universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Genetic structure of cheetah (*Acinonyx jubatus*) populations through genome-wide double-digest restriction-site associated DNA (ddRAD) sequencing"

verfasst von / submitted by

Julia Zumbroich, B.Sc.

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

Master of Science (M.Sc.)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 829
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Evolutionsbiologie
Betreut von / Supervisor:	Dr. med.vet. Pamela Burger

1. Acknowledgements

I would like to thank my parents; for making everything possible, Pamela Burger; for being the best supervisor I could have asked for, Jean Elbers; for all the explanations, Bogi, Franz, Gopi, Maria, Sara and Steve; for answering all my stupid questions and being the coolest lab group ever, Frank Zachos; for being a part of this, and my friends; for all the support.

2. Table of content

3. Zusammenfassung

Zu der Art *Acinonyx jubatus* (Gepard) gehörten bis vor einigen Jahren fünf Unterarten, wovon vier über Afrika verteilt waren, und nur eine, vom Aussterben bedrohte Unterart (*A. venaticus*) blieb in Asien erhalten und ist heute ausschließlich im Iran vorhanden. Wegen der bekannten, relativ kurzen Divergenzzeiten zwischen einigen dieser Unterarten wurde seitens der IUCN Cat Specialist Group diskutiert, die Unterarten auf zwei zu reduzieren, und unter den Namen *A. j. jubatus* für den südlichen, und *A. j. venaticus* für den nördlichen Geparden zu vereinen. Dieser Ansatz macht eine neue, größer angelegte Evaluierung der phylogeografischen Struktur notwendig. Die Probensammlung des Forschungsinstituts für Wildtierkunde und Ökologie der Veterinärmedizinischen Universität Wien vereint moderne Proben mit historischen Proben, die das ursprüngliche Habitat des Geparden widerspiegeln, und ermöglicht somit diese Evaluierung. Angelehnt an frühere Studien lautete unsere Hypothese, dass die beiden südlichen afrikanischen Unterarten, *A. j. jubatus* und *A. j. raineyii*, eine ähnliche genetische Abstammung aufweisen, wohingegen zwischen afrikanischen und asiatischen Populationen größere Divergenz zu erwarten ist. Um diese Hypothese zu testen haben wir 180 Individuen ausgewählt, die die historische Verbreitung des Geparden repräsentieren, und ihre genetische Vielfalt mittels ddRAD Sequenzierung analysiert. Trotz der vergleichsweise niedrigen Anzahl gemeinsamer SNPs unterstützen unsere Analysen deutlich die Diversität der bestehenden Unterarten, weshalb wir unbedingt empfehlen, die letzten verbleibenden asiatischen Geparden nicht mit den nordafrikanischen *A. j. soemmeringii* zusammen zu legen. Die Ergebnisse dieses Projekts haben unmittelbare Anwendbarkeit für den Erhalt der Geparden in Afrika und Asien.

4. Abstract

The cheetah (*Acinonyx jubatus*) species once consisted of five subspecies; four were distributed over Africa, but only one critically endangered population (*A. j. venaticus*) remained in Asia, now exclusively present in Iran. Due to short divergence times between some of the subspecies, there have been discussions by the IUCN Cat Specialist Group to collapse them into only two remaining subspecies, the northern *A.j.venaticus*, and the southern *A.j.jubatus.* The ongoing discussion made a re-evaluation of the phylogeographic structure on a larger scale necessary. A globally distributed sample set including historic specimens stored at the Research Institute of Wildlife Ecology, Vetmeduni Vienna, made this re-evaluation possible. Based on previous studies, we hypothesised that the two southern African subspecies, *A. j. jubatus* and *A. j. raineyii*, share similar genetic ancestry, while we expected long-term divergence between Asiatic and African populations. To test this hypothesis, the genetic diversity of 180 individuals covering the historic distribution range of the cheetah was re-evaluated using double-digest restriction-site associated DNA (ddRAD) sequencing. Despite a low amount of shared SNPs between individuals, our analyses clearly support the split between the existing subspecies; therefore we strongly recommend not collapsing the northern African subspecies with the Asiatic population. The results of this project have immediate applicability on the current conservation and management strategies of cheetah populations throughout Africa and Asia.

5. Introduction

5.1 The Cheetah

The cheetah, *Acinonyx jubatus* (Schreber, 1775) is listed as "vulnerable" by the International Union for Conservation of Nature (IUCN) (17). It used to occupy a habitat ranging from India in the south and east, to the north until Southern Kazakhstan and Russia, westwards through the Middle East and the Arabian Peninsula (35), and stretching over most northern and sub-Saharan Africa (37, 17, 21). Today, the cheetah is estimated to sustain on only 9% of its former habitat (see figure 1), with no more than 7100 individuals (16) and a declining population trend (17).

The Asiatic cheetah, *Acinonyx jubatus venaticus* (Griffith, 1821), is rated as "critically endangered" by the International Union for Conservation of Nature (IUCN) (17). Current studies estimate the last existing population, located in Iran, to consist of only 40 to 70 individuals (cf. 2, 36, 49) in 17 areas (20). Recently, these areas have been found to divide in two landscapes, which have no connection between each other (23, 36). In Iran, the cheetah has been protected since 1959, and the Asiatic cheetah has not been encountered outside of Iran since 1982 (21). Since 2001, the "Conservation of the Asiatic Cheetah Project (CACP)", a joint project of the Iranian government and the United Nations Development Program (UNDP) is conducted in so far two phases, aiming towards a holistic conservation of the cheetah, its habitat, and its prey (8, 49). But also the African cheetah subspecies*, A. j. hecki* (Hilzheimer, 1913), *A. j. soemmeringii* (Fitzinger 1855), *A. j. raineyii* (Heller, 1913) and *A. j. jubatus* (Schreber, 1775), have suffered serious decline within the last 50 years (5, 12, 16).

Hence, to save viable populations, the IUCNs Cat Specialist Group proposed to plan on merging the cheetah's subspecies into three, or even just two remaining subspecies. In Southern Africa, for *A. j. raineyii* and *A. j. jubatus*, this has already been realised; these former two subspecies are now combined as *A. j. jubatus*. The three remaining subspecies might be merged into just one, overall "Northern" group, incorporated under the name *A. j. venaticus*.

Here, we emphasise the importance of keeping *A. j. venaticus*, and also *A. j. hecki*, as independent subspecies from the still quite viable population of *A. j. soemmeringii*, as former studies showed the Asiatic cheetah to be more closely related to the Southern African *A. j. jubatus*, and the Western *A. j. hecki*, then to the East African *A. j. soemmeringii* (6, 11, 34, 41).

Figure 1 provides an overview of the cheetah's current range, and additional supposed occurrences. Within the past decades, the known present range of the cheetah increased, but this is not due to possible range recovery or expansion, but to an increased research effort and more efficient survey methods (21).

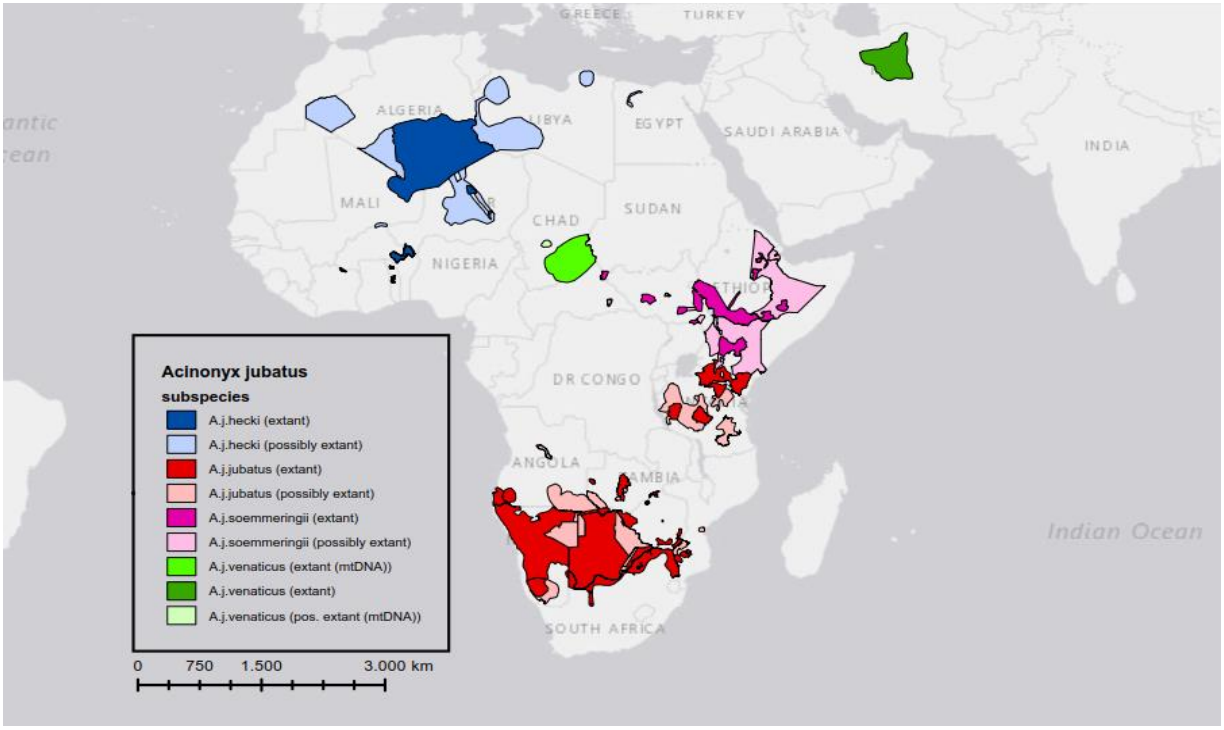


Figure 1: Distribution of the four remaining cheetah subspecies, plus further supposed habitat patches. The populations in Chad were ranked among *A. j. venaticus*, based on the clustering from earlier studies on mtDNA (11). Distribution data provided by Durant et al. (17).

5.2 Reduced Representation Genome Sequencing

After previous studies have mostly been conducted analysing mitochondrial DNA (mtDNA) and /or a limited number of nuclear microsatellite loci (11), we utilised the collection for a broader scale sequencing of nuclear DNA at a genome-wide level. This was done to further support and extent the foregone studies, as nuclear genome-wide SNPs are known to have higher resolving power for in-depth population structure and phylogeographic analysis than mtDNA (47).

The ddRAD sequencing is an especially suitable method for reduced whole-genome representation, as it allows large numbers of individuals to be shotgun sequenced in parallel. This approach results in a large number (thousands) of SNPs evenly spread over the whole genome (18), without the high requirements of full genome sequences, as e.g. cost or server capacity (4, 14).

The ddRAD method (40) is an advancement of the RAD method (18), where the sample DNA is randomly sheared with a restriction enzyme and then the fragments are sequenced. In ddRAD, two different enzymes are used; one cutting at a rare cut site, the other at a common cut site (see figure 2). Like this, more fragments are cut, which enlarges the number of possible shared SNPs, and minimises the server capacity necessary for the analysis, as the single fragments are shorter.

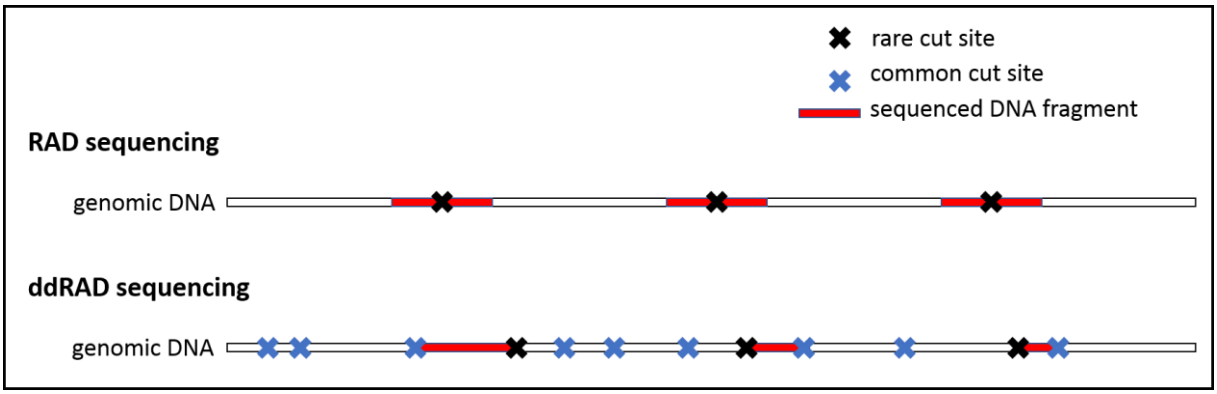


Figure 2: Mechanisms of RAD vs. ddRAD sequencing. By using two enzymes, ddRAD can significantly increase the SNP output while minimising the necessary fragment length. Graph modified after Peterson et al. (40).

6. Aims and Hypothesis

The population structure and subspecies status of the different Asiatic and African cheetah populations are still a matter of debate. Building on the results of previous studies, we hypothesise that the two Southern African subspecies, *Acinonyx jubatus jubatus* and *Acinonyx jubatus raineyii*, do not show large population differentiation. However, we expect long-term divergence between the other three cheetah subspecies, specifically between the African subspecies *A. j. soemmeringii* and *A. j. hecki*, and the Asiatic *A. j. venaticus*.

The aim of this thesis is to clarify the population structure of *Acinonyx jubatus* and its recognised subspecies, not only for recent populations and their present distribution, but also throughout its historic range. To establish the genetic relationship between subspecies, we sequenced single nucleotide polymorphisms (SNPs) over the entire genome in modern and up to 200 years old cheetah samples originating from 26 African and ten Asian countries.

7. Material & Methods

7.1 Sample Collection

For this study, we extracted 196 samples, of which 125 originated from the Rex' Foundations collection, from Lena Godsall Bottriell and Paul Bottriell, UK. This collection of cheetah samples represents the historic distribution from 31 countries, and includes samples from within the last 200 years. Additionally, 38 historic samples from Charruau et al. (11), and 22 fresh tissue samples, as well as nine blood samples from a collection from the University of Lisbon (C. Fernandes), were extracted. Please see Appendix 12.2 for the CITES (Convention on International Trade in Endangered Species) of all utilised samples. The specimen consisted mainly of small bones, e.g. nasal bones, but also a variety of other tissues, such as frozen muscle tissue, hair, skin, nipples, toepads, and diluted blood samples. As outgroups, we extracted one *Lynx lynx* and one *Puma concolor* sample from the Pathology Collection at Vetmeduni Vienna. With our choice of samples, we covered as comprehensively as possible the global distribution the cheetah used to have before anthropogenic influences started forcing it back (2, 5, 6). After quality control (see 7.3), 180 of the 196 extracted samples were sent for sequencing (for full list of samples, see Appendix, table A1).

7.2 DNA Extraction

7.2.1 Historic DNA

The 165 historic DNA samples were extracted in the cleanroom-laboratory at the Research Institute of Wildlife Ecology at Vetmeduni Vienna, using a modified version of the Rohland et al. (2010) protocol (45). The samples were centrifuged rather than placed on a vacuum manifold, and to enhance the purity of the DNA output, the residual ethanol was evaporated for 2 min. before eluting the DNA pellet in TE buffer in two steps, eluting 25 µl each.

7.2.2 Modern DNA

The 22 modern tissue samples, which we received from C. Fernandes, University of Lisbon, were extracted following the protocol for tissue from the Quiagen DNeasy blood and tissue kit. The nine diluted blood samples from the same group were extracted with the Analytik Jena Blood Kit. As for the outgroups, *P. concolor* was received as DNA extraction, while *L. lynx* was extracted alongside the historic samples, according to Rohland et al. (45).

