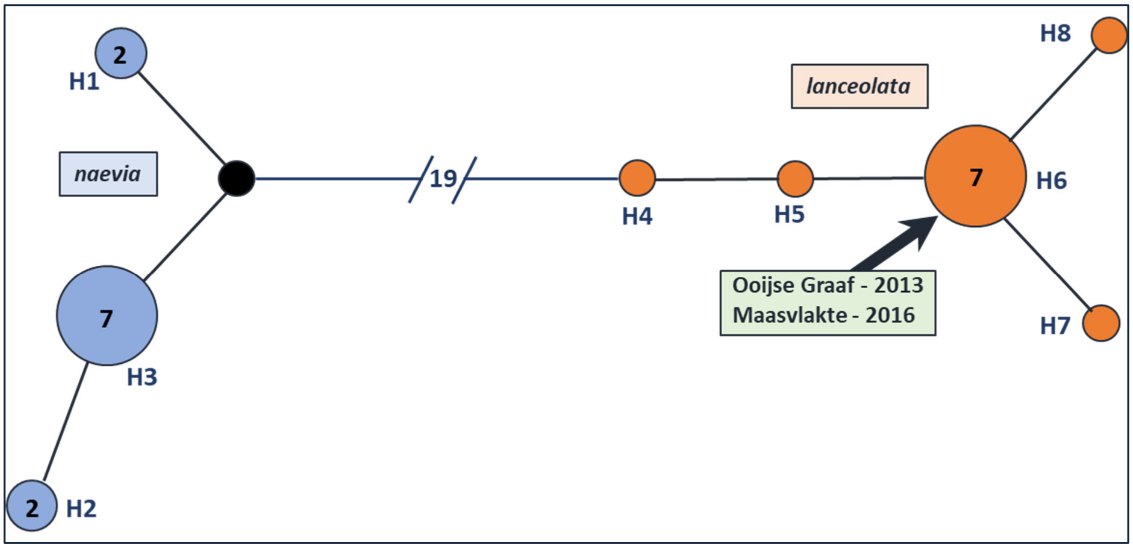
Table 4. Haplotype numbers as used in the accompanying MJ-networks.

		Gene (fragment)						
		CYTB	LDH		MYO		ODC	
Code	Taxon	Haplotype	Haplotype1	Haplotype2	Haplotype1	Haplotype2	Haplotype1	Haplotype2
From GenBank								
LU0A	Locustella naevia	H1	H3	H4	H1	H3	H1	H4
LU0B	Locustella naevia	-	H1	H2	-	-	-	-
LU0C	Locustella naevia	-	-	-	-	-	H1	H2
LU0D	Locustella naevia	-	-	-	-	-	H3	H3
LU0E	Locustella naevia	-	-	-	H1	H1	-	-
New								
LU01	Locustella naevia	H2	H5	H5	H1	H1	H1	H5
LU02	Locustella naevia	H3	H6	H6	H1	H1	H1	H5
LU03	Locustella naevia	H1	H5	H5	H1	H1	H6	H6
LU04	Locustella naevia	H3	H5	H5	H2	H2	H7	H8
LU05	Locustella naevia	H3	H7	H7	H3	H1	H1	H1
LU06	Locustella naevia	H3	H8	H8	H1	H1	-	-
LU07	Locustella naevia	H3	H9	H10	H1	H1	H9	H9
LU08	Locustella naevia	H3	H11	H11	H3	H3	H10	H10
LU09	Locustella naevia	H2	H6	H12	H1	H1	H10	H11
LU10	Locustella naevia	H3	H8	H8	H1	H1	H10	H10
From GenBank								
LL0A	Locustella lanceolata	H4	H14	H14	H4	H4	H12	H13
LL0B	Locustella lanceolata	H5	-	-	-	-	-	-
LL0C	Locustella lanceolata	H6	-	-	-	-	-	-
LL0D	Locustella lanceolata	H7	-	-	-	-	-	-
LL0E	Locustella lanceolata	H6	-	-	-	-	-	-
LL0F	Locustella lanceolata	-	H13	H13	-	-	-	-
LL0G	Locustella lanceolata	-	H14	H14	-	-	-	-
LL0H	Locustella lanceolata	-	H14	H14	-	-	-	-
LL0K	Locustella lanceolata	-	-	-	-	-	H14	H15
LL0L	Locustella lanceolata	-	-	-	H4	H4	-	-
New								
LL01	Locustella lanceolata	H6	H15	H15	H5	H5	H14	H15
LL02	Locustella lanceolata	H6	H14	H14	H5	H5	H16	H17
LL03	Locustella lanceolata	H6	-	-	-	-	-	-
LL04	Locustella lanceolata	H6	H16	H16	H5	H5	H18	H18
LL05	Locustella lanceolata	H6	H17	H17	H5	H5	H12	H12
LL06	Locustella lanceolata	H8	H17	H17	H5	H5	H12	H19

