universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Phylogenetic relationships of the goshawk *Accipiter* [*gentilis*] superspecies"

verfasst von / submitted by

Min Jenny Chai BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 831
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Zoologie
Betreut von / Supervisor:	Dr. Elisabeth Haring, Privatdoz.

Acknowledgments

I would first like to thank my supervisor Dr. Elisabeth Haring for guiding and supporting me through the whole master project. You were always there when I needed help, whether by mail, phone or in person. The discussions we had always left me with new ideas and thoughts. They not only contributed to this project but influenced me to think more like a scientist.

Further, I would like to thank Florian Kunz for his co-supervision and his time and support during the entire project, especially during the laboratory work and bioinformatic phase. You calmed me down, when some things did not work as planned and gave me advice how problems can be solved. Then I would like to express my deepest thanks to Dr. Martin Schwentner for introducing me to the ddRAD laboratory work and helping me with the bioinformatic work as well. You always had time if I was lost in the laboratory or in front of the computer and you did not hesitate to come across to help. I really appreciated your dedication.

Special thanks go to Dr. Martin Kapun. Without him the whole bioinformatic analyses would not have been possible. Your ddRAD pipeline helped me a lot and I don't want to imagine if your bioinformatic knowledge was not available. Even with the simplest questions, you always had time and understanding. Thank you so much for your patience.

Further, I would like to thank Dr. Frank Zachos and Dr. Luise Kruckenhauser for their ideas and help. Due to Covid-19 the first meetings were online, however, these meetings with you and the people mentioned above were very important and helpful for the start of this project.

And then I would like to thank all the people at the Central Research Laboratories department. You have welcomed me so warmly and you were always there if I needed help. I had the best time during my project, and I want to thank you so much for the lovely environment. Especially I want to thank Julia Schindelar, who introduced me to the basic laboratory rules and helped me with some important steps during my project.

My last expressions of thanks go to my family and friends. You have supported me endlessly and always reminded me to take some time out.

Abstract

The focus of the present study is the Northern Goshawk (*Accipiter gentilis*), a member of the family Accipitridae. *Accipiter gentilis* has been placed within the *Accipiter* [*gentilis*] superspecies complex together with *Accipiter meyerianus, Accipiter melanoleucus* and *Accipiter henstii.* Currently, 10 subspecies are distinguished within *Accipiter gentilis*. In a previous analysis of mitochondrial sequences by Kunz et al. (2019), was a deep split within the *Accipiter gentilis* clade into a Nearctic and a Palearctic clade. Both clades were not closely related in these analyses. The Palearctic clade was more closely related to *A. meyerianus* and these two in turn were more closely related to the Old-World taxa, *A. henstii* and *A. melanoleucus.* To test whether the mitochondrial phylogeny is also reflected in nuclear markers, we analyzed the intron 7 of the *β-fibrinogen* (*β-fibint 7*) gene*,* which is a commonly used nuclear marker in birds, for 42 samples. Furthermore, a reduced-representation genome sequencing by ddRADseq was performed to investigate the relationships among the genus *Accipiter* on a genome-wide scale*.* Our study indicates that *β-fibint 7* is not a suitable marker as there were only minimal differences among taxa. Conversely, bioinformatic analyses of the ddRADseq data from 58 samples support the pattern previously detected with mitochondrial markers showing that there is a split within the Holarctic *A. gentilis* into two clades. Taken together, our findings would indicate to split the Holarctic *A. gentilis* into two species. Importantly, however, an outgroup must be included into the phylogenetic analyses before solid decisions can be made.

Zusammenfassung

Im Mittelpunkt dieser Masterarbeit steht der Habicht (*Accipiter gentilis*), ein Mitglied der Familie der Habichtartigen (Accipitridae). *Accipiter gentilis* gehört zusammen mit *Accipiter meyerianus, Accipiter melanoleucus* and *Accipiter henstii* zum *Accipiter* [*gentilis*] Superspezies Komplex. Derzeit werden 10 Unterarten innerhalb von *Accipiter gentilis* unterschieden. In einer früheren Analyse mitochondrialer Sequenzen durch Kunz et al. (2019) wurde eine tiefe Spaltung innerhalb der *Accipiter gentilis* Klade in eine nearktische und eine paläarktische Klade festgestellt. Beide Kladen waren in diesen Analysen nicht eng miteinander verwandt. Die paläarktische Klade ist näher mit *A. meyerianus* verwandt und diese zwei wiederum näher mit den Altweltlichen Taxa, *A. henstii* und *A. melaoleucus*. Um zu prüfen, ob sich die mitochondriale Phylogenie auch in den Kernmarkern widerspiegelt, haben wir bei 42 Individuen das Intron des *β-Fibrinogen (β-fibint 7)* analysiert, welches ein häufig verwendeter Kernmarker ist, insbesondere bei Vögeln. Zusätzlich zur Analyse von *β-fibint 7* wurde eine Genomsequenzierung mittels ddRADseq durchgeführt, um die Beziehungen innerhalb der Gattung *Accipiter* weiter zu untersuchen. *β-fibint 7* ist kein geeigneter Marker, da es nur minimale Unterschiede zwischen den Taxa aufzeigte. Die Ergebnisse aus der ddRADseq wurden aus 58 Proben gewonnen und bioinformatische Analysen zeigten, dass es innerhalb des *Accipiter* [*gentilis*] Superspezies Komplex eine Aufspaltung in zwei Kladen, eine nearktische und eine paläarktische Klade, gibt. Zusammen mit der mitochondrialen Phylogenie deuten die Ergebnisse darauf hin, dass die holarktische *A. gentilis* in zwei Arten aufgeteilt werden sollte. Wichtig ist jedoch, dass eine Aussengruppe in die phylogenetischen Analysen einbezogen wird, bevor solche Entscheidungen getroffen werden können.

Table of Contents

1 Introduction

1.1 The family Accipitridae

Generally, members of Accipitridae inhabit various environments, from rainforests to arctic tundra. While some taxa, e.g., the snake eagles (Circaetinae) are only found in the Old-World, others like the sea eagles (Haliaeetinae) are globally distributed (Lerner and Mindell, 2005). The family of Accipitridae is very taxon-rich, with up to 14 subfamilies, 65 genera, and 231 species (Dickinson, 2003; Stresemann and Amadon, 1979). Knowledge of this large family including phylogenetic relationships among taxa, as well as genetic differentiation within them is needed for effective conservation programs. Awareness of species diversity can ensure research-guided species protection measures. Members of Accipitridae are particularly well suited for this, since they are ecologically sensitive and thus valuable indicators of habitat quality for conservation programs (Lerner and Mindell, 2005). The focus of the present study is the Northern Goshawk (*Accipiter gentilis*), a member of the family Accipitridae. They are widely distributed in temperate to boreal forests in the Northern Hemisphere. These medium-sized diurnal raptors forage long distances for their prey (Squires and Kennedy, 2006). Northern Goshawks are not continuously distributed across the landscape because their distribution depends on presence and availability of suitable habitat (Hanski and Simberloff, 1997).

In the genus *Accipiter*, where adaptive convergence may have led to similar plumage patterns in distantly related taxa, molecular analyses were attempted to reveal phylogenetic relationships within this group (Riegner, 2008). Besides the potential for the before mentioned adaptive convergence, high morphological variation found in *Accipiter gentilis* resulted in description of different subspecies (Table 1, Figure 1). Furthermore, several studies established sufficient evidence that the harriers of the genus *Circus* (Lacépède, 1799) are also part of the *Accipiter* clade (Griffiths et al., 2007; Lerner et al., 2008; Mindell et al., 1999; Nagy and Tökölyi, 2014; Oatley et al., 2015). Therefore, the genus *Accipiter* is currently not considered monophyletic (Mindell et al., 2018).

1.2 The Northern Goshawk *Accipiter gentilis*

The British Ornithologists' Union (BOU;Helbig et al., 2002) describe the terminology superspecies as monophyletic groups of allo- and semispecies. Moreover, allo- and semispecies are more closely related than normally species are. Allopatric species have an allopatric distribution area, whereas semispecies have a stable hybrid zone. The Northern

Goshawk is one of four allospecies in the goshawk *Accipiter* [*gentilis*] superspecies (Amadon, 1966).

Besides the Northern Goshawk *Accipiter* [*gentilis*] *gentilis* (Linnaeus, 1758), there are the Black Sparrowhawk *A.* [*gentilis*] *melanoleucus* A. Smith, 1830, Meyer's Goshawk *A.* [*gentilis*] *meyerianus* (Sharpe, 1878) and Henst's Goshawk *A.* [*gentilis*] *henstii* (Schlegel, 1873). According to Dickinson and Remsen (2013) and del Hoyo and Collar (2014) 10 subspecies are currently recognized in *A. gentilis* (Figure 1): *A. gentilis gentilis* (Linnaeus, 1758), *A. gentilis buteoides* (Menzbier, 1882), *A. gentilis albidus* (Menzbier, 1882), *A. gentilis marginatus* (Piller & Mitterpacher, 1783), *A. gentilis arrigonii* (O. Kleinschmidt, 1903), *A. gentilis schvedowi* (Menzbier, 1882) and *A. gentilis fujiyamae* (Swann & Hartert, 1923) in Eurasia and *A. gentilis atricapillus* (A. Wilson, 1812), *A. gentilis laingi* (Taverner, 1940) and *A. gentilis apache* van Rossem, 1938 in North America.

Table 1: List of allospecies and subspecies of the *Accipiter* [*gentilis*] superspecies.

Allospecies of *Accipiter* [*gentilis*] superspecies	
	Subspecies
A. [gentilis] gentilis	*A. g. albidus*
	A. g. apache
	A. g. arrigonii
	A. g. atricapillus
	A. g. buteoides
	A. g. fujiyamae
	A. g. gentilis
	A. g. laingi
	A. g. marginatus
	A. g. schvedowi
A. [gentilis] melanoleucus	*A. m. melanoleucus*
	A. m. temminckii
A. [gentilis] meyerianus	monotypic
A. [gentilis] henstii	monotypic

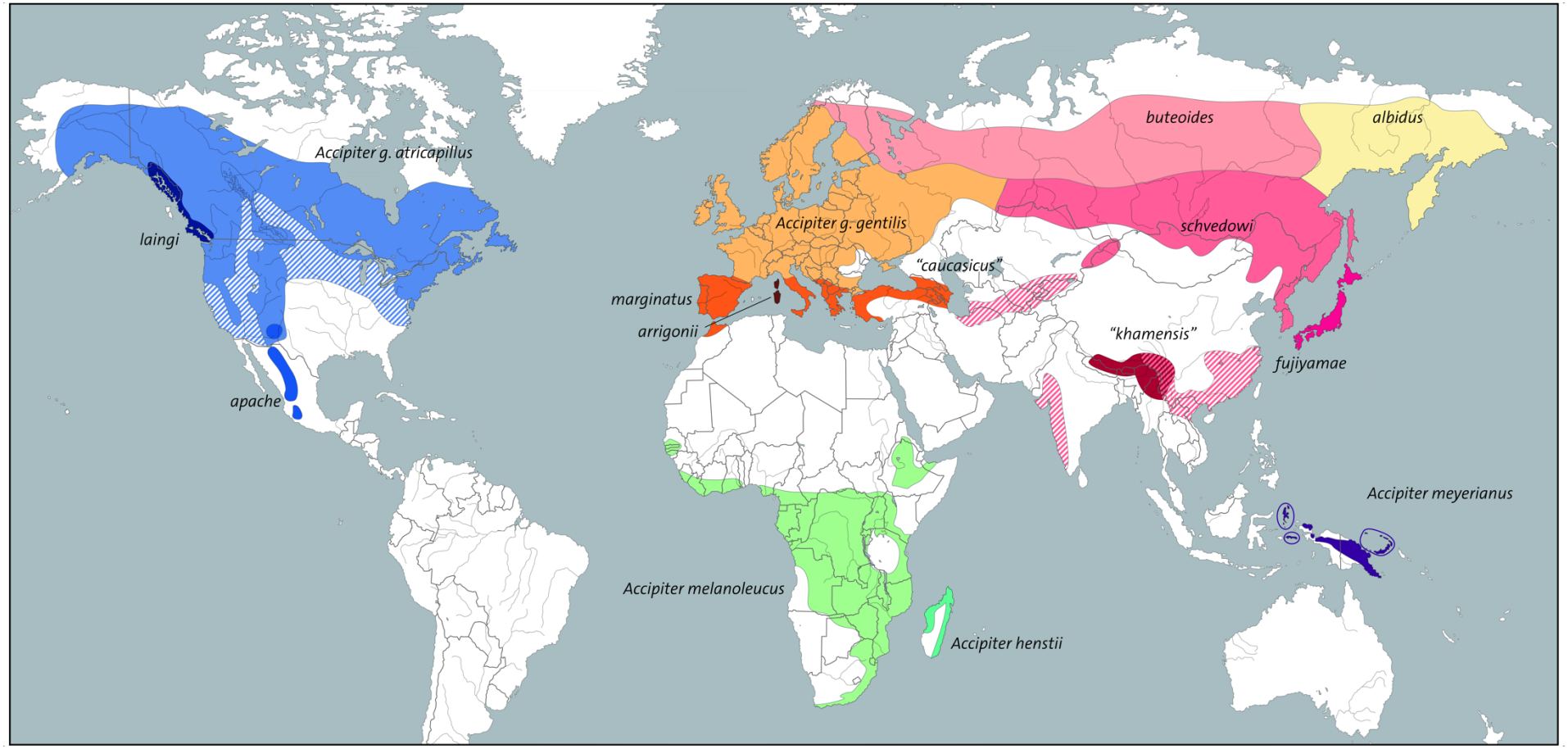

Figure 1: Geographical distribution of the *Accipiter* [*gentilis*] superspecies. Subspecies of *A. gentilis* according to del Hoyo and Collar (2014) as well as *A. melanoleucus*, *A. henstii*, and *A. meyerianus* are indicated by coloration. Additionally, formerly recognized *A. g. caucasicus* and *A. g. khamensis* are included. Filled areas represent all-year-round occurrence; hatched areas indicate temporary winter migration. Modified from Kunz et al. (2019).

1.3 Phylogenetic DNA markers: Nuclear DNA vs. mitochondrial DNA

To resolve phylogenetic trees, it is necessary that the studied nucleotide sequences must have enough variability within and between species to guarantee an appropriate resolution of the phylogenetic tree (Song et al., 2016). In animals, mitochondrial genes are haploid, maternal inherited, lack recombination and have a conserved gene organization (gene order) (Gupta and Bhardwaj, 2015). Due to the favorable characteristics mentioned above, mitochondrial DNA (mtDNA) sequences were considered as useful for phylogenetic analyses and different mtDNA regions are suitable markers especially in closely related taxa (Piganeau et al., 2004; Rokas et al., 2003). In the past, phylogenetic reconstructions of vertebrate taxa have been predominantly based on mtDNA. The substitution rate in the mitochondrial genome was reported as 5-10 times higher compared to the substitution rate in the nuclear genome (ncDNA), which is due to lower efficiency of DNA repair pathways in mtDNA (Haag-Liautard et al., 2008). Higher substitution rate in mtDNA varies in different animal taxa and depends on the animal taxa that are investigated in. In insects, the ratio of substitution rate in mtDNA over ncDNA has been reported to vary between 2 and 6. In comparison to vertebrates, like in birds, the ratio is on average above 20 (Allio et al., 2017). The early diversification of Neoaves includes all extant birds, except the lineages leading to Paleognathae and Galloanseres (Jarvis et al., 2014). This diversification happened between 55 and more than 90 MYA, which was then followed by a rapid radiation over a timespan of 5 to 10 Myr (Mindell et al., 1999). The rapid radiation resulted in conflicting phylogenetic analyses of Neoaves (Ericson et al., 2006; Feduccia, 1995). To disentangle rapid speciation events in Neoaves it is suggested to include other gene sequences besides mtDNA in phylogenetic analyses as well. The maternal inheritance of mtDNA only allows to assess shared mtDNA haplotype in the maternal family line (Gupta and Bhardwaj, 2015). Therefore, phylogenetic analyses of mtDNA might not coincide with phylogenetic analyses of ncDNA. A phylogeny based on mtDNA should therefore be complemented by a gene tree based on nuclear sequences.