7.3 Pre-sequencing Quality Control

7.3.1 NanoDrop

As quality control of the extracted DNA, we measured all samples with the peqlab NanoDrop 2000c microvolume spectrophotometer, to make sure the DNA level for each individual was sufficient for the sequencing requirements of a minimum of 5 µg DNA.

7.3.2 Qubit

As additional quality control on the first 111 of samples, we used the ThermoFisher Invitrogen Qubit 2.0 fluorometer, to reassure that the amount of DNA in each sample was sufficient.

7.3.3 Gel-electrophoresis

For a few selected batches, we used gel electrophoresis to visualize the integrity of the extracted DNA before sending it for sequencing. For this, we made 0,8% agarose gels of either 50 ml or 100 ml, depending on the number of samples, and ran them on 120 V for 30 minutes.

7.4 Double-digest Restriction-Site Associated DNA (ddRAD) Sequencing

For ddRAD sequencing, we sent 180 samples (see Appendix, table A1), including the two outgroups *Puma concolor* and *Lynx lynx*, to IGA Technologies, Udine, Italy. The ddRAD method is an especially suitable method for whole-genome representation, as it allows large numbers of individuals to be shotgun sequenced, and gives an evenly spread representation of the whole genome (18), without the high requirements of full genome sequences, as e.g. cost or server capacity (4, 14). For this method, the genomic DNA is digested with not only one, but two restriction enzymes, which in this case were SphI and HindIII. Due to this double digestion, sequencing cost and required time can be significantly reduced, and SNP output increased (4, 40). The sequencing was performed following the Peterson et al. (40) protocol.

The targeted amount of reads per individual sample is 2,5 million reads; however, due to low quality of the historic samples, only six of our sampled individuals reached this value. The company's usual cutoff to call SNPs is 1 million reads, however, for our samples the company's bioinformatician agreed to use 0,5 million reads per sample. Even with this lower threshold we only reached 0,5 million reads for 17 samples, including the *Puma concolor* (see figure 3; 8.4). Most likely due to the high fragmentation of the DNA in the historic samples, the cut sites of the restriction enzymes, SphI and HindIII, were not as frequently present in the low quality museum samples as in the modern samples. This resulted in a lower number of digested DNA fragments and consequently in less sequencing reads.

Similarly, some of the modern DNA samples, i.e. the nine diluted blood samples, just didn't have the adequate quality or quantity.

7.5 Preliminary Sequence Filtering & Quality Control

7.5.1 Raw Sequence Data Filtering

The pipeline for raw data quality control and filtering was developed by Dr. J.P. Elbers, postdoc in the Conservation and Population Genetics group at the Research Institute of Wildlife Ecology, Vetmeduni Vienna. The SNPs were called under his guidance and the complete analysis protocol is provided in Appendix 12.3.

In brief, we received the demultiplexed fastq reads by IGA technologies. The first quality control and adapter trimming were performed using the stacks (10) function process_radtags. To map the trimmed reads to the cheetah reference genome (15, Genbank accession number GCF_001443585.1), we used the Burrows-Wheeler-Aligner (BWA_MEM) (32), with default settings. Then, pstacks was used to analyse the alignment and determine the RAD-loci based on the reference genome. Moreover, in this step, the minimal number of reads per SNP is determined, which in our analysis was set to three. Subsequently, cstacks was used to create a catalogue of stacks for each individual, and sstacks to compare these individual stacks and find overlapping loci. Afterwards, the SNPs were called using the stacks function "populations", which creates a VCF-file, and VCF-tools (13) was used to filter for only biallelic SNPs. The eventuating plink-file was then used for the following steps of the quality filtering.

7.5.2 Quality Filtering of the Individual Genotypes

Due to the intrinsic problem to call a sufficient number of SNPs throughout the complete data set, we decided on a multi-step approach. This gave us the opportunity to control each filtering step individually, and to adjust settings to prevent the elimination of too many individuals. Using Plink 1.9 (42, 43), we filtered the data set consisting of 88 individuals and a total of 50316 SNPs, to retain a maximum number of SNPs covered in as many individuals as possible. First, we filtered for "individual missingness" (--mind 0.999) of 99,9%, which resulted in 29 individuals showing at least 0,1% of the 50316 SNPs. This step removes individuals from the sample set which do not show the sufficient genotype rates, meaning that of the represented SNPs they so not cover the desired share. Here, it was performed to keep as many individuals as possible during the beginning of the filtering. Next, we filtered for genotype missingness (--geno 0.25) of 25% to keep only those SNPs which were covered in 75% of the 29 individuals. During this step, the number of SNPs was reduced to 1377. Filtering for a minor allele frequency of 1% (--maf 0.01) stringently removed possible sequencing errors, which did not exclude any further individuals. Finally, we filtered for individual missingness a second time, this time excluding seven individuals with more than

50% of missing genotypes (mind 0.5). One additional individual was removed due to close relatedness based on pedigree analysis.

These stringent filtering steps resulted in a final data set (A) consisting of 21 cheetahs and 207 SNPs, which was used for the consecutive analyses of population structure and genetic differentiation (see 7.6) as well as population genetic summary statistics (see 7.8). For the purpose of phylogenetic analyses (see 7.7) we created a second dataset (B) by adding two outgroup species, *Puma concolor* and *Lynx lynx*, with a total of 881 SNPs.

By relaxing the filtering parameters (--mind 0.999, --geno 0.25, --maf 0.01) we generated a third, explorative data set (C) containing 27 individuals and 207 SNPs, but with a higher (> than 50%) individual missingness of genotypes. This dataset included six additional individuals, partly originating from the *A. j. soemmeringii* subspecies range (for the full dataset, see Appendix, table A2). The dataset was only used to explore the separate clustering of this subspecies with a higher number of (less covered) samples (see 7.6). To further study population substructure within *A. j. soemmeringii,* which represented the largest number of successfully sequenced cheetah samples, we created a final sub-dataset (D) with 15 individuals and 201 SNPs. The four different datasets are summarised in Table 1.

Set	No. of SNPs	No. of individuals	Content of sets
A	207	21	Cheetahs, stringent filtering
B	881	23	Cheetahs plus two outgroups
C	207	27	Cheetahs, relaxed filtering
D	201	15	*A. j. soemmeringii* only

Table 1: Sets of SNPs and individuals used for population structure and phylogenetic analyses.

7.5.3 Relatedness between individuals

We used Plink 1.9 (42, 43) to test for close relatedness between individuals. A symmetric identity-by-state (IBS) matrix was created (--cluster --matrix) for all pairs of individuals and samples with a proportion of share alleles higher than 0.95 should be excluded from consecutive population genetic analyses.

7.6 Population Structure & Genetic Differentiation

For necessary conversions to create specific input files for the population structure and phylogenetic analyses, the PGDSpider 2.1.0.3 data conversion tool (33) was used. R packages were implemented in R-Studio v. 1.0.143 (46).

7.6.1 Population Structure and Admixture Analysis

To gain a first overview about potential genetic structure between the cheetah populations, we first utilised the R package Adegenet (29, 30) and performed a Principal Component Analysis (PCA). The complementary R package Pegas (39) was installed on R-Studio (46) as well. An additional visualisation of clustering between individuals was done with a multi-dimensional scaling (MDS) plot using the distance matrix (1-IBS) from Plink.

We further clarified population structure by Bayesian-model based clustering implemented in FastStructure (44), which tests for the most likely number of ancestral populations (K) within individuals. In the data sets A (21 cheetahs, 207 SNPs) and D (*A. j. soemmeringii* only), we applied 10 iterations for multiple choices of K from two to five. To find the best number of model components, which explain structure in the dataset, we used the python script chooseK.py implemented in the software with default parameters.

Subsequently, we ran Admixture (3) with 10 iterations for each K (1 – 5) to detect admixture within single individuals of different populations and cheetah subspecies. The most likely number of clusters was indicated by the lowest cross validation (CV) error given for each K. The resulting Q-estimates, which provide the individuals' levels of belonging to different ancestral populations and of admixture, were imported into R-Studio and visualised as bar plots.

7.6.2 Genetic Differentiation

To estimate the pairwise genetic differences between the three detected populations in the 21 cheetahs (dataset A), we calculated population pairwise F_{ST} values using Arlequin (20). This fixation index, a theoretical value between 0 and 1, indicates the differentiation between populations by calculating how much of the population's differences can be explained by genetic population structure. Thus, a value of 0 indicates complete panmixia, no differentiation between populations, and a value of 1 indicates completely separated populations.

Because of the small sample size for *A. j. venaticus* and *A. j. jubatus*, and the low explanatory power for F_{ST} values for less than five individuals, we additionally created a matrix of pairwise genetic differences in Plink 1.9 (--distance-matrix).

7.7 Phylogenetic Analyses

The evolutionary relationship between individuals and populations was represented in the framework of Adegenet and Splitstree (22), using Neighbor-Joining tree and network (NeighborNet) analyses. NeighborNets provide more general graphs than single phylogenetic trees, which might not be appropriate representations under more complex models of evolution, e.g. involving gene flow, gene loss, or recombination. Neighbor-Joining trees and NeighborNets were created with (23 individuals, 881 SNPs) and without outgroups (21 cheetahs, 207 SNPs).

7.8 Population Genetic Summary Statistics

The calculation of observed (H_O) and expected (H_E) heterozygosities was performed with Adegenet, while we used Plink (--het) to estimate the inbreeding coefficient (F_{IS}) based on the observed versus expected number of homozygous genotypes. These parameters were calculated over all 21 cheetahs and separately for the *A. j. soemmeringii* population (n = 15), while the populations of *A. j. jubatus* (n = 2) and *A. j. venaticus* (n = 3) unfortunately contained too few individuals for reliable estimates. Significant differences between H_O and H_E, as well as significance of the F_{IS}-values were calculated with a one sample t-test in R-Studio. The calculations of heterozygosity provide a depiction of the population's genetic diversity. With a H_O to H_E ratio <1, genetic variance is lower than expected, which might hint to inbreeding in the population. With a ratio of H_O to H_E >1, the population shows a higher than statistically expected variance, which indicates an outbred, "healthy" population with no signs of inbreeding. The F_{IS}-value indicates whether a population contains a higher number of homozygous individuals than expected under Hardy-Weinberg equilibrium, therefore, a positive value indicates higher than expected homozygosity, which might be due to inbreeding.

All commands and bioinformatics analytical steps are provided in Appendix 12.3.

8. Results

8.1 Sample Collection

Due to the diverse conditional states of the samples, influenced by age, treatments for conservation, and storage in the museums, the DNA quality was, in some cases severely, affected. Despite special aDNA extraction protocols, less than 10% of the samples produced sufficient data applicable for population genetic analyses. The successfully sequenced samples unfortunately were not distributed equally over the sampled historic range of the cheetah, which limited the explanatory power of this study.

8.2 DNA Extraction

8.2.1 Historic DNA

Of the 165 historic cheetah samples we send for sequencing, we received SNPs for 80, and, after filtering, could include eleven (13 including outgroups) into the final analyses.

8.2.2 Modern DNA

Of the 31 modern DNA samples, we retrieved sufficient DNA quality and quantity for only 16, which we sent for sequencing, and successfully called SNPs for all of them. The low output from the DNA extraction was mostly due to the bad quality of the highly diluted blood samples, while the tissue samples yielded good results (see Appendix, table A1).

8.3 Pre-Sequencing Quality Control

8.3.1 NanoDrop

We used NanoDrop to control for sufficient DNA quality and quantity for ddRAD sequencing, and most of the samples were in an acceptable range of values (For complete list of NanoDrop results, see Appendix, table A1). Those with values ranging under the mark of 1,40 for the 260/280 ratio or below 20 ng/µl of DNA concentration were either re-extracted or substituted with another individual from the same country, if no more specimen was available.

8.3.2 Qubit

For a number of 112 samples the NanoDrop could be confirmed with the qubit measurement. The complete list of Qubit results is given in the Appendix (table A1).

8.3.3 Gel-electrophoresis

To additionally check for the DNA integrity in the extracted samples, we ran three gels from selected batches of the extractions. As to be expected from historical specimens, we mostly detected highly fragmented DNA smears showing a general presence of DNA.

8.4 Double-digest Restriction-Site Associated DNA (ddRAD) Sequencing

We received the demultiplexed fastq reads of the sequenced ddRAD loci from the Italian company IGA technologies. Unfortunately, only a fraction of the 180 individuals reached one million reads, the usual cut-off for calling SNPs, and only about 30 samples reached a sufficient number of reads for further analyses. The distribution of the number of reads per sample is presented in figure 3.

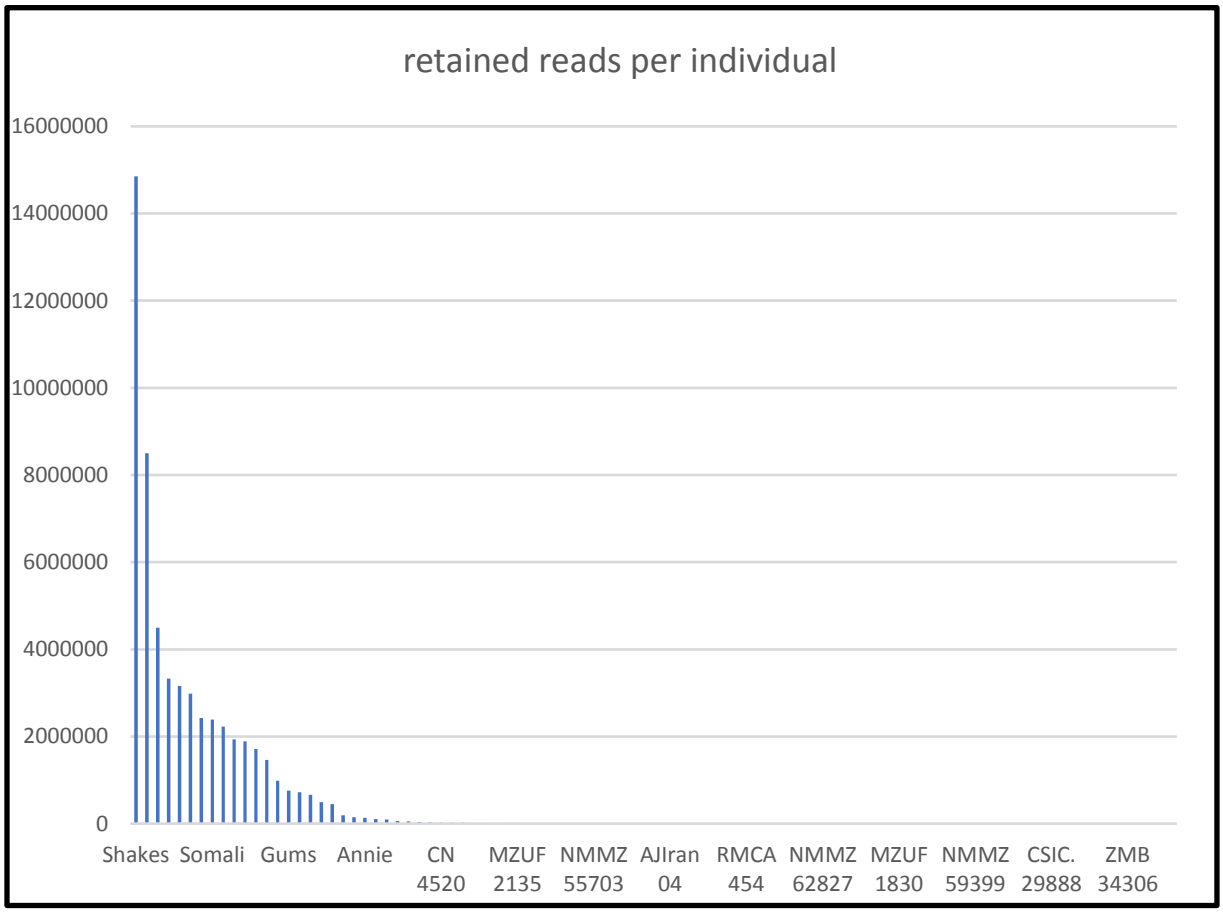


Figure 3: Retained reads per individual after sequencing. One bar representing one individual, with exemplary individuals named on the x-axis, and number of reads on the y-axis. Graph provided by IGA technologies.

