The genome assembly and genetic diversity of the laggar falcon (*Falco jugger*)

MSc thesis by Bob van Strien

The biodiversity and whole genome assembly of the Laggar Falcon (*Falco jugger*)

MSc thesis by Bob van Strien – 1042354

ABG80436

Supervised by Richard Crooijmans & Mirte Bosse

Wageningen University,

Wageningen

Table of content

1 Abstract 4
2 Introduction 4
2.1 background 4
2.2 Project aim 7
3 Methods 7
3.1 Sampling and ethics 7
3.2 DNA isolation and genome sequencing 8
3.3 Nanopore long-read sequencing 8
3.4 Illumina short-read sequencing 8
3.5 *F. jugger* reference genome assembly 8
3.6 Short-read alignment to reference genome 8
3.7 Variant calling 9
3.8 Construction of phylogenetic tree with other falcon species 9
3.9 Principal component analysis (PCA) 9
3.10 Genetic diversity, Relatedness and inbreeding (ROH) 9
3.11 Historic and current effective population size 9
4 results 10
4.1 Genome assembly and comparative genomics 10
4.2 Phylogenetics and demographic history 12
4.3 PCA, inbreeding and kinship within the captive laggar population 13
4.4 Historic trend in effective population size 16
5 Discussion 17
5.1 De novo genome assembly of the laggar falcon 17
5.2 Phylogenetics of the hierofalcons 18
5.3 Genetic health of the breeding population 18
5.4 Trend in effective population size (Ne) 19
5 Conclusion 20
6 Acknowledgements 21
References 21
Supplementary information 23

1 Abstract

The laggar falcon (*Falco jugger*) is an extremely understudied bird species native to the middle-east and India classified as Near Threatened. Still, falconers have observed a population decline, expectedly due to anthropogenic pressures. Project Lugger has started a European captive breeding program of the laggar falcon, but information of relatedness and origin of the birds is lacking. Therefore, this study aimed to assess the genetic diversity, phylogeny and demographic history of the captive breeding population. Additionally, to expand the genomic toolbox and investigate overall genome structure, a new reference genome was assembled for the first time. Whole genome sequences of 18 captive animals from 4 European countries were analyzed. Overall 4,7 million single nucleotide variants were detected after quality control, and several structural variants after genome comparison with the gyrfalcon, including a potential inversion on chromosome 3. Overall, the breeding population exhibited low inbreeding and high genetic variability, but also a high pairwise mean-kinship. The mitochondrial phylogenetic tree showed a split of the laggar falcon in two different clades, each being more related to a different sister-species. However, these results rely on the alignment software that was used. The estimations of past effective population sizes (Ne) were extremely low for a Near Threatened species and resembled those of Critically Endangered species. However, the declining trend in Ne did not align with the anthropogenic threats, which emerged relatively recently. This research especially highlights the need for a reliable studbook to prevent further inbreeding.

Keywords; laggar falcon, hierofalcon, reintroduction, breeding program, conservation genomics, inbreeding, phylogeny, population decline, genetic diversity

2 Introduction

2.1 background

Birds (Aves) are one of the most diverse evolutionary lineages and include over 10,000 species. Falcons belong to the genus *Falco* within the relatively small family Falconidae. Due to several radiation periods in the late Miocene period (7.5mya), falcons experienced a rapid diversification, fueled by the extension of open landscapes and C4 grasslands [Fuchs et al., 2015]. The family includes 39 species which can be roughly categorized into four groups: kestrels, hierofalcons, peregrine falcons and hobbies [Wink, 2018].

The hierofalcon clade consists of 4-5 species and have evolved around 420,000 years ago, whereafter they radiated around the world and now distribute a significant part of the globe. This also means that the different species are under different threats and pressures. However, despite their wide distribution, the species are phenotypically and ecologically very similar (Table 1).

- Laggar falcon: monotypic (Afghanistan, Indian subcontinent, Myanmar). There is some speculation about the existence of a subspecies in Myanmar, although this has not been studied yet. The laggar falcon is categorized as Near Threatened, with a decreasing population size of 10,000 to 19,900 individuals.
- Lanner falcon: this species includes five proposed subspecies; *F. b. feldeggii* (Mediterranean Europe), *F. b. erlangeri* (Morocco, Tunisia, Algeria, Libya, S Spain), *F. b. tanypterus* (Egypt), *F. b. abyssinicus* (Ethiopia, Somalia, Eritrea, Togo) and *F. b. biarmicus* (South Africa). The five subspecies have been proposed due to their different phenotypes, but a genetic structure is yet to be found. Globally, lanner falcons are categorized as Least Concern, but the European population faces a decline due to pollution, illegal killing and hunting, leading to its classification as Endangered.

- Gyrfalcon: this species comprises four subspecies with various distributions; *F. r rusticolus* (Scandinavia and N Russia); *F. r. uralensis* (E Russia); *F. r. islandus* (Iceland); *F. r. candicans* (Greenland). The gyrfalcon is currently categorized as Least Concern, with a stable population size ranging from 12,600 to 55.300 individuals, and is facing no significant extinction threats.
- Saker falcon: it includes two confirmed subspecies; *F. c. cherrug* (E Europe, SW Russia, Kazakhstan) and *F. c. milvipes* (SE Russia, Mongolia, China). Recently, Petrov et al., 2023, found significant genetic structure between the two populations. The saker falcon is categorized as Endangered, with a decreasing population size of 12,200 to 29,800 individuals.
- Altai falcon: now, altai falcons are perceived as a color morph subspecies of the saker falcon. But recent evidence suggests that the altai falcons are a distinct species. Their conservation status remains unknown. Al Ajli et al., 2023, distinguished two different morphs based on their morphological similarities to the saker and gyrfalcon; the saker-like and gyr-like altai falcon. however, only the saker-like altai falcon showed genetic distinctiveness.

Table 1: Representation of the interspecific phenotypical differences of hierofalcons [ebird.org]. Note that due to its recent discovery, a picture of the Altai falcon is not included

				
Laggar Falcon – They have a slender build compared to other Hierofalcons. Adults can be defined by their brown upperparts, rusty orange crown and a bold dark eye-stripe that extends to the nape. They are also solid brown at the thighs.	**Lanner Falcon –** They are large and powerful. Adults can be defined by their grey upperparts and rusty orange hindcrown and nape. They are also palish brown at the thighs.	**Gyrfalcon –** They are large, powerful and have a sturdier build compared to other Hierofalcons. Adults can be defined by their long tail and broad wings. Their color ranges from brown to grey and even white. They also have a more diffuse facial pattern and silvery undersides on their flight feathers.	**Saker Falcon –** They are large and powerful. Adults can be defined by their tail, which extends beyond the wingtips, their brown upperparts and their thinner eyelines.	**Altai falcon -** They are large and powerful, and are differentiated from the saker, by their darker brown color, and a bluish tinge on the upper tail covert. Their beaks and feet are slimmer compared to the gyrfalcon.

Nittinger et al., 2005, propose and African origin for hierofalcons, due to the high mitochondrial diversity observed in the lanner falcon. The divergence of the hierofalcons began with the formation and expension of the Sahara desert. The population north of the desert then spread from the Mediterranean to the whole Eurasian continent. The gyrfalcon radiated to North America and colonized the northern parts of Eurasia, including Scandinavia, Mongolia and Russia, while the laggar falcon evolved in Afghanistan, Myanmar, and the Indian subcontinent. The saker falcon emerged from the southern population, eventually colonizing central Asia and parts of Europe via the east-African coast. Those remaining in Africa evolved into the lanner falcon [Pomichal et al., 2014].

Modern genomic analyses allow for a better understanding of the evolution and genetic diversity of species, by investigating their genome sequence and structure. Most bird species have a karyotype containing around

80 chromosomes, of which 7-10 pairs are large- medium-sized, 30 to 33 pairs are microchromosomes, and the sex chromosomes W and Z. But the karyotype of falcons is remarkably different from this pattern. The chromosome number is much lower and ranges from 20 pairs in the merlin (*Falco columbarius*) to 26 in the kestrel (*Falco tinnunculus*). The hierofalcons have 24 chromosome pairs, of which two are the sex chromosomes.

A recent study that used DNA information to compare the saker and gyrfalcon, was able to distinguish the two species based on both whole mitogenomes and W-chromosomes. The hierofalcons started to diverge from the peregrine falcon 2.77 million years ago. Then, they diverged into two separate clades, one comprising the lanner and (saker-like) altai falcon, and the other comprising the saker and gyrfalcon [Al-Ajli et al., 2023]. Hierofalcons started radiating 422,000 years ago, and in turn, the saker and gyrfalcon diverged 109,000 years ago (Figure 1). Though, the position of the lanner falcon and especially the laggar falcon within this evolutionary framework remains unclear.