In many taxonomic groups, nuclear introns can act as feasible alternative markers to mtDNA (Dool et al., 2016; Jarvis et al., 2014). Nuclear introns are promising candidates for phylogenetic analyses and resolving relationships in closely related species because they may evolve at an appropriate rate to resolve phylogenetic relationships among closely-related avian taxa (Prychitko, 2003). Even though introns were “generally believed to evolve too rapidly and too erratically to be of much use in phylogenetic reconstruction” (Kupfermann et al., 1999), there are systematic studies in vertebrate groups using introns, which show that introns of nuclear genes are useful for phylogenetic analysis (Prychitko and Moore, 1997; Kupfermann et al., 1999; Mundy and Kelly, 2001; Weibel and Moore, 2002; Moyle, 2004). Introns are generally easy to amplify by PCR because they are usually flanked by conserved

exon sequences that can be used as targets for PCR primers (Prychitko, 2003). Nevertheless, it may happen that nuclear introns do not resolve phylogenetic relationships among recently evolved species because the nuclear intron does not harbor sufficient genetic variation hence no differences are detected yet between the investigated species. Based on the study from Prychitko and Moore (2003), the composition of the bases of the nuclear encoded *β-fibint 7* varies slightly over the diversity of birds and it serves as an informative genetic marker for various avian taxa independent of the divergence time (Prychitko, 2003; Prychitko and Moore, 1997). Prychitko and Moore (1997, 2000) showed that the *β-fibint 7* can be used as phylogenetic marker for woodpeckers and it also successfully used to resolve phylogenetic relationship in other bird taxa (Johnson and Clayton, 2000; Moyle, 2004; Weibel and Moore, 2002). Past studies generally had a small sample size or investigated only a single molecular marker, which was often based on mtDNA. For example Sibley and Ahlquist examined only eight genera in their DNA hybridization study (Sibley et al., 1990). Phylogenetic studies based on mtDNA sequences had limited taxon sampling and only looked on specific Accipitridae subspecies (Gamauf et al., 2005; Gamauf and Haring, 2004; Helbig et al., 2005; Kruckenhauser et al., 2004). Lerner and Mindell (2005) showed in their molecular study phylogenetic relationships for birds of prey in the family Accipitridae based on one nuclear intron, *β-fibint 7*, and two mitochondrial genes, NADH dehydrogenase subunit 2 and cytochrome-*b*. Other than that, no molecular study has used ncDNA to examine Accipitridae subgroups, hence no studies have been published specifically dedicated to the genus *Accipiter* using the nuclear DNA *β-fibint 7*.

In 2019 Kunz and colleagues extended these previous efforts and investigated a large geographic sample spanning 156 specimens. In 2019 Kunz and colleagues found a deep split within *A. gentilis* into two monophyletic groups, a Nearctic clade with three subspecies and a Palearctic clade with seven subspecies based on mtDNA. Moreover, the topology of the mitochondrial tree excluded the possibility that the two clades were belonging to the same species. Furthermore, Geraldes et al. (2019) showed a differentiation between the Palearctic and Nearctic groups as well by analyzing SNPs using genotyping-by-sequencing of high-quality genetic samples. The vocalization between the Nearctic and Palearctic groups differs as well, which is shown in Sangster et al. (2022).

In the present study, I extended this effort by adding a ncDNA marker, *β-fibint 7*, to perform phylogenetic analysis on the goshawk A. [*gentilis*] superspecies, to test if nuclear and mitochondrial markers show similar phylogenetic signals. Studies including the family Accipitridae have mostly included only *A. gentilis* (Johnsen et al., 2010; Lerner et al., 2008; Nagy and Tökölyi, 2014). There are no previously published molecular studies based on nuclear markers in combination with genome sequencing that include representatives of nearly all *Accipiter* [*gentilis*] subspecies.

1.4 Double-digest restriction-site associated DNA sequencing

Genetic information is crucial to resolve phylogenetic relationships among individuals and taxa and are essential for studies of behavior as well as ecology of wild organisms (Thrasher et al., 2018). To study phylogenetic relationships in detail and to measure genetic diversity, variable types of genetic markers to represent DNA sequences in the genome are needed. Nuclear microsatellite loci were discovered in the 1980s and have revolutionized studying genetic differentiation in closely related taxa. Microsatellite loci are repeated sequences of up to six bases, which are evenly distributed throughout the whole genome (Li et al., 2002; Morin et al., 2004). They are suitable molecular markers for studies of close phylogenetic relationships because of their high mutation rates, which lead to highly polymorphic sequences at each locus. Besides using microsatellite loci for distinguishing closely related taxa, microsatellite loci are attractive markers for parentage studies, population genetics and individual genotyping (Kaiser et al., 2017). Nevertheless, there are some drawbacks by using microsatellite-based methods, like higher development costs or to find species-specific primers (Miah et al., 2013). Besides microsatellite loci, single nucleotide polymorphisms (SNPs) are powerful for studies of parentage, relatedness and overall phylogeny and population structure. SNPs have a genome-wide distribution, are highly abundant and variable across the genome. These characteristics make SNPs suitable markers for studies in genomic evolution. SNPs are normally biallelic and have lower mutation rates per generation (10^{-8}-10^{-9}) compared to microsatellites which are often multiallelic and have higher mutation rates per generation (10^{-4}) (Brumfield et al., 2003; Ellegren, 2004). As a consequence higher numbers of SNPs than microsatellites are mandatory to obtain sufficient data and achieve enough statistical power for relationship studies (Ball et al., 2010).

Next-generation sequencing (NGS) is a powerful tool, which allows for sampling enough SNPs across genomes to create meaningful data in a cost and time effective manner. NGS in combination with restriction-site associated sequencing (RADseq) is capable to discover thousands of markers across any genomes (Davey et al., 2011). RADseq enables processing high throughput generated by next-generation sequencing in a simple and cost-effective way. Other than whole genome sequencing, RADseq only targets a subset of the genome without prior knowledge of the genome. RADseq is especially interesting for studies of non-model organisms without prior genetic information on the taxonomic group. It is a method that uses restriction enzymes to cut DNA into fragments. Restriction enzymes have the characteristic that they only cut at their specific cut site. This will produce fragments with a characteristic overhang, which can be ligated to other fragments with the complement sequence, hence those overhangs are also called sticky-ends. Only fragments with these sticky-ends are later ligated to modified Illumina adapters which can bind to an Illumina flow

cell and are then sequenced (Davey and Blaxter, 2010). This process enables to target only a subset of the genome, therefore allowing for a higher depth of coverage per locus (Andrews et al., 2016).

In recent years RADseq has been the basis of several methods. Nowadays, RADseq refers to several related techniques, which differ in type and number of restriction enzymes used and in the applied size selection method. Those different protocols have their advantages and disadvantages, which are summarized by Andrews and colleagues (Andrews et al., 2016). One modified technique of the RADseq protocol, named double-digest restriction-site associated DNA sequencing (ddRADseq), was introduced in 2012 by Peterson and colleagues (Peterson et al., 2012). As the name already implies, ddRADseq is based on a RADseq protocol that uses a double digest approach with two different restriction enzymes. As already mentioned above restriction enzymes have specific cut sites. In ddRADseq restriction enzymes cut DNA at different cut sites with different frequencies, hence one restriction enzyme is called common cutter and the other restriction enzyme rare cutter. Subsequently, only fragments which have cut sites from both restriction enzymes are ligated to both Illumina adapters and can thus be sequenced. Furthermore, fragments need to undergo a tight size selection step to be selected into the final library, which will then be sequenced. Library size selection enables a reduced genome complexity which, in the end, makes it possible to study genome-wide genetic variation without any prior genomic knowledge (Peterson et al., 2012). High number of individuals or samples are required per population for comparative analyses with high statistical power. Double digest restriction site-associated DNA sequencing is nowadays widely used to explore phylogenetic questions in non-model organisms (Ba et al., 2017; Janjua et al., 2020; Schwentner and Lörz, 2021).

1.5 Aim of the study

The aim of this study was to test, whether the mtDNA-based phylogeny of species relationships in the *Accipiter* [*gentilis*] superspecies from Kunz et al. (2019) also is reflected in the nuclear DNA? The following questions were asked to test this phylogeny:

1. Are the two marker systems (*β-fibint 7*, ddRAD) or one of them informative for resolving the phylogeny of *Accipiter* [*gentilis*] superspecies?
2. Are the nuclear results in congruence with the phylogenetic tree based on mitochondrial data from Kunz et al. (2019)? Do they contradict each other?
3. Are the four allospecies of the *Accipiter* [*gentilis*] superspecies differentiated in the generated phylogenetic trees based on nuclear markers?

4. Are the Holarctic *A. gentilis* specimens forming one Nearctic clade and one Palearctic clade?
5. Is it possible to gain enough data (whether with *β-fibint 7* or ddRADseq) when sequencing mainly museum material?
6. What taxonomic interpretations can further be made with the data generated?

2 Methods

2.1 Specimens analyzed

In total 96 specimens (muscle tissue *n* = 35, feathers *n* = 8, footpads *n* = 47, skin *n* = 6) were analyzed. For the ddRADseq analysis, all 96 specimens were included while 42 specimens were selected for the analysis of *β-fibint 7.* From the four allospecies all taxa were included except *A. meyerianus* and from the ten subspecies of *A. gentilis* all subspecies were included except *A. g. fujiyamae* and *A. g. apache* (Appendix, Table S1). All tissue samples were obtained from previous work by Kunz et al. (2019) and had been obtained from museum collections and private collectors. They have been stored in AE buffer in -80 °C freezers in the Natural History Museum of Vienna. The age of specimens ranged from 1881 to 2015.

2.2 DNA extraction and DNA concentration

For all 96 samples extracted DNA was already available from previous work (see above). To have an overview of the range of DNA concentrations of the already extracted DNA, a Qubit dsDNA HS (double stranded DNA high sensitivity) assay with the Invitrogen Qubit 3.0 Fluorometer (Invitrogen, Thermo Fisher Scientific, Massachusetts, USA) was done for all 96 samples. Volumes of all 96 samples were measured by micropipettes. Out of the total 96 samples, 44 had too low DNA concentrations and therefore new DNA extractions were necessary. Pieces of toepad, feather quills or skin tissue of approximately one to 4-10 mm^2 were cut into smaller pieces and DNA was isolated using the QIAMP Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol with the following modifications: samples were incubated in lysis buffer at 56 °C overnight and finally eluted in 50 µl nuclease free water (see below). A negative control with no tissue sample was carried out simultaneously through the whole DNA extraction process to allow testing of contaminations in chemicals used for extraction. Elution was done two times separately, each time in 50 µl in a separate tube, to get as much DNA out as possible. Elution in nuclease free water should be less obstructive to desiccate and resuspend the DNA samples later than to

elute (as frequently done) in AE buffer. Since the AE buffer includes salts, salt concentration would be too high after desiccation and resuspension in small volumes would have been more difficult. Concentration of DNA of the newly extracted 44 DNA tissue samples was again measured with a Qubit dsDNA HS assay. Since DNA concentrations were generally still rather low, it was necessary to increase concentration, which was done using a desiccator (see chapter 2.6).

2.3 PCR amplification

2.3.1 PCR of the *β*-fibrinogen intron 7 sequence

The complete intron 7 of the nuclear encoded *β*- fibrinogen gene (*β-fibint 7)* was analyzed. For this task, primers for the conserved exons flanking the intron were taken from literature (Prychitko and Moore, 1997) and slightly modified to fit better to the conserved flanking exons of *Accipiter* (Table 2). Primer AccFib1+ was either combined with AccFib2- or with AccFib3- to obtain fragments of 1073 bp and 1005 bp, respectively (Figure 2). Degradation of aDNA made it impossible for ten samples to amplify the one of the large fragments as one piece. Therefore, for these ten samples the complete *β-fibint 7* marker sequence was generated by amplification of maximal four overlapping fragments (285 to 368 bases in length) the sequences of which were combined later. For this purpose, nine primers were used in varying combinations with different annealing temperatures (Figure 2, Table 3). Primers for the four overlapping fragments were designed using the software Amplify 4 1.0 (https://engels.genetics.wisc.edu/amplify/). Primers (Microsynth, Balgach, Switzerland) were received lyophilized and were dissolved with nuclease free water to a starting concentration of 100 µM. From this solution, aliquots of 50 µM were prepared with nuclease free water and stored at -20 °C.

Before starting with samples of fragmented *Accipiter* DNA, a gradient PCR was used on two blood samples of *A. gentilis* with good DNA quality to determine the optimum annealing temperature for the designed primers. Amplification was performed using 1 µl of template DNA in a 25 µl PCR reaction volume, containing 0.1 µl TopTaq polymerase (5 U/µl), 0.5 µl dNTP mix (10mM), 0.25 µl of each primer (50 µM), 2.5 µl 10x buffer and 20.4 µl AD. PCRs were done using an Eppendorf Mastercycler nexus gradient (Eppendorf, Hamburg, Germany). A negative control of the DNA extraction (without DNA sample) was included in each PCR to check for contamination during extraction process. Additionally, PCR reactions without template DNA were used as negative controls to check for contaminations of PCR reagents. PCR products were checked by agarose gel electrophoresis. PCR products were cleaned-up with QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and sequenced bidirectionally

using the same PCR primers as for the amplification at Microsynth. Chromatograms were checked and raw sequences were edited, assembled and aligned using BioEdit version 7.2.3 (Hall, 1999).

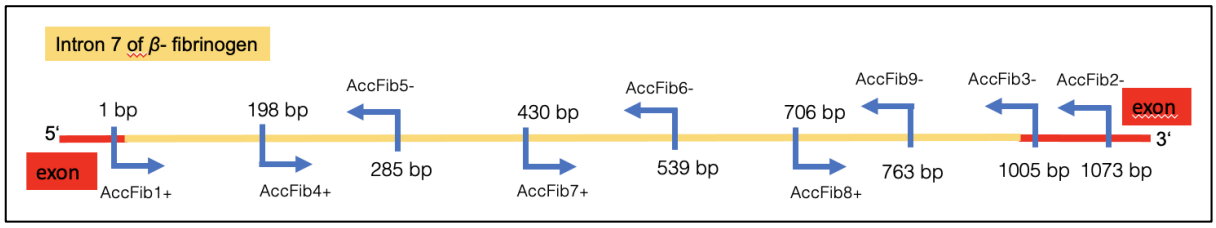

Figure 2: *β-fibint 7* (yellow) with its primers (blue arrows) and exons (red).

Table 2: Name of primers, primer sequences, primer references and notes.