8.5 Preliminary Sequence Filtering & Quality Control

8.5.1 Raw Sequence Data Filtering

After raw data filtering, we received a vcf-file of all 96 individuals, with which we continued the filtering process.

8.5.2 Quality Filtering of the Individual Genotypes

With the received 96 individuals, we created the four different sets (A-D) represented in table 1. Unfortunately, the quality filtering eliminated all *A. j. hecki* individuals, as well as *A. j. raineyii*, and reduced *A. j. venaticus* and *A. j. jubatus* to under five individuals each, depending on the set.

8.5.3 Relatedness between individuals

In general, the relatedness values varied between 64% and 92%. Strikingly, one Iranian cheetah (IR1) showed an exceptional position in the dataset, being closely related (89%) to a Somalian cheetah (SO10); in general IR1 showed the lowest relatedness values between all pairwise comparisons. The three pairs showing a high relatedness above 90%; IR3&IR4, SO3&SO4, and SO7&SO12, each originated from one mutual subspecies (see figure 4).

	IR1	IR2	IR3	IR4	NM1	SD1	SD2	SO1	SO2	SO3	SO4	SO5	SO6	SO7	SO8	SO9	SO10	SO11	SO12	SO13	ZA1
IR1	1	0,705	0,665	0,638	0,691	0,71	0,741	0,815	0,75	0,786	0,755	0,732	0,746	0,705	0,737	0,696	0,888	0,777	0,714	0,772	0,67
IR2	0,705	1	0,893	0,887	0,788	0,774	0,799	0,783	0,791	0,794	0,799	0,766	0,766	0,745	0,772	0,742	0,755	0,764	0,753	0,791	0,761
IR3	0,665	0,893	1	0,901	0,763	0,749	0,766	0,75	0,751	0,751	0,756	0,754	0,758	0,734	0,749	0,715	0,727	0,739	0,742	0,778	0,742
IR4	0,638	0,887	0,901	1	0,766	0,751	0,768	0,762	0,758	0,758	0,771	0,746	0,756	0,742	0,756	0,732	0,734	0,746	0,749	0,761	0,739
NM1	0,692	0,788	0,763	0,766	1	0,761	0,79	0,769	0,775	0,766	0,786	0,768	0,749	0,754	0,763	0,759	0,761	0,744	0,756	0,768	0,843
SD1	0,71	0,774	0,749	0,751	0,761	1	0,893	0,815	0,818	0,821	0,831	0,851	0,823	0,806	0,808	0,794	0,794	0,799	0,828	0,813	0,751
SD2	0,741	0,799	0,766	0,768	0,79	0,893	1	0,828	0,8357	0,831	0,845	0,857	0,849	0,819	0,824	0,79	0,821	0,829	0,845	0,843	0,763
SO1	0,815	0,783	0,750	0,762	0,769	0,815	0,828	1	0,850	0,847	0,847	0,833	0,833	0,801	0,823	0,814	0,876	0,837	0,823	0,830	0,748
SO2	0,750	0,791	0,751	0,758	0,775	0,818	0,836	0,850	1	0,850	0,877	0,857	0,814	0,800	0,843	0,834	0,831	0,843	0,807	0,843	0,773
SO3	0,786	0,794	0,751	0,758	0,766	0,821	0,831	0,847	0,850	1	0,919	0,833	0,814	0,800	0,853	0,810	0,850	0,838	0,821	0,833	0,749
SO4	0,755	0,799	0,756	0,771	0,786	0,831	0,845	0,847	0,877	0,919	1	0,845	0,810	0,805	0,835	0,811	0,825	0,855	0,823	0,855	0,768
SO5	0,732	0,766	0,754	0,746	0,768	0,851	0,857	0,833	0,857	0,833	0,845	1	0,816	0,792	0,826	0,817	0,814	0,826	0,800	0,826	0,756
SO6	0,746	0,766	0,758	0,756	0,749	0,823	0,848	0,833	0,814	0,814	0,810	0,816	1	0,792	0,797	0,773	0,824	0,831	0,824	0,865	0,737
SO7	0,705	0,745	0,734	0,742	0,754	0,806	0,819	0,801	0,800	0,800	0,805	0,792	0,792	1	0,812	0,783	0,785	0,783	0,911	0,797	0,737
SO8	0,737	0,772	0,749	0,756	0,763	0,808	0,824	0,823	0,843	0,853	0,835	0,826	0,797	0,812	1	0,798	0,809	0,874	0,829	0,821	0,756
SO9	0,696	0,742	0,715	0,732	0,759	0,794	0,790	0,814	0,834	0,810	0,811	0,817	0,773	0,783	0,798	1	0,795	0,783	0,815	0,778	0,746
SO10	0,888	0,755	0,727	0,734	0,761	0,794	0,821	0,876	0,831	0,850	0,825	0,814	0,824	0,785	0,809	0,795	1	0,819	0,802	0,829	0,739
SO11	0,777	0,764	0,739	0,746	0,744	0,799	0,829	0,837	0,843	0,838	0,855	0,826	0,831	0,783	0,874	0,783	0,819	1	0,800	0,821	0,756
SO12	0,714	0,753	0,742	0,749	0,756	0,828	0,845	0,823	0,807	0,821	0,823	0,800	0,824	0,911	0,829	0,815	0,802	0,800	1	0,814	0,739
SO13	0,772	0,791	0,778	0,761	0,768	0,813	0,843	0,830	0,843	0,833	0,855	0,826	0,865	0,797	0,821	0,778	0,829	0,821	0,814	1	0,771
ZA1	0,670	0,761	0,742	0,739	0,843	0,751	0,763	0,748	0,773	0,749	0,768	0,756	0,737	0,737	0,756	0,746	0,739	0,756	0,739	0,771	1

Figure 4: Relatedness matrix, generated with Plink 1.9.

8.6 Population Structure & Genetic Differentiation

To review the population structure of our samples, we ran several population structure analyses, and tested for population clustering and genetic differentiation.

8.6.1 Population Structure and Admixture Analysis

The Principal Component Analysis (PCA) performed on the 21 cheetahs and 207 SNPs (set A), showed a clear demarcation between the three subspecies, which we visualised first as a Scatterplot (figure 5) including the samples' names. For a clearer representation of the clusters, we made a Colorplot (figure 6), which displays the genetic relationship not only by spatial distances but also by similar colors. In general, the genetic distances between the three populations were low, as the 20 retained eigenvalues only presented 20% of the total variation.

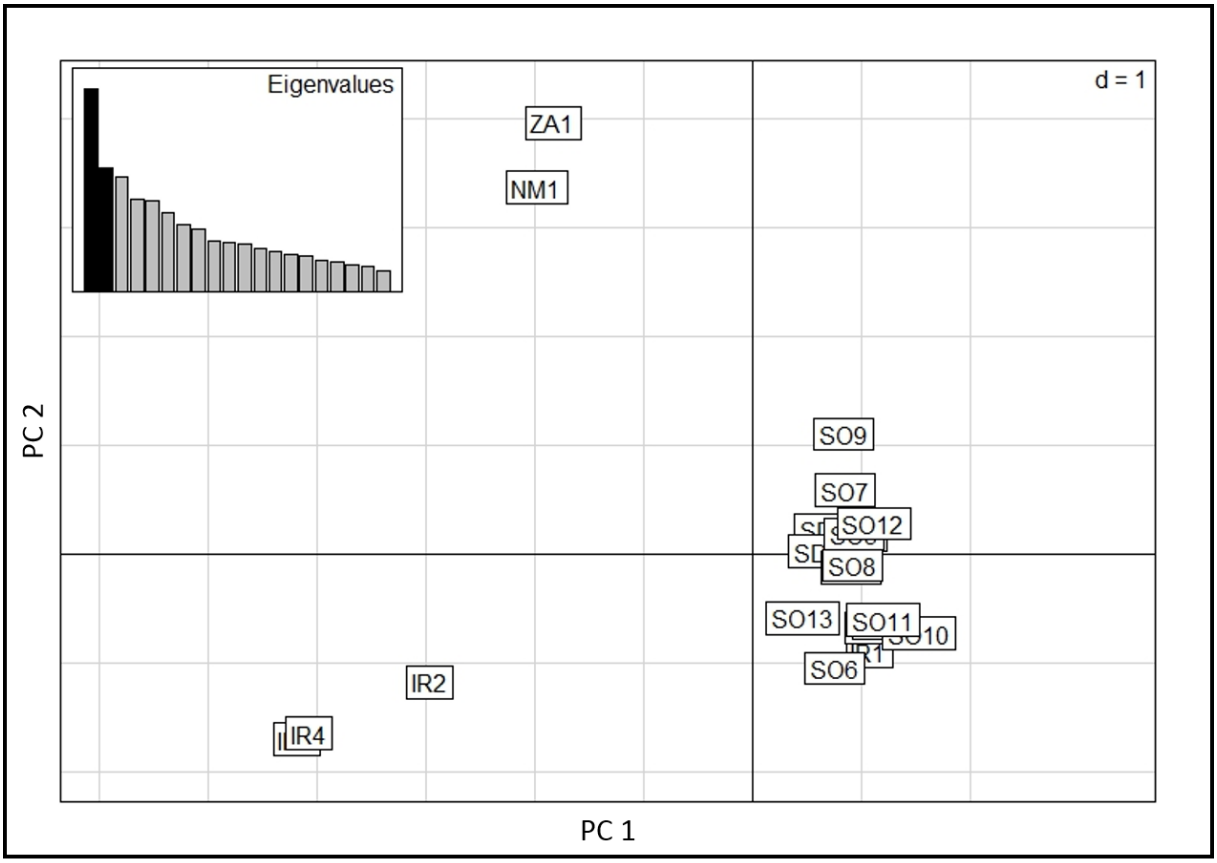


Figure 5: Scatterplot of the PCA for set A (21 individuals, 207 SNPs), representing the two major Principal Components. The retained 20 Eigenvalues represent 20% variation.

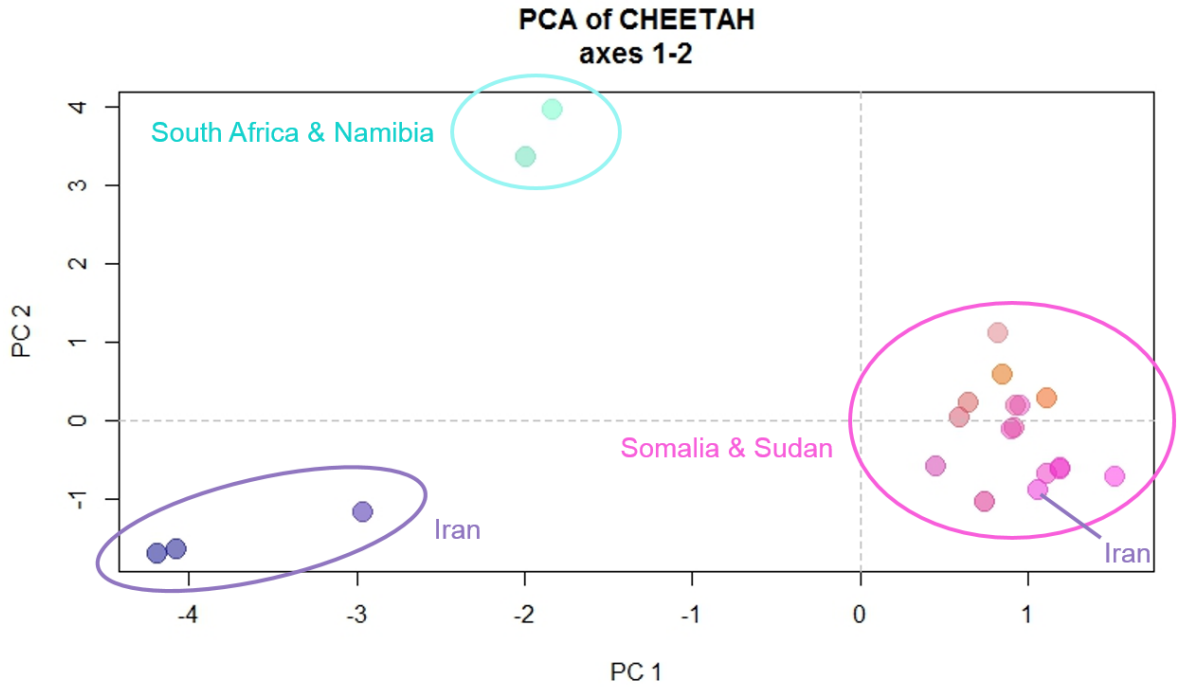


Figure 6: Colorplot of the PCA performed for set A, with clear distinction between the three subspecies.

To confirm the clustering of the samples into the three subspecies *A. j. jubatus*, *A. j. soemmeringii* and *A. j. venaticus,* we added additional six samples, most from the original habitat of *A. j. soemmeringii*, and made a new Scatterplot (figure 7) and Colorplot (figure 8), including 27 individuals and 881 SNPs (dataset C). It further hedges the clustering shown in figures 5 and 6, but noticeably, the individual from Zimbabwe clusters with the *A. j. soemmeringii* population. This is most likely due to the high allowed missingness of this sample set, and resulting in a biased representation of this individual's genome.

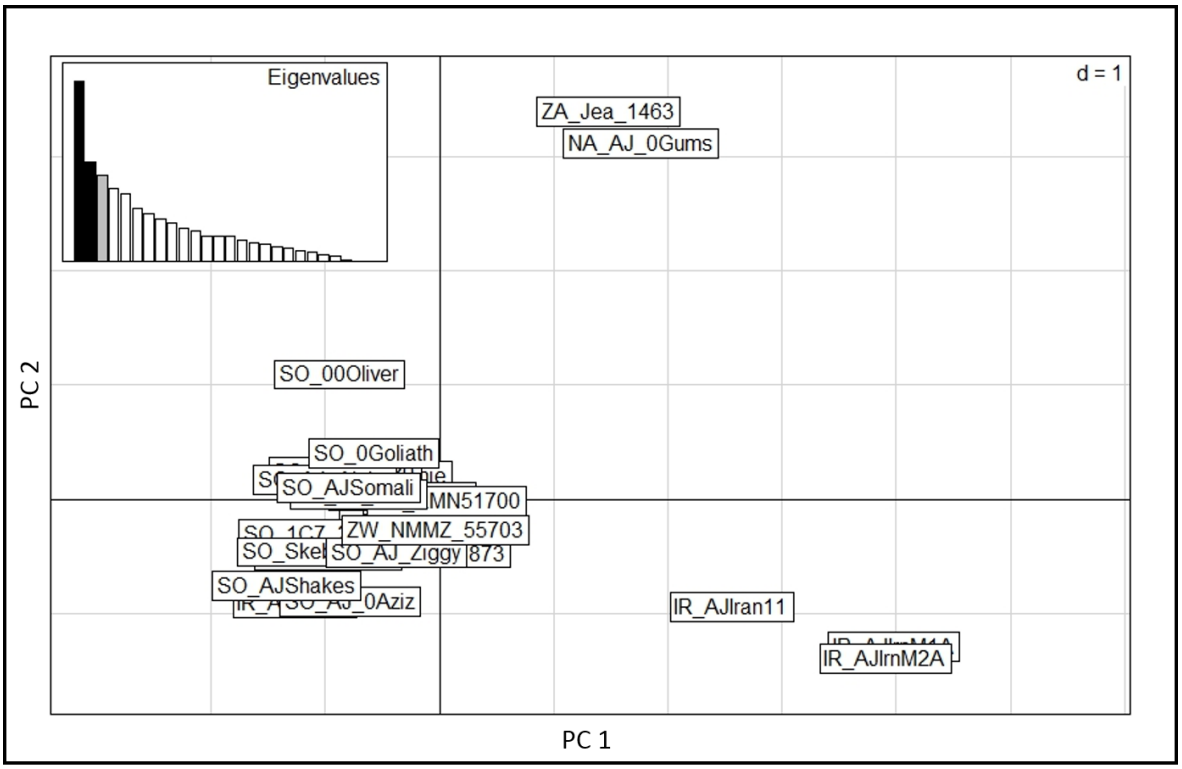


Figure 7: Scatterplot visualizing the Principal Component Analysis conducted for dataset C with 27 individuals.