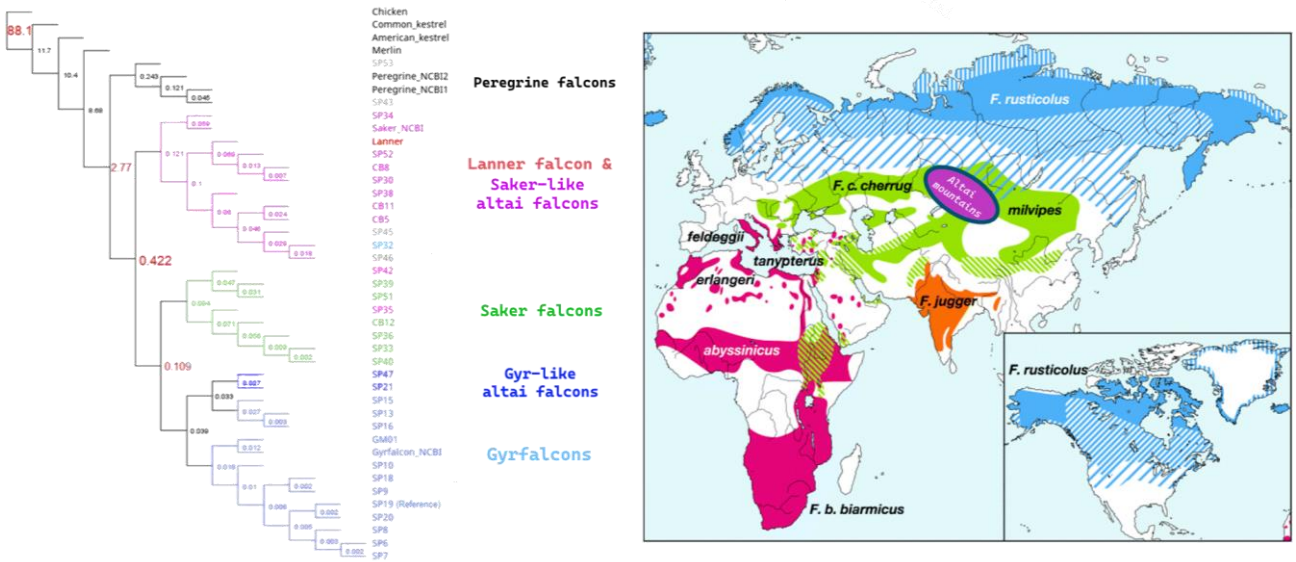

Figure 1: the left figure shows a chronogram of 42 falcons based on their whole mitogenome [Al-Ajli et al., 2023]. The right figure is a distribution map of the hierofalcon clade [Nittinger et al., 2005], showing the gyrfalcon (light blue), saker falcon (light green), lanner falcon (dark red) and laggar falcon (orange). It also shows the (rough) location of the Altai mountains (purple), the habitat of the altai falcon. Both figures contain minor edits.

Having a large and stable population size, the gyr and lanner falcon are doing relatively well in the wild. On the other hand the saker falcon did suffer of a significant population decline since 1990, mainly due to electrocutions on power lines and decreased prey availability. A similar population decline is observed with, the laggar falcon. The decline is caused by pollution (DDT pesticides, heavy metals), and the loss of their primary prey species – the spiny-tailed lizard (*Saara hardwickii*), which contributes up to 80% of their diet [IUCN, 2023; Mori et al., 2019]. Additionally, a significant part of the population decline in both the laggar and saker can be linked to illegal trapping and trade for falconry, a threat that emerged somewhere in the last century. Herein, the smaller, undesirable laggar is used as bait to lure the stronger and more desirable saker. Poor conditions during transit cause a high rate of mortality [Bailey et al., 2000].

Despite IUCN's classification of Near Threatened, the western falconry community are of opinion that the laggar population decrease estimated by IUCN is an underestimation based on too little scientific research. This opinion has been supported by field observations of local falconers in Pakistan, although it has never been professionally measured or reported on a large scale. However, it is clear that within hierofalcons, the laggar falcon is the least well-known and relatively understudied. This results in scarce data and limited information for conservation use. While the status of the lanner, saker and gyrfalcon is supported by recent scientific

literature, the assessment of the laggar has been severely outdated since the 1970s, possibly due to the difficulties of obtaining research permits in the laggar's distribution range. A review in 2019 found that the laggar has only been the subject of four peer-reviewed studies, in contrast to the 40-100 for each of the other hierofalcons [Buechley et al., 2019]. Such a short list of literature can result in difficulty with species management, conservation assessment and predicting future movement, population decline and suitable habitat [Sutton et al., 2020].

Due to this uncertainty, a captive laggar falcon breeding program was established in Europe as a preventive measure by Project Lugger. They started with a founder population of unknown size, of which all founders originated from the same region in Pakistan. The current population size is 100 individuals. Given the small population size, the breeding population should be carefully managed to avoid the occurrence of inbreeding. Since inbreeding is linked to infertility, inbred falcons should not be reintroduced to the wild. They could provide an extra source of competition to an already vulnerable wild population, while not reproducing and contributing to the future generations. Due to a lack of population management, there is no recorded pedigree, and as a result no insight on inbreeding in the population. Additionally, due to their similar looks, hierofalcons can be difficult to distinguish. This could result in accidental hybridization events in the breeding population. If this is the case, a proportion of the breeding population is no longer representative of the wild population. Therefore, there is a need for assessment of the captive and wild laggar falcon population.

2.2 Project aim

This research will focus on the genomics of the laggar falcon (*Falco jugger*). The overall aim is to assess the reintroduction potential of the breeding program. I also assembled the first laggar falcon reference genome. After genome assembly, this research will assess the viability and reintroduction potential of the breeding program in three steps: 1 – What is the evolutionary relationship between the laggar falcon and the remaining hierofalcon species? To test for taxonomic distinctiveness between the falcons in the breeding program and the other hierofalcon species, I will construct a phylogenetic tree based on whole mitogenome sequences. 2 – What is the genetic health of the laggar falcon breeding population? Because of management issues and the small founder size, it is expected that some- if not all falcons will be (recently) inbred, and that some of the falcons will be related. 3 – What is the estimated trend in population size of the laggar falcon in the wild? Since the estimated population trends are outdated, it is unknown whether a breeding program is truly necessary. Because there is no ecological data, I will study the demographic history of the wild population in Pakistan, with a genetic approach.

3 Methods

3.1 Sampling and ethics

A total of 18 EDTA-blood samples were collected by individual falcon breeders in the Netherlands (5), Belgium (4), France (6) and Germany (3) (Table 2). Sampling of the animals was performed by a veterinarian and were leftovers from rest products of veterinary research required by law (CITES). No "Centrale Commissie Dierproeven" (CCD) permit was needed after consulting an Animal Welfare Officer (AWO) of the ethical commission within Wageningen University.

Table 2: information about the 18 selected laggar falcon samples

Sample name	Country of origin	Cites ID	Sex	Hatching date (d.m.y)
Lug_001	Belgium	zgg14.0200530	Female	04.04.20
Lug_002	Belgium	5530npws.w	Female	04.05.20
Lug_003	Belgium	zg12.0120771	Female	10.15.12
Lug_004	Belgium	bof236v12608003	Female	12.03.08

Lug_005	France	F22 120 002 AVF 0310	Unknown	01.04.22
Lug_006	France	F 19 12 002 AVF 03 10	Unknown	06.04.19
Lug_007	France	F 17 12 001 AVF 80 58	Male	17.03.17
Lug_008	France	1206d435avf 250228730006477	Male	21.03.08
Lug_009	France	UOF 5 0520 F11 12 AA 0008	Female	29.05.11
Lug_010	France	F12 12 0001 AVF 3060	Male	27.02.12
Lug_011	Germany	b13.0g190006	Unknown	13.04.19
Lug_012 [assembly sample]	Germany	g140150b140	Female	14.04.14
Lug_013	Germany	b13.0g190003	Unknown	24.03.19
Lug_014	Netherlands	nl2950bec12.010005	Male	09.04.10
Lug_015	Netherlands	NL2449BEC12.010001	Female	Unknown
Lug_016	Netherlands	F11-02-11137	Male	24.03.03
Lug_017	Netherlands	nl2338bec12.012013	Female	24.04.12
Lug_018	Netherlands	wfg044l21unkf	Unknown	25.04.21

3.2 DNA isolation and genome sequencing

The blood samples were sent to Gendika B.V. (Netherlands) for DNA isolation with the DNeasy Blood and Tissue kit, according to its official protocol (QIAGEN, UK).

3.3 Nanopore long-read sequencing

The sample of one individual (Lug_012) was then sequenced using the Oxford Nanopore PromethION 24 machine (PC24B117), with multiple loadings, which were then merged for further analysis. After merging, the average coverage was 26X.

3.4 Illumina short-read sequencing

Library preparation and whole-genome sequencing took place at Novogene (United Kingdom). The 18 samples were sequenced for short-reads using Illumina NovaSeq6000 technology at 10X coverage, resulting in paired-end reads with a 150bp length.

3.5 *F. jugger* reference genome assembly

The Nanopore output was first trimmed using porechop (--discard_middle), to remove the adapters on the end of the long-reads. Then, they were assembled using Flye (--nano-raw). Following, the assembly was scaffolded using longstitch (ntLink-arks) and polished with the Illumina short-read data, using Polca. The assembly statistics were obtained using the Python code get_assembly_stats.py [wur Github], and are published in Table 2. Further, the genome completeness was assessed using BUSCO, with the BUSCO lineage aves_adb10. To detect major structural variants, the new assembly was aligned to the reference genome of the saker, lanner and gyrfalcon using minimap2 (-cx asm5). After this, the alignment was visualized using R [pafCoordsDotPlotly.R, WUR Github]. The 23 scaffolds longer than 15Mb were used to represent the whole-genome in various further analyses. These scaffolds accounted for roughly 75% of the whole-genome.

3.6 Short-read alignment to reference genome

The obtained short-reads of 18 falcons were mapped against the newly assembled reference genome. The reference genome was first indexed with bwa-mem2 and then the short-reads were mapped using bwa-mem2, after which duplicates were removed by samblaster. Then, the bamfiles were sorted and indexed using samtools and assessed for quality by qualimap.

3.7 Variant calling

Next, the bamfiles were used to detect SNP variants with Freebayes (--use-best-n-alleles 4 --min-base-quality 10 --min-alternate-fraction 0.2 --haplotype-length 0 --ploidy 2 --min-alternate-count 2). Low quality SNPs were filtered out using vcftools (-f 'QUAL > 20') and were then indexed using tabix. This resulted in an unfiltered set of 5,752,949 variable sites. For the quality control, this set was filtered for a sequence depth between 4x and 30x, a minor allele frequency higher than 0.02 (~1/36, or one minor allele per 18 samples) and missing data in more than 20% of the samples (vcftools –minDP 4 –maxDP 30 –maf 0.02 --max-missing 0.8), resulting in a set of 4,733,350 high-quality variable sites. Another set of linkage disequilibrium (LD) pruned SNPs was also made using plink (indep-pairwise 50 10 0.1), which resulted in a set of 138,706 SNPs for further analyses.