Name of primer	Sequence 5'-3'	Reference	Note
AccFib1+	GAAAACAGGAACAATGACAATTC	Prychitko and Moore, 1997	Modified
AccFib2-	CAATAGTATCTGCCATTGGGGT	Prychitko and Moore, 1997	Modified
AccFib3-	GGAGCACTGTTTTCTTGGAT	Prychitko and Moore, 1997	Modified
AccFib4+	CAAGTTACCAGCCAAATGTC	Present study	
AccFib5-	GGTTGTGGAGCAGCACTAAC	Present study	
AccFib6-	ACTTTACAACTGAGCTCCTG	Present study	
AccFib7+	GGTACTCACTCCAGTAACAC	Present study	
AccFib8+	GTAATTGTAGTTGTCAATCAGC	Present study	
AccFib9-	TTTAGCTGCAGCTCTTTGGC	Present study	

Table 3: Primer combinations, size of amplified PCR fragments and primer pair annealing temperature.

Primer combinations			
Forward primer	**Reverse primer**	**Fragment size (bp)**	**Annealing temperature (°C)**
AccFib1+	AccFib2-	1073	55.8
AccFib1+	AccFib3-	1005	54.4
AccFib1+	AccFib5-	285	58
AccFib8+	AccFib2-	368	58
AccFib4+	AccFib6-	342	52
AccFib7+	AccFib9-	334	56
AccFib8+	AccFib3-	300	55

2.4 TOPO TA Cloning (Topoisomerase TA based cloning)

Amplification of the second fragment 2 (AccFib4+/AccFib6-, 342 bp) of *β-fibint 7* was performed to separate length variable alleles amplified from single individuals causing partially unreadable sequences. For this task, 1 µl of template DNA in a 25 µl PCR reaction volume, containing 12.5 µl of Phusion Mastermix (New England BioLabs, Ipswich, MA), 11 µl of nuclease free water and 0.25 µl of each primer (50 µM). A negative control during PCR was done to check for contamination of PCR reagents. The Phusion Mastermix contains a proofreading polymerase instead of a Taq polymerase to minimize incorrect base pairing during PCR. PCR cycling conditions were: Initial denaturation at 98 °C for 30 sec, denaturation at 98 °C for 10 sec, annealing at 64 °C for 30 sec, extension at 72 °C for 30 sec, final extension at 72 °C for 10 min and hold at 10 °C. Annealing temperature of primers was calculated by the web application NEB Tm calculator (https://tmcalculator.neb.com/#!/main). PCR was checked by agarose gel electrophoresis on a 2% agarose gel. Gel electrophoresis runs were done with 80V for 40 min.

The next day, a fresh 2% agarose gel was prepared, and its tub was irradiated with UV light for 11 min. Then the PCR product from the day before was loaded on the gel and gel electrophoresis runs were done with 80V for 45 min. Afterwards, the PCR product was cut out of the gel with a flamed spatula and the enclosed DNA was purified with the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) following manufacturer´s protocol except for the final elution step, where DNA was eluted with 17 µl nuclease free water instead of 50 µl buffer EB (10mM Tris-Cl, pH 8.5). Before starting to add the TOPO vector, a ligation step containing dATPs was necessary to add an adenine overhang to the purified PCR products because the Phusion DNA polymerase does not produce adenine overhangs but only blunt end PCR products. Overhangs are needed for the TOPO vector, which has 3´ thymine overhangs, so that it can easily ligate to the PCR product. Ligation reaction containing TOPO vector was done on ice and contained 2 µl purified PCR product, 0.5 µl salt solution and 0.5 µl TOPO vector 2.1. It was gently mixed and incubated for 5 min at room temperature. Then 20 µl of *E. coli* suspension (TOP10 bacterial strain from Invitrogen TOPO™ TA Cloning™ Kits) was carefully transferred into a new 2 ml tube to which 2 µl of ligation reaction was added. This tube was incubated for 30 min at room temperature. After heat shocking the cells for 40 seconds at 42 °C they were transferred back on to ice. Then 300 µl of pre-warmed S.O.C. medium was added to the cells to obtain maximal transformation efficiency. Then the tube was incubated in a mini oven at 37 °C for one-hour, with slow rotating overhead. In the meantime, an already prepared X-Gal solution, which was prepared by mixing 25 mg X-Gal and 1 ml dimethylformamide together, was spread with glass-spatulas on Luria-Bertani (LB) agar plates. These LB agar plates were prepared by weighing and mixing 20 g Peptone, 10 g Yeast

Extract and 20 g NaCl together and dissolving it altogether in 2000 ml distilled water. This 2000 ml solution was then divided in 4 x 500 ml bottles and 7.5 g agar was added to each bottle. Subsequently all four bottles were autoclaved and cooled down to 50 °C before 1500 µl ampicillin was added. Then the LB plates were poured. These Plates were incubated at 37 °C to completely dry the X-Gal solution. From the transformation 250 µl was spread with a glass-spatula on the first LB agar plate and 70 µl on the second LB agar plate. This difference in volume is to make sure, that at least on one plate well-spaced colonies are present. The plates were then incubated overnight at 37 °C.

To check for successful transformation of the desired Amplicon, a “Colony PCR” was performed in a 25 µl reaction volume, containing 21.4 µl nuclease free water, 2.5 µl 10x buffer, 0.5 µl dNTPs (10mM), 0.25 of each primer (universal primers M13 21+ and M13 29-, both 10 µM) and 0.1 µl TopTaq polymerase (5 U/µl). With autoclaved and flamed toothpicks, a selected white colony from the transformation plates was touched and transferred to a new plate. Then the same toothpick was washed into the PCR tube. The used toothpick was burnt again and then discarded. This transfer of the colonies was repeated with all the chosen colonies (first time 12 colonies, second time 14 colonies). Subsequently, the plate with the selected colonies was incubated at 37 °C and the PCR tubes were put into the cycler with following PCR cycling conditions: Initial denaturation at 94 °C for 3 min, [denaturation at 94 °C for 30 sec, annealing at 50 °C for 30 sec, extension at 72 °C for 60 sec] x 30, final extension at 72 °C for 7 min and hold at 4 °C. Then PCR products were run on a 1.5% agarose gel with 80 V for 30 min. An appropriate nutrient medium of 2 ml was poured into small, autoclaved glass eprouvettes. Colonies, which produced after colony PCR a correct fragment size in the gel electrophoresis were chosen to be transferred into the eprouvettes with autoclaved and flamed toothpicks. Eprouvettes were incubated at 37 °C overnight in an incubator shaker (GFL Gesellschaft Fuer Labprtec™ 3032). The next day these liquid cultures were transferred into 2 ml tubes. They were washed and eluted in 60 µl nuclease free water following QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany).

As a final test to make sure the fragment was inserted correctly into the plasmid an EcoRI digestion (two EcoRI restriction sites flank the insertion site in the pCR 2.1 TOPO vector) was done, followed by a 1.5% gel electrophoresis. After successful EcoRI digestion, fragments were sent to Microsynth for sequencing (Microsynth, Balgach, Switzerland).

2.5 ddRAD method

The ddRAD protocol largely followed the protocol by Peterson et al. (2012), with some modifications according to Schwentner and Lörz (2021). DNA concentration of each DNA sample was quantified on a Qubit 3.0 Fluorometer (Invitrogen, Thermo Fisher Scientific, Massachusetts, USA) beforehand.

2.5.1 Library Preparation

For the first 48 samples, which ended up in libraries A, B, C and D, each sample was digested with 1.5 µl fastdigest MspI (100,000 U/ml), 1.5 µl fastdigest EcoRI (20,000 U/ml) and 3 µl 10x buffer (all Thermo Fisher) and incubated for restriction digestion for two hours at 37 °C. At this point each sample had a volume of 30 µl (start volume was 24 µl). For the next 48 samples, which ended up in libraries E, F, G and H, each sample was digested with 0.5 µl fastdigest MspI (100,000 U/ml), 0.5 µl fastdigest EcoRI (20,000 U/ml), 3 µl 10x buffer (all Thermo Fisher) and 2 µl nuclease free water and incubated for restriction digestion for two hours at 37 °C. Like for the first 48 samples, these samples had at this point a volume of 30 µl as well. Subsequently, pools were cleaned with the AmpliClean Cleanup Kit (Nimagen, Nijmegen, The Netherlands) where magnetic beads were used to bind the DNA following the instructions and by using 1.5x volume of the digested DNA (30 µl), which means 45 µl magnetic beads. Each sample was eluted with 21 µl H_2O, of which 19 µl were transferred for adapter ligation. Several specimens were digested twice and one specimen three times and each of these replicates was treated like an independent specimen by receiving its own barcode and index combination. The reason and explanation for this process is explained later in chapter 3.2.1.

To prepare the adapters, which have specific barcodes, one needs 10x Annealing Buffer (AB), which consisted of 500 mM NaCl and 100 mM Tris-Cl. To prepare the adapter stocks, adapter oligos with a concentration of 100 µM per oligo were combined. This happened by combining complementary adapter oligos (100 µM each, Eurofins Genomics, Ebersberg, Germany). Specifically, 100 µl of adapter stocks in a concentration of 40 µM were made by mixing 40 µl of one oligo with 40 µl of the complement oligo and 10 µl 10x AB and 10 µl H_2O. Then the stocks were diluted in 1x AB to the desired concentration, 2 µM for EcoRI and 31.6 µM for MspI in an end volume of 200 µl and 100 µl respectively. Four random nucleotides were added following the barcodes to the MspI adapters to allow downstream detection and removal of PCR duplicates (Franchini et al., 2017; Schwentner and Lörz, 2021). For the ligation reaction of these adapters to the samples, one needs 19 µl digested DNA, 3 µl 10x T4 ligase buffer, 2 µl ligase, 3 µl MspI adapter and 3 µl EcoRI adapter per sample. After adding

to the samples, they were incubated at 22 °C for 1h, heat killed at 65 °C for 10 min and then cooled down at 2 °C per 90 s to 10 °C. At this point, digested samples were individually barcoded with a unique adapter. In the following steps, samples were either processed in batches of eight or sixteen, which were pooled prior to PCR amplification. Either eight or sixteen specimens with the same starting DNA concentration shared the index added by PCR but received individual barcodes during adapter ligation (Table S1).

The targeted gene fragment was excised using BluePippin (1.5% dye free cassettes with marker L; Biozym, Hessisch Oldendorf, Germany), selecting a fragment size between 280-480 bp. Each size selected library was then amplified in four separate PCR reactions to reduce PCR amplification biases and pooled subsequently. The PCR for one reaction comprised 0.3 µl of each primer (10 nM), 5 µl of NEBNext Q5 HotStart HiFi PCR Master Mix (all New England BioLabs, Ipswich, Massachusetts, USA) and 4.4 µl of the pool after BluePippin size selection. The 2-step PCR program consisted of 98°C initial denaturation and 12 cycles of 98°C for 10 sec, 72°C for 30 sec plus a final elongation step at 65°C for 45 sec and a final extension step of 5 min at 65 °C. PCR primers were available with eight different 8 bp indices each and combined to add a unique index combination for each pool (Table S1).
After PCR, the four replicates per library were pooled together and cleaned with AmpliClean (Nimagen, Nijmegen, The Netherlands) magnetic beads (1x volume of PCR reaction) following the instructions, which means adding 40 µl magnetic beads to one pool and eluting with 22 µl nuclease free water, of which 20 µl were transferred as final library. Qubit measurements were done on an Invitrogen Qubit 3.0 Fluorometer. Desired fragment size length (~460 bp) was assessed on a Tapestation (D1000 ScreenTape; Agilent, Santa Clara, USA). All libraries were mixed at equal concentration into a final pool of 10 nM in 30 µl. This was done with the help of the pooling calculator from Illumina (https://support.illumina.com/help/pooling-calculator/pooling-calculator.htm). Paired-end sequencing with 150 bp each was carried out on one Illumina HiSeq4000 lane by Macrogen (Seoul, South Korea).

2.5.2 Data analysis

Macrogen pre-demultiplexed the data into respective pools (based on provided index combinations). All reads were quality trimmed with Trim Galore! v.0.6.2. (https://github.com/FelixKrueger/TrimGalore) to remove sequences with quality phred scores < 30 and to filter trimmed reads on their sequence length of 130 bp using four cores. PCR duplicates were removed by ‘clone_filter’ from STACKS v.2.59 (Rochette et al., 2019) before demultiplexing pools into the eight single libraries. Further on, ‘process_radtags’ from STACKS was applied and libraries were demultiplexed into 96 single individuals based on the barcode combinations using default parameters. As there is a reference genome of *A. gentilis*

gentilis available (August et al., 2022), the reference-based alignment method was applied. Demultiplexed individuals were mapped to the reference genome using BWA version 0.7.13 (Li and Durbin, 2009), which as well converted the output files to required .bam files. Further on, SAMtools version 1.12 (Danecek et al., 2021) was used to sort the generated .bam files. For the reference-based method, 'gstacks' was the first program executed with the number of threads set to 50. It creates loci from the aligned and paired end reads. After 'gstacks' in STACKS was done, the 'population' program was run in STACKS, which calculated the required population genetic statistic like the SNP data file. The chosen parameters in the population program step only allowed for loci to be present if they were in ≥ 65% of the individuals. The generated .vcf file of the SNP data from the population program was converted into a nexus file based on the R script (from Martin Kapun on GitHub, https://github.com/capoony/Trochulus_ddRAD). Subsequently, MEGA X version 10.0.5 (Kumar et al., 2018) was used for phylogenetic analyses. Mean genetic p distances between and within groups were calculated using pairwise deletion option in MEGA X version 10.0.5.
The phylogenetic trees (all midpoint rooted) were calculated with the following settings: Settings for No. of Bootstrap Replications: 100; for Gaps/Missing Data Treatment: partial deletion; Site Coverage Cutoff: 80%.
FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) was used for tree visualization.

2.6 Increasing the concentration using a Desiccator

For ten samples, DNA concentrations were too low (< 10 pg/µl) to start a ddRADseq library preparation (Table 4). Even after new DNA extractions, the optimal starting concentration of 6.25 ng/µl was never reached. Concentrating the DNA by a desiccator was necessary for those samples. As a desiccant calcium chloride (110.99 g/mol) was used. Desiccation was done without vacuum because the drying process of the samples should happen slowly to guarantee a consistent loss of volume. In an ideal situation, the samples do not dry up completely in the desiccator and are taken out at an end volume of 25-30 µl. After measurements and calculations, samples lost approximately 2 µl of volume every hour. With that knowledge one can predict how long each sample needs in the desiccator to reach the desired end volume. Depending on the sample volumes in the beginning, samples were stored in the desiccator for different time spans, ranging from 33 hours to 4 days. Subsequently, samples were taken out and spinned down. If desired volume was lower than 25-30 µl the samples were supplemented with nuclease free water to the desired volume of 30 µl. Then the samples were incubated in a thermomixer at 50 °C and 300 rpm (revolutions per minute) for one hour to facilitate an even distribution and resolution of DNA in nuclease free water.

Afterwards, resuspended samples were incubated at room temperature overnight and end concentration was measured the next day with Qubit 3.0 Fluorometer (Table 4).

Table 4: Samples concentrated by desiccator with initial and final concentrations in ng/µl.