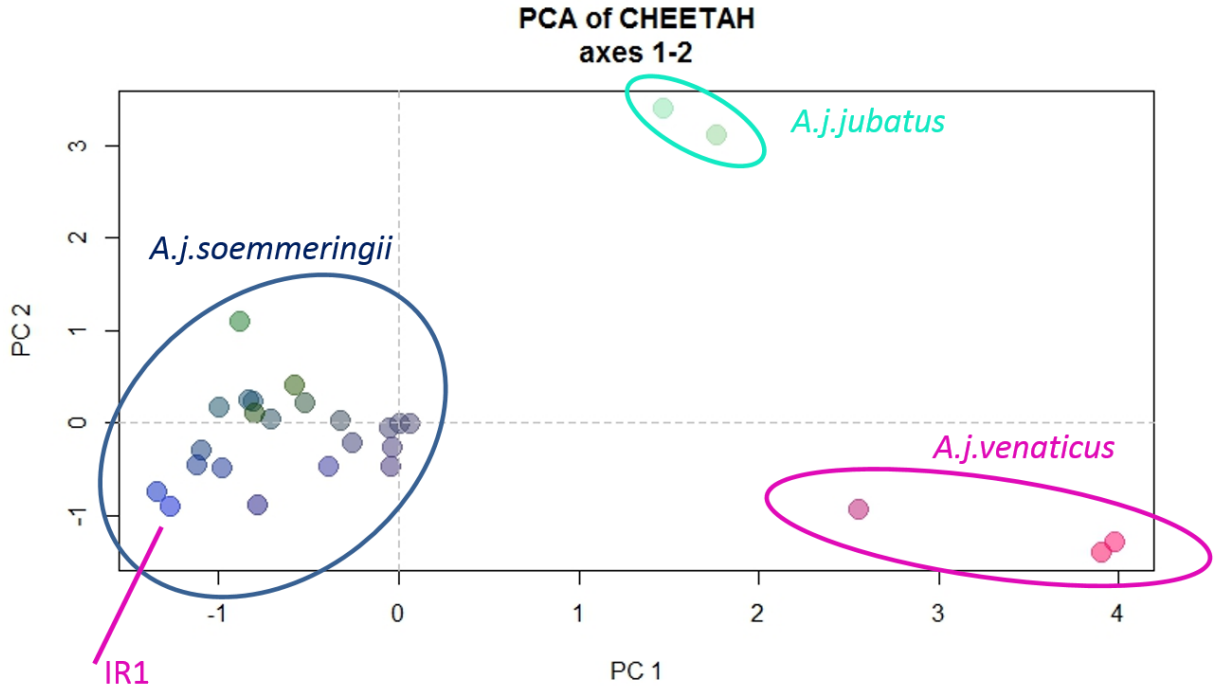


Figure 8: Colorplot of the PCA performed for set C (27 individuals, 207 SNPs), representing the geographic distributions.

Notably, the Iranian cheetah (IR1), which showed high relatedness (89%) with a Somalian individual, consistently clustered with the *A. j. soemmeringii* group. To clarify whether this result might be an artefact due to high individual missingness of genotypes, we re-filtered dataset A for a maximum genotype missingness of 4% (--geno 0.04) and excluded all SNPs with low representation; hence we kept only 82 SNPs covered in 96% of the 21 individuals. With this smaller dataset we still retrieved the same clustering solutions in the PCA as IR1 continued grouping with the *A. j. soemmeringii* individuals (data not shown).

The analysis of set D with only *A. j. soemmeringii* individuals was solely conducted to detect potential structure within this subspecies. It clearly depicts how the population does not show major divergence (figure 9), independent from the individuals' countries of origin, which was either Sudan or Somalia.

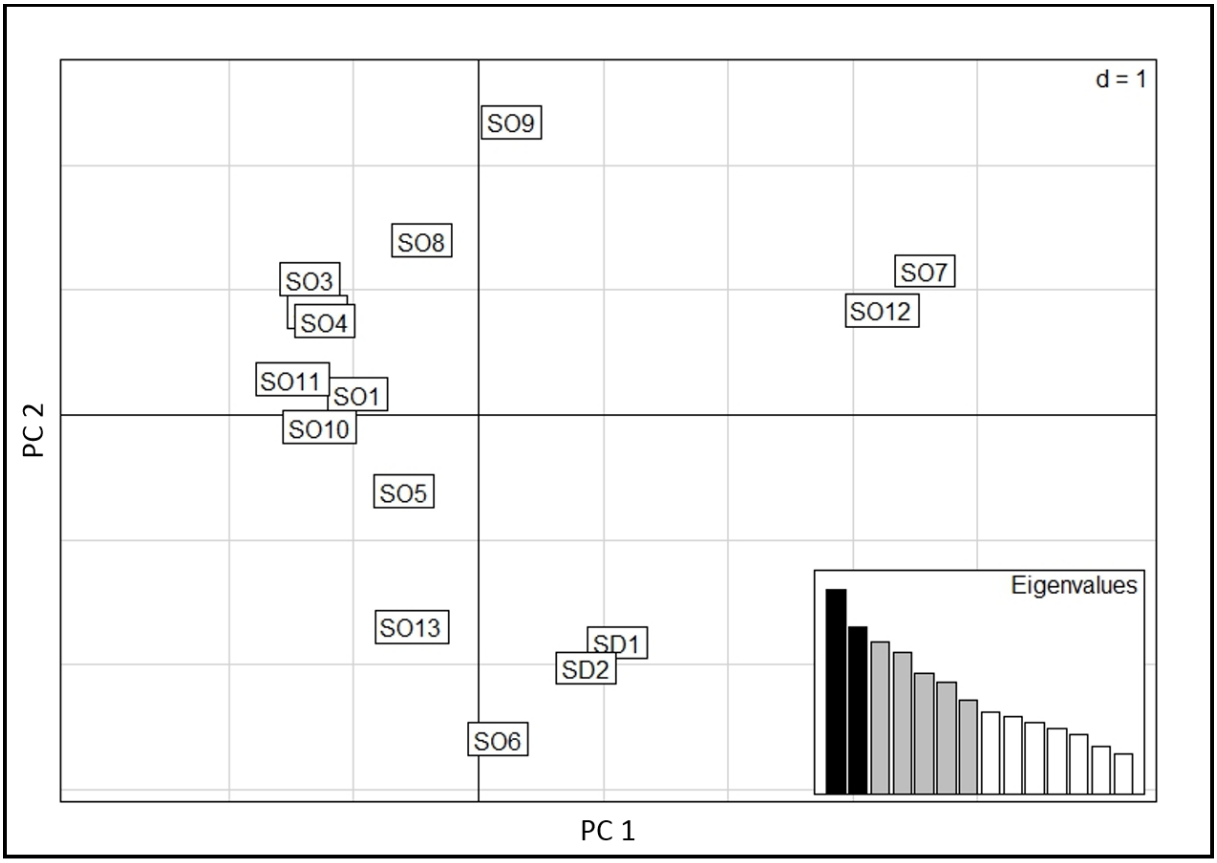
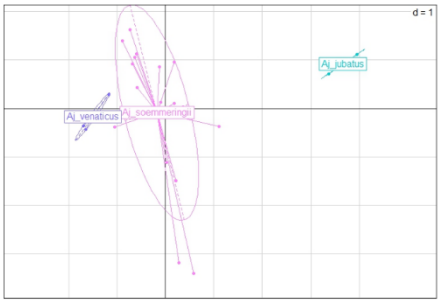
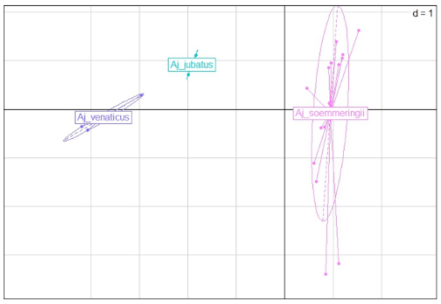


Figure 9: Scatterplot of the PCA performed for set D. The population shows only slight divergences, but these seem to be independent of the individual's country of origin. For this analysis, 7 eigenvalues were kept.

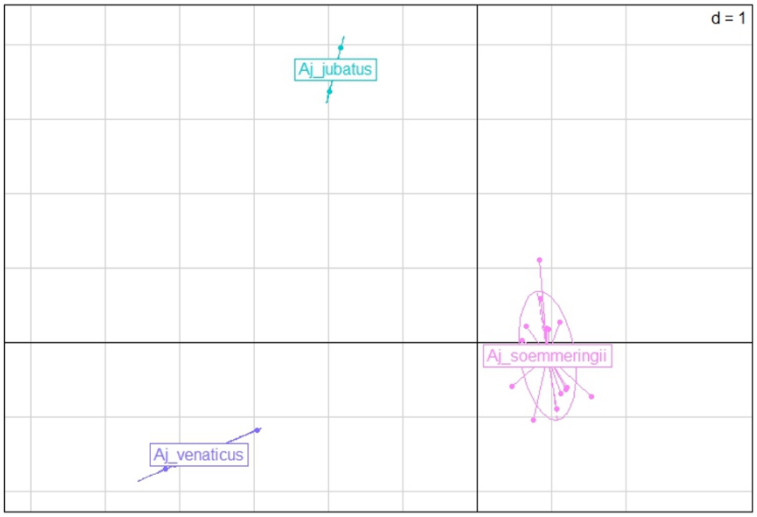
Additionally, we created s.class plots of all combinations of the first three principal components for the main set A, including 21 individuals and 207 SNPs. S.class plots display two quantitative variables (Principal Components) each for predefined clusters within the sample set. These three graphs depict well the different clustering options, depending on analysis settings and retained principal components. However, only the combination of PC2 and PC3 gives a slightly shifted clustering, compared to all other plots shown so far (see figure 10).



s.class plot of PC2 & PC3



s.class plot of PC1 & PC3



s.class plot of PC1 & PC2

Figure 10: s.class plots for set A. Combinations of PC 1 and PC2, as well as PC1 and PC3, support the closeness of *A. j. venaticus* to *A. j. jubatus*. Only inspecting the combination of PC2 and PC3 gives a slightly different image, with *A. j. soemmeringii* and *A. j. venaticus* clustering closely together.

Finally, for dataset A, we created a multi-dimensional scaling plot (MDS plot), which is designed to visualise the similarities between all given samples. As in earlier analyses, the depiction of principal components 1 and 2 represents the quite distinct clustering of the three subspecies, and also gives an impression of closeness between individuals of the same population (see figure 11).

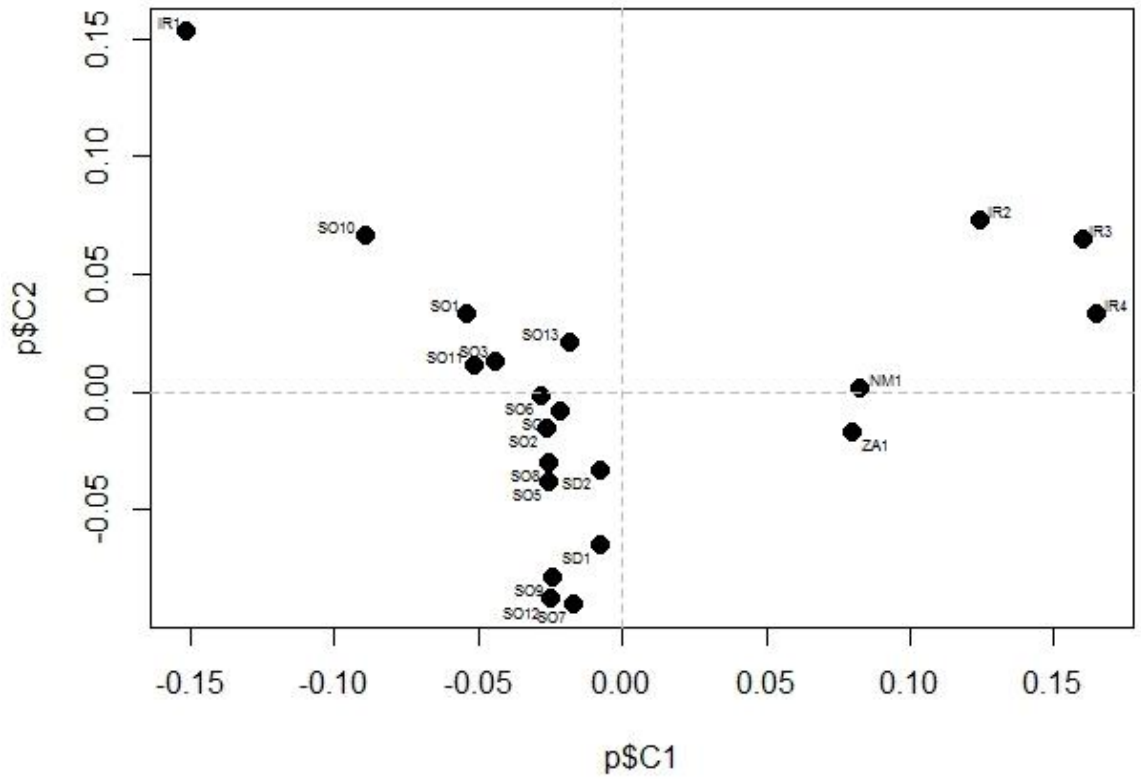


Figure 11: Multi-Dimensional Scaling plot created with plink 1.9 for 21 individuals, 207 SNPs. Representing the principal components 1 & 2, the clusters recover the three subspecies of *A. jubatus*.

The FastStructure runs for the stringently filtered dataset A, as well as the for the *A. j. soemmeringii* population (dataset D), defined two and one ancestral populations (K), respectively, as the best number of model components, which explain structure in the dataset. As we consecutively ran Admixture for a more in-depth representation, FastStructure outputs were not plotted.

For a more detailed analysis and visualisation of admixture levels within individuals and between populations, the results given by Admixture, which distinctly pointed to two clusters (smallest CV error of 0.62), were plotted. The resulting bar plots (figure 12) show a precise differentiation between the individuals of *A. j. soemmeringii* and the residual samples, combining *A. j. jubatus* and *A. j. venaticus*. However, the *A. j. jubatus* samples show some individual admixture with *A. j. soemmeringii*. When we assumed three ancestral populations (K = 3), *A. j. jubatus* and *A. j. venaticus* were clearly differentiated. The graphs for K=4 and K=5 depicted structuring within *A. j. soemmeringii*, which was further tested with the dataset D (figure 13). What is striking is that only a few individuals show considerable signs of individual admixture.