3.8 Construction of phylogenetic tree with other falcon species

The mitochondrial consensus sequence of each individual was extracted using samtools consensus, where nucleotides with a base quality lower than 20, and a depth lower than 4 were changed to N (-min-BQ 20 -d 4). The control region was excluded from the analysis, because of its high variability, which could interfere with the phylogenetic signal. I then aligned the laggar mitogenomes to the alignment file of Al-Ajli et al., 2023 (Figure 1) using two different alignment software's; MUSCLE and clustalW. The chicken (NCBI accession KT626858.1) was used at an outgroup. The other species included in the analysis were the common kestrel, American kestrel, merlin, peregrine, saker, lanner, gyrfalcon and two morphs of altai falcon (saker-like and gyr-like). Then, a neighbor-joining phylogenetic tree was constructed using Phylip, while ignoring gaps in the alignment.

3.9 Principal component analysis (PCA)

The LD-pruned dataset was used to create a PCA using plink (--pca –allow-extra-chr) and was then visualized in R (pca.R github wur). Additionally, to prevent gender bias, the sex chromosomes were first removed from the dataset. The individual luggers were also assigned to different populations depending on their country of origin, to check for genetic structure between these populations.

3.10 Genetic diversity, Relatedness and inbreeding (ROH)

To check for unexpected kinship, I performed a relatedness analysis using vcftools (--relatedness2). Here, a relatedness φ of >0.354, [0.177, 0.354], [0.0884, 0.177] and [0.0442, 0.0884] corresponds to identical twins, 1st, 2nd and 3rd degree relationships. Following, the mean kinship (MK) was calculated for each falcon, based on the pairwise kinship between that individual and all the remaining falcons in the sample population. To assess the overall genetic diversity of the population, heterozygosity analyses were carried out on the filtered SNP-dataset. This was done using vcftools, to get the nucleotide diversity (π) over a sliding window of 50kb (--window-pi 50000). Then, the whole-genome of each individual was checked for runs of homozygosity (ROHs), using bcftools and accounting for GT errors (--roh -G 30).

Then, the inbreeding coefficient (Froh) was calculated based on the bcftools output, from formula 1:

$$F_{ROH} = \sum k_{length}(ROH_k)/L \qquad (1).$$

Froh corresponds to the percentage of the whole-genome consisting of ROHs. Here, k is the total length of the identified ROHs for each individual, and L is the total length of the whole-genome (~1.2e10^9). The minimum length of ROHs was set to 200kb. Additionally, Fis statistics were calculated using the pruned SNP dataset of all the scaffolds longer than 15Mb in size, and excluding scaffolds aligning to the sex chromosomes. Then the results of the window-pi and ROH detection were plotted on these scaffolds, using Circos.

3.11 Historic and current effective population size

To gain insight into the demographic history of the breeding population, I used the program SNeP. This program estimates historical effective population size trajectories through linkage disequilibrium. Unlike

several other methods such as PSMC, SneP can be applied to larger sample sizes. Therefore, it is more accurate at estimating recent population size history. The filtered SNP data of the 23 largest scaffolds were used as input and the recombination rate was set to 4.6e-009, with the Sved & Feldman (1973) recombination rate modifier. The r2 values were also adjusted for a sample size of 18 (./SneP_111 -recrate 4.6e-009 -svedf -samplesize 18). The other parameters were kept at default. Additionally, since SneP may underestimate very recent Ne, I estimated current Ne using the LD method in NeEstimator v2.1, along with jackknife confidence intervals. The latter was performed on the filtered SNP data of scaffold1, containing 37,333 SNPs. To visualize slight changes in the slope of the trend, NeS analysis was performed over the estimations by SneP. The slope of each linking pair of neighboring Ne estimates was calculated and normalized with the median of the two preceding slope values, according to formula 2 and 3, where Sn is the slope of the Nth pair of neighboring Ne estimates:

$$NeS_n = (S_n - X_n)(1 + X_n)^{-1} \qquad (2)$$

$$X_n = med\{S_n, S_{n+1}, S_{n+2}\} \qquad (3)$$

4 results

4.1 Genome assembly and comparative genomics

The nanopore sequencing resulted in 1.02 million generated reads, with an estimated amount of 16.76Gbp. The long-reads were assembled in Flye, followed by a round of polishing using the illumina short-reads. This resulted in N50 and L50 values of 41.0 Mb and 8 Mb, respectively and the assembly comprised 407 scaffolds, ranging from 112.4Mbp to 476bp. Highly consistent with the estimated genome size of other hierofalcons, the full length of the assembled genome is 1.203Gbp. To assess the completeness of the laggar assembly, I conducted a BUSCO analysis, which identified 8,091 out of 8,338 complete genes, corresponding to 97% of the genes included in aves_adb10. 96.5% were single copies, and 0.5% were duplicates. 200 BUSCOs were missing and 0.6% were fragmented (table 3).

Table 3: the assembly statistics and BUSCO results of the new assembly, compared to the gyrfalcon reference genome.

	F. jugger	***F. rusticolus***
Genome size	1.2 Gbp	1.2 Gbp
Number of scaffolds	407	132
Scaffold L50	8	6
Scaffold N50	41.0 Mbp	91.1 Mbp
Number of contigs	530	767
Contig L50	10	24
Contig N50	36.3 Mbp	15.3 Mbp
Genome coverage	26x	44.3x
	BUSCO – 8,338 genes	
Complete BUSCOs	97.0%	97.0%
Complete single-copy	96.5%	96.4%
Complete duplicated	0.5%	0.6%
Missing	2.4%	2.4%
Fragmented	0.6%	0.6%

21 scaffolds were larger than 15Mb in size, representing ~75% of the whole-genome (Figure 2) and covering large proportions of the larger chromosomes 1 to 15 and chromosome Z. Next, this selection of scaffolds was used to calculate the historic effective population size of the breeding population.

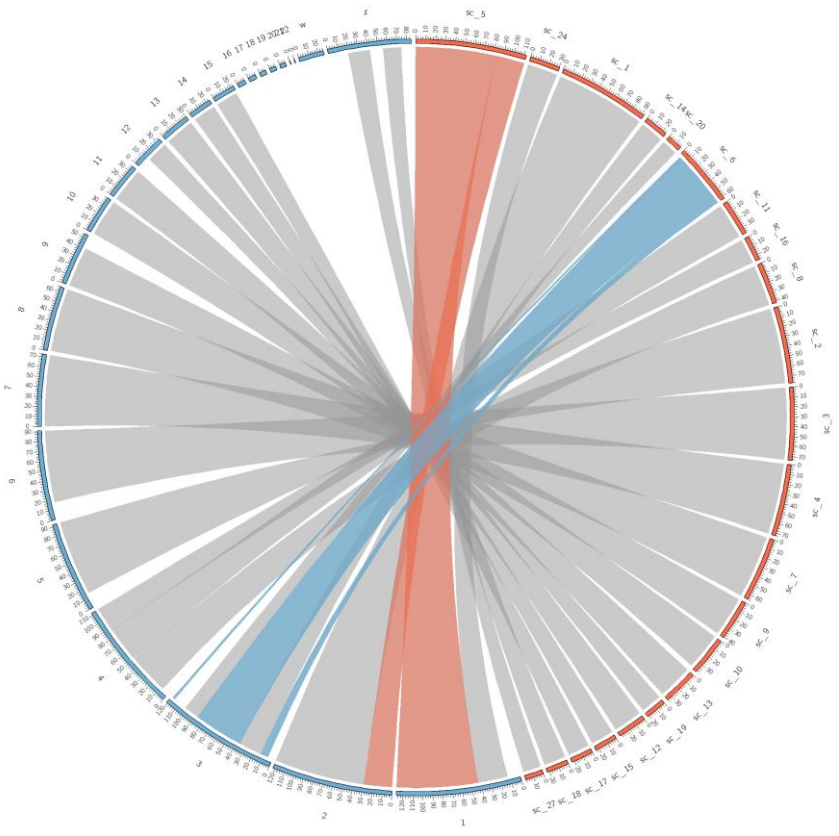

Figure 2: circus plot of the 21 largest scaffolds of the laggar assembly. The red lines on the outer layer tagged sc_1 to sc_27 represent the assembly scaffolds. The blue lines tagged 1 to Z represent the chromosomes of the gyrfalcon. The bands show the alignment of each scaffold to its subsequent chromosome. Scaffold5 (red) and scaffold6 (blue) have been highlighted.

The comparison of the laggar assembly to the gyrfalcon reference genome, depicted in the minimap2 dotplot (Figure 3), highlights some structural variants, mainly on chromosome 3 (NC_051189.1). The most remarkable are the two putative transpositions at the chromosome tips, roughly 5-8 Mb in size. The high amount of fragmentation of the W chromosome can be explained by the female reference sample (Lug_012, WZ haplotype). Another interesting finding is the fact that scaffold5 contains regions of both chromosome 1 and 2 of the gyrfalcon genome.