Sample ID	Initial DNA concentration [ng/µl]	Final DNA concentration [ng/µl]
A.henstii_134	0.838	3.16
A.henstii_136	3.5	9.26
A.henstii_138	Too low	0.186
A.henstii_139	1.23	4.54
A.mel.temminckii_146	2.14	2.08
A.mel.temminckii_148	0.868	5.4
A.meyerianus_150	0.206	1.67
A.meyerianus_151	Too low	0.742
A.meyerianus_152	1.49	2.18
A.meyerianus_153	Too low	2.7

3 Results

3.1 Analysis of the nuclear marker Intron 7 of the *β*-fibrinogen gene

Initially 42 specimens were selected for the nuclear marker Intron 7 of the β-fibrinogen gene (*β-fibint 7*) analysis. DNA concentrations for some samples were very low ranging from 0.11 ng/µl to 52.5 ng/µl for the first eluate and 0.07 ng/µl to 0.88 ng/µl for the second eluate. The complete *β-fibint 7* sequence (length 894 – 905 bp) could not be obtained for all specimens. We were unable to obtain any sequences from *A. gentilis fujiyamae* because PCR products for both samples showed a blank agarose gel electrophoresis. Consequently, no samples were sequenced for this subspecies. Out of the 42 samples included in the PCRs (using the various primer combinations), from 25 samples at least two fragments of the *β-fibint 7* marker sequence could be amplified and successfully sequenced, and for 14 samples the complete sequences of *β-fibint 7* were successfully obtained (Table 5).

Overall, *β-fibint 7* sequences showed minimal differences, which are mentioned below, between samples. Therefore, a phylogenetic tree reconstruction of the nuclear marker *β-fibint 7* was not made. The three samples of *A. melanoleucus* showed one base pair difference compared to other samples at position 135 bp (G instead of T). Furthermore, one of them A.mel.melanoleucus_142 has in addition three autapomorphies (all A to G, 809 bp, 923, 1012 (in exon)). Finally, A.mel.temminckii_146 and A.henstii_137 share one substitution (A to G) at position 297 bp. In addition to the substitutions, some length polymorphisms were observed

in some of the sequences. An indel (insertion/deletion) of 4 bp was observed at position 413 bp. While six individuals were homozygous possessing these 4 bp (the remaining individuals either were homozygous for the deletion (A.gen.albidus_119, A.gen.buteoides_82, A.gen.buteoides_85, A.gen.gentilis_7, A.gen.gentilis_9, A.gen.gentilis_11, A.gen.gentilis_12, A.gen.marginatus_58, A.gen.marginatus_70, A.gen.marginatus_78, A.gen.schvedowi_111) or heterozygous (A.gen.gentilis_8, A.gen.gentilis_10, A.gen.arrigonii_53)).

In A.gen.albidus_118 the indel at position 413 bp was 11 bp long (instead of 4 bp) and an additional indel of 7 bp at position 287 bp was observed. To separate length variable alleles in A.gen.albidus_118 and to resolve the resulting unreadable sequences TOPO TA cloning (described in chapter 2.4) was done. After TOPO TA cloning those indels revealed to be either homozygous for deletions or to be homozygous for insertions at position 287 bp and 413 bp, respectively.

Table 5: Overview of sequenced fragments from all 25 samples, in green are samples with complete sequences.

Sample ID	Sequence						
	AccFib1+ AccFib2-	AccFib1+ AccFib3-	AccFib1+ AccFib5-	AccFib4+ AccFib6-	AccFib7+ AccFib9-	AccFib8+ AccFib3-	AccFib8+ AccFib2-
A.gen.ablbidus_118		√					
A.gen.ablbidus_119	√						
A.gen.arrigonii_51			√		√		
A.gen.arrigonii_53			√		√		√
A.gen.atricapillus_128	√						
A.gen.atricapillus_129	√						
A.gen.buteoides_82		√					
A.gen.buteoides_85				√	√		
A.gen.gentilis_7	√						
A.gen.gentilis_8	√						
A.gen.gentilis_9		√					
A.gen.gentilis_10		√					
A.gen.gentilis_11	√						
A.gen.gentilis_12		√					
A.gen.laingi_131			√	√	√		√
A.gen.marginatus_58			√		√		
A.gen.marginatus_70			√	√	√		
A.gen.marginatus_78	√						
A.gen.schvedowi_111			√	√	√		
A.henstii_133			√		√		
A.henstii_137	√						
A.mel.melanoleucus_142			√				
A.mel.temminckii_143			√				
A.mel.temminckii_146			√		√		√
A.meyerianus_152			√	√			

3.2 Nuclear ddRAD sequences

The DNA concentrations for the 96 samples ranged from 1.03 – 292 ng/µl after measuring with micropipettes. After new extractions each sample had a DNA concentration between 110.4–199.92 ng/µl DNA and was brought to 24 µl with nuclease free water. Three samples had a starting concentration between 4.6 and 5.0 ng/µl, 14 samples of 5.2 ng/µl, 16 samples of 5.3 ng/µl, 48 samples of 6.25 ng/µl and 16 samples of 8.33 ng/µl.

After library preparation and sequencing the data was sent back. The raw data consisted of eight pools (8 to 16 individuals in each pool), demultiplexed according to their barcode combinations. The STACKS reference-based pipeline detected a total of 4365 assembled loci

and 4179 SNPs, shared by at least 65 % of the 58 individuals. Out of the 96 individuals in the initial sampling set, 38 had to be removed for the final analysis, because the quality was either below a phred score of 30 or STACKS had problems with phasing alleles correctly leading to loci with high numbers of uncalled bases.

3.2.1 Phylogenetic analyses of nuclear ddRAD sequences

Individuals used as replicates during library preparation, sequencing, and analysis are almost identical in the final tree reconstruction (see Figure 3 and Figure 4: *A. melanoleucus melanoleucus*_141 and *A. melanoleucus temminckii*_146). This control by replicates showed that no biases were generated during library preparation or bioinformatic analyses. Phylogenetic trees from Maximum Likelihood and Neighbor Joining analyses showed the same topology. Furthermore, the bootstrap values between the Maximum Likelihood and Neighbor Joining trees are very similar. Molecular genetic analyses of nuclear ddRAD data supported Holarctic *A. gentilis* to split into two clades. Nearctic *A. gentilis* forms one clade and is sister to the other clade, which includes Palearctic *A. gentilis*, *A. henstii* and *A. melanoleucus*. Palearctic *A. gentilis* forms a clade as well and is sister to a clade comprising *A. henstii* and *A. melanoleucus.* The bootstrap support for these sister relationships is maximal with 100. In the phylogenetic tree *A. melanoleucus* is sister to *A. henstii.* However, one specimen, A. melanoleucus temminckiiI_146, appears inside *A. henstii*, which will be discussed in chapter 4.3.3.

Mean genetic p distances between and within groups were calculated using pairwise deletion option in MEGA X version 10.0.5. The uncorrected mean *p* distance between Nearctic and Palearctic *A. gentilis* is 18.7 %, while distances between Palearctic *A. gentilis* and *A. henstii* and *A. melanoleucus* are 15.05 % and 13.83 %, respectively. The mean *p* distance within the Palearctic clade is 0.86 %, within the Nearctic clade 1.13 %, within *A. henstii* 0.88% and within *A. melanoleucus* 1.79% (A.meltemminckii_146 and A.meltemm_146 were excluded for calculation of mean *p* distance, for explanation see chapter 4.3.3) (Table 6).

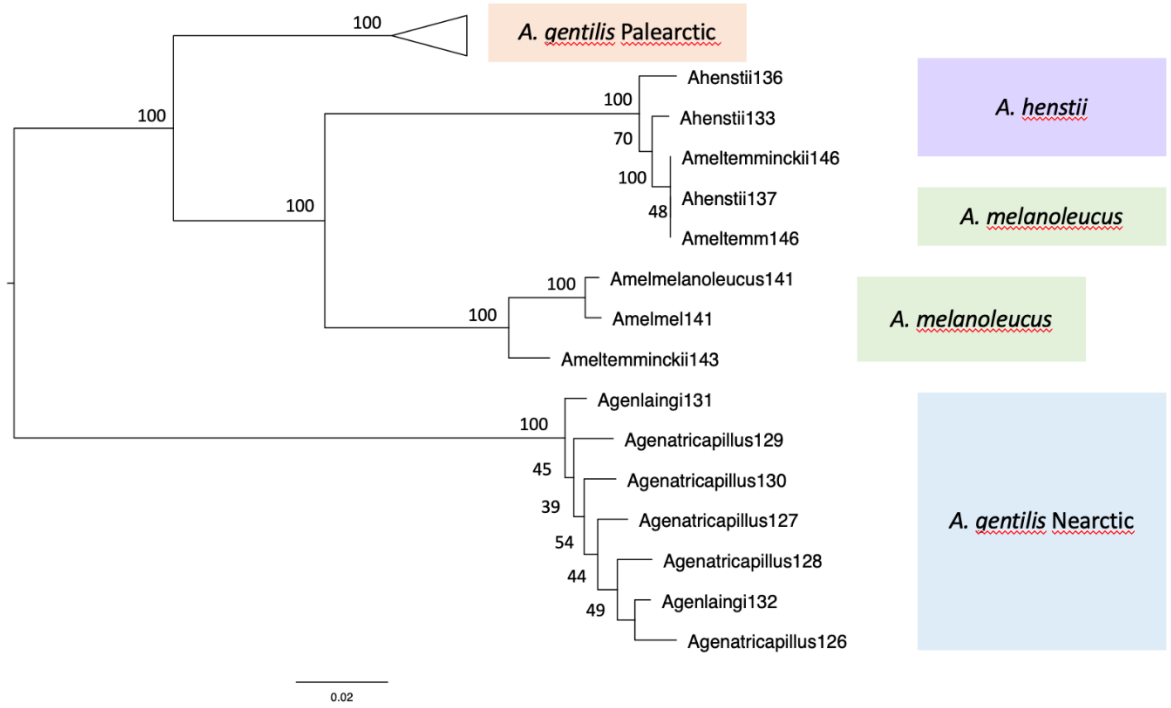

Figure 3: Maximum Likelihood tree (midpoint rooting) based on nuclear ddRAD data for the *Accipiter* [*gentilis*] superspecies. Settings for Gaps/Missing Data Treatment: partial deletion; Site Coverage Cutoff: 80%. Node support (bootstrap values in %) is given at the nodes.

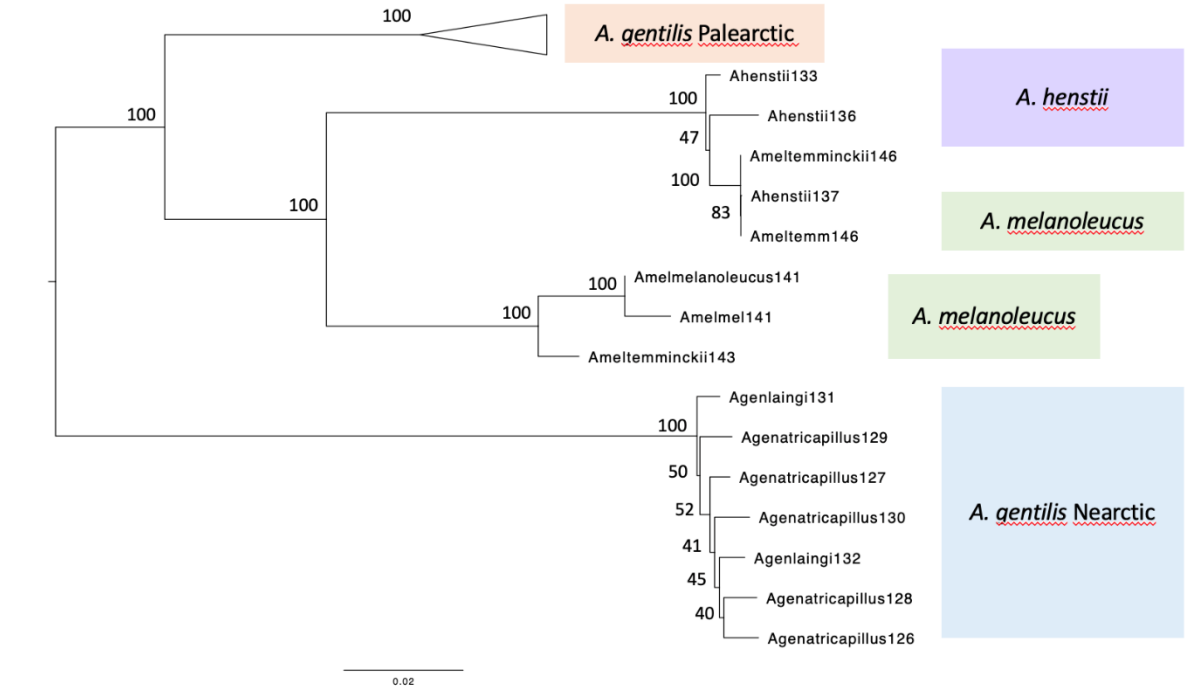

Figure 4: Neighbor Joining tree (midpoint rooting) based on nuclear ddRAD data for the *Accipiter* [*gentilis*] superspecies. Settings for Gaps/Missing Data Treatment: partial deletion; Site Coverage Cutoff: 80%. Node support (bootstrap values in %) is given at the nodes.

Table 6: Uncorrected mean p distances calculated from the nuclear ddRAD data as well as mitochondrial sequences. Below diagonal: nuclear - between group mean distances with standard deviations; diagonal: within group mean distances. Above diagonal in blue: mitochondrial *control* region - between group mean distances with standard deviations (taken from Kunz et al. 2019).

	A. gentilis Nearctic	*A. gentilis* Palearctic	*A. henstii*	*A. melanoleucus*
A. gentilis Nearctic	0.0113	0.043 ± 0.008	0.049 ± 0.009	0.044 ± 0.008
A. gentilis Palearctic	0.1870 ± 0.006	0.0086	0.028 ± 0.006	0.026 ± 0.006
A. henstii	0.2272 ± 0.0067	0.1504 ± 0.0061	0.0088	0.030 ± 0.007
A. melanoleucus	0.2174 ± 0.0069	0.1383 ± 0.0050	0.0736 ± 0.0030	0.0179

According to Kunz et al. (2019), Palearctic *A. gentilis* have a high genetic similarity. Thus, in Figure 3 and 4 the Palearctic clade was shown collapsed, otherwise the tree would have been not readable. Therefore, the Palearctic clade is shown again by itself and not collapsed as a Neighbor Joining tree and Maximum Likelihood tree (Figure 5 and Figure 6). Both trees present a similar topology and the bootstrap values between both trees are very similar. Nevertheless, there are differences, while Agenbuteoides80 and Agenbuteoides82 in the Neighbor Joining tree show some differences, there are identical in the Maximum Likelihood tree. Same with Agenmarginatus79 and Agenarrigonii45.

In both trees two clades are forming: One clade consists of the subspecies *A. gentilis schvedowi*, *A. gentilis buteoides* and *A. gentilis albidus* and the other clade the subspecies *A. gentilis gentilis, A. gentilis marginatus* and *A. gentilis arrigonii*. Nevertheless, there is one specimen, *A. gentilis gentilis*_15, located inside the first mentioned clade instead of the second mentioned one.