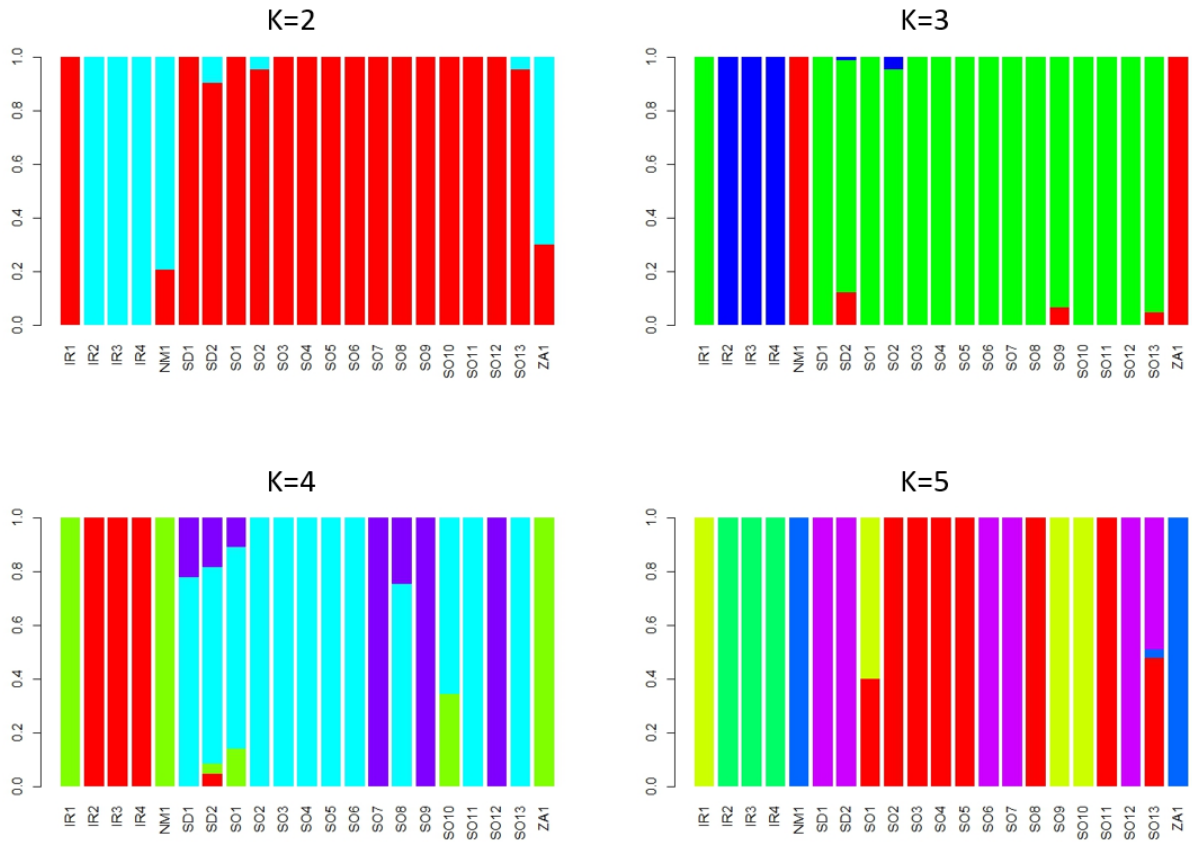


Figure 12: Admixture plots for set A (21 individuals, 207 SNPs). Clustering for cases K=2 – K=5.

Although the smallest CV error (0.57) within the *A. j. soemmeringii* population was calculated for K = 1, supporting no clear genetic differentiation within the group, further analyses with increasing numbers of K ≤ 5 revealed subtle population structure and increasing admixture levels. The observed clusters did not correspond with the country of origin of the samples.

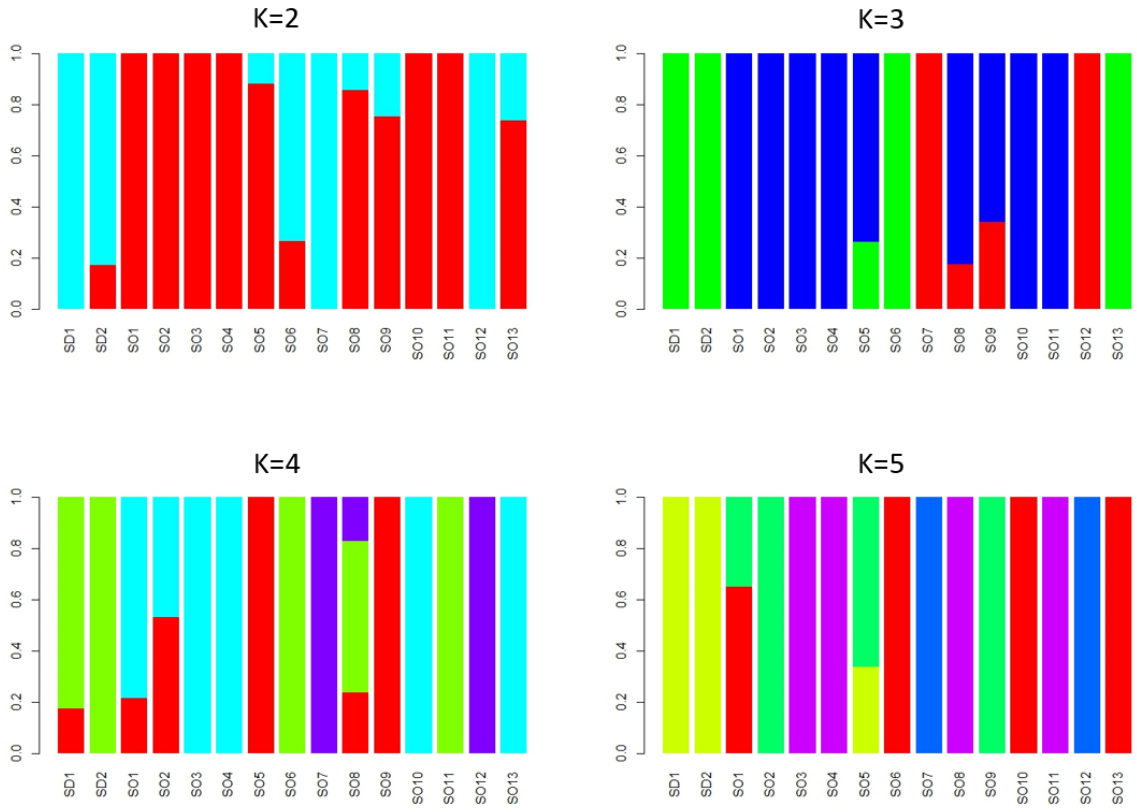
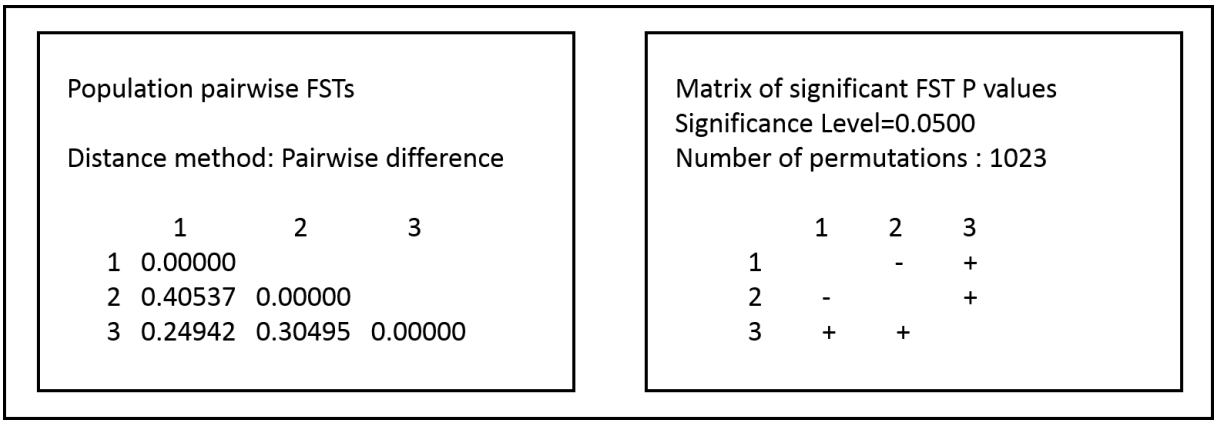


Figure 13: Admixture plots for set D (*A. j. soemmeringii* only). Clustering for cases K=2 – K=5.

8.6.2 Genetic Differentiation

Pairwise F_{ST}-values were calculated for the set including 21 individuals and 207 SNPs (set A) using Arlequin, to better evaluate the clustering and closeness of relationships between the three subspecies. As can be seen in figure 14, around 40% of variation between populations 1 and 2 (*A. j. venaticus* and *A. j. jubatus*, respectively) is explained by population structure, whereas only 25% variation between population 1 and 3 (*A. j. venaticus* and *A. j. soemmeringii*), and 30% between 2 and 3 (*A. j. jubatus* and *A. j. soemmeringii*) are explained by population structure alone. However, these results cannot stand on their own, as population sizes were too low to reliably estimate F_{ST} values, and significance is only given for the *A. j. soemmeringii* calculations. Therefore, we also calculated the individual pairwise genetic distances (figure 14).



Population pairwise FSTs

Distance method: Pairwise difference

	1	2	3
1	0.00000		
2	0.40537	0.00000	
3	0.24942	0.30495	0.00000

Matrix of significant FST P values
Significance Level=0.0500
Number of permutations : 1023

	1	2	3
1		-	+
2	-		+
3	+	+	

Figure 14: Population pairwise F_{ST}s and significance. Population 1 = *A. j. venaticus*, 2 = *A. j. jubatus*, 3 = *A. j. soemmeringii*

The supporting pairwise distance matrix (figure 15) underlines the previous analyses, showing the lowest distances exist between individuals belonging to the same subspecies, while the highest distances occur between *A. j. soemmeringii* and either of the other two subspecies.

	IR1	IR2	IR3	IR4	NM1	SD1	SD2	SO1	SO2	SO3	SO4	SO5	SO6	SO7	SO8	SO9	SO10	SO11	SO12	SO123	ZA1
IR1	0	0,295	0,335	0,362	0,308	0,290	0,259	0,185	0,250	0,214	0,245	0,268	0,254	0,295	0,263	0,304	0,112	0,223	0,286	0,228	0,330
IR2	0,295	0	0,107	0,113	0,212	0,226	0,201	0,217	0,209	0,206	0,201	0,234	0,234	0,255	0,228	0,258	0,245	0,236	0,247	0,209	0,239
IR3	0,335	0,107	0	0,099	0,237	0,251	0,234	0,250	0,249	0,249	0,244	0,246	0,242	0,266	0,251	0,285	0,273	0,261	0,258	0,222	0,258
IR4	0,362	0,113	0,099	0	0,234	0,249	0,232	0,238	0,242	0,242	0,229	0,254	0,244	0,258	0,244	0,268	0,266	0,254	0,251	0,239	0,261
NM1	0,308	0,212	0,237	0,234	0	0,239	0,210	0,231	0,225	0,234	0,214	0,232	0,251	0,246	0,237	0,241	0,239	0,256	0,244	0,232	0,157
SD1	0,290	0,226	0,251	0,249	0,239	0	0,107	0,185	0,182	0,179	0,169	0,149	0,177	0,194	0,192	0,206	0,206	0,201	0,172	0,187	0,249
SD2	0,259	0,201	0,234	0,232	0,210	0,107	0	0,172	0,164	0,169	0,155	0,143	0,152	0,181	0,176	0,210	0,179	0,171	0,155	0,157	0,237
SO1	0,185	0,217	0,250	0,238	0,231	0,185	0,172	0	0,150	0,153	0,153	0,167	0,167	0,199	0,177	0,186	0,124	0,163	0,177	0,170	0,252
SO2	0,250	0,209	0,249	0,242	0,225	0,182	0,164	0,150	0	0,150	0,123	0,143	0,186	0,200	0,157	0,166	0,169	0,157	0,193	0,157	0,227
SO3	0,214	0,206	0,249	0,242	0,234	0,179	0,169	0,153	0,150	0	0,081	0,167	0,186	0,200	0,147	0,190	0,150	0,162	0,179	0,167	0,251
SO4	0,245	0,201	0,244	0,229	0,214	0,169	0,155	0,153	0,123	0,081	0	0,155	0,190	0,195	0,165	0,189	0,175	0,145	0,177	0,145	0,232
SO5	0,268	0,234	0,246	0,254	0,232	0,149	0,143	0,167	0,143	0,167	0,155	0	0,184	0,208	0,174	0,183	0,186	0,174	0,200	0,174	0,244
SO6	0,254	0,234	0,242	0,244	0,251	0,177	0,152	0,167	0,186	0,186	0,190	0,184	0	0,208	0,203	0,227	0,176	0,169	0,176	0,135	0,263
SO7	0,295	0,255	0,266	0,258	0,246	0,194	0,181	0,199	0,200	0,200	0,195	0,208	0,208	0	0,188	0,217	0,215	0,217	0,089	0,203	0,263
SO8	0,263	0,228	0,251	0,244	0,237	0,192	0,176	0,177	0,157	0,147	0,165	0,174	0,203	0,188	0	0,202	0,191	0,126	0,171	0,179	0,244
SO9	0,304	0,258	0,285	0,268	0,241	0,206	0,210	0,186	0,166	0,190	0,189	0,183	0,227	0,217	0,202	0	0,205	0,217	0,185	0,222	0,254
SO10	0,112	0,245	0,273	0,266	0,239	0,206	0,179	0,124	0,169	0,150	0,175	0,186	0,176	0,215	0,191	0,205	0	0,181	0,198	0,171	0,261
SO11	0,223	0,236	0,261	0,254	0,256	0,201	0,171	0,163	0,157	0,162	0,145	0,174	0,169	0,217	0,126	0,217	0,181	0	0,200	0,179	0,244
SO12	0,286	0,247	0,258	0,251	0,244	0,172	0,155	0,177	0,193	0,179	0,177	0,200	0,176	0,089	0,171	0,185	0,198	0,200	0	0,186	0,261
SO13	0,228	0,209	0,222	0,239	0,232	0,187	0,157	0,170	0,157	0,167	0,145	0,174	0,135	0,203	0,179	0,222	0,171	0,179	0,186	0	0,229
ZA1	0,330	0,239	0,258	0,261	0,157	0,249	0,237	0,252	0,227	0,251	0,232	0,244	0,263	0,263	0,244	0,254	0,261	0,244	0,261	0,229	0

Figure 15: Pairwise distance matrix, created with Plink 1.9.

8.7 Phylogenetic Analyses

To visualise the phylogenetic relations between the samples, Splitstree was used to create both Neighbor Joining trees and NeighborNets. While the first gives an easier to interpret image and clear splits representing the most parsimonious setting, the latter shows all possible connections between all individuals as a net, representing the various options within a phylogenetic tree.

Trees and Networks were constructed for dataset A, including 21 cheetahs and 207 SNPs, to closely illustrate the relationships between the individuals of the three subspecies. In dataset B, the two outgroups *L. lynx* and *P. concolor* were included to root the tree/ network. Notably, *A. j. venaticus* was identified as the subspecies closest to this root.