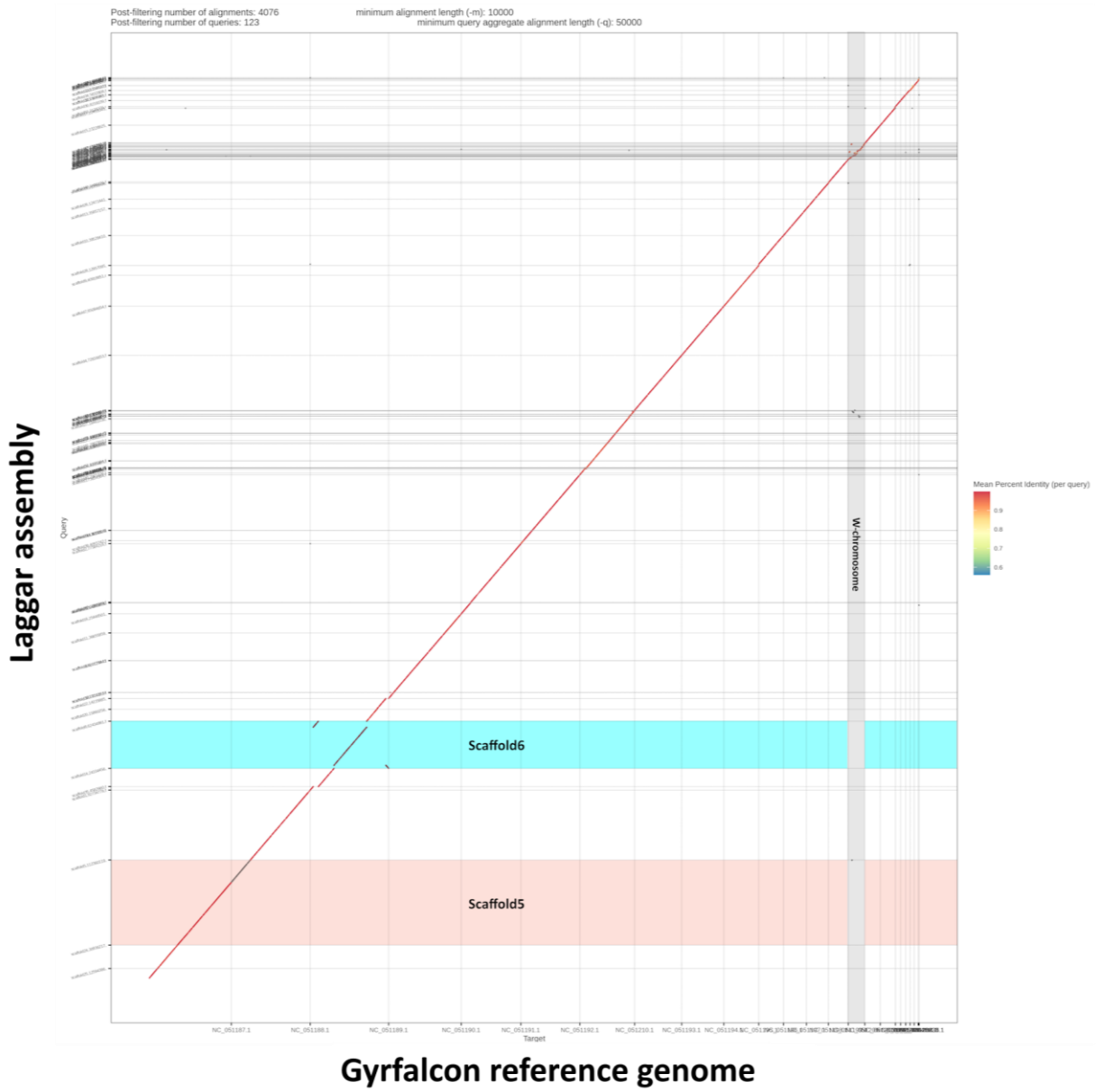

Figure 3: dotplot of the new laggar falcon genome assembly, against the reference genome of the gyrfalcon. Some scaffolds have been colored to highlight structural variants. Namely, scaffold5 (red), scaffold6 (blue) and the fragmented W chromosome (grey).

4.2 Phylogenetics and demographic history

First the consensus sequence of the laggar mitogenomes were extracted from scaffold204, which had a length of exactly 18kb. They were then aligned to the mitogenomes of the remaining hierofalcons and several other species [Al-Ajli et al., 2023]. Consistent with the previously constructed trees, the hierofalcon clade is closely related to the peregrine falcon. The hierofalcon group was further divided into two main clades; clade 1 with the majority of the saker-like altai falcons and the lanner falcon, and clade 2 with the saker falcon, gyrfalcon and gyr-like altai falcon. Contrasting to the expectation, the laggar breeding population is not distinguished in a monophyletic group. It is divided into two different clades, with clade 1 clustering with the saker-like altai falcons and lanner falcon. It should be noted that, likely due to the hierofalcon's close relationship, the results of the phylogenetic tree greatly depended on the alignment software that was used. Figure 4 below shows the results of the ClustalW alignment. The results of the MUSCLE alignment can be found in the appendix (Figure S3).

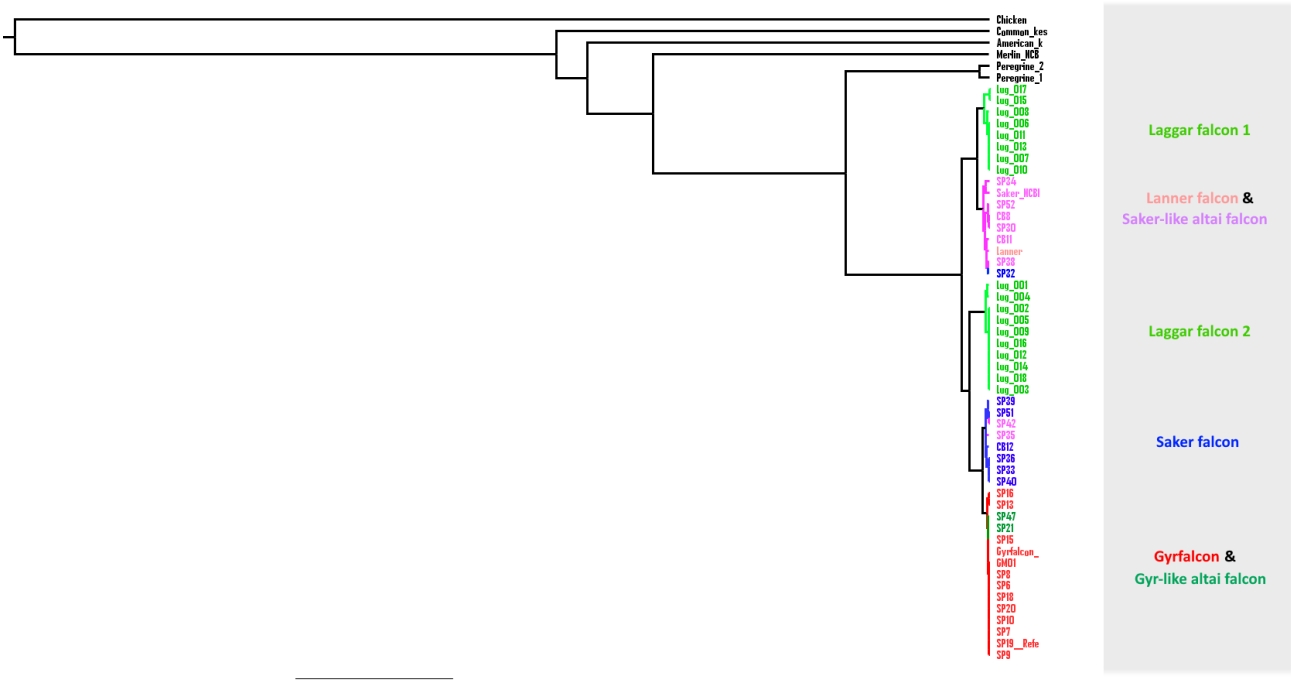

Figure 4: Neighbor-joining phylogenetic tree, based on the whole mitogenome (excluding the control region) of 55 falcons, with the chicken as an outgroup, generated using Phylip.

4.3 PCA, inbreeding and kinship within the captive laggar population

The short-read sequences were mapped to the laggar reference genome, with an average of 80,216,089 short-reads, a mean coverage of 9.723 and a mean mapping quality of 45.53 (Table 4).

A Principal Component Analysis (PCA) was performed on the pruned SNP dataset of all 18 falcons using plink. Figure 5 illustrates the results of the PCA, revealing some structure within the breeding population. Notably, three clades can be observed, with PC1 and PC2 accounting for 12.4% and 9.75% of the total variance respectively. Birds from within France and within Germany form separate clusters. For Belgium and the Netherlands, the observed clustering does not align with the countries of origin, suggesting that other factors are contributing to the population structure.

The inbreeding coefficients based on number of expected and observed homozygotic sites (Fis) was -0.033 on average (max 0.090, min -0.0751). A negative inbreeding coefficient could suggest occurrences of outbreeding. The Fis statistics are depicted in Table 4, along with their subsequent value of mean kinship. The mean total genome length covered by ROHs comprised 87.1Mbp, with a maximum of 243.0Mbp, and a minimum of 13.9Mbp. The mean number of ROHs per falcon was 155.2, of which 19.4% had a length of 1Mbp or higher. The mean Froh considering ROHs longer than 100kb was 0.073 (max 0.203, min 0.012). Fis and Froh values were strongly correlated (r2 = 0.668, p = <0.001), therefore suggesting that the fraction of the genome under ROH can be used as a reliable measure of inbreeding.

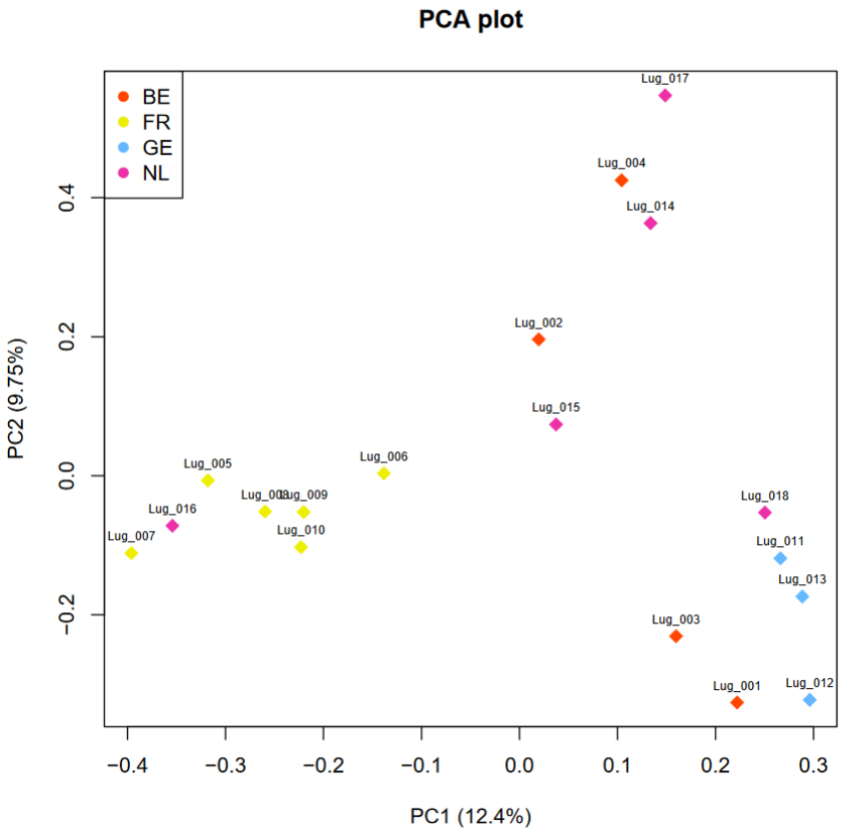

Figure 5: PCA plot of the laggar falcon breeding population has been categorized according to their country of origin; Belgium (BE, red), France (FR, yellow), Germany (GE, blue), Netherlands (NL, purple).