Please note that LDH, MYO, and ODC are autosomal gene fragments. As a consequence, individuals can – but not must - have two different sequence haplotypes. With the sequencing method used by us, these two different haplotypes can easily be distinguished. Grey boxes in this table indicate fragments for which an individual had two identical sequence haplotypes.

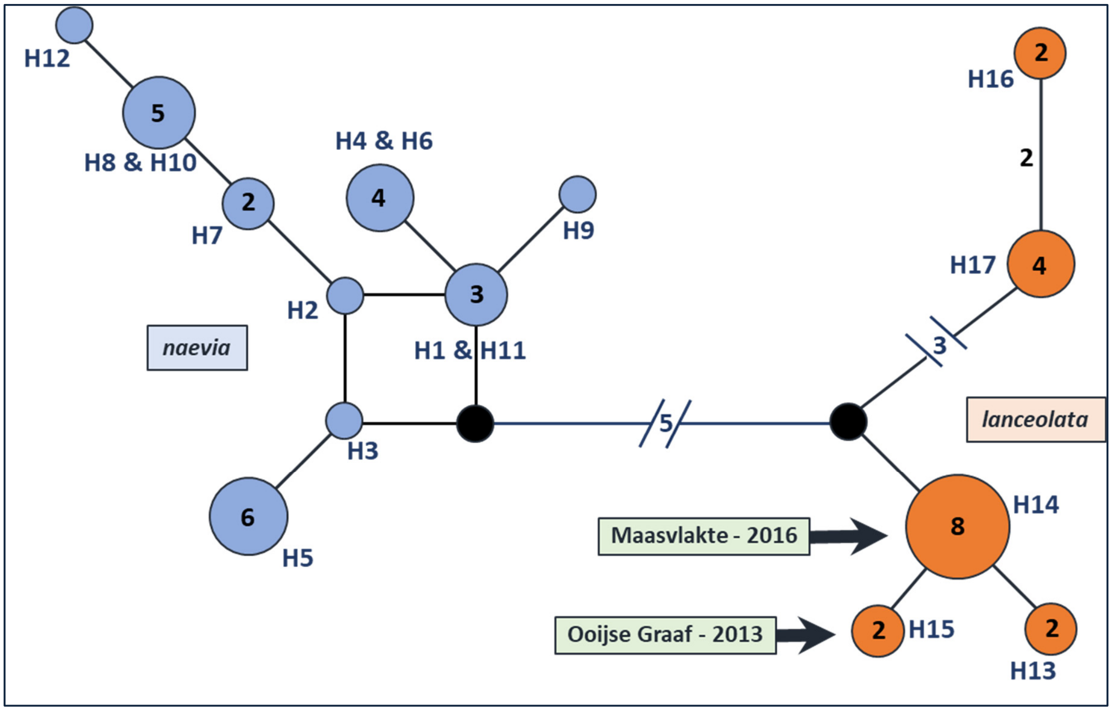
Figure 1. Median Joining Network of the mtDNA cytochrome b (cytB) gene fragment (350 bp.) of Grasshopper warbler *Locustella naevia* and Lanceolated warbler *Locustella lanceolata*.