The different phylogenetic trees received (figure 16) depict some variation, but overall support the earlier analyses. All of them show a clear clustering between the three subspecies, as well as a split between *A. j. venaticus* and *A. j. jubatus* against *A. j. soemmeringii*. Also, in both sets including the outgroups, the root is located closest to the Iranian samples.

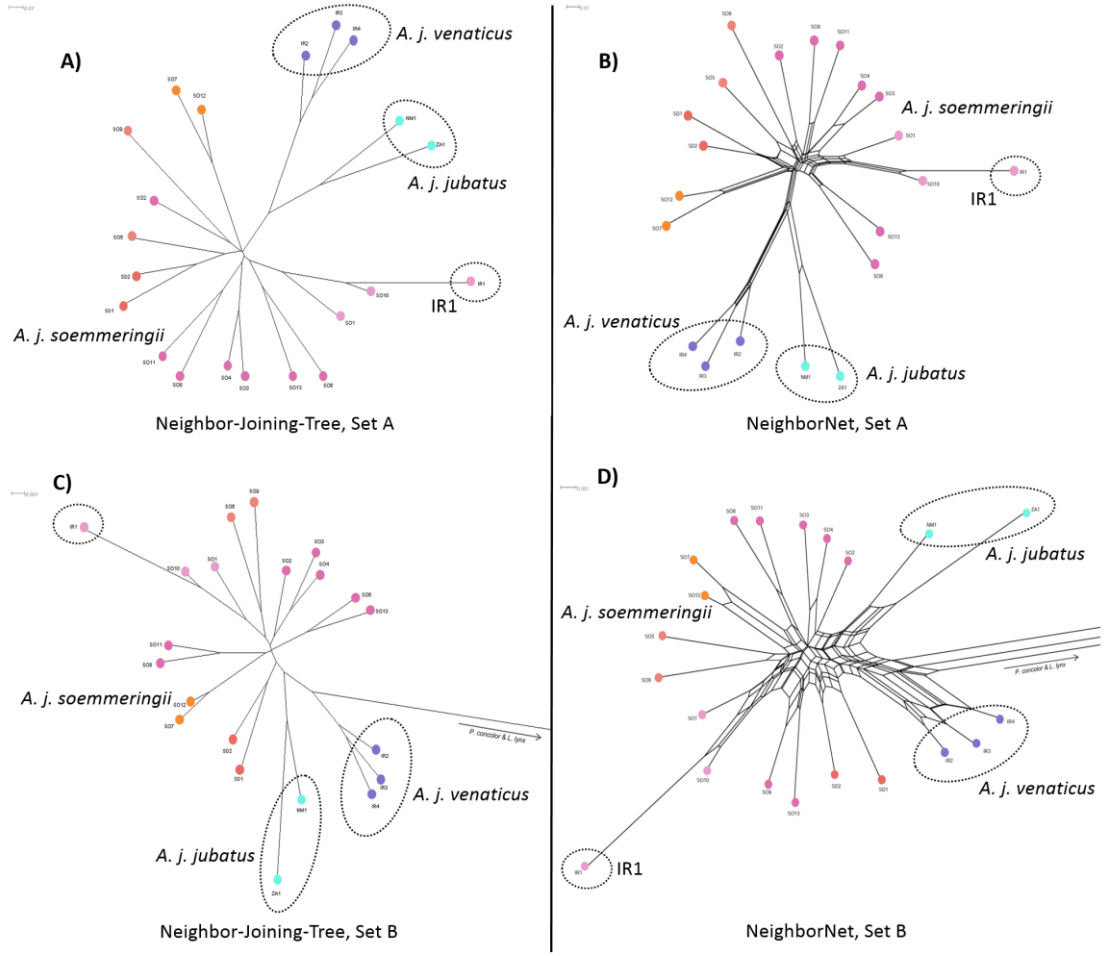


Figure 16: NeighborNets and Neighbor-Joining-Trees for Sets A and B, generated with Splitstree. NeighborNets (B, D) illustrating all possible phylogenetic relations, Neighbor-Joining-Trees (A, C) the most parsimonious.

8.8 Population Genetic Summary Statistics

To test for possible signs of low heterozygosity, and coherently inbreeding, Adegenet was used to calculate the observed vs. expected heterozygosity (H_o/H_E) and the FIS-value. Due to the low numbers of individuals in both the *A. j. venaticus* and *A. j. jubatus* subspecies, these calculations could only be done for the overall set and *A. j. soemmeringii*.

As can be seen in figure 17, the calculation of H_o/H_E highly significantly (p-value = 1.023^{-15}) shows a reduction of H_o compared to H_E. Also, the values indicate a higher diversity within *A. j. soemmeringii* than within the overall set, thus indicating a lower diversity within *A. j. jubatus* and/or *A. j. venaticus*. The same is suggested by the F_{IS}-values of 0.25 (overall) and 0.22 (*A. j. soemmeringii*), which show that the cheetahs have more homozygous genotypes than expected, hinting to a high relatedness or inbreeding.

Heterozygosity	**FIS-value**	**FIS-value**
A. soemmeringii	*A. soemmeringii*	Overall
H_o mean 0.1721245		
Standard deviation Ho 0.1628038	One Sample t-test	One Sample t-test
H_e mean 0.1953633		
Standard deviation He 0.1721251	t = 5.2721, df = 15, p-value = 9.389e-05	t = 5.7602, df = 20, p-value = 1.23e-05
	95 percent confidence interval:	95 percent confidence interval:
Overall	0.1325860 0.3125465	0.1571493 0.3355840
H_o mean 0.1662865	sample estimates:	sample estimates:
Standard deviation Ho 0.1332865	mean of x	mean of x
H_e mean 0.2180195	FIS-value 0.2225662	FIS-value 0.2463667
Standard deviation He 0.1557611		

Figure 17: Heterozygosity (H_o/H_E) and F_{IS}-values for *A. j. soemmeringii* and entire set A (overall).

9. Discussion

We analysed the genetic diversity and phylogenetic relationship of the cheetah, *A. jubatus*, and its subspecies using genome-wide ddRAD sequencing.

9.1 Limitations in Sample collection and DNA extraction

A lot of the samples originated from various sources all over the historic cheetah distribution. Taking into consideration that some of these samples were up to 200 years old, their history could not always be fully reconstructed. Even if so, conditions and early preservation attempts from both museums and private collectors were often not beneficial for the DNA quality. Therefore, even after being collected in modern museums and treated for preservation, the expected DNA output and quality was low, which made the extraction with the comparably intricate Rohland et al. (45) protocol for ancient DNA extraction necessary.

During the extraction of the 165 historic samples, these limiting factors became clear. While expecting a usually brownish liquid after the 48-hour digestion, a compound of dissolved sample material and extraction buffer, for some samples this liquid was red, green, brightly yellow, or even purple. Also, when taking different sample materials into account, these colours visualised the complex conditions within the collection.

For the 31 modern samples, more precisely the nine blood samples, we decided on the specialised blood extraction kit in the hope of being able to extract from the highly diluted samples of low volume, but were not successful.

9.2 Population Structure, Admixture, Genetic Differentiation, and Diversity

Despite working with sets containing only between 201 and 884 SNPs, and 15 to 27 individuals, most of our analyses scored congruent results for population structure and relatedness. The tested relatedness was in an expected range, after already filtering out one of the two South African individuals, which were closely related on pedigree analysis. Usually, a relatedness (identity-by-descend) of 95% is seen as a threshold indicating siblings or parent-offspring, above which unbiased analyses of population structure are unfeasible. Our values mainly varied within the 70% and 80% range, with the maximum at 92% between two individuals within the *A. j. soemmeringii* population. These values indeed indicate rather high levels of relatedness within the remaining cheetah populations, especially compared to other large mammals, but are not unexpected. The cheetah has been known to have a rather low divergence already since the 1980s (38). The results for observed and expected heterozygosity, as well as F_{IS}-values, support these findings with a high significance. As

expected, they point towards a higher level of heterozygotes within the *A. j. soemmeringii* subspecies, opposed to the overall sample set. These results match the findings that smaller populations, as *A. j. venaticus*, react more sensitively to external disturbances (31).

Of course, with the low number of remaining individuals, especially in the western African *A. j. hecki* populations, and the Asiatic *A. j. venaticus*, introduction of fresh genes from a viable genepool might be the key to a successful subspecies conservation. With less than 70 Asiatic individuals left (2, 36, 49), taking immediate action to preserve the subspecies and its genetic identity is inevitable. As earlier studies have shown (28), genetic rescue of a population does not necessarily have to correspond with loss of regional adaptations, but can be successfully conducted by carefully choosing the most suitable population. While we cannot be certain concerning the suitability of *A. j. hecki*, the Southern African subspecies *A. j. jubatus* is both viable and closely related to *A. j. venaticus*. This qualifies as the ideal population to enhance the Asiatic cheetah's genetic viability by introducing individuals into the critically endangered subspecies. Therefore, our main research question; whether or not a clustering of *A. j. venaticus* with *A. j. soemmeringii* might aid the species' recovery and conservation, is more topical than ever, and can be clearly negated.

The only factors suggesting a closer relatedness between *A. j. soemmeringii* and *A. j. venaticus* are the PCA's principal components no.2 and no.3, which in combination only explain about 4% of the variation. Despite the generally low value of single eigenvalues in all set's PCAs, there was no change in the results when retaining three, eight, or 20 eigenvalues. Hence, our results are reliable in the means of consistency and variation between different sets of SNPs, individuals, and analysis settings.

The same is valid for the comparisons between phylogenetic trees and networks using the Adegenet and Splitstree softwares. Moreover, we find consistency between FastStructure and Admixture results, and the different measures of variability retained with Arlequin and Adegenet. What is congruent within all these analyses is the validation of our primary hypothesis; that the cheetah, despite having comparably low genetic diversity, shows a distinct segregation between the tested subspecies. This is especially true for *A. j. venaticus* opposed to *A.j.soemmeringii*, as we hypothesized in this study, and which has been detected already in foregone studies (11, 34, 41).

For the detected outlier IR1, all conducted analyses show a clear clustering with the *A. j. soemmeringii* subspecies, far from the *A. j. venaticus* individuals. Hence, we concluded that this single individual might have been wrongly labelled in its history or was given as a present to the Iranian museum, where it was sampled. However, it might not originate from this country.

The results of this thesis respond to several national action plans (1, 8, 25, 26, 27) and conference meetings (7) supposing further genetic studies to clarify the exact genetic structure of *Acinonyx jubatus*. Therefore, these results can be used for a direct application towards cheetah conservation, and improve future conservation attempts.

10. Outlook

To expand this study, and find further, more detailed answers to the initial research question, our group will continue working with the exceptional sample set collected by Lena Godsall Bottriell and Paul Bottirell from the Rex Foundation, UK. As a next step, the remaining DNA, combined with further, new extractions, will be re-analysed. For this, we will use hybridization capture (hyRAD) (48) using custom baits designed for the RAD loci developed in this study, and do the library preparation in-house to quality control each step of the protocol. This set will include for example newly received samples from *A. j. jubatus* populations.

Concerning conservation efforts, our results will make an impact on future attempts, in terms of subspecies clustering and possible introduction of individuals to other populations. We contradict the approach to cluster *A. j. venaticus* , *A. j. hecki*, and *A. j. soemmeringii* into one subspecies. As for conservation, anthropogenic factors, and human-carnivore conflict in particular have proven to be more severe than genetic factors (9, 36, 50). Thus, the management of human-carnivore conflict should be one of the main aims of wildlife conservation. This could be improved e.g. by focusing on the conservation of whole, intact guilds rather than individual species (5, 50), to obtain viable habitats and food chains. Prey depredation is known to be one of the main threats endangering the survival of carnivores in the wild (17, 36). Like this, the impact on farm animals should be minimised, and hence the acceptance of large carnivores within the communities increased. Additionally, transboundary collaborations for conservation will be necessary (17), as the cheetah is an especially large ranging predator, and to protect a guild properly, the conservation area should not be smaller than the largest range of a single species within it (50).

For *A. j. venaticus* in particular, protecting it and its genetic distinctiveness should be a major concern (11). However, with a current disjunct population of less than 70 individuals, this might turn out to be challenging.

Regarding possible future studies, further analyses including *A. j. hecki*, and maybe the former subspecies of *A. j. raineyii*, could give interesting insight into cheetah phylogeny. For this, more modern samples would be needed, which are difficult to sample, as populations decline. Also, as earlier studies (i.e. 11) have already suggested, a large-scale project with better sample and SNP coverage would be desirable.

11. References

1) Action for Cheetahs in Kenya (ACK) mission (2013) Year End Report 2013 - Cheetah Conservation and Human Impact in Kenya

2) Ahmadi, M., Balouchi, B.N., Jowkar, H. et al. (2017) Combining landscape suitability and habitat connectivity to conserve the last surviving population of cheetah in Asia. *Divers. Distrib*. 2017; 23:592–603.

3) Alexander, D.H., and Lange, K. (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC bioinformatics*, *12*(1), 246.

4) Baird, N.A., Etter, P.D., Atwood, T.S. et al., (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3, e3376.

5) Belbachir, F., De Smet, K., (2017) Algerian Cheetah´s Habitat – The Cheetah in the Algeria´s Central Sahara. Sandscript Issue 22, p.8.

6) Belbachir, F., Pettorelli, N., Wacher, T., Belbachir-Bazi, A., Durant, S. M. (2015) Monitoring rarity: The critically endangered Saharan cheetah as a flagship species for a threatened ecosystem. *PLoS One*, *10*, e0115136.

7) Berzins, R., Belbachir, F. (Eds) (2007) *Compte-rendu de la deuxième réunion de l'Observatoire du Guépard en Régions d'Afrique du Nord (OGRAN), 20-25 Novembre 2006, Tamanrasset, Algérie*. Paris: Société Zoologique de Paris (SZP). vi + 36 pp.

8) Breitenmoser, U., Alizadeh, A., Breitenmoser-Würsten, C. (2009) Conservation of the Asiatic Cheetah, its Natural Habitat and Associated Biota in the I.R. of Iran, Iranian Cheetah Society Terminal Evaluation Report.

9) Castro-Prieto, A., Wachter, B., Sommer, S. (2011) Cheetah paradigm revisited: MHC diversity in the world's largest free-ranging population. *Molecular Biology and Evolution* 28(4): 1455 1468

10) Catchen, J., Hohenlohe, P., Bassham, S., Amores, A. and Cresko, W. (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology.

11) Charruau, P., Fernandes, C., Orozco-terWengel, P. *et al.* (2011) Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: evidence for long-term geographic isolates. *Molecular Ecology*, 20, 706–724

12) Constantini, D., Wachter, B., Melzheimer, J., Czirják, G.Á. (2017) Socioecological and environmental predictors of physiological stress markers in a threatened feline species. Conserv Physiol 5(1): cox069

13) Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R., Lunter, G., Marth, G., Sherry, S.T., McVean, G., Durbin, R. and 1000

Genomes Project Analysis Group (2011) *The Variant Call Format and VCFtools*, Bioinformatics.

14) Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., Blaxter, M.L. (2013) Special features of RAD sequencing data: implications for genotyping. Mol. Ecol. 22 (11), 3151–3164.

15) Dobrynin, P., Liu, S., Tamazian, G., Xiong, Z., Yurchenko, A.A., Krasheninnikova, K. et al. (2015) Genomic legacy of the African cheetah, Acinonyx jubatus. Genome biol. BioMed Central. 2015; 16:277.