Table 4: values of inbreeding coefficients Fis & Froh, along with the mean kinship of each sample to the rest of the sample population.

	Mapping statistics			**Inbreeding & kinship**		
Sample name	# reads	Mean coverage	Mean mapping quality	F_{is}	F_{roh}	Mean Kinship
Lug_001	67,779,940	8.193	46.1	-0.056	0.075	0.033
Lug_002	74,271,034	8.6956	46.23	-0.028	0.048	0.020
Lug_003	101,375,602	12.320	46.39	**0.087**	**0.203**	-0.025
Lug_004	81,023,23	9.8073	46.16	-0.044	0.064	0.035
Lug_005	69,908,588	8.4907	46.24	-0.071	0.072	0.049
Lug_006	73,007,670	8.8336	46.06	-0.055	0.062	0.039
Lug_007	72,443,438	8.8134	43.62	-0.036	0.076	0.037
Lug_008	76,331,356	9.2475	43.62	-0.053	0.012	0.037
Lug_009	92,241,956	11.1815	43.6	-0.063	0.016	0.041
Lug_010	99,729,668	12.0495	46.12	-0.011	0.044	0.067
Lug_011	75,681,930	9.1662	46.27	-0.075	0.068	0.052
Lug_012	66,592,234	8.0476	46.31	-0.056	0.072	0.034
Lug_013	70,359,566	8.5013	46.2	0.009	0.074	0.017
Lug_014	68,715,928	8.3878	43.77	0.016	0.074	0.011
Lug_015	92,808,634	11.2924	46.4	-0.077	0.044	0.047
Lug_016	120,914,308	14.7354	43.73	0.029	0.081	0.061
Lug_017	79,817,856	9.7308	46.22	**0.091**	**0.179**	-0.023
Lug_018	61,693,802	7.5263	46.51	-0.090	0.041	0.055
Average	80,216,089	9.723	45.53	-0.033	0.073	0.033

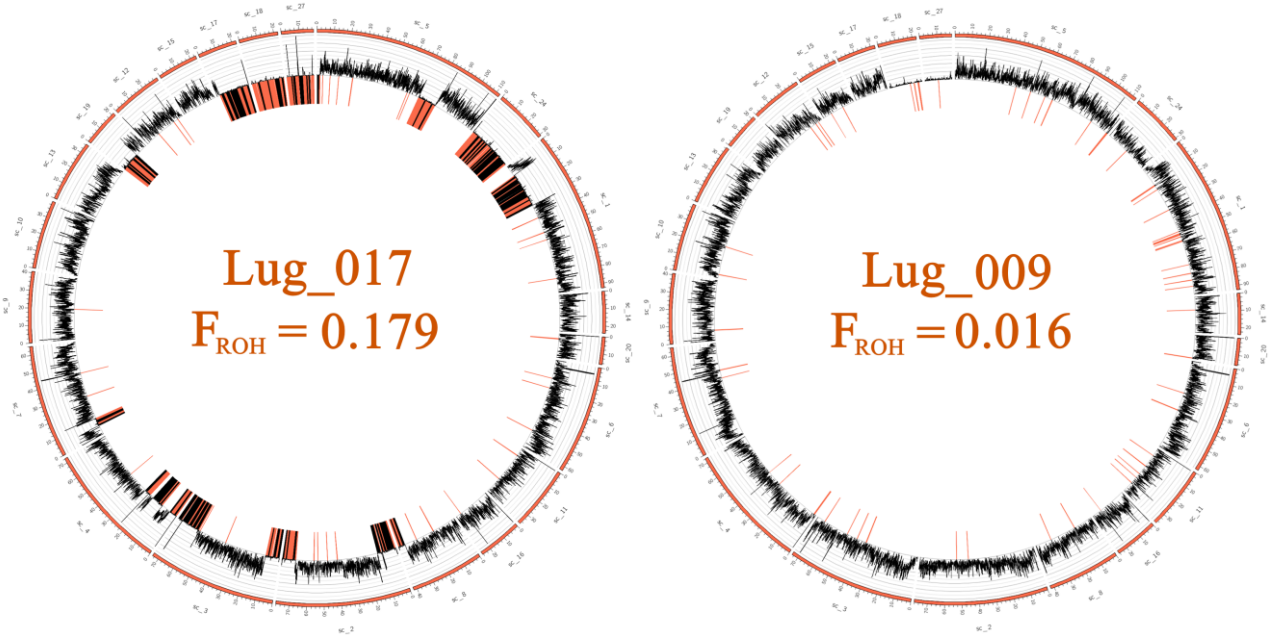

Figure 6: Circos plots showing (outer circle inward) the 21 largest scaffolds, windowed pi and the ROH analyses on Lug_017 (left) and Lug_009 (right). Windowed pi peaks with a higher value than 0.05 were cut off for better visibility. The inner layer shows ROHs larger- (black) and smaller than 2Mbp (red) in length.

After calculation of the inbreeding coefficients, two samples with a high (Lug_017) and low (Lug_009) Froh were chosen for visualization of the ROHs. BCFtools managed to detect the ROHs very well, as can be seen in Figure 6. Long regions of very low nucleotide diversity correspond with ROHs. The high Froh in some individuals indicate that they are inbred. However, in Lug_017, a large number of small ROHs seem to be clustering together. It is therefore likely that e.g. sequencing errors have resulted in a false-positive case of heterozygosity, and were therefore taken by BCFtools as separate ROHs. Thus, these large segments are most likely one ROH, as opposed to multiple smaller ones. This suggests that the parents of Lug_017 are closely related [Ceballos et al., 2018].

According to the relatedness analysis, many falcons share 1st,2nd and 3rd degree relationships, conflicting with the assumption of an unrelated sample population. The related pairs are depicted in the kinship Table below (Table 5).

Table 5: kinship Table of the 18 laggar falcons. 1st, 2nd and 3rd degree relationships are highlighted with red, orange and yellow respectively.

	Lug_001	Lug_002	Lug_003	Lug_004	Lug_005	Lug_006	Lug_007	Lug_008	Lug_009	Lug_010	Lug_011	Lug_012	Lug_013	Lug_014	Lug_015	Lug_016	Lug_017	Lug_018
Lug_001																		
Lug_002																		
Lug_003																		
Lug_004																		
Lug_005																		
Lug_006																		
Lug_007																		
Lug_008																		
Lug_009																		
Lug_010																		
Lug_011																		
Lug_012																		
Lug_013																		
Lug_014																		
Lug_015																		
Lug_016																		
Lug_017																		
Lug_018																		

4.4 Historic trend in effective population size

The estimate of the current effective population size with the molecular coancestry method in NeEstimator was very small (4.7; 95% confidence interval [CI] = 4.6-4.8). Similarly, the linkage disequilibrium model also estimated a small Ne (11.2; 95% CI = 6.9-19.0). To study the demographic history of the laggar falcon, I performed a SNeP analysis on the 23 largest scaffolds. My analysis showed a steady decline in Ne over time, with no significant fluctuations or bottlenecks over 100 generations. At the most recent point, 13 generations ago, the Ne was 100. Assuming a generation time of 6 years [Zhan et al., 2013], this time point corresponds to 78 years ago. These results can be found in figure 7, combined with the estimation of the current Ne. Compared to estimates further back in time (e.g., 54 generations ago: Ne = 403; 98 generations ago: Ne = 721), this is a great reduction in Ne. To further visualize subtle changes in the slope of the decrease, an Ne slope analysis was used to investigate the direction of Ne changes. Here, a negative NeS value corresponds to a decrease in slope, and therefore an increasing loss of diversity. As seen in figure 8, there are changes in the slope even though the Ne trend looks totally straight.

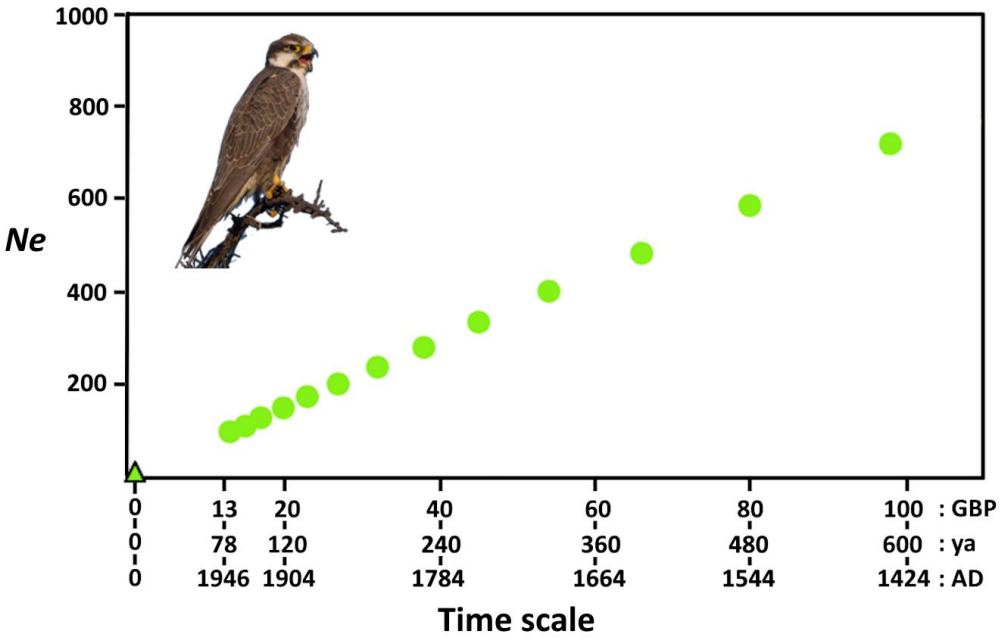

Figure 7: the historic effective population size of the breeding population with three different time scales; generations before present (GBP), years ago (ya) and the year AD. The dots represent the results of the SNeP analysis up to 98 generations before present. The triangle represents the result of the NeEstimator LD method.