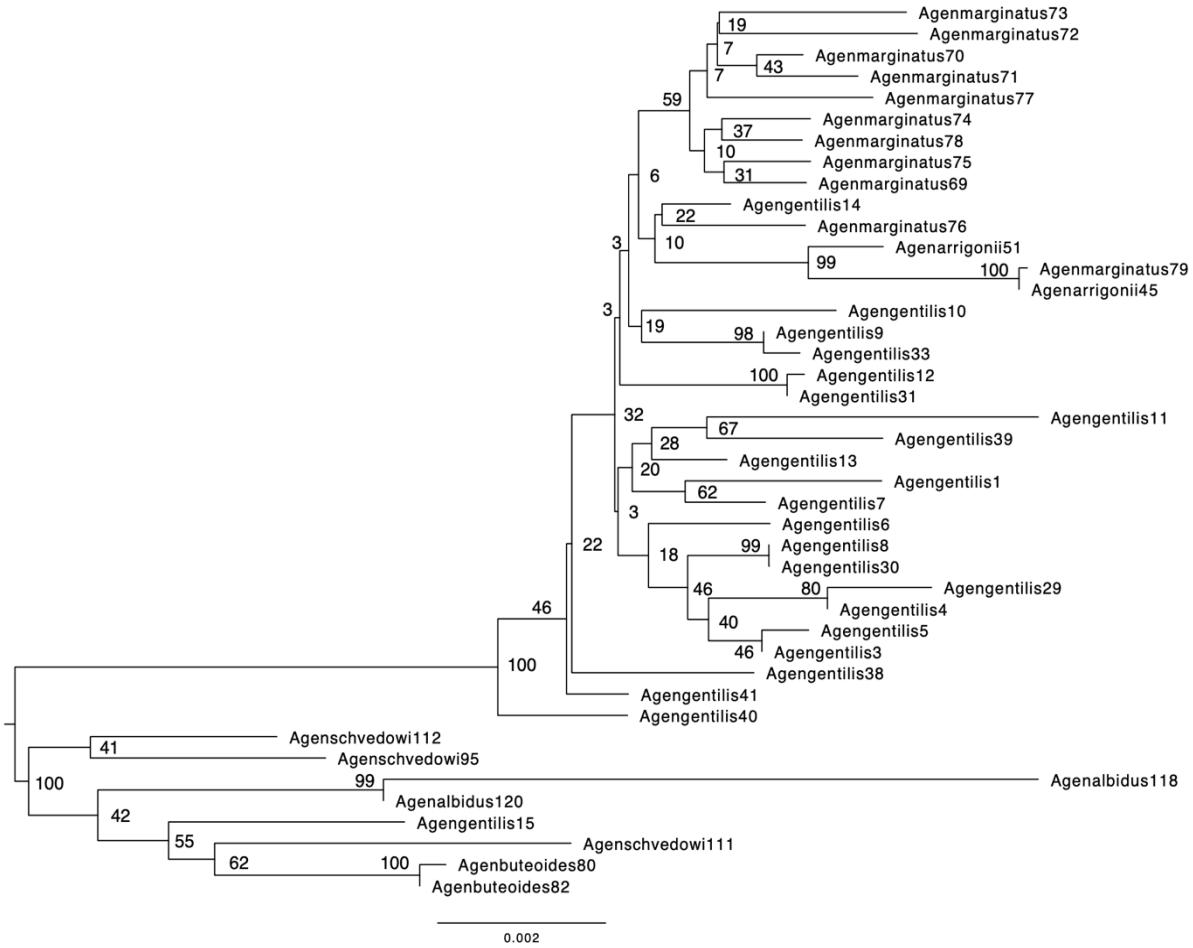

Figure 5: Neighbor Joining tree (midpoint rooting) based on nuclear ddRAD data for Palearctic *A gentilis*. Settings for Gaps/Missing Data Treatment: partial deletion; Site Coverage Cutoff: 80%.

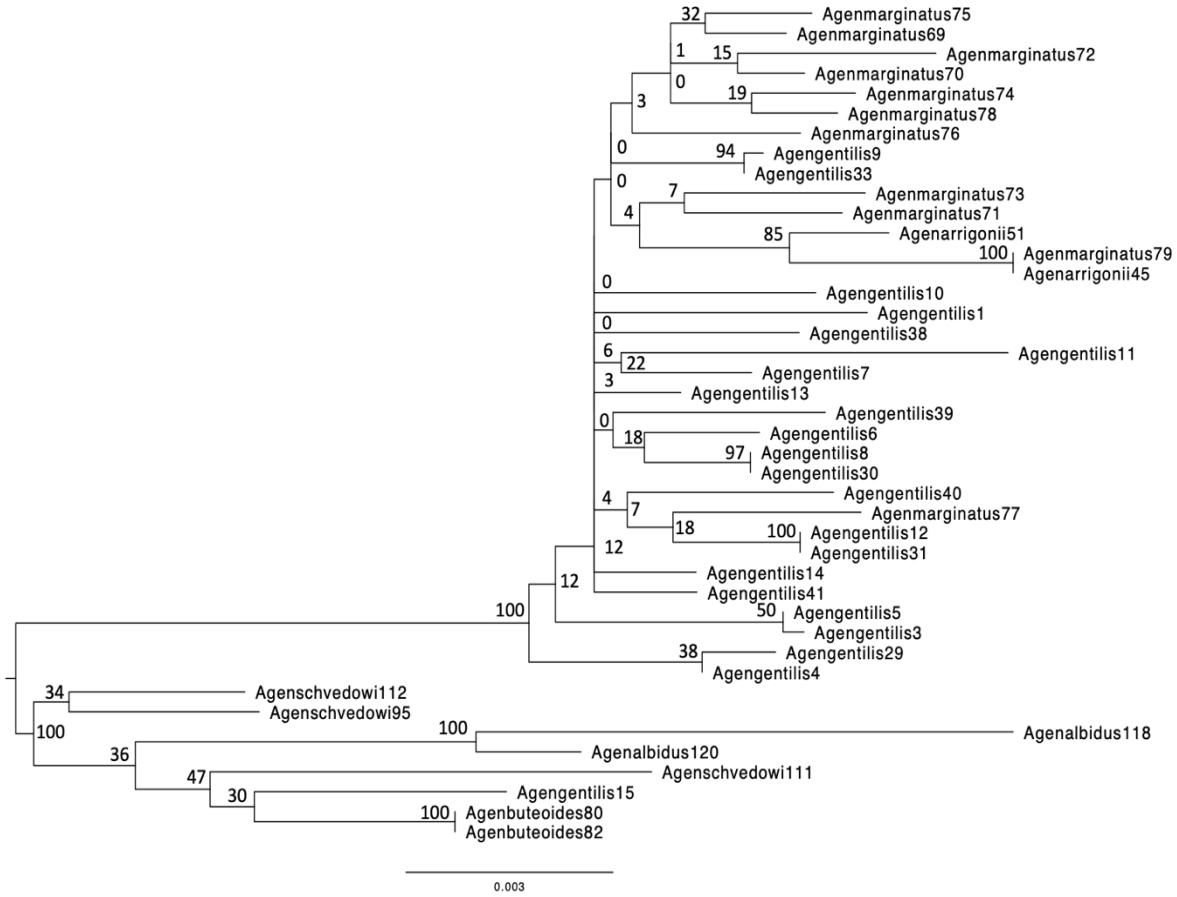

Figure 6: Maximum Likelihood (midpoint rooting) based on nuclear ddRAD data for Palearctic *A gentilis*. Settings for Gaps/Missing Data Treatment: partial deletion; Site Coverage Cutoff: 80%.

4 Discussion

In chapter 1.5 six questions have been asked regarding the present study. These questions are listed below with short answers followed by detailed discussion of the various topics (4.2, 4.3).

1. Are the two marker systems (*β-fibint 7*, ddRAD) or one of them informative for resolving the phylogeny of *Accipiter* [*gentilis*] superspecies?
   The nuclear marker *β-fibint 7* was not very informative because it only showed minimal differences between the sequenced samples. The data of the ddRADseq were much more informative, because 4179 SNPs between the samples have been generated.
2. Are the nuclear results in congruence with the phylogenetic tree based on mitochondrial data from Kunz et al. (2019)? Do they contradict each other?
   Even though, the data of the nuclear marker *β-fibint 7* only showed minimal differences, there are single substitutions, which support the genomic ddRAD data. Hence, the results from both methods support the mitochondrial data from Kunz et al. (2019).
3. Are the four allospecies of the *Accipiter* [*gentilis*] superspecies differentiated in the generated phylogenetic trees based on nuclear markers?
   With the genomic ddRAD data, the three allospecies, *A. gentilis, A. henstii, A. melanoleucus,* could be differentiated in the phylogenetic trees. Due to low DNA concentrations of the samples from the allospecies *A. meyerianus* this specimen could not be included in the analysis for the ddRADseq, hence no statement could be drawn for this specimen. The data of the *β-fibint 7* analysis did not provide enough information to give a realistic statement about the allospecies.
4. Are the Holarctic *A. gentilis* specimens forming one Nearctic clade and one Palearctic clade?
   Yes, the Holarctic *A. gentilis* specimens are separated in two clades as in the mitochondrial tree.
5. Is it possible to gain enough data (whether with *β-fibint 7* or ddRADseq) when sequencing mainly museum material?
   Yes, it is possible to gain enough data for both methods, although not all samples worked sufficiently.
6. What taxonomic interpretation can further be made with the data generated?
   Splitting *A. gentilis* into two species *A*. [*gentilis] gentilis* and *A.* [*gentilis*] *atricapillus* appears as the most reasonable solution.

4.1 Limitations of the study

A drawback of the present study is that no outgroup was included during the ddRAD library preparation. The inclusion of samples from an outgroup (e.g., *Accipiter nisus*, as used by Kunz et al. 2019), has simply been forgotten. Furthermore, not all taxa of the *A.* [*gentilis*] superspecies were included because *A. meyerianus, A. g. fujiyamae* and *A. g. apache* could not be included. For the latter mentioned subspecies, *A. g. apache,* no museum material was available. Thus, no DNA extractions could be done for this specimen. The other two specimens, *A. meyerianus* and *A. g. fujiyamae,* could not be included because DNA concentrations of all measured samples were too low to start a useful library preparation for ddRAD. Moreover, it is not advantageous that *A. meyerianus* is not included in the ddRAD analysis because it would have been interesting to analyze based on genomic data if the Palearctic *A. gentilis* would nest within *A. meyerianus* as in the study from Kunz et al. (2019). Furthermore, our sample sizes for the allospecies and subspecies were not the same. For example, the sample size for the Palearctic *A. gentilis* was much bigger than for the Nearctic *A. gentilis*. Hence, the generated number of SNPs for each allospecies should be viewed with caution. Another problem in the present study was the insufficient DNA quality in some samples. This could be due to the old museum materials. If DNA extractions would not have been done with museum material but with fresh samples, the DNA quality would have been better and the PCR amplification and sequencing of the nuclear marker *β-fibint 7* would have work better (discussed in detail below). Moreover, the better DNA quality would also have contributed to better PCR amplification and sequencing for ddRADseq and the number of uncalled bases would have been lower.

4.2 Intron 7 of *β*-fibrinogen gene

In the present study we established well adapted primers binding perfectly to the nuclear marker *β-fibint 7* in *Accipiter.* Nevertheless, in some of the specimens not all fragments could be amplified. Since the primers were specifically designed to fit the *Accipiter β-fibint 7*, it is more likely that the quality of the DNA was not sufficient in those samples. Working with ancient DNA (aDNA) bears the risk that the DNA is too fragmented. Thus, in our case it seems that fragment sizes >300 bp were generally too large and could not be obtained in all cases. This result suggests that it would have been better to aim at even lower amplicon sizes (maximally 250 bp) to increase the amplification success. Nevertheless, even though only 14 complete sequences were obtained it was sufficient to reveal the low variation of this marker sequence in *Accipiter* (comprising the ten taxa). Although only an incomplete *β-fibint 7*

sequence could be obtained for several specimens, the high sequence similarity (identical in most cases) was found in all fragments. However, at least, for all three *A. melanoleucus* specimens *β-fibint 7* showed one base pair difference to the other specimens. Even though it is just one base pair difference, this indicates that *A. melanoleucus* is slightly different from the other specimens in the marker *β-fibint 7*. Furthermore, one specimen of the three *A. melanoleucus*, *A. melanoleucus*_146, shares one substitution with *A. henstii*_137. This substitution supports the genomic data generated by ddRADseq, because the similarity between those two specimens is also present there (chapter 4.3.2) and in the mitochondrial tree from Kunz et al. (2019) as well.

Despite *β-fibint 7* is a well-studied marker in avian phylogenetics that was used in several groups (Fain and Houde, 2004; Hackett et al., 2008; Johnsen et al., 2010; Prychitko, 2003), this marker was insufficient to resolve or to add much information on the phylogenetic relationships within the *A.* [*gentilis*] superspecies complex.

4.3 Nuclear ddRAD sequences

4.3.1 Challenges with the library preparation

During library preparation for ddRADseq the difficulties of aDNA became evident as well. Not only is aDNA highly fragmented and made amplification of longer fragments by PCR more difficult as mentioned above, but DNA concentrations were often too low for ddRAD library preparation. As already mentioned in chapter 2.5, a minimal DNA concentration is necessary (in our case 4.6 ng/µl in a volume of 24 µl) to start ddRAD library preparation. Even after new DNA extractions from museum material the required DNA concentrations have not been reached for some samples after measuring with Qubit dsDNA HS assay. Therefore, not all taxa of the *A.* [*gentilis*] superspecies complex could be included in the present study because of low DNA concentrations. Out of the four allospecies (*A. gentilis, A. melanoleucus, A. henstii, A. meyerianus*) *A. meyerianus* was not included and out of the ten subspecies of *A. gentilis* two, namely *A. gentilis apache* and *A. gentilis fujiyamae*, were not included into the ddRAD analysis.

Challenges regarding the concentration also occurred in the final PCR. To avoid the overrepresentation of some fragments in the final indexed library, PCR cycles were held at a minimum (we run 10-12 cycles). But for some libraries the DNA concentration were too low after PCR, and they had to be repeated with more cycles (14 cycles). The trade-off between running too many cycles in the PCR to avoid overrepresented fragments and ending up with too low DNA concentration after PCR became evident. This was challenging to estimate and

decide because the whole process leaves one in uncertainty until after sequencing and bioinformatic work is done.

4.3.2 Phylogeny

4.3.2.1 Non-monophyly of *A. gentilis*

The nuclear ddRAD data support the phylogenetic relationships of the *A.* [*gentilis*] superspecies based on mitochondrial sequences (Kunz et al., in 2019). In both phylogenetic hypotheses (nuclear and mitochondrial), Holarctic *A. gentilis* formed two clades, one Palearctic clade and one Nearctic clade. Although the Palearctic and Nearctic clades are currently considered as belonging to one species, the Palearctic clades was more closely related to the Old-World taxa (*A. melanoleucus, A. henstii*). This was visible in the trees as well as when comparing the genetic distances with those from Kunz et al. (2019) (Table 6). In the mitochondrial tree this resulted in the paraphyly of *A. gentilis,* even though the bootstrap values were only moderate. In contrast, it must be considered that the nuclear tree lacked an outgroup. Therefore, the non-monophyly of *A. gentilis* cannot be deduced from the genomic data obtained in the present study, although Nearctic and Palearctic *A. gentilis* are split into two distinct clades. To facilitate a good comparison between the topologies of the mitochondrial tree (Figure 7) from Kunz et al. (2019) and the phylogenetic trees generated in the present study, the latter were midpoint rooted. Although the two topologies must be interpreted with caution, there is at least no discrepancy between mitochondrial and nuclear trees. The distinctness of clades and the genetic distances are in agreement (Table 6).