16) Durant, S.M., Mitchell, N., Groom, R., Pettorelli, N., Ipavec, A., Jacobson, A.P., et al. (2016) The global decline of cheetah Acinonyx jubatus and what it means for conservation. Proc Natl Acad Sci. 2016; 114: 201611122.

17) Durant, S.M., Mitchell, N., Ipavec, A., Groom, R. (2015) Acinonyx jubatus. The IUCN red list of threatened species 2015: e.T219A50649567.

18) Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., Cresko, W.A. (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol Biol.:772: 157–78.

19) Excoffier, L., Lischer, H.E.L. (2010) Arlequin suit ver 3.5: A new series of programs to perform population genetics analysis under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

20) Farhadinia, M.S., Eslami, M., Hobeali, K., Hosseini-Zavarei, F., Gholikhani, N. & Tak Tehrani, A. (2014). Status of Asiatic cheetah in Iran: A country-scale assessment. Report to People's Trust for Endangered Species. Tehran: Iranian Cheetah Society.

21) Farhadinia, M.S., Hunter, L.T., Jourabchian, A., Hosseini-Zavarei, F., Akbari, H., Ziaie, H., Schaller, G.B. and Jowkar, H., (2017) The critically endangered Asiatic cheetah *Acinonyx jubatus venaticus* in Iran: a review of recent distribution, and conservation status. Biodiversity and Conservation, 1-20.

22) Huson, D.H. and Bryant, D. (2006) Application of Phylogenetic Networks in Evolutionary Studies, Mol. Biol. Evol., 23(2):254-267.

23) Iranian Cheetah Society (ICS) (2017) Final Report 2017. http://www.wildlife.ir/wp-content/uploads//sites/1/2017/10/ICS_Cheetah_Final_Report.pdf

24) IUCN CatSG (Cat Specialist Group) (2017) Revised taxonomy of the Felidae, Cat News Special Issue Nr.11.

25) IUCN/SSC (2015) Regional Conservation Strategy for the Cheetah and African Wild Dog in Southern Africa, IUCN, Gland, Switzerland

26) IUCN/SSC (2007) Regional Conservation Strategy for the Cheetah and African Wild Dog in Eastern Africa, IUCN, Gland, Switzerland

27) IUCN/SSC (2012) Regional Conservation Strategy for the Cheetah and African Wild Dog in West, North and Central Africa, IUCN, Gland, Switzerland

28) Johnson, W.E., Onorato, D.P., Roelke, M.E., Land, E.D., Cunningham, M., Belden, R.C., et al. (2010) Genetic restoration of the Florida panther. Science. 2010;329:1641–5.

29) Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405

30) Jombart, T. and Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics.

31) Keller, L.F., Waller, D.M., (2002) Inbreeding effects in wild populations. Trends Ecol. Evol. 17, 230–241.

32) Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60.

33) Lischer, H.E.L. and Excoffier, L. (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28: 298-299.

34) Machado, A. (2015) Phylogeography and population structure of the cheetah (*Acinonyx jubatus*): new insights from museum samples. Master thesis, University of Lausanne, Vetmeduni Vienna.

35) Mallon, D.P. (2007) Cheetahs in Central Asia: A historical summary. Cat News 46:4-7.

36) Moqanaki, E., & Cushman, S. (2016) All roads lead to Iran: Predicting landscape connectivity of the last stronghold for the critically endangered Asiatic cheetah. *Animal Conservation*, *20*, 29–41.

37) Nowell, K., Jackson, P. (1996) North Africa and Southwest Asia, Cheetah. In: Nowell, K., Jackson, P., editors. Wild cats: Status survey and conservation action plan. Gland, Switzerland: IUCN/SSC Cat Specialist Group; p 41-44.

38) O'Brien, S.J., Wildt, D.E., Goldman, D. *et al.* (1983) The cheetah is depauperate in genetic variation. *Science*, 221, 459–462.

39) Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26: 419-420.

40) Peterson, B.K., Weber, J.N., Kay, E.H. et al. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One, 7, e37135.

41) Preier, L. (2016) Neutral and adaptive genetic diversity in snow leopard and cheetah. Master thesis, University of Natural Resources and Applied Life Sciences Vienna, Vetmeduni Vienna.

42) Purcell, S. (2007) PLINK version 1.9. http://pngu.mgh.harvard.edu/purcell/plink/

43) Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. and Sham, P.C. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

44) Raj, A., Stephens, M. and Pritchard, J.K. (2014) fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets, Genetics 197:573-589.

45) Rohland, N., Siedel, H. & Hofreiter, M. (2010) A rapid column-based ancient DNA extraction method for increased sample throughput. – Molecular Ecology Resources, 10: 4, 677–683.

46) RStudio Team (2015). RStudio: Integrated Development for R. RStudio Inc., Boston

47) Springer, M.S., Debry, R.W., Douady, C. et al. (2001) Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Molecular Biology and Evolution, 18, 132–143.

48) Suchan, T., Pitteloud, C., Gerasimova, N.S., et al. (2016) Hybridization capture using RAD Probes (hyRAD), a new tool for performing genomic analyses on collection specimens. Plos One 11:3.

49) United Nations Development Program (UNDP) (2016) Conservation of the Asiatic Cheetah Project: Phase II

50) Winterbach, H., Winterbach, C., Somers, M., Hayward, M. (2013). Key factors and related principles in the conservation of large African carnivores. Mammal Review, 43, 89–110.

12. Appendix

12.1 Additional Tables

Extraction no.	Lab Code	Sample Name	Country of Origin	Nano Drop	Nucleic Acid	Qubit
				260/280	(ng/µl)	µg/ml
1	428	ZMB 34306	TZ	1,49	32,4	2,66
2	431	ZMB 56306	TZ	1,46	185,7	6,82
3	430	ZMB 56302	TZ	1,53	192,7	16,34
4	429	ZMB 56287	TZ	1,53	145,1	29,8
5	432	ZMB 56309	TZ	1,49	182,5	29,8
6	343	SMF 58993	EG	1,43	132	2
7	336	RMCA 454	CD	1,52	39,8	5,92
8	334	RMCA 22347	CD	1,46	170,3	2,12
9	333	RMCA 19237 (1)	CD	1,44	102,7	<2
10	335	RMCA 22390	CD	1,48	134,8	5,84
11	327	RMCA 1236	CD	1,45	165,7	3,32
12	345	SMNS 18941	EG	1,47	67,2	10,02
13	7	RMNH.MAM.51700	SD	1,63	50,5	24,8
14	14	ZMA.MAM.959	KE	1,46	62,6	<2
15	16	ZMA.MAM.9405	KE	1,45	84	<2
16	18	RMNH.MAM.16519	KE	1,49	65,6	<2
17	21	RMNH.MAM.51704	SN	1,77	15,1	2,92
18	22	ZMA.MAM.9410	SD	1,64	13,9	<2
19	23	RMNH.MAM.51699	IN	1,5	158,7	12,4
20	24	RMNH.MAM.41212	TZ	1,48	55,5	<2
21	25	NMW.Z.1912.014.4	KE	1,63	158,3	39,8
22	45	ID8436	SD / SO	1,65	94,4	32,6
23	46	ID1873	SD / SO	1,8	140,4	59,8
24	47	ID8337	SD / SO	1,53	23,6	<2
25	49	ID8952	SD / SO	1,49	79,9	<2
26	72	MNHN 2006-460	TD	1,4	113,7	<2
27	74	CSIC.19316	EH	1,39	34,1	<2
28	76	CSIC.29889	EH	1,78	347,7	180,8
29	78	CSIC.29951	DZ	1,46	117,4	<2
30	79	CN 4520	AF	1,82	366,4	240
31	82	NMSA 864	ZW	1,47	225,9	17,7
32	88	KM 14432	NA	1,78	48,3	15,54
33	98	BMNH ZD 1912.5.10.1	CD	1,64	196,3	34,2
34	103	BMNH ZD 1905.5.9.12	AO	1,66	378,8	142,8
35	110	ZMB MAM 56276	TN	1,45	557,5	3
36	111	ZMB MAM 1219	SD	1,44	102,5	<2
37	119	ZMB MAM 56305	TZ	1,48	624,9	8,5
38	125	ZMB MAM 56122	JO	1,72	415	187
39	128	BMNH 1939.3665	KE	1,5	112,9	8,8
40	130	ZMMU S-49247	TM	1,62	205,3	47,2
41	135	ZIN 2808	TM	1,45	53,5	8,94

42	139	ZIN 23893	TM	1,52	156,3	8,68
43	140	ZIN 24830	TM	1,43	210,3	2,28
44	27	RCSOM/A 118.6	TZ	1,75	279,5	110
45	80	CN 3493	KE	1,44	55,2	11,1
46	176	MCZ 16714	SY	1,45	155,6	6,1
47	182	HUJI M. 4038	SY	1,44	85,6	<2
48	183	HUJI M. 4039	SY	1,39	211,6	<2
49	186	MSNG 4296	ER	1,43	188,4	4,7
50	187	MSNG 17832	ER	1,41	177,3	4,5
51	188	MSNG 209	ER	1,68	158,1	106
52	189	MSNG 35224	ET	1,45	123,3	<2
53	190	MSNG 33059a	ET	1,56	128,1	23,6
54	191	MSNG 33059b	ET	1,61	195,7	52,4
55	194	MZUF 1664	SO	1,51	122,1	19,48
56	195	MZUF 1830	SO	1,49	83,2	8,62
57	196	MZUF 1850	SO	1,38	131,6	<2
58	197	MZUF 2135	SO	1,43	105,8	5
59	198	MZUF 15501	AO	1,4	94,6	<2
60	207	ZMUZ 10092	IN	1,41	108,5	<2
61	208	ZMUZ 17214	CM	1,39	189,3	4,68
62	209	ZMUZ 11579	ET	1,41	100,3	2,34
63	141	CCEC40000366	SY	1,48	86,7	<2
64	243	AMNH 80619	AO	1,51	81,8	<2
65	244	AMNH 80618	AO	1,45	60,7	2,02
66	245	AMNH 80865	AO	1,49	89,5	<2
67	246	AMNH 119657	MZ	1,5	70,1	<2
68	235	AMNH 100309	IN	1,46	35,5	2,62
69	307	AJIran 04	IR	1,48	80,7	3,64
70	308	AJIran 05	IR	1,48	71,9	7,3
71	309	AJIran 06	IR	1,49	92,3	2,02
72	310	AJIran 07	IR	1,44	65,1	<2
73	311	AJIran 08	IR	1,5	166,1	31,2
74	312	AJIran 09	IR	1,46	51	<2
75	313	AJIran 10	IR	1,46	66,7	<2
76	314	AJIran 11 "Narita"	IR	1,85	501	<2
77	411	#4304 La Palmyre	AE	1,77	191,2	129,2
78	437	BMNH 1981.2531		1,48	80,9	<2
79	442	Aju DJI 1 P1	DJ	1,62	129,4	31
80	444	Aju DJI 3 P1	DJ	1,78	292,2	194
81	445	Aju DJI 4 P1	DJ	1,83	50,4	34,4
82	446	Aju DJI 5 P1	DJ	1,71	246,7	121,4
83	447	Aju DJI 6 P1	DJ	1,72	119,9	61,8
84	448	Aju DJI 7 P1	DJ	1,59	138	27
85	350	Tigger	KE	1,72	14,2	3,02
86	317	Claudia	KE	1,5	167,4	<2
87	322	NMW 12072	LY	1,44	53,3	2,34
88	320	NMW 12070	LY	1,82	777,3	407

89	441	NMW ST 554	LY	1,39	126,2	8,76
90	271	YM CH 1	NG	1,48	188,6	<2
91	315	AJIran M1A	IR	1,9	112,2	96,2
92	316	AJIran M2A	IR	1,87	157	112,4
93	109	ZMB.MAM 56127	BF	1,48	74,6	28,4
94	179	TAUZM M.3132	KE	1,48	153,2	<2
95	65	MNHN. A-7920	IN	1,61	25,9	4,02
96	86	KM 14430	NA	1,79	193,9	127,4
97	91	CF.REX	BW	1,46	154,2	4,78
98	118	ZMB.MAM 72847	SD	1,41	148,8	36,8
99	20	ZMA.MAM.9406	KE	1,44	134,8	4,02
100	94	BMNH ZD 1905.5.13.1	TD	1,45	129,2	3,36
101	250	AMNH 54352	TZ	1,51	99,1	4,88
102	247	AMNH 119656	MZ	1,49	172,5	2,4
103	193	MZUF 532	DZ	1,49	186	2,44
104	178	TAUZM M.2819	IQ	1,48	106,1	3,24
105	116	ZMB MAM 56121	LY	1,5	200,2	5,9
106	117	ZMB MAM 13076	LY	1,5	100,7	5,24
107	412	HZM 2.26502	OM	1,78	283	103,4
108		*Lynx lynx*	outgroup	1,84	145,7	<2
109	70	MNHN 1921-142	AO	3,06	2,4	<2
110	15	ZMA.MAM.9501	KE	1,26	23	4,1
111	115	ZMB MAM 79221	IN	1,38	188,9	11,32
112	CO_3_L	*Puma concolor*	outgroup			not tested
113	84	RBIVY.REX.1	BW	1,43	79	not tested
114	154	NMMZ 56281	BW	1,55	185,1	not tested
115	156	NMMZ 62826	BW	1,6	28,2	not tested
116	161	NMMZ 62827	BW	1,55	104	not tested
117	248	AMNH 169109	BW	1,78	229,9	not tested
118	249	AMNH 169108	BW	1,79	340,8	not tested
119	97	BMNH ZD 1905.8.13.1	MW	1,44	112,1	not tested
120	123	ZMB MAM 82181	NA	1,32	194	not tested
121	126	ZMB MAM 91304	NA	1,48	35,8	not tested
122	108	BMNH 1932.12.27.1	NG	1,43	33,3	not tested
123	134	ZIN 1891	RU	1,42	53,3	not tested
124	138	ZIN 1133	RU	1,45	149,5	not tested
125	129	ZMMU S-47281	TM	1,38	351,6	not tested
126	131	ZMMU S-51711	TM	1,47	171,9	not tested
127	132	ZMMU S-51898	TM	1,46	118,4	not tested
128	75	CSIC. 29888	EH	1,66	196,6	not tested
129	92	RBIVY.REX.2	ZA	1,48	250,8	not tested
130	95	BMNH ZD 1890.2.26.1	ZA	1,5	148,8	not tested
131	96	BMNH ZD 1886.1.25.1	ZA	1,75	33,7	not tested
132	105	BMNH ZD 1981.742	ZA	1,46	37,4	not tested
133	106	BMNH ZD 1981.743	ZA	1,35	64	not tested
134	142	NMMZ 56280	ZW	1,43	50,7	not tested
135	143	NMMZ 56241	ZW	1,48	137,8	not tested