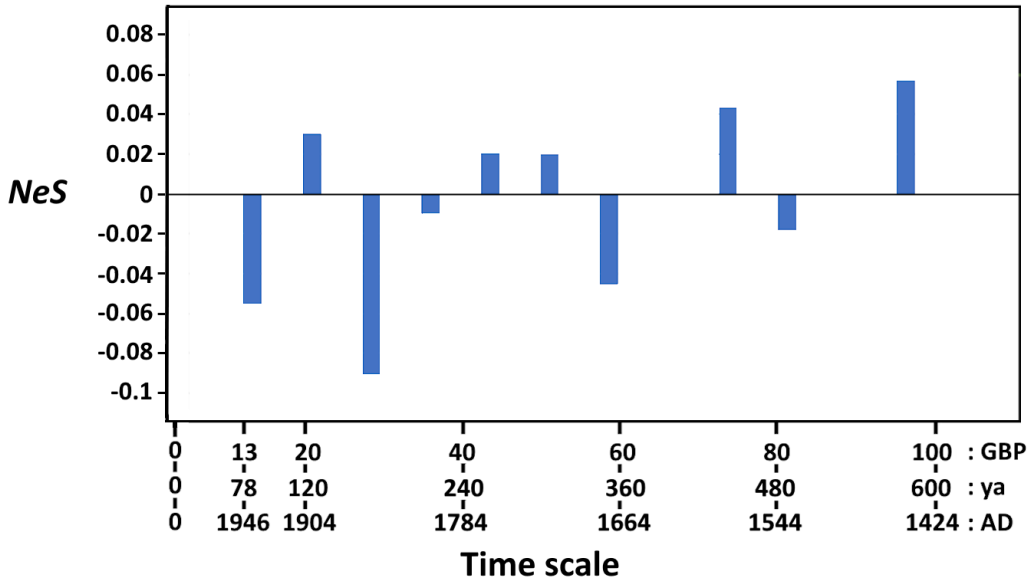

Figure 8: Ne slope (NeS) calculation between 13 and 100 generations before present. Bars larger than 0 correspond with a positive change in slope, while bars below 0 correspond with negative change.

5 Discussion

5.1 De novo genome assembly of the laggar falcon

The general aim of my thesis was to assess the reintroduction potential of the European laggar falcon breeding population. First off, this work completed the collection of available reference genomes for all the hierofalcon species. After the lanner falcon [NCBI, PRJNA842826], gyrfalcon [NCBI, PRJNA561988] and saker falcon [PRJNA842831], this research assembled the reference genome of the laggar falcon. With scaffold N50 and L50 values of 41.0Mbp respectively, the assembly represents comparatively weaker statistics than those of the gyrfalcon reference genome. But given the lower sequencing depth of 26x compared to 44x, and sole use of Nanopore long-reads, it can be considered a high quality reference. The alignment to the gyrfalcon reference genome highlights only a few structural variants (SVs), as is expected with the very recent split between the two species. The first major SV was the apparent merging of a large section of chromosome 2, to chromosome 1 on scaffold5. As of yet, it is uncertain whether this merge is a true merge or caused by a misassembly. This could provide an interesting topic for further research, since such chromosomal rearrangements have not yet been observed in studies comparing hierofalcons [Joseph et al., 2018; Justin et al., 2022].

The second and third SV were two small putative translocations, located on scaffold6 which corresponds to chromosome 3. While SVs like these are common, it is also possible that the alignment software (minimap2) has mapped the different segments in such a way that it visually aligns better, but is actually reversed. Since these SVs are both based on the same scaffold, it might be more likely that there is one large variant in the center, instead of two small variants [Li, 2018]. In this scenario, the large segment in-between the translocations would be reversed. Such inversions are common in birds, for example the great tit (*Parus major*) exhibits a large inversion of 64Mb on chromosome 1A [Da Silva et al., 2019]. Other examples are the ruff (*Calidris pugnax*) and the common quail (*Coturnix coturnix*) [Kupper et al., 2016; Sanchez-Donoso et al., 2022]. However, the inversion in the laggar falcon assembly could also be the result of technical errors in the assembly process, so further research is needed to shed more light onto this matter.

5.2 Phylogenetics of the hierofalcons

The first step to assess the breeding population viability was to find the phylogenetic relationship between the breeding population and the other hierofalcons. In line with recent findings, the saker, gyrfalcon and saker-like altai falcon (SLF) each form their own monophyletic group, except for a few exceptions. Namely, two of the SLFs (SP42 & SP35) cluster as saker, and one saker (SP32) clusters as SLF. The one lanner clusters with the saker-like altai falcon clade and the gyr-like altai falcons (GLF) cluster together within the gyrfalcon cluster. These results exhibit a striking similarity to those of Al-Ajli et al., 2023, which is to be expected since I used the same mitogenomes to construct the tree.

Unlike expected, the captive laggar falcons do not cluster together in one monophyletic group. They are divided into two different clades, one of which clusters with the lanner and SLF (Clade 1), and the other forming its own (Clade 2). The latter of which, assuming the diversion times of Al-Ajli et al., 2023, split from the sakers, gyrfalcons and GLF somewhere between 0.422 and 0.109 MYA. In other phylogenetic studies that involved the laggar falcon, they are most closely related to gyrfalcons and saker, and relatively much further from the lanner [Wink et al., 2004]. Therefore, I expect the falcons in clade 2 to be true laggar falcons, because they follow the patterns that have been observed in the past.

There could be several reasons for this observed split between clade 1 and 2. The first could be that one of the founders was not a true laggar falcon, but a lanner falcon or a SLF. Namely, the Fis values below zero, that were observed within the breeding population cannot solely be explained by active avoidance of inbreeding, they can also be caused by outbreeding or hybridization [Kearns et al., 2022; Hristova et al., 2018]. However, with their geographic distribution ranges not overlapping with the laggar falcon, this possibility is unlikely. If one of the founding falcons was not a true laggar, it would have most likely been a saker falcon, since they share their breeding ground with the laggar in Pakistan [Nittinger et al.,2007]. The constructed tree does not support this hypothesis. Though it should be noted that previous studies only included just one or two laggars, and were mostly constructed based on mitochondrial markers, as opposed to whole mitogenomes.

Apart from hybridization, the results were also impacted by the type of alignment software. Namely, in a second trial in MEGA11, using the MUSCLE alignment software instead of CLUSTALW, the results were very similar, except the laggar falcons were not separated (Figure S3). The MUSCLE software though, did not cluster either the saker falcons or the GLF together, but as part of the gyrfalcon clade. Also, they exhibited a closer relationship to the lanner and SLF clade, which is not in line with the results of Wink et al., 2004. For this reason, I chose to base my conclusions on the CLUSTALW tree, which divides the laggar. Also, the PCA which was based on a large proportion of the autosomal genome, did not cluster the individuals of clade 1 and 2 together. Perhaps, a tree which is based on larger segments of autosomal DNA, e.g., will give better insight.

No matter the reason for the split, further research to confirm the presence or absence of hybridization within the breeding program is important. Hybrid offspring should not be reintroduced into the wild, because it could cause outbreeding depression to a species that is already under pressure in the wild. Depending on their amount of hybridization, hybrid falcons should be removed from the breeding program.

5.3 Genetic health of the breeding population

The second step to assess the breeding population viability was to examine their genetic health, based on inbreeding, kinship and overall genetic diversity. According to the studbook, all samples were supposed to be unrelated. However, our study showed a high amount of kinship within the sampled 18 laggar falcons of the breeding population. Therefore, these results show the kind of consequences a breeding population can have, with poor management and a lacking studbook to keep track of family lineages. The relatedness scores went as high as the 1st degree, meaning a sibling- or parent-offspring relationship [Danecek et al., 2014], but most of the discovered relationships were of the 2nd or 3rd degree. A total of three 1st degree relationships were found (Table 5), of which one pair shares a very close hatching date, pointing to a hatchling relationship

(Lug_013 & Lug_011). The remaining two pairs did not share a close hatching date, therefore suggesting either a parent-offspring- or a sibling (but not hatchling) relationship. All 1st degree relationships occurred between falcons of the same country, while the 2nd and 3rd degree relationships could not be explained by the country of origin. It could be that the breeders missed the instructions to only provide unrelated falcons. This is likely the case for the 1st degree relationships, since breeders often know the close relationships of their falcons.

Even though kinship values were higher than expected, the mean inbreeding coefficient Fis was negative (-0.033), meaning that the observed homozygosity was lower than expected under Hardy-Weinberg equilibrium. Negative Fis values point to high genetic diversity and an active avoidance of inbreeding within a population [Wright, Sewall, 1965]. This result shows that even though there is a severely lacking studbook, inbreeding is still avoided effectively in most cases. This means that inbreeding has not found its way into the breeding population yet, despite some exceptions. However, even though Fis values at the moment are low, further inbreeding is imminent, because the pairwise kinship values are worryingly high for a selection that was supposed to be unrelated.