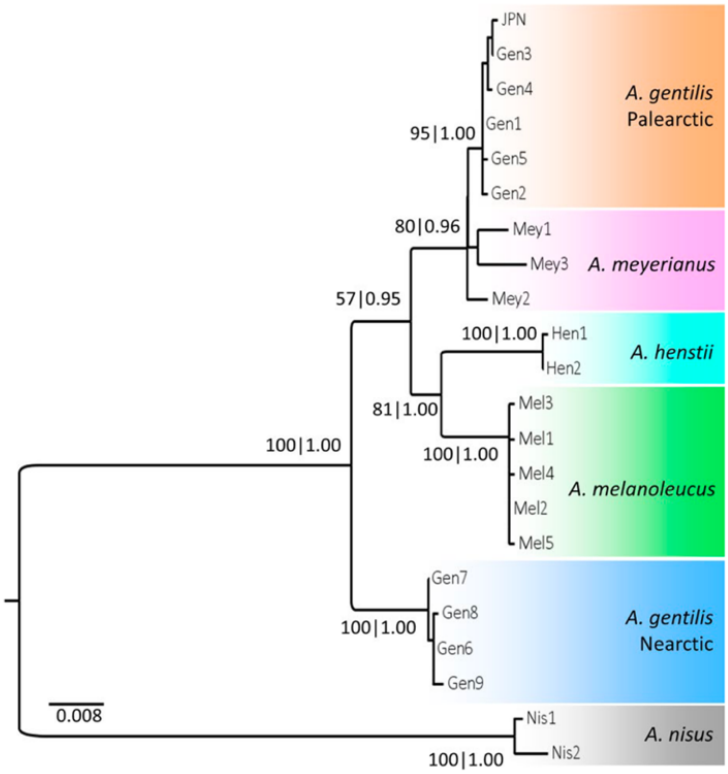

Figure 7: Maximum Likelihood tree based on the concatenated *cytochrome b* and *control region* dataset for the *Accipiter [gentilis]* superspecies and *Accipiter nisus* as outgroups. Node support is given as bootstrap values (%) and posterior probabilities (floating point) (Kunz et al., 2019).

4.3.2.2 Taxonomic considerations

In the literature Holarctic *A. gentilis* is specified as one of four allospecies in the *Accipiter* [*gentilis*] superspecies concept (Amadon, 1966). However, our data showed that Holarctic *A. gentilis* forms two clades, where the Palearctic clade is more closely related to the Old-World taxa (*A. melanoleucus* and *A. henstii*) than to Nearctic *A. gentilis*. Together with the results from the mitochondrial sequences, Holarctic *A. gentilis*, based on the current taxonomy, is most probably non-monophyletic. Even though one allospecies (*A. meyerianus*) and two *A. gentilis* subspecies (*A. gentilis apache*, *A. gentilis fujiyamae*) were absent in our dataset, a taxonomic revision would seem necessary. The most straightforward solution would be to split Holarctic *A. gentilis* into two species as already Kunz et al. (2019) and Sangster et al. (2022) suggested. This would result in five instead of four allospecies within the *Accipiter* [*gentilis*] superspecies: *A.* [*gentilis*] *gentilis*, *A.* [*gentilis*] *atricapillus*, *A.* [*gentilis*] *meyerianus*, *A.* [*gentilis*] *melanoleucus* and *A.* [*gentilis*] *henstii.* In this context, Sangster et al. (2022) listed the following subspecies for these two species: (1) *Accipiter gentilis* (Eurasian goshawk) including *A. g. gentilis* (Linnaeus, 1758), *A. g. buteoides* (Menzbier, 1882), *A. g. albidus* (Menzbier, 1882), *A. g. schvedowi* (Menzbier, 1882), *A. g. fujiyamae* (Swann & Hartert, 1923), *A. g. marginatus*

(Piller and Mitterpacher, 1783), and *A. g. arrigonii* (O. Kleinschmidt, 1903). (2) *Accipiter atricapillus* (American goshawk) including *A. a. atricapillus* (A. Wilson, 1812), *A. a. laingi* (Taverner, 1940) and *A. a. apache* van Rossem, 1938.

The other possibility, to lump the whole *Accipiter* [*gentilis*] superspecies complex into one species, *A. gentilis* does not appear reasonable because of the differences in several aspects. Besides the mitochondrial analysis of Kunz et al. (2019) and the genomic results presented here, there are further studies which underline the differentiation of the Palearctic and Nearctic groups. Geraldes et al. (2019) showed a differentiation between the Palearctic and Nearctic groups by analyzing SNPs using genotyping-by-sequencing of high-quality genetic samples. Moreover, Sangster (2022) showed that this differentiation was also visible in vocalization, where Nearctic and Palearctic groups showed differences in the duration of call-notes. Sangster (2022) also mentioned differences in the plumage pattern between *A. g. atricapillus* and *A. g. gentilis* individuals. While the coloration of the upperparts and the upper-wings in *A. g. atricapillus* is pure grey or blue-grey, in *A. g. gentilis* it is more brownish-grey. Besides that, the head patterns of *A. g. atricapillus* is much more contrasting to their upperparts, because *A. g. atricapillus* has much darker crowns and ear-coverts than *A. g. gentilis*. In contrast, *A. g. atricapillus* has much paler underparts than *A. g. gentilis*, which are dark brown there (Cramp and Simmons, 1980; Ferguson-Lees and Christie, 2001; Wattel, 1973).

4.3.2.3 Palearctic clade

A closer look on the Palearctic clade is presented as Maximum Likelihood and Neighbor Joining trees (Figure 5 and Figure 6) including only individuals from the Palearctic. Two groups became apparent: (1) including *A. gentilis buteoides, A. gentilis schvedowi, A. gentilis albidus* and (2) including the more western subspecies *A. gentilis gentilis, A. gentilis marginatus, A. gentilis arrigonii.* Thus, these two groups roughly reflect the geographic distribution of subspecies. Possibly, the Ural Mountains might act as a geographic barrier, which separates these subspecies groups from each other. Still, one sample, A.gengentilis15, clustered inside of the first mentioned clade instead of the second one. By checking the genomic sequence generated by ddRADseq there were many uncalled bases in this sample. Hence, the possibility, that the position of this sample in the tree was artificial and not trustworthy cannot be ruled out. Furthermore, in the trees of the Palearctic clade (Figure 5 and Figure 6) A.genalbidus118 showed an extremely long branch. Other than the sample A.gengentilis15, the ddRAD sequence of this sample did not contain many uncalled bases. Hence, the position of this sample in the tree could be of artificial but unclear source.

4.3.3 Discrepancies concerning *A. melanoleucus*

Another point to discuss in the present study is the phylogenetic position of A.mel.temminckii_146 and A.mel.temm_146 in the Maximum Likelihood tree and Neighbor Joining tree (Figure 3 and Figure 4). Both represent sequence data from the same specimen (but different ddRAD runs) and, moreover, sample Mel5 in Kunz et al. (2019) also represents this single specimen (Figure 7). Thus, both DNA samples, the first one extracted by Kunz et al. (2019) and the second one extracted in the present study were taken from the same specimen (here indicated by the same sample number 146). They were treated as two different samples by barcoding them differently during preparation of the ddRAD library. The same approach was done with A.mel.melanoleucus_141 and A.mel.mel_141. This step was performed to test the accuracy of the laboratory work and the reproducibility of the laboratory bioinformatic pipeline.

In the trees (Figure 3 and Figure 4) each pair of samples mentioned above (A.mel.melanoleucus_141/ A.mel.mel_141 and A.mel.temminckii_146/ A.mel.temm_146) clustered together. The sequences of A.mel.melanoleucus_141 and A.mel.mel_141 differed slightly, which was also visible in the tree. Those minimal differences were not surprising, given that during PCR, different fragments were amplified by chance and later sequenced. Hence, the samples were not exactly at the same position in the tree, even though we are comparing the data of the same specimen. As one would expect, the samples of A.mel.temminckii_146 and A.mel.temm_146 were identical in the tree. However, surprisingly, they were also identical with the sequence of another specimen, namely A.henstii_137. This contradicts the results from Kunz et al. (2019), where in the mitochondrial tree Hen1 (i.e., A. henstii_137) and Mel5 (i.e., A.mel.temm_146/ A.mel.temminckii_146) did not cluster together. In that tree, the latter clustered with the other *A. melanoleucus* specimens. To explain this discrepancy, one could assume interchanged samples during one step of the laboratory work (or even sampling at the museum collection) or incorrect identification. Still, one must also consider biological causes. Unfortunately, it was not possible to repeat the mitochondrial analysis with the two markers *cytochrome b* (*cyt-b*) and *control region* (*CR*), because neither DNA extractions nor tissue material of those samples were left.However, with the ddRAD data we could obtain mitochondrial sequences of those samples and tried to clarify this discrepancy. By comparing the SNPs from the mitochondrial data, we could identify most shared SNPs between A. henstii_137 and A.mel.temm_146/ A.mel.temminckii_146 (which differentiate them from all other studied specimens), even though they are belonging to different species. Further, the comparison of the mitochondrial SNPs of *A. melanoleucus* specimens again revealed A.mel.temminckii_146/ A.mel.temm_146 to be distinct from the others, A.mel.melanoleucus_141/ A. melmel_141 and A.mel.temminckii_143. The latter three

samples were rather similar as expected. Considering the mitochondrial sequences from the ddRAD data, one could assume that hybridization and introgression of the mitochondrial genome into *A. melanoleucus* might be the reason, although the geographic distribution of the two taxa is quite far distant from each other (*A. henstii* occurs in Madagascar and *A. m. temminckii* in central and western Africa). However, this would explain its similarity to A. henstii_137 in the mtDNA but the clear assignment to *A. henstii* in the nuclear ddRAD data contradict the scenario of mitochondrial introgression. Moreover, this scenario does not explain why in the mitochondrial tree from Kunz et al., (2019) A. henstii_137 and A.mel.temm_146/ A.mel.temminckii_146 do not cluster together.

To summarize there are two unresolved discrepancies: (1) The sample A.mel.temminckii_146, which was identified as *A. m. temminckii* appears inside *A. henstii* with ddRAD data. (2) The mitochondrial sequences obtained by ddRAD contradict those of Kunz et al. (2019). The possibility of interchanged samples in the laboratory is quite low, since the two extractions of A.mel.temm_146 and A.mel.temminckii_146 were done in two separate studies and both samples are still clustering together in the nuclear tree based on ddRAD (Figure 3 and Figure 4). One possibility could be a confusion of samples after the first extraction, but this is presently hard to prove since there is no tissue left. Currently, we are trying to obtain more information on these two specimens. They are from the Natural History Museum in Paris, and we asked for additional tissue samples of them.

5 Conclusion

The attempt to resolve the phylogenetic relationships in the *Accipiter* [*gentilis*] superspecies based on one single nuclear marker, *β-fibint 7,* was not achievable because only minimal differences between the studied samples could be found. Hence, no conclusions for the relationships between the specimens could be drawn based on *β-fibint 7*. Nevertheless, by using a reduced-representation genome sequencing, like ddRADseq used here, many SNPs were generated. Based on the phylogenetic analyses using these SNPs, we showed that the Holarctic *A. gentilis* formed two clades, one Nearctic clade and one Palearctic clade. This supports the genomic results from Geraldes et al. (2019) and the mitochondrial results from Kunz et al. (2019). Moreover, Sangster et al. (2022) support these genetic results mentioned above with their vocalization study as well.

In summary, the present study based on reduced-representation genomes from almost all taxa of the *A.* [*gentilis*] superspecies, provides important information on the phylogeny of *Accipiter.* Further investigations on genomic level including the three missing taxa, *A. meyerianus, A. g. fujiyamae* and *A. g. apache* as well as an outgroup*,* would be helpful to revise the taxonomy in the *A.* [*gentilis*] superspecies.

6 Data availability

Our bioinformatic pipeline is available on https://github.com/mineli8/Accipiter_ddRAD. Most of the analysis pipeline is based on https://github.com/capoony/Trochulus_ddRAD (access on request). Sequences of *β-fibint 7* were uploaded in GenBank.

References

1. Allio, R., Donega, S., Galtier, N., Nabholz, B., 2017. Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker. Mol. Biol. Evol. 34, 2762–2772. https://doi.org/10.1093/molbev/msx197
2. Amadon, D., 1966. The Superspecies Concept. Taylor Francis Ltd Soc. Syst. Biol. 15, 245–249. https://doi.org/10.2307/sysbio/15.3.245
3. Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17, 81–92. https://doi.org/10.1038/nrg.2015.28
4. August, K., Davison, M., Bortoluzzi, C., Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium, 2022. The genome sequence of the northern goshawk, Accipiter gentilis (Linnaeus, 1758). Wellcome Open Res. 7, 122. https://doi.org/10.12688/wellcomeopenres.17821.1
5. Ba, H., Jia, B., Wang, G., Yang, Y., Kedem, G., Li, C., 2017. Genome-Wide SNP Discovery and Analysis of Genetic Diversity in Farmed Sika Deer (*Cervus nippon*) in Northeast China Using Double-Digest Restriction Site-Associated DNA Sequencing. G3 GenesGenomesGenetics 7, 3169–3176. https://doi.org/10.1534/g3.117.300082
6. Ball, A.D., Stapley, J., Dawson, D.A., Birkhead, T.R., Burke, T., Slate, J., 2010. A comparision of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (Taeniopygia guttata) 15.
7. Brumfield, R.T., Beerli, P., Nickerson, D.A., Edwards, S.V., 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18, 249–256. https://doi.org/10.1016/S0169-5347(03)00018-1
8. Cramp S, Simmons KEL (1980) Handbook of the birds of Europe, the Middle East and North Africa: The birds of the Western Palearctic. Vol. II: Hawks to Bustards. Oxford, UK: Oxford University Press.
9. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008
10. Davey, J.W., Blaxter, M.L., 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics 9, 416–423. https://doi.org/10.1093/bfgp/elq031
11. Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510. https://doi.org/10.1038/nrg3012
12. del Hoyo, J., & Collar, N. J. (2014). *HBW and birdlife international illustrated checklist of the birds of the world. Vol. 1: Non-passerines*. Barcelona, Spain: Lynx Edicions.
13. Dickinson, E.C. (Ed.), 2003. The howard and moore complete checklist of the birds of the world, Life of the past. Indiana University Press, Bloomington.
14. Dickinson, E. C., & Remsen, J. V. (2013). *The Howard and Moore complete checklist of the birds of the world. Vol. 1: Non-passerines*. Eastbourne, UK: Aves Press
15. Dool, S.E., Puechmaille, S.J., Foley, N.M., Allegrini, B., Bastian, A., Mutumi, G.L., Maluleke, T.G., Odendaal, L.J., Teeling, E.C., Jacobs, D.S., 2016. Nuclear introns outperform mitochondrial DNA in inter-specific phylogenetic reconstruction: Lessons from horseshoe bats (Rhinolophidae: Chiroptera). Mol. Phylogenet. Evol. 97, 196–212. https://doi.org/10.1016/j.ympev.2016.01.003
16. Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5, 435–445. https://doi.org/10.1038/nrg1348
17. Ericson, P.G.P., Anderson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Källersjö, M., Ohlson, J.I., Parsons, T.J., Zuccon, D., Mayr, G., 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biol. Lett. 2, 543–