136	144	NMMZ 59257	ZW	1,48	133	not tested
137	145	NMMZ 17633	ZW	1,55	138,7	not tested
138	146	NMMZ 5733	ZW	1,71	139,7	not tested
139	147	NMMZ 63600	ZW	1,62	132	not tested
140	148	NMMZ 3992	ZW	1,44	112,6	not tested
141	149	NMMZ 56283	ZW	1,73	279,5	not tested
142	150	NMMZ 65168	ZW	1,46	131,5	not tested
143	151	NMMZ 55703	ZW	1,59	148,3	not tested
144	152	NMMZ 63599	ZW	1,81	127,4	not tested
145	153	NMMZ 55706	ZW	1,57	180,9	not tested
146	155	NMMZ 5052	ZW	1,47	65,9	not tested
147	162	NMMZ 58685	ZW	1,46	91,5	not tested
148	163	NMMZ 59399	ZW	1,68	227,3	not tested
149	164	NMMZ 56284	ZW	1,46	139	not tested
150	165	NMMZ 4043	ZW	1,52	91,6	not tested
151	166	NMMZ 53017	ZW	1,45	71,7	not tested
152	167	NMMZ 57344	ZW	1,47	104,5	not tested
153	168	NMMZ 58688	ZW	1,44	126,7	not tested
154	170	NMMZ 64768	ZW	1,49	91,3	not tested
155	171	NMMZ 58698	ZW	1,49	113,1	not tested
156	172	NMMZ 59314	ZW	1,46	163,1	not tested
157	173	NMMZ 56969	ZW	1,45	88,5	not tested
158	169	NMMZ 58686	ZW	1,44	155,7	not tested
159	102	BMNH 1935.9.1.134	ZM	1,43	196,5	not tested
160	104	BMNH ZD 1932.9.1.123	ZM	1,47	101,3	not tested
161	113	ZMB MAM 27664	ZM	1,5	38,7	not tested
162	158	NMMZ 8832	ZM	1,49	134,4	not tested
163	301	Jeanette/1463	ZA	1,87	287,3	not tested
164	300	Hilda/1921	ZA	1,89	453,8	not tested
	396	Mahdi	SO	2,04	0,6	not tested
	438	Sheik Butti 15	SO	1,14	3,6	not tested
	386	Sleeky	TD	1,96	1,1	not tested
	385	Pumba	ET	1,29	1,3	not tested
	439	Zikomo	SO	1,82	1,1	not tested
	440	Zara	SO	7,74	0,4	not tested
	378	Sheba	NA	no	material	not tested
	373	Li	SO	6,83	1	not tested
	451	00-01C6-EC3E	SO	1,94	1,1	not tested
165	368	Ziggy	SO	1,94	108,7	not tested
	379	Dhaid	SD	3,06	2,2	not tested
	372	Scally	SO	2,19	2,9	not tested
166	449	00 01C8 2420	SO	1,91	30,5	not tested
167	383	Aisha	SO	1,87	91,1	not tested
168	376	Niswa	SD	1,89	47,5	not tested
169	450	00 01C7 2276	SO	1,87	289,4	not tested
	377	Abu	NA	1,66	5,8	not tested
170	371	Shakes	SO	1,86	21,4	not tested

171	375	Annie	SD	1,9	34	not tested
172	381	Olive	SO	1,86	27,5	not tested
173	369	Oliver	SO	1,97	42,2	not tested
174	374	Aziz	SO	1,97	38,5	not tested
	370	Marley	SO	2,45	9,8	not tested
175	395	Skebenga	SO	1,89	58,8	not tested
	394	Tokalosh	SO	2,11	22	not tested
	393	Sangoma	SO	2,05	22,4	not tested
176	391	Somali	SO	1,93	74	not tested
177	392	Goliath	SO	1,95	39,1	not tested
178	384	Gums	NA	1,9	79	not tested
179	388	00-01C7-0706	SO	1,96	57,1	not tested
180	390	00-01C7-F346	SO	1,88	159,2	not tested

Table A1: List of 196 extractions, listing running extraction number, lab code, individual sample name, country of origin, and Nano Drop, as well as Qubit results. Samples with no extraction number were not send for sequencing.

No.	Lab Code	Sample Name	Country of Origin	Final Sample Name
				for sets
13	7	RMNH.MAM.51700	Sudan (SD)	**SD3**
23	46	ID1873	Sudan or Somalia	**SDSO1**
33	98	BMNH ZD 1912.5.10.1	D.R. Congo (CD)	**CD1**
73	311	AJIran 08	Iran (IR)	**IR1**
76	314	AJIran 11 "Narita"	Iran (IR)	**IR2**
80	444	Aju DJI 3 P1	Djibouti (DJ)	**DJ1**
91	315	AJIran M1A	Iran (IR)	**IR3**
92	316	AJIran M2A	Iran (IR)	**IR4**
96	86	KM 14430	Namibia (NA)	**NM2**
143	151	NMMZ 55703	Zimbabwe (ZW)	**ZW1**
163	301	Jeanette/1463	South Africa (ZA)	**ZA1**
165	368	Ziggy	Somalia (SO)	**SO13**
166	449	00 01C8 2420	Somalia (SO)	**SO4**
167	383	Aisha	Somalia (SO)	**SO5**
168	376	Niswa	Sudan (SD)	**SD2**
169	450	00 01C7 2276	Somalia (SO)	**SO3**
170	371	Shakes	Somalia (SO)	**SO10**
171	375	Annie	Sudan (SD)	**SD1**
172	381	Olive	Somalia (SO)	**SO8**
173	369	Oliver	Somalia (SO)	**SO9**
174	374	Aziz	Somalia (SO)	**SO6**
175	395	Skebenga	Somalia (SO)	**SO11**
176	391	Somali	Somalia (SO)	**SO12**
177	392	Goliath	Somalia (SO)	**SO7**
178	384	Gums	Namibia (NA)	**NM1**
179	388	00-01C7-0706	Somalia (SO)	**SO1**
180	390	00-01C7-F346	Somalia (SO)	**SO2**

Table A2: List of all samples included in the final sets for analysis. For easier identification and visualisation in plots, samples were given a combination of country code and running number.

12.2 CITES

Samples were imported under the following CITES numbers:

- AT 16-E-0753
- 16SG006329CR
- 15JP001990/TE
- 11US761881/9
- AT 15-E-1769
- D79/DFF

Additionally, we exchanged samples between CITES certified institutes listed under the following numbers:

- Forschungsinstitut für Wildtierkunde und Ökologie, Vetmeduni Vienna, Österreich **AT031**
- Centro de Biologia Ambiental Faculdade de Ciências da Universidade de Lisboa, Portugal **PT004**
- Koret School Veterinary Medicine, Hebrew University of Jerusalem, Israel **IL002**
- Museo Civico di Storia Naturale Giacomo Doria, Genova, Italy **IT019**
- Natural History Museum "La Specola", Florence, Italy **IT008**
- Museum of Evolution, Uppsala, Sweden **SE011**
- Zoologisches Museum der Universität Zürich, Switzerland **CH005**
- Natural History Museum of Geneva, Switzerland **CH004**
- Naturhistorisches Museum Basel, Switzerland **CH002**
- Naturmuseum Sankt Gallen, Switzerland **CH033**
- Field Museum of Natural History, Chicago, USA **US012**
- Natural History Museum, London, England **GB001**
- Botswana National Museum, Gaborone, Botswana **BW004**
- Tel Aviv University, Israel **IL001**
- Museum National d'Histoire Naturelle, Paris, France **FR75A**
- Natural History Museum Denmark, Copenhagen, Denmark **DK003**
- Kwa-Zulu Natal Museum, Pietermaritzburg, South Africa **ZA025**
- Amathole Museum, King William's Town, South Africa **ZA020**

12.3 Code

```
# Cheetah ddRAD data analysis

================================================================================

raw data preparation

================================================================================

## setup directories

mkdr /genetics/Burger/cheetah-ddrad-all/

## download the demulitiplexed data

cd /genetics/Burger/cheetah-ddrad-all/

wget -r -nH -nd -np --user pamela_burger \

--password 6NkA57n7kHAVkrFB https://support.igatech.it/sequences-export/634-ddRAD_Dolichopis_Burger_Vienna/delivery_01122017/

# get cheetah genome assembly

# note this step has already been done

cd /genetics/Burger/DROM_ddRAD/

wget ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/443/585/GCF_001443585.1_aciJub1/GCF_001443585.1_aciJub1_genomic.fna.gz

## make BWA index

/usr/bin/bwa index -a bwtsw \

-p GCF_001443585.1_aciJub1_genomic.fna \

GCF_001443585.1_aciJub1_genomic.fna.gz

# rename reads

cd /genetics/Burger/cheetah-ddrad-all/

mkdir original

cp *.fq.gz /genetics/Burger/cheetah-ddrad-all/original/

mv [sample number].fq.gz [collection sample name].1.fq.gz

ls *.1.fq.gz |perl -pe "s/.1.fq.gz//g" > samples

cd /genetics/Burger/CHEETAH_ddRAD/rawfastq/

mkdir /genetics/Burger/CHEETAH_ddRAD/trimmedfastq/

================================================================================
```

Stacks & VCFtools

==========

#STACKS

#1.Trim cheetah ddrad reads2_parallel with process_radtags

trim reads and make sure they have the restriction enzyme sites

cd /genetics/Burger/cheetah-ddrad-all/

mkdir trimmedfastq

parallel --jobs 75 \

'process_radtags \

-i gzfastq \

-1 {}.1.fq.gz \

-2 {}.2.fq.gz \

-q -c -r \

--renz_1 sphI \

--renz_2 hindIII --retain_header \

-o trimmedfastq/' < samples

#BWA

#2.Map trimmed cheetah ddrad reads with bwa

cd /genetics/Burger/cheetah-ddrad-all/trimmedfastq/

while read i;do

/usr/bin/bwa mem -t 75 \

/genetics/Burger/CHEETAH_ddRAD/GCF_001443585.1_aciJub1_genomic.fna \

${i}.1.1.fq.gz \

${i}.2.2.fq.gz | samtools view -Sb -F 4 - > ../pstacks/${i}.bam

done < ../samples

#PStacks

#3.Process mapped reads with pstacks

cd /genetics/Burger/cheetah-ddrad-all/pstacks/

x=1

while read i;do

```
pstacks -t bam -p 75 -f ${i}.bam -o ./ -i $x -m 3

x=$((x+1))

done < ../samples

#CStacks

#4.Make a catalog of stacks with cstacks -n 2

cd /genetics/Burger/cheetah-ddrad-all/pstacks/

ls *.bam | perl -pe "s/(.+).bam/\1\tONE/g" | grep -v
"CD_RMCA_454\|AO_AMNH_80618\|TM_ZIN_2808\|NA_ZMB_MAM_82181\|LY_NMW_ST
_554\|ZW_NMMZ_55706" > popmap.txt

cstacks -p 75 -n 2 -b 2 \

-M ./popmap.txt \

-P ./

#SStacks

#5.Compare catalog to individuals -n 2

cd /genetics/Burger/cheetah-ddrad-all/pstacks/

sstacks -p 75 -b 2 \

-M ./popmap.txt \

-P ./

#6.Create vcf file -n 2

cd /genetics/Burger/cheetah-ddrad-all/pstacks/

# b = Batch ID to examine when exporting from the catalog

# P = path to the directory containing the Stacks files

# p = minimum number of populations a locus must be present in to process a locus

# m = minimum stack depth required for individuals at a locus

# r = minimum percentage of individuals in a population required to process a locus for that
population

# --fasta = output fasta sequences

# --vcf = output variant call format file

populations -t 75 -P ./ -b 2 -p 1 \

--write_random_snp \

--max_obs_het 0.5 \

-M popmap.txt -m 6 -r 0 --vcf --fasta
```

```
#VCFTools

#7.VCFtools -n 2 cheetahs and outgroups

cd /genetics/Burger/cheetah-ddrad-all/pstacks/

/usr/local/bin/vcftools --min-alleles 2 --max-alleles 2 --vcf batch_2.vcf --recode --out
batch_2

/usr/local/bin/vcftools --vcf batch_2.recode.vcf --missing-indv

awk '$5>0.95' FS='\t' OFS='\t' out.imiss|cut -f 1 > remove.txt

/usr/local/bin/vcftools --vcf batch_2.recode.vcf --remove remove.txt --recode --max-missing
0.75 --out CHEETAH_and_outgroups_0.95_missingness

/usr/local/bin/vcftools --vcf CHEETAH_and_outgroups_0.95_missingness.recode.vcf --plink

mv out.ped CHEETAH_and_outgroups_0.95_missingness.ped

mv out.map CHEETAH_and_outgroups_0.95_missingness.map

#7.VCFtools -n 2 only cheetahs

cd /genetics/Burger/cheetah-ddrad-all/pstacks/

/usr/local/bin/vcftools --min-alleles 2 --max-alleles 2 --vcf batch_2.vcf --recode --out
batch_2

/usr/local/bin/vcftools --vcf batch_2.recode.vcf --missing-indv

awk '$5>0.75' FS='\t' OFS='\t' out.imiss|cut -f 1 > remove.txt

echo 'Lynx' >> remove.txt

echo 'CO_3_L' >> remove.txt

/usr/local/bin/vcftools --vcf batch_2.recode.vcf --remove remove.txt --recode --max-missing
0.75 --out CHEETAHs_only_0.75_missingness

/usr/local/bin/vcftools --vcf CHEETAHs_only_0.75_missingness.recode.vcf --plink

mv out.ped CHEETAHs_only_0.75_missingness.ped

mv out.map CHEETAHs_only_0.75_missingness.map

================================================================================

================================================================================

Filtering

================================================================================

#FILTER THE PED AND MAP FILE (OUTPUT FROM VCFtools)

#FILTERING STEPS IN PLINK

#CHEETAH_PUMA
```

```
#./plink --file /Applications/plink_mac/Cheetah-ddrad-raw-snps --mind 0.999 --make-bed
--recode --out /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999

#PLINK v1.90b5 64-bit (14 Nov 2017)           www.cog-genomics.org/plink/1.9/

##  (C) 2005-2017 Shaun Purcell, Christopher Chang   GNU General Public License v3

#Logging to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.log.

#Options in effect:

#  --file /Applications/plink_mac/Cheetah-ddrad-raw-snps

#--make-bed

#--mind 0.999

#--out /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999

#--recode

#16384 MB RAM detected; reserving 8192 MB for main workspace.

#.ped scan complete (for binary autoconversion).

#Performing single-pass .bed write (50316 variants, 90 people).

#--file: /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999-temporary.bed +

#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999-temporary.bim +

#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999-temporary.fam written.

#50316 variants loaded from .bim file.

#90 people (0 males, 0 females, 90 ambiguous) loaded from .fam.

#Ambiguous sex IDs written to

#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.nosex .

#59 people removed due to missing genotype data (--mind).

#IDs written to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.irem .

#Using 1 thread (no multithreaded calculations invoked).

#Before main variant filters, 31 founders and 0 nonfounders present.

#Calculating allele frequencies... done.

#Total genotyping rate in remaining samples is 0.505185.

#50316 variants and 31 people pass filters and QC.

#Note: No phenotypes present.

#--make-bed to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.bed +

#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.bim +
```

```
#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.fam ... done.
#--recode ped to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.ped +
#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999.map ... done.
#PLINK v1.90b5 64-bit (14 Nov 2017)           www.cog-genomics.org/plink/1.9/
#  (C) 2005-2017 Shaun Purcell, Christopher Chang   GNU General Public License v3
#Logging to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25.log.
#Options in effect:
#--file /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999
#--geno 0.25
#--recode
#--out /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25
#16384 MB RAM detected; reserving 8192 MB for main workspace.
#.ped scan complete (for binary autoconversion).
#Performing single-pass .bed write (50316 variants, 31 people).
#--file:
#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25-temporary.bed +
#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25-temporary.bim +
#  /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25-temporary.fam
#written.
#50316 variants loaded from .bim file.
#31 people (0 males, 0 females, 31 ambiguous) loaded from .fam.
#Ambiguous sex IDs written to
#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25.nosex .
#Using 1 thread (no multithreaded calculations invoked).
#Before main variant filters, 31 founders and 0 nonfounders present.
#Calculating allele frequencies... done.
#Total genotyping rate is 0.505185.
#49432 variants removed due to missing genotype data (--geno).
#884 variants and 31 people pass filters and QC.
#Note: No phenotypes