Another important point to consider is the potential for adaptation to captivity and the European climate, which may have caused a selective pressure [Williams et al., 2009]. Genetic adaptation to captivity has already been demonstrated in mice, fish, insects and amphibians [Heath et al., 2003; Kraaijeveld-Smit et al., 2006; Lewis et al., 2001; Lacy et al., 2013]. When reintroduced, captive-born falcons may experience poor adaptation to their new wild environment. In fact, captive reintroduction programs show a lower success rate of 38-50%, compared to the 71-75% success rate of translocations of wild-caught individuals that were never in captivity [Wolf et al., 1996; Griffith et al., 1989]. Williams et al., 2009 concluded that the best strategy to minimize genetic adaptation to captivity were to reduce the number of generations that a species will spend in captivity. However, there are other strategies that have a variable rate of success, depending on the species. Also, translocations are not recommended for highly endangered wild populations, as this may cause further negative effects for the population that remains. So, depending on the wild status of the laggar falcon, it might be interesting to look into the possibility of translocation, in contrast to reintroduction of captive-bred falcons. Given the complicated political situation of Pakistan and other middle- to south-east Asian countries, translocation might be difficult to arrange. However, translocation of other species in these areas is not uncommon. The houbara (*Chlamydotis undulata*) for example, are captively bred for reintroduction, but also translocated. Although, as opposed to the laggar falcon, houbarras are a prized prey species for falconry practice in the middle-east. Their conservation program is heavily funded and supported by several cooperating governments. So further education might be needed to increase awareness of the laggar falcon, and in turn the potential for approval of translocation.

5.4 Trend in effective population size (Ne)

The third step to assess the breeding population was to gain insight over their changes in population size over time, which I estimated using NeEstimator and SNeP, which used linkage disequilibrium (LD) to estimate effective population size over time. The population size of the laggar falcon back in the 1970s (~50 years ago) was estimated to be 10,000-25,000 pairs, which strongly contrasts to the most recent SNeP estimation of 100 individuals (13 generations before present, 78 years ago). However, it is not uncommon to find a substantial difference in Ne and population census size (Nc) [Palstra & Ruzzante, 2008].

Even though falconry in the middle east has existed for over 5000 years, the explosive export of falcons has only emerged in the last century [Bailey et al., 2000]. As one of the main threats to the status of the wild laggar falcon, this trade was expected to have a significant impact on Ne. However, the steady decline of the estimated Ne trend does not align with the emergence of illegal falcon trade. Therefore these results suggest that illegal falcon trade did not directly affect Ne. Nevertheless, the Ne of 13 generations ago approaches values which were found in several critically endangered bird species [Kearns et al., 2022]. Given this,

translocating laggar falcons to Pakistan may not be such a good idea, since this could result in detrimental effects to the population that remains.

The current Ne is estimated to be 11.2 by the LD method of NeEstimator. These values predict a tremendous loss of genetic diversity over time. Therefore, the breeding program should seriously consider a more refined breeding strategy to increase the Ne. For example, Ne can be increased by skewed sex ratios, inefficient use of fertile individuals, genetic overrepresentation and variance in family size [Frankham et al., 2004;]. All of the latter, are variables which can be managed by human intervention. This is especially true since the falcon breeder can pick a suitable match between individuals within the breeding program, as opposed to breeders of species that live in groups and/or with a sexual hierarchy, such as several sheep or fish species (group) or ungulate and social predators (hierarchy).

It is important to note that these estimates could have been influenced by the sample size of the population. Whereas most other studies used a sample of 50-60, I used only 18. It is likely that the results for Ne will keep increasing as additional laggars are added to the sample size. Until eventually, sample size will no longer influence the results. Additionally, inbreeding has been shown to strongly increase the rate of decline. Given the fact that inbreeding was present within the samples, this may have caused an underestimation of the Ne over time. So, additional sampling is needed to ensure that the estimations are indicative.

5 Conclusion

To conclude, this study has resulted in a new, high quality reference genome of the laggar falcon, a species that has been severely understudied for several decades compared to other falcon species. Hopefully, this new assembly will encourage further genetic research on the laggar falcon. For the first time, the laggar falcon has been included in a large scale phylogenetic analysis. My findings are strikingly similar to those of Al-Ajli et al., 2023, exhibiting clear monophyletic groups for saker falcons, gyrfalcons and saker-like altai falcons. It would be particularly interesting though, to compare the selected samples to wild laggar falcons, and include wild laggars into the phylogenetic tree, since the results suggest potential for hybridization within the breeding program.

Taken together, my findings provide valuable insights into the status of the European laggar falcon breeding program. As of now, inbreeding values are low and genetic diversity is high, yet the mean kinship of the population is concerning. The very low value of current effective population size foreshadows a substantial loss of genetic diversity in the near future. Fortunately, there are several ways to increase effective size. Therefore, I especially highlight the need for a studbook, given the unexpected high values of pairwise kinship. A reliable studbook will give a better overview of suitable mates, and which falcons are over- and underrepresented within the population. Which is important information, that the breeders will need to prevent any further inbreeding and minimize loss of genetic diversity. The patterns in historic effective population size align with other bird species, which are critically endangered. However, there is a need for more sampling to make sure that these results are not underestimated due to a small sampling size. Additionally, further sampling of wild laggar falcons may give a better understanding of its status in the wild.

Finally, I encourage further genetic and ecological research into the status of the laggar falcon in the wild. Because research of wild laggars will give the best insight on 1; if a reintroduction program is actually needed to support the wild population, and 2; if it is needed, would reintroduction truly be the better strategy compared to translocation? Because if the laggar falcon breeding program is not viable for the long term, translocation might be a suitable alternative.

6 Acknowledgements

With this thesis, I am following up on my internship. This thesis has taught me a lot about the world of falconry and its role in raptor conservation. First of all, I would like to thank Dr. Richard Crooijmans (Wageningen University, Animal Breeding and Genomics) and Dr. Mirte Bosse (Wageningen University, Animal Breeding and Genomics) for guiding me into the world of genetics and for their invaluable support. I would further like to thank Mrs. Tula Stapert (overseas coordinator, Project Lugger) and Dr. Farooq Al-Ajli (Internship provider, Project Lugger) for their supervision, dedication, and encouragement.

A big thank you to all the people working with Project Lugger, including founder and chair Bob Dalton, Jean-Michel Defoirdt (Belgium partner), Dr. Dominique Fisher (German partner), Leon Arends (Dutch partner), and Nicola Géli (French partner). They all generously donated money and supplied Laggar Falcons to be sequenced for the project. Anique Kappe from Gendika, the Dutch lab that collaborated and separated the DNA from the blood for the project, also deserves special mention. Without any of these people, this project would not have been possible.

The funding for the genome sequencing came from:

Tula Stapert (Dutch falconer)

L. Sterken (Dutch falconer)

M. Laarakker(Dutch falconer)

Ulli and Stefanie Jasser (German falconers)

Zayed AL- Maadeed (Qatari falconer)

Orde (Dutch falconers club)

References

Bailey, T., Launay, F., & Sullivan, T. (2000, July). Health issues of the international trade of falcons and bustards in the Middle East: the need for regional monitoring and regulation. In *Proceedings of the II International Conference on the Saker Falcon and Houbara Busard, Mongolia* (pp. 1-4).

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics*, *19*(4), 220-234.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Da Silva, V. H., Laine, V. N., Bosse, M., Spurgin, L. G., Derks, M. F., van Oers, K., ... & Groenen, M. A. (2019). The genomic complexity of a large inversion in great tits. *Genome Biology and Evolution*, *11*(7), 1870-1881.

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2004). *A primer of conservation genetics*. Cambridge University Press.

Fuchs, J., Johnson, J. A., & Mindell, D. P. (2015). Rapid diversification of falcons (Aves: Falconidae) due to expansion of open habitats in the Late Miocene. *Molecular phylogenetics and evolution*, *82*, 166-182.

Griffith, B., Scott, J. M., Carpenter, J. W., & Reed, C. (1989). Translocation as a species conservation tool: status and strategy. *Science*, *245*(4917), 477-480

Heath, D. D., Heath, J. W., Bryden, C. A., Johnson, R. M., & Fox, C. W. (2003). Rapid evolution of egg size in captive salmon. *science*, *299*(5613), 1738-1740.

Hristova, D. G., Tanchev, S. G., Velikov, K. P., Gonchev, P. G., & Georgieva, S. J. (2018). Single nucleotide polymorphism of the growth hormone (GH) encoding gene in inbred and outbred domestic rabbits. World Rabbit Science, 26(1), 49-55.

Joseph, S., O'Connor, R. E., Al Mutery, A. F., Watson, M., Larkin, D. M., & Griffin, D. K. (2018). Chromosome level genome assembly and comparative genomics between three falcon species reveals an unusual pattern of genome organisation. *Diversity*, *10*(4), 113.

Justin J S Wilcox, Barbara Arca-Ruibal, Jaime Samour, Victor Mateuta, Youssef Idaghdour, Stéphane Boissinot, Linked-Read Sequencing of Eight Falcons Reveals a Unique Genomic Architecture in Flux, *Genome Biology and Evolution*, Volume 14, Issue 6, June 2022, evac090, https://doi.org/10.1093/gbe/evac090

Kearns, A. M., Campana, M. G., Slikas, B., Berry, L., Saitoh, T., Cibois, A., & Fleischer, R. C. (2022). Conservation genomics and systematics of a near-extinct island radiation. Molecular Ecology, 31(7), 1995-2012.

Kraaijeveld-Smit, F. J., Griffiths, R. A., Moore, R. D., & Beebee, T. J. (2006). Captive breeding and the fitness of reintroduced species: a test of the responses to predators in a threatened amphibian. *Journal of Applied Ecology*, *43*(2), 360-365.

Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., ... & Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. Nature genetics, 48(1), 79-83

Lacy, R. et al. (2013). Evolution of peromyscus leucopus mice in response to a captive environment. PloS one, 8, e72452

Lewis, O. T., & Thomas, C. D. (2001). Adaptations to captivity in the butterfly Pieris brassicae (L.) and the implications for ex situ conservation. *Journal of Insect Conservation*, *5*, 55-63.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-3100.

Miller, C. S., Petrunenko, Y. K., Goodrich, J. M., Hebblewhite, M. A. R. K., Seryodkin, I. V., & Miquelle, D. G. (2011). Translocation a success, but poaching remains a problem for Amur tigers. *Cat News*, *55*, 22-25.

Mori, D., Vyas, R., & Kini, S. (2019). Monitoring a nest of Laggar Falcons Falco jugger.

Nittinger, F., Gamauf, A., Pinsker, W., Wink, M., & Haring, E. (2007). Phylogeography and population structure of the saker falcon (Falco cherrug) and the influence of hybridization: mitochondrial and microsatellite data. Molecular Ecology, 16(7), 1497-1517.