547. https://doi.org/10.1098/rsbl.2006.0523
18. Fain, M.G., Houde, P., 2004. PARALLEL RADIATIONS IN THE PRIMARY CLADES OF BIRDS. Evolution 58, 2558–2573. https://doi.org/10.1111/j.0014-3820.2004.tb00884.x
19. Feduccia, A., 1995. Explosive Evolution in Tertiary Birds and Mammals. Science 267, 637–638. https://doi.org/10.1126/science.267.5198.637
20. Ferguson-Lees J, Christie DA (2001) Raptors of the World. Christopher Helm, London, 992 pp
21. Franchini, P., Monné Parera, D., Kautt, A.F., Meyer, A., 2017. quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. Mol. Ecol. 26, 2783–2795. https://doi.org/10.1111/mec.14077
22. Gamauf, A., Gjershaug, J.-O., Røv, N., Kvaløy, K., Haring, E., 2005. Species or subspecies? The dilemma of taxonomic ranking of some South-East Asian hawk-eagles (genus *Spizaetus*). Bird Conserv. Int. 15, 99–117. https://doi.org/10.1017/S0959270905000080
23. Gamauf, A., Haring, E., 2004. Molecular phylogeny and biogeography of Honey-buzzards (genera Pernis and Henicopernis). J. Zool. Syst. Evol. Res. 42, 145–153. https://doi.org/10.1111/j.1439-0469.2004.00250.x
24. Geraldes, A., Askelson, K.K., Nikelski, E., Doyle, F.I., Harrower, W.L., Winker, K., Irwin, D.E., 2019. Population genomic analyses reveal a highly differentiated and endangered genetic cluster of northern goshawks (*Accipiter gentilis laingi*) in Haida Gwaii. Evol. Appl. 12, 757–772. https://doi.org/10.1111/eva.12754
25. Griffiths, C.S., Barrowclough, G.F., Groth, J.G., Mertz, L.A., 2007. Phylogeny, diversity, and classification of the Accipitridae based on DNA sequences of the RAG-1 exon. J. Avian Biol. 38, 587–602. https://doi.org/10.1111/j.0908-8857.2007.03971.x
26. Gupta, A., Bhardwaj, A., 2015. Mitochondrial DNA- a Tool for Phylogenetic and Biodiversity Search in Equines. J. Biodivers. Endanger. Species 01. https://doi.org/10.4172/2332-2543.S1-006
27. Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., Keightley, P.D., 2008. Direct Estimation of the Mitochondrial DNA Mutation Rate in Drosophila melanogaster. PLoS Biol. 6, e204. https://doi.org/10.1371/journal.pbio.0060204
28. Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.-L., Harshman, J., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Steadman, D.W., Witt, C.C., Yuri, T., 2008. A Phylogenomic Study of Birds Reveals Their Evolutionary History. Science 320, 1763–1768. https://doi.org/10.1126/science.1157704
29. Hall, T., 1999. BIOEDIT: A USER-FRIENDLY BIOLOGICAL SEQUENCE ALIGNMENT EDITOR AND ANALYSIS PROGRAM FOR WINDOWS 95/98/ NT.
30. Hanski, I., Simberloff, D., 1997. The Metapopulation Approach, Its History, Conceptual Domain, and Application to Conservation, in: Metapopulation Biology. Elsevier, pp. 5–26. https://doi.org/10.1016/B978-012323445-2/50003-1
31. Helbig, A.J., Knox, A.G., Parkin, D.T., Sangster, G., Collinson, M., 2002. Guidelines for assigning species rank†: Guidelines for assigning species rank. Ibis 144, 518–525. https://doi.org/10.1046/j.1474-919X.2002.00091.x
32. Helbig, A.J., Kocum, A., Seibold, I., Braun, M.J., 2005. A multi-gene phylogeny of aquiline eagles (Aves: Accipitriformes) reveals extensive paraphyly at the genus level. Mol. Phylogenet. Evol. 35, 147–164. https://doi.org/10.1016/j.ympev.2004.10.003
33. Janjua, S., Peters, J.L., Weckworth, B., Abbas, F.I., Bahn, V., Johansson, O., Rooney, T.P., 2020. Improving our conservation genetic toolkit: ddRAD-seq for SNPs in snow leopards. Conserv. Genet. Resour. 12, 257–261. https://doi.org/10.1007/s12686-019-01082-2
34. Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Li, J.,

Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S., Zavidovych, V., Subramanian, S., Gabaldon, T., Capella-Gutierrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford, M.W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F., Sheldon, F.H., Brumfield, R.T., Mello, C.V., Lovell, P.V., Wirthlin, M., Schneider, M.P.C., Prosdocimi, F., Samaniego, J.A., Velazquez, A.M.V., Alfaro-Nunez, A., Campos, P.F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D.M., Zhou, Q., Perelman, P., Driskell, A.C., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, F.K., Jonsson, K.A., Johnson, W., Koepfli, K.-P., O'Brien, S., Haussler, D., Ryder, O.A., Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren, H., Alstrom, P., Edwards, S.V., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320–1331. https://doi.org/10.1126/science.1253451

35. Johnsen, A., Rindal, E., Ericson, P.G.P., Zuccon, D., Kerr, K.C.R., Stoeckle, M.Y., Lifjeld, J.T., 2010. DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. J. Ornithol. 151, 565–578. https://doi.org/10.1007/s10336-009-0490-3
36. Johnson, K.P., Clayton, D.H., 2000. Nuclear and Mitochondrial Genes Contain Similar Phylogenetic Signal for Pigeons and Doves (Aves: Columbiformes). Mol. Phylogenet. Evol. 14, 141–151. https://doi.org/10.1006/mpev.1999.0682
37. Kaiser, S.A., Taylor, S.A., Chen, N., Sillett, T.S., Bondra, E.R., Webster, M.S., 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. Mol. Ecol. Resour. 17, 183–193. https://doi.org/10.1111/1755-0998.12589
38. Kruckenhauser, L., Haring, E., Pinsker, W., Riesing, M.J., Winkler, H., Wink, M., Gamauf, A., 2004. Genetic vs. morphological differentiation of Old World buzzards (genus Buteo, Accipitridae). Zool. Scr. 33, 197–211. https://doi.org/10.1111/j.0300-3256.2004.00147.x
39. Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096
40. Kunz, F., Gamauf, A., Zachos, F.E., Haring, E., 2019. Mitochondrial phylogenetics of the goshawk *Accipiter* [*gentilis*] superspecies. J. Zool. Syst. Evol. Res. 57, 942–958. https://doi.org/10.1111/jzs.12285
41. Kupfermann, H., Satta, Y., Takahata, N., Tichy, H., Klein, J., 1999. Evolution of Mhc–DRB Introns: Implications for the Origin of Primates. J. Mol. Evol. 48, 663–674. https://doi.org/10.1007/PL00006510
42. Lerner, H.R.L., Klaver, M.C., Mindell, D.P., 2008. MOLECULAR PHYLOGENETICS OF THE BUTEONINE BIRDS OF PREY (ACCIPITRIDAE). The Auk 125, 304–315. https://doi.org/10.1525/auk.2008.06161
43. Lerner, H.R.L., Mindell, D.P., 2005. Phylogeny of eagles, Old World vultures, and other Accipitridae based on nuclear and mitochondrial DNA. Mol. Phylogenet. Evol. 37, 327–346. https://doi.org/10.1016/j.ympev.2005.04.010
44. Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
45. Li, Y.-C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E., 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review: MICROSATELLITE EVOLUTION. Mol. Ecol. 11, 2453–2465. https://doi.org/10.1046/j.1365-294X.2002.01643.x
46. Miah, G., Rafii, M., Ismail, M., Puteh, A., Rahim, H., Islam, Kh., Latif, M., 2013. A

Review of Microsatellite Markers and Their Applications in Rice Breeding Programs to Improve Blast Disease Resistance. Int. J. Mol. Sci. 14, 22499–22528. https://doi.org/10.3390/ijms141122499
47. Mindell, D.P., Fuchs, J., Johnson, J.A., 2018. Phylogeny, Taxonomy, and Geographic Diversity of Diurnal Raptors: Falconiformes, Accipitriformes, and Cathartiformes, in: Sarasola, J.H., Grande, J.M., Negro, J.J. (Eds.), Birds of Prey. Springer International Publishing, Cham, pp. 3–32. https://doi.org/10.1007/978-3-319-73745-4_1
48. Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C., Yuri, T., 1999. Interordinal Relationships of Birds and Other Reptiles Based on Whole Mitochondrial Genomes. Syst. Biol. 48, 138–152. https://doi.org/10.1080/106351599260490
49. Morin, P.A., Luikart, G., Wayne, R.K., the SNP workshop group, 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216. https://doi.org/10.1016/j.tree.2004.01.009
50. Moyle, R.G., 2004. Phylogenetics of barbets (Aves: Piciformes) based on nuclear and mitochondrial DNA sequence data. Mol. Phylogenet. Evol. 30, 187–200. https://doi.org/10.1016/S1055-7903(03)00179-9
51. Mundy, N.I., Kelly, J., 2001. Phylogeny of lion tamarins (Leontopithecus spp) based on interphotoreceptor retinol binding protein intron sequences. Am. J. Primatol. 54, 33–40. https://doi.org/10.1002/ajp.1010
52. Nagy, J., Tökölyi, J., 2014. Phylogeny, Historical Biogeography and the Evolution of Migration in Accipitrid Birds of Prey (Aves: Accipitriformes). Ornis Hung. 22, 15–35. https://doi.org/10.2478/orhu-2014-0008
53. Oatley, G., Simmons, R.E., Fuchs, J., 2015. A molecular phylogeny of the harriers (Circus, Accipitridae) indicate the role of long distance dispersal and migration in diversification. Mol. Phylogenet. Evol. 85, 150–160. https://doi.org/10.1016/j.ympev.2015.01.013
54. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7, e37135. https://doi.org/10.1371/journal.pone.0037135
55. Piganeau, G., Gardner, M., Eyre-Walker, A., 2004. A Broad Survey of Recombination in Animal Mitochondria. Mol. Biol. Evol. 21, 2319–2325. https://doi.org/10.1093/molbev/msh244
56. Prychitko, T.M., 2003. Alignment and Phylogenetic Analysis of beta-Fibrinogen Intron 7 Sequences Among Avian Orders Reveal Conserved Regions Within the Intron. Mol. Biol. Evol. 20, 762–771. https://doi.org/10.1093/molbev/msg080
57. Prychitko, T.M., Moore, W.S., 1997. The Utility of DNA Sequences of an Intron from the b-Fibrinogen Gene in Phylogenetic Analysis of Woodpeckers (Aves: Picidae) 12. https://doi.org/10.1006/mpev.1997.0420
58. Riegner, M.F., 2008. PARALLEL EVOLUTION OF PLUMAGE PATTERN AND COLORATION IN BIRDS: IMPLICATIONS FOR DEFINING AVIAN MORPHOSPACE. The Condor 110, 599–614. https://doi.org/10.1525/cond.2008.8503
59. Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28, 4737–4754. https://doi.org/10.1111/mec.15253
60. Rokas, A., Ladoukakis, E., Zouros, E., 2003. Animal mitochondrial DNA recombination revisited. Trends Ecol. Evol. 18, 411–417. https://doi.org/10.1016/S0169-5347(03)00125-3
61. Sangster, G., 2022. The taxonomic status of Palearctic and Nearctic populations of northern goshawk Accipiter gentilis (Aves, Accipitridae): New evidence from vocalisations. Vertebr. Zool. 72, 445–456. https://doi.org/10.3897/vz.72.e85419
62. Schwentner, M., Lörz, A., 2021. Population genetics of cold-water coral associated

Pleustidae (Crustacea, Amphipoda) reveals cryptic diversity and recent expansion off Iceland. Mar. Ecol. 42. https://doi.org/10.1111/maec.12625
63. Sibley, C.G., Ahlquist, J.E., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution, Alfred P. Sloan Foundation. Yale University Press.
64. Song, S.-N., Tang, P., Wei, S.-J., Chen, X.-X., 2016. Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans. Sci. Rep. 6, 20972. https://doi.org/10.1038/srep20972
65. Stresemann, E., Amadon, D., 1979. Order Falconiformes. In: Peters, J.L. (Ed.), Check-list of the Birds of the World. Harvard University Press, Cambridge, MA, pp. 271–425.
66. Squires, J.R., Kennedy, P.L., 2006. Northern goshawk ecology: An assessment of current knowledge and information needs for conservation and management 55.
67. Thrasher, D.J., Butcher, B.G., Campagna, L., Webster, M.S., Lovette, I.J., 2018. Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. Mol. Ecol. Resour. 18, 953–965. https://doi.org/10.1111/1755-0998.12771
68. Wattel J (1973) Geographical differentiation in the genus Accipiter. Nuttall Ornithological Club, Cambridge, MA, 231 pp. https://doi. org/10.5962/bhl.title.155250
69. Weibel, A.C., Moore, W.S., 2002. A Test of a Mitochondrial Gene-Based Phylogeny of Woodpeckers (Genus Picoides) Using an Independent Nuclear Gene, β-Fibrinogen Intron 7. Mol. Phylogenet. Evol. 22, 247–257. https://doi.org/10.1006/mpev.2001.1062

Appendix

Table S1: Specimen list with origin, tissue type, barcoded adapters and indices for ddRADseq as well as information on the *β-fibint 7* analysis. Tissue type: t: muscle tissue, fp: footpad, s: skin, f: feather. Museum voucher ID: AMNH: American Museum of Natural History, MCNB: Natural Sciences Museum of Barcelona, NHRM: Swedish Museum of Natural History, NMW: Natural History Museum Vienna, ZFMK: Zoological Research Museum Alexander Konig, ZMB: Museum für Naturkunde Berlin. ✓ in green: complete fragments of β-fibint 7, ✓: at least one fragment obtained for this sample, X: PCR done for this sample but no *β-fibint 7* could be obtained for this sample.

Sample ID	Year	Museum voucher ID	Tissue type	Locality	barcoded EcoRI adapter	barcoded MspI adapter	Index i7	Index i5	Library	β-fibint 7
A.gen.gentilis_1	2003	NMW_682	t	AUT, Lower Austria	GCTGA	GGATA	CAAGGAAC	AGGCGAAG	E	
A.gen.gentilis_3	2007	NMW_1071	t	AUT, Vienna	GTCCG	CGTCG	TTGCTACC	GGCTCTGA	D	
A.gen.gentilis_4	2008	NMW_1074	t	AUT, Lower Austria	GCTGA	CTGTC	CAAGAGGT	CCTATCCT	C	
A.gen.gentilis_5	2010	NMW_1108	t	AUT, Lower Austria	CGAAT	TATAC	CAAGAGGT	CCTATCCT	C	
A.gen.gentilis_6	2012	NMW_2667	t	AUT, Upper Austria	ACTTC	ACTTC	TTGCCGTT	TAATCTTA	F	
A.gen.gentilis_7	2011	NMW_2698	t	AUT, Lower Austria	GTCCG	GTCCG	TTGCCGTT	TAATCTTA	F	✓
A.gen.gentilis_8	2013	NMW_2699	t	AUT, Burgenland	CTGCG	GTCCG	TTGCTACC	GGCTCTGA	D	✓
A.gen.gentilis_9	2005	NMW_2700	t	AUT, Carinthia	CGAAT	TATAC	TTGCTACC	GGCTCTGA	D	✓
A.gen.gentilis_10	2012	NMW_2730	t	AUT, Carinthia	GTCCG	GTCCG	CAACTGCA	TATAGCCT	A	✓
A.gen.gentilis_11	2010	NMW_2731	t	AUT, Vienna	TATAC	TATAC	CAACTGCA	TATAGCCT	A	✓
A.gen.gentilis_12	2009	NMW_2732	t	AUT, Vienna	GAGAT	GGATA	TTGCTACC	GGCTCTGA	D	✓
A.gen.gentilis_13	2012	NMW_2799	t	AUT, Vienna	GGCTC	CGTCG	CAAGAGGT	CCTATCCT	C	