0



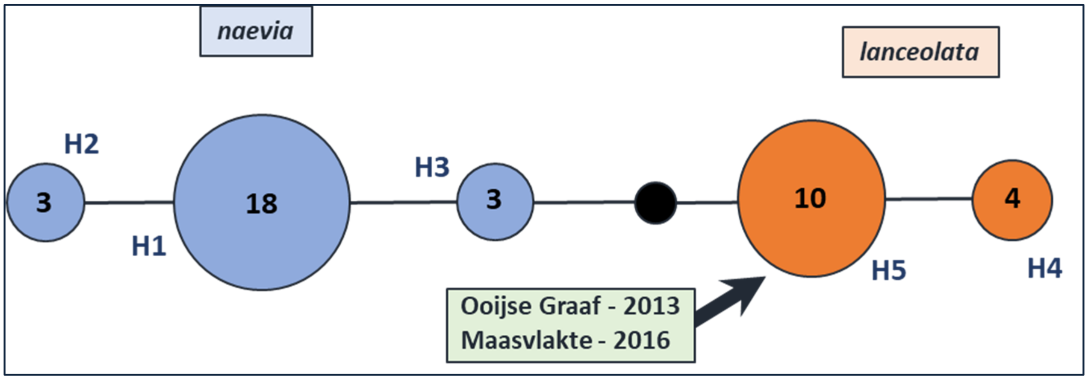
The two Dutch birds (Ooijse Graaf and Maasvlakte) are indicated with a green box and black arrow. Every filled circle represents a unique cytB sequence, of which a fragment of 350 base pairs was sequenced. Numbers (H1 – H8) adjacent to the circles indicate unique sequences (or haplotypes). These numbers are also shown in Table 4. The varying diameter of each circle is a rough indication of its frequency in the total dataset (n=22), and the numbers inside each circle are the number of times observed. Circles without an inside number represent haplotypes that were only seen once. The small solid-black circles represent inferred (not observed) haplotypes that are necessary for building this network. Short lines without an adjacent number connecting two circles indicate a single sequence difference between two haplotypes. Otherwise, the numbers adjacent to lines indicate the number of different positions (out of a total of 350) between two haplotypes. *L. naevia* specific haplotypes are shown in blue. *L. lanceolata* haplotypes are shown in orange. These two taxa can easily be distinguished by means of this short mtDNA fragment. They differ in at least 19 out of 350 positions, and no haplotypes are shared.

Figure 2. Median Joining Network of the autosomal lactate dehydrogenase (LDH) gene fragment (249 bp.) of Grasshopper warbler *Locustella naevia* and Lanceolated warbler *Locustella lanceolata*.



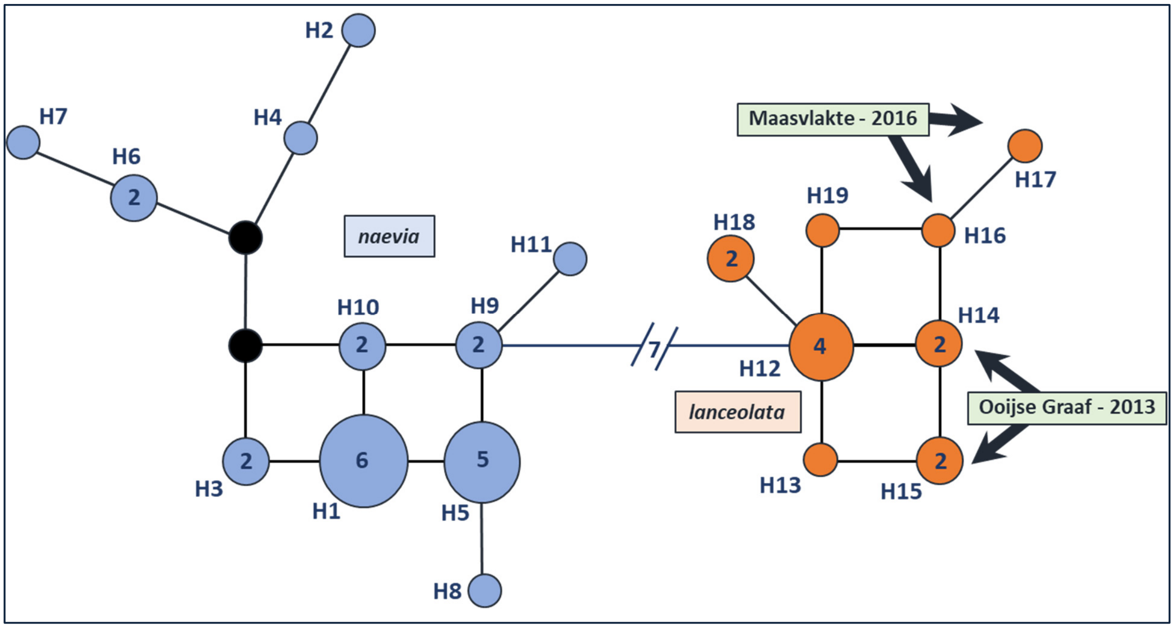
The two Dutch birds (Ooijse Graaf and Maasvlakte) are indicated with a green box and black arrow. Every filled circle represents a unique LDH sequence, of which a fragment of 249 base pairs was sequenced. Numbers (H1 – H17) adjacent to the circles indicate unique sequences (or haplotypes). These numbers are also shown in Table 4. The varying diameter of each circle is a rough indication of its frequency in the total dataset (n=42), and the numbers inside each circle are the number of times observed. Circles without an inside number represent haplotypes that were only seen once. The small solid-black circles represent inferred (not observed) haplotypes that are necessary for building this network. Short lines without an adjacent number connecting two circles indicate a single sequence difference between two haplotypes. Otherwise, the numbers adjacent to lines indicate the number of different positions (out of a total of 249) between two haplotypes. *L. naevia* specific haplotypes are shown in blue. *L. lanceolata* haplotypes are shown in orange. These two taxa can easily be distinguished by means of this short autosomal DNA fragment. They differ in at least 5 out of 249 positions, and no haplotypes are shared. Note that NetWork, to reduce complexity, sometimes combines two highly similar sequence haplotes, see HT4 & HT6 and HT8 & HT10. Also note that the total number of haplotypes (n=42) is twice the number of individuals (n=21) because LDH is an autosomal gene fragment.

Figure 3. Median Joining Network of the autosomal myoglobine (MYO) gene fragment (234 bp.) of Grasshopper warbler *Locustella naevia* and Lanceolated warbler *Locustella lanceolata*.



The two Dutch birds (Ooijse Graaf and Maasvlakte) are indicated with a green box and black arrow. Every filled circle represents a unique MYO sequence, of which a fragment of 234 base pairs was sequenced. Numbers (H1 – H5) adjacent to the circles indicate unique sequences (or haplotypes). These numbers are also shown in Table 4. The varying diameter of each circle is a rough indication of its frequency in the total dataset (n=38), and the numbers inside each circle are the number of times observed. Circles without an inside number represent haplotypes that were only seen once. The small solid-black circles represent inferred (not observed) haplotypes that are necessary for building this network. Short lines without an adjacent number connecting two circles indicate a single sequence difference between two haplotypes. Otherwise, the numbers adjacent to lines indicate the number of different positions (out of a total of 234) between two haplotypes. *L. naevia* specific haplotypes are shown in blue. *L. lanceolata* haplotypes are shown in orange. These two taxa can easily be distinguished by means of this short autosomal DNA fragment. They differ in at least 2 out of 234 positions, and no haplotypes are shared. Note that the total number of haplotypes (n=38) is twice the number of individuals (n=19) because MYO is an autosomal gene fragment.

Figure 4. Median Joining Network of the autosomal ornithine decarboxylase (ODC) gene fragment (240 bp.) of Grasshopper warbler *Locustella naevia* and Lanceolated warbler *Locustella lanceolata*.



The two Dutch birds (Ooijse Graaf and Maasvlakte) are indicated with a green box and black arrow. Every filled circle represents a unique ODC sequence, of which a fragment of 240 base pairs was sequenced. Numbers (H1 – H19) adjacent to the circles indicate unique sequences (or haplotypes). These numbers are also shown in Table 4. The varying diameter of each circle is a rough indication of its frequency in the total dataset (n=38), and the numbers inside each circle are the number of times observed. Circles without an inside number represent haplotypes that were only seen once. The small solid-black circles represent inferred (not observed) haplotypes that are necessary for building this network. Short lines without an adjacent number connecting two circles indicate a single sequence difference between two haplotypes. Otherwise, the numbers adjacent to lines indicate the number of different positions (out of a total of 240) between two haplotypes. *L. naevia* specific haplotypes are shown in blue. *L. lanceolata* haplotypes are shown in orange. These two taxa can easily be distinguished by means of this short autosomal DNA fragment. They differ in at least 2 out of 234 positions, and no haplotypes are shared. Note that the total number of haplotypes (n=38) is twice the number of individuals (n=19) because ODC is an autosomal gene fragment.