Petrov, R., Lazarova, I., Yarkov, D., Andonova, Y., & Dimitrova, S. (2023). First biochemical comparison between saker falcon subspecies falco cherrug cherrug and falco cherrug milvipes. Journal of Raptor Research, 57(3), 405–412. https://doi.org/10.3356/JRR-22-28

Sanchez-Donoso, I., Ravagni, S., Rodríguez-Teijeiro, J. D., Christmas, M. J., Huang, Y., Maldonado-Linares, A., ... & Vila, C. (2022). Massive genome inversion drives coexistence of divergent morphs in common quails. Current Biology, 32(2), 462-469.

Vinicius H da Silva, Veronika N Laine, Mirte Bosse, Lewis G Spurgin, Martijn F L Derks, Kees van Oers, Bert Dibbits, Jon Slate, Richard P M A Crooijmans, Marcel E Visser, Martien A M Groenen, The Genomic Complexity of a Large Inversion in Great Tits, *Genome Biology and Evolution*, Volume 11, Issue 7, July 2019, Pages 1870–1881, https://doi.org/10.1093/gbe/evz106

Wink, M. (2018). Phylogeny of Falconidae and phylogeography of Peregrine Falcons. *Ornis Hungarica*, *26*(2), 27-37.

Williams, S. E., & Hoffman, E. A. (2009). Minimizing genetic adaptation in captive breeding programs: a review. *Biological conservation*, *142*(11), 2388-2400.

Wright, S. (1965). The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. Evolution, 19(3), 395–420. https://doi.org/10.2307/2406450

Wolf, C. M., Griffith, B., Reed, C., & Temple, S. A. (1996). Avian and mammalian translocations: update and reanalysis of 1987 survey data. *Conservation biology*, *10*(4), 1142-1154.

Zhan, X. et al. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nat. Genet*. **45**, 563–566 (2013).

Supplementary information

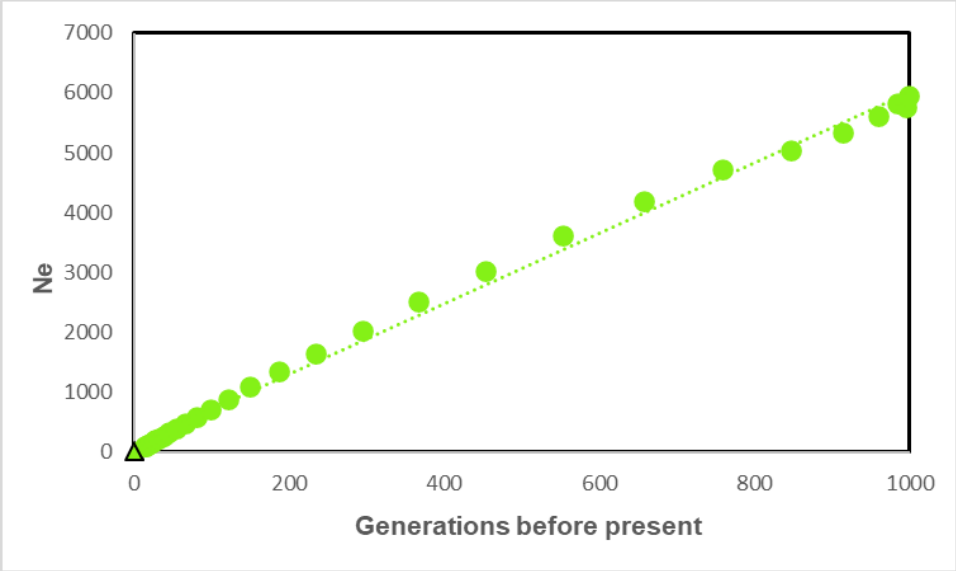

Figure S1: the estimated Ne over time, for the full range of SNeP estimates.

Table S1: overview of 1st, 2nd and 3rd degree relationships within the breeding population.

Sample	1	2	3

Lug_001		12,11	18,13,15,10
Lug_002			4,15,11
Lug_003			
Lug_004		17,18	15,11,14,2,5
Lug_005	7	10,8,6	16,9,15,4
Lug_006		10,5	9,15
Lug_007	5	10,8,16	9
Lug_008		7,16,5,9	10
Lug_009		16,10,8	7,5,15
Lug_010		7,6,9,5,16,15	8,11,1,18
Lug_011	13	1,12	18,10,15,4,2
Lug_012	18	1,11	13,15
Lug_013	11		1,12,18
Lug_014			17,4,15,4,18
Lug_015		10	18,6,14,1,11,2,5,9,12
Lug_016		9,7,8,10	5
Lug_017		**4**	**14**
Lug_018	12	4	11,1,15,13,10,14
Population			

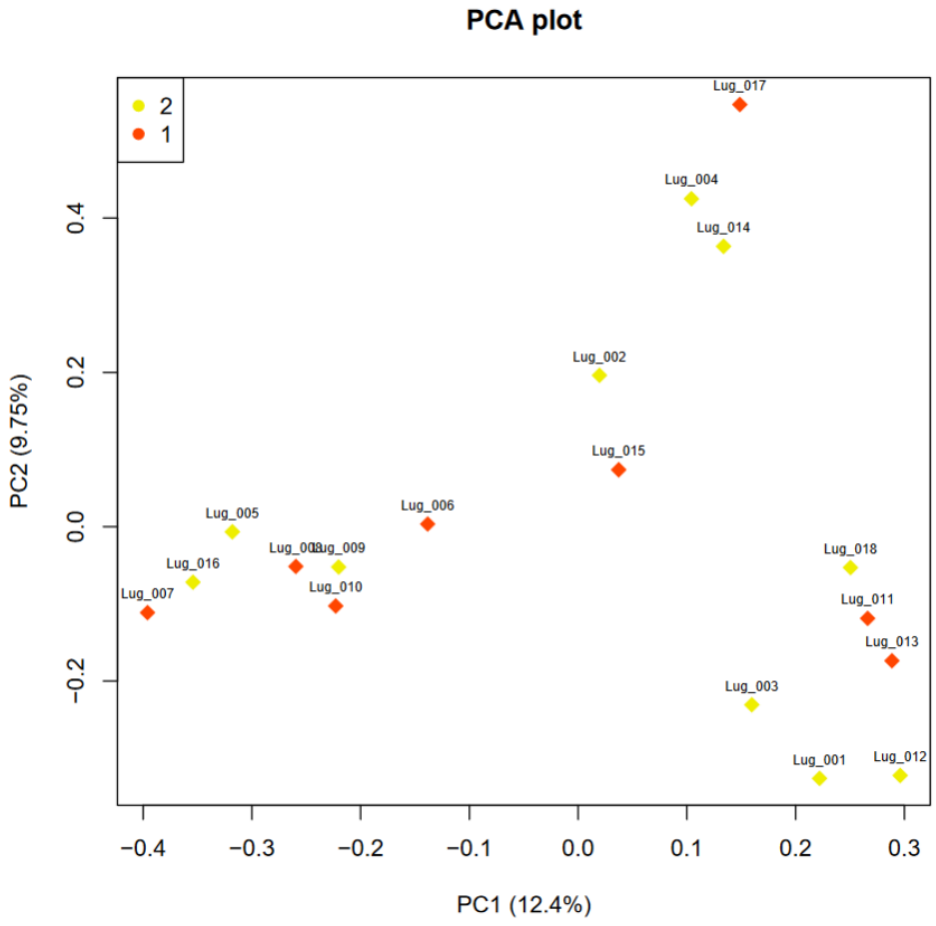

Figure S2: the PCA results, colored by the 2 different clades

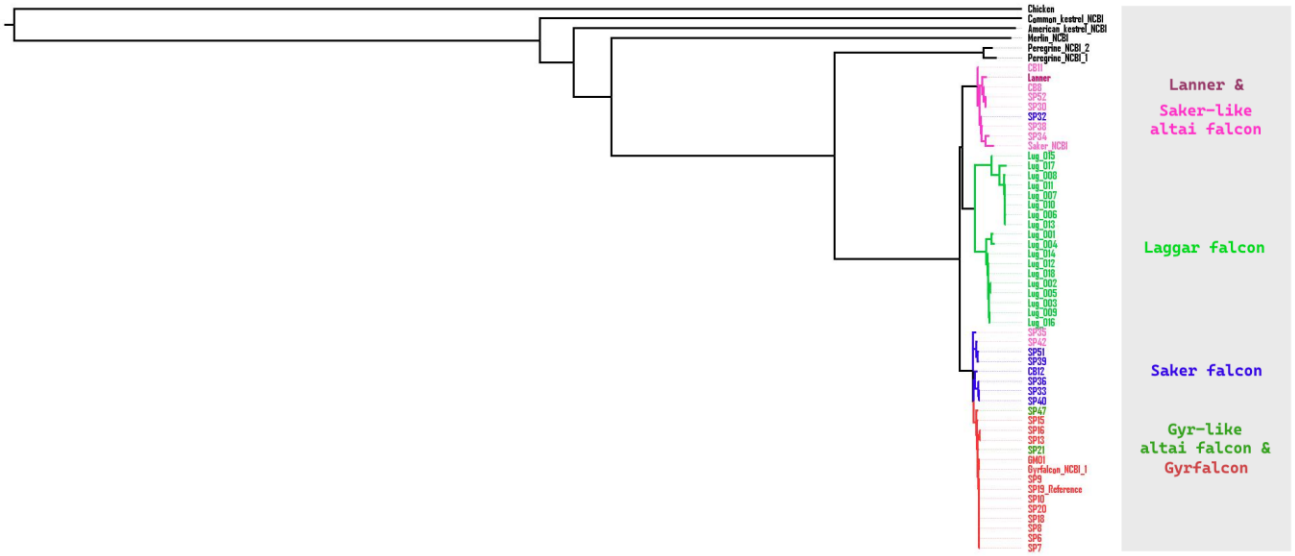

Figure S3: phylogenetic tree that was aligned using MUSCLE