A.gen.gentilis_14	2013	NMW_2799	t	AUT, Upper Austria	CTGTC	CTGTC	TTGCCGTT	TAATCTTA	F	
A.gen.gentilis_15	2015	NMW_98023	fp	AUT, Vienna	TATAC	TATAC	CAAGGAAC	AGGCGAAG	E	X
A.gen.gentilis_17	1924	NMW_73543	fp	AUT, Styria	TATAC	ACTTC	TTGCTACC	GGCTCTGA	D	
A.gen.gentilis_18	1924	NMW_73544	fp	AUT, Styria	CTGTC	GGATA	CAAGAGGT	CCTATCCT	C	
A.gen.gentilis_21	1922	NMW_56709	fp	HUN, NA	TAGTA	AAGGA	TTGCTACC	GGCTCTGA	D	
A.gen.gentilis_23	1927	NMW_73546	fp	SVN, Municipality of Litija	TCACG	ACTTC	CAAGAGGT	CCTATCCT	C	
A.gen.gentilis_24	1937	NMW_73548	fp	AUT, Styria	ACGGT	TCCGG	CAAGAGGT	CCTATCCT	C	
A.gen.gentilis_29	2008	NMW_95261	fp	AUT, Burgenland	GGATA	CTGTC	CAAGAGGT	CCTATCCT	C	
A.gen.gentilis_30	2013	NMW_96894	fp	AUT, Burgenland	ACGGT	TCCGG	TTGCTACC	GGCTCTGA	D	
A.gen.gentilis_31	2012	NMW_96758	fp	AUT, Carinthia	CGTCG	CGTCG	TTGCCGTT	TAATCTTA	F	
A.gen.gentilis_33	2005	NMW_96895	fp	AUT, Carinthia	CTGTC	CTGTC	CAAGGAAC	AGGCGAAG	E	
A.gen.gentilis_38	2010	NHRM_20106309	t	SWE, Stockholm County	GGATA	GGATA	TTGCCGTT	TAATCTTA	F	
A.gen.gentilis_39	2013	NHRM_20146058	t	SWE, Gotland County	AATTA	AAGGA	TTGCCGTT	TAATCTTA	F	
A.gen.gentilis_40	2012	NHRM_20126855	t	SWE, Kalmar County	TAGTA	TCCGG	CAAGGAAC	AGGCGAAG	E	
A.gen.gentilis_41	2011	NHRM_20116141	t	SWE, Halland, County	CGAAT	ACTTC	CAAGGAAC	AGGCGAAG	E	
A.gen.arrigonii_45	2015	Private collection	f	ITA, Sardinia	CGAAT	ACTTC	TTGCCGTT	TAATCTTA	F	X
A.gen.arrigonii_51	1902-1925	ZFMK	fp	ITA, Sardinia	ACTTC	ACTTC	CAACTGCA	TATAGCCT	A	✓
A.gen.arrigonii_51	1902-1925	ZFMK	fp	ITA, Sardinia	TCACG	ACTTC	TGGACTAC	GTACTGAC	H	✓
A.gen.arrigonii_53	1902–1925	ZFMK	fp	ITA, Sardinia	TATAC	ACTTC	CAAGAGGT	CCTATCCT	C	✓
A.gen.arrigonii_54	1901	ZFMK_7779	s	ITA, Sardinia	ACGGT	AAGGA	CAAGGAAC	AGGCGAAG	E	
A.gen.marginatus_57	2003	NMW_94458	s	GRC, Eastern Macedonia and Thrace	TCACG	ACTTC	TTGCTACC	GGCTCTGA	D	X

A.gen.marginatus_58	1903	ZMB_29411	fp	RUS, Northern Caucasian Federal District (No ISO 3166-2 subdivision)	GCTGA	CTGTC	TTGCTACC	GGCTCTGA	D	✓
A.gen.marginatus_59	1903	ZMB_29420	fp	RUS, Northern Caucasian Federal District (No ISO 3166-2 subdivision)	TAGTA	AAGGA	CAAGAGGT	CCTATCCT	C	
A.gen.marginatus_62	1904	ZFMK	fp	RUS, Kabardino-Balkaria	TCCGG	AAGGA	TTGCTACC	GGCTCTGA	D	
A.gen.marginatus_64	1952	ZFMK_52664	fp	ESP, Castile and León	ACGGT	AAGGA	TTGGAAGG	ATAGAGGC	B	
A.gen.marginatus_64	1952	ZFMK_52664	fp	ESP, Castile and León	CTGCG	GTCCG	TGGACTAC	GTACTGAC	H	
A.gen.marginatus_66	1938	ZFMK_38526	s	ESP, Castile and León	CTGCG	CGTCG	TTGCCGTT	TAATCTTA	F	
A.gen.marginatus_67	1938	ZFMK_38526	s	ESP, Castile and León	CTGCG	CGTCG	CAAGGAAC	AGGCGAAG	E	
A.gen.marginatus_68	1950	ZFMK_527	fp	ESP, Castile and León	ACTTC	TATAC	TTGCTACC	GGCTCTGA	D	
A.gen.marginatus_69	2010	MCNB_20101118-T	t	ESP, Catalonia	GCTGA	GGATA	TTGCCGTT	TAATCTTA	F	
A.gen.marginatus_70	1997	MCNB_970602-T	t	ESP, Catalonia	ACGGT	AAGGA	TTGCCGTT	TAATCTTA	F	✓
A.gen.marginatus_71	2001	MCNB_20010476-T	t	ESP, Catalonia	TAGTA	TCCGG	TTGCCGTT	TAATCTTA	F	
A.gen.marginatus_72	1995	MCNB_990473-T	t	ESP, Catalonia	TCCGG	TCCGG	TTGCCGTT	TAATCTTA	F	
A.gen.marginatus_73	1998	MCNB_990473-T	t	ESP, Catalonia	TCACG	TATAC	CAAGGAAC	AGGCGAAG	E	
A.gen.marginatus_74	2002	MCNB_20020739-T	t	ESP, Catalonia	GAGAT	CTGTC	CAAGGAAC	AGGCGAAG	E	
A.gen.marginatus_75	2001	MCNB_20010744-T	t	ESP, Catalonia	TATAC	TATAC	TTGCCGTT	TAATCTTA	F	
A.gen.marginatus_76	1998	MCNB_980818-T	t	ESP, Catalonia	GGCTC	GTCCG	TTGCCGTT	TAATCTTA	F	
A.gen.marginatus_77	1998	MCNB_980998-T	t	ESP, Catalonia	GGCTC	CGTCG	TTGCTACC	GGCTCTGA	D	
A.gen.marginatus_78	2005	MCNB_20060271-T	t	ESP, Catalonia	TCACG	TATAC	TTGCCGTT	TAATCTTA	F	✓
A.gen.marginatus_79	2016	Private collection	f	GRC, Eastern Macedonia and Thrace	CGAAT	ACTTC	TTGGAAGG	ATAGAGGC	B	
A.gen.buteoides_80	2004	NMW_604	t	RUS, Zabaykalsky Krai	GTCCG	CGTCG	CAAGAGGT	CCTATCCT	C	X

A.gen.buteoides_82	2004	NMW_94534	s	RUS, Siberia (no ISO 3166-2 subdividsion)	GGATA	GGATA	CAACTGCA	TATAGCCT	A	✓
A.gen.buteoides_82	2004	NMW_94534	s	RUS, Siberia (no ISO 3166-2 subdividsion)	GCTGA	CTGTC	TGGACTAC	GTACTGAC	H	
A.gen.buteoides_83	1881	NMW_56703	fp	POL, Silesian Voivodeship	TCCGG	TCCGG	CAAGGAAC	AGGCGAAG	E	
A.gen.buteoides_83	1881	NMW_56703	fp	POL, Silesian Voivodeship	-	-	-	-	-	X
A.gen.buteoides_85	1941	ZMB_29434	fp	TKM, Mary	ACTTC	ACTTC	CAAGGAAC	AGGCGAAG	E	✓
A.gen.schvedowi_95	2014-2015	Private collection	f	CHN, North (no ISO 3166-2 subdivision)	GAGAT	CTGTC	TTGCCGTT	TAATCTTA	F	
A.gen.schvedowi_95	2014-2015	Private collection	f	CHN, North (no ISO 3166-2 subdivision)	-	-	-	-	-	X
A.gen.schvedowi_100	NA	NMW_56701	fp	RUS, NA	CGTCG	CGTCG	CAAGGAAC	AGGCGAAG	E	
A.gen.schvedowi_105	NA	MNHN_1922-2	fp	RUS, Western Siberia (no ISO 3166-2 subdivision)	CGTCG	GTCCG	CAAGAGGT	CCTATCCT	C	
A.gen.schvedowi_106	1907	ZMB_29400	fp	RUS, Republic of Kalmykia	CTGCG	GTCCG	CAAGAGGT	CCTATCCT	C	X
A.gen.schvedowi_106	1907	ZMB_29400	fp	RUS, Republic of Kalmykia	CGAAT	TATAC	TGGACTAC	GTACTGAC	H	
A.gen.schvedowi_107	1907	ZMB_186	fp	RUS, Altai Republic	TCCGG	AAGGA	CAAGAGGT	CCTATCCT	C	
A.gen.schvedowi_108	1906	ZMB_186	fp	RUS, Altai Republic	AATTA	AAGGA	CAAGGAAC	AGGCGAAG	E	
A.gen.schvedowi_110	1905	ZMB_29416	fp	KGZ, NA	CTGCG	CGTCG	TTGGAAGG	ATAGAGGC	B	X
A.gen.schvedowi_111	1905	ZMB_29430	fp	KAZ, Almaty Region	GAGAT	CTGTC	TTGGAAGG	ATAGAGGC	B	✓
A.gen.schvedowi_112	1938	ZMB_29433	fp	TKM, Mary	AATTA	TCCGG	CAAGAGGT	CCTATCCT	C	
A.gen.schvedowi_112	1938	ZMB_29433	fp	TKM, Mary	-	-	-	-	-	X
A.gen.schvedowi_116	1938	AMNH_307795	fp	MMR, Kachin State	GTCCG	GTCCG	CAAGGAAC	AGGCGAAG	E	
A.gen.albidus_118	2014	NMW_2595	t	RUS, Chukotka Autonomous Okrug	AATTA	AAGGA	CAACTGCA	TATAGCCT	A	✓
A.gen.albidus_119	1972	NMW_73603	fp	RUS, Kamchatka Krai	GGATA	CTGTC	TTGCTACC	GGCTCTGA	D	✓
A.gen.albidus_119	1972	NMW_73603	fp	RUS, Kamchatka Krai	TAGTA	AAGGA	TGGACTAC	GTACTGAC	H	

A.gen.albidus_120	1971	NMW_73604	fp	RUS, Kamchatka Krai	CGTCG	GTCCG	TTGCTACC	GGCTCTGA	D	
A.gen.fujiyamae_123	NA	AMNH_532445	fp	JPN, Tochigi	TCCGG	AAGGA	CAAGTCGT	CAGGACGT	G	
A.gen.fujiyamae_124	NA	AMNH_532446	fp	JPN, Saitama	TATAC	ACTTC	CAAGTCGT	CAGGACGT	G	
A.gen.atricapillus_126	2015	NMW_3201	t	USA, Maine	AATTA	TCCGG	TTGCTACC	GGCTCTGA	D	
A.gen.atricapillus_127	2015	NMW_3200	t	USA, Maine	GAGAT	GGATA	CAAGAGGT	CCTATCCT	C	
A.gen.atricapillus_128	2015	NMW_3206	t	USA, Maine	CGTCG	CGTCG	CAACTGCA	TATAGCCT	A	
A.gen.atricapillus_128	2015	NMW_3206	t	USA, Maine	GGCTC	CGTCG	TGGACTAC	GTACTGAC	H	✓
A.gen.atricapillus_129	2015	NMW_3189	t	USA, Maine	CTGTC	CTGTC	CAACTGCA	TATAGCCT	A	
A.gen.atricapillus_129	2015	NMW_3189	t	USA, Maine	GAGAT	GGATA	TGGACTAC	GTACTGAC	H	✓
A.gen.atricapillus_130	NA	Private collection	f	CAN, British Columbia	GGCTC	GTCCG	CAAGGAAC	AGGCGAAG	E	
A.gen.laingi_131	2012	AMNH_839337	fp	USA, Alaska	ACTTC	TATAC	CAAGAGGT	CCTATCCT	C	✓
A.gen.laingi_132	NA	Private collection	f	CAN, British Columbia	TCCGG	TCCGG	CAACTGCA	TATAGCCT	A	
A.gen.laingi_132	NA	Private collection	f	CAN, British Columbia	-	-	-	-	-	X
A.henstii_133	1929	MNHN_1932-3790	fp	MDG, Toamasina Province	CTGTC	GGATA	TTGCTACC	GGCTCTGA	D	✓
A.henstii_136	1930	MNHN_1932-717	fp	MDG, Toamasina Province	GCTGA	GGATA	TTGGAAGG	ATAGAGGC	B	X
A.henstii_136	1930	MNHN_1932-717	fp	MDG, Toamasina Province	GTCCG	CGTCG	CAAGTCGT	CAGGACGT	G	
A.henstii_137	1973	MNHN_1973-547	fp	MDG, Toamasina Province	GGCTC	GTCCG	TTGGAAGG	ATAGAGGC	B	
A.henstii_137	1973	MNHN_1973-547	fp	MDG, Toamasina Province	ACGGT	TCCGG	TGGACTAC	GTACTGAC	H	✓
A.henstii_139	1924	AMNH_532447	fp	MDG, Toamasina Province	GGATA	CTGTC	CAAGTCGT	CAGGACGT	G	
A.mel.mel_141	1959	NMW_98126	fp	TZA, Kilimanjaro Region	CTGTC	GGATA	CAAGTCGT	CAGGACGT	G	
A.mel.melanoleucus_141	1959	NMW_98126	fp	TZA, Kilimanjaro Region	TCACG	TATAC	TTGGAAGG	ATAGAGGC	B	

A.mel.melanoleucus_141	1959	NMW_98126	fp	TZA, Kilimanjaro Region	-	-	-	-	-	X
A.mel.melanoleucus_142	1920	NMW_1485	f	KEN, Bungoma County/Trans-Nzoia County	-	-	-	-	-	✓
A.mel.temm_146	1971	MNHN_1982-215	fp & f	GAB, Ogooué-Ivindo Province	AATTA	TCCGG	CAAGTCGT	CAGGACGT	G	
A.mel.temminckii_143	1957	MNHN_1961-821	fp	BEN, Borgou Department	TAGTA	TCCGG	TTGGAAGG	ATAGAGGC	B	✓
A.mel.temminckii_144	1946	MNHN_1947-521	fp	GAB, Estuaire Province	-	-	-	-	-	X
A.mel.temminckii_145	1970	MNHN_1982-214	fp	GAB, Ogooué-Ivindo Province	-	-	-	-	-	X
A.mel.temminckii_146	1971	MNHN_1982-215	fp & f	GAB, Ogooué-Ivindo Province	GGATA	GGATA	CAAGGAAC	AGGCGAAG	E	✓
A.mel.temminckii_146	1971	MNHN_1982-215	fp & f	GAB, Ogooué-Ivindo Province	CGTCG	GTCCG	CAAGTCGT	CAGGACGT	G	
A.mel.temminckii_148	NA	AMNH_532459	fp	UGA, NA	ACTTC	TATAC	CAAGTCGT	CAGGACGT	G	
A.meyerianus_150	1927	AMNH_220666	fp	SLB, Western Province	-	-	-	-	-	X
A.meyerianus_151	1933	AMNH_333704	fp	PNG, East New Britain	-	-	-	-	-	X
A.meyerianus_152	1972	AMNH_838066	fp	PNG, Morobe Province	-	-	-	-	-	✓
A.meyerianus_153	1958	NHRM_566998	fp	IDN, Papua	-	-	-	-	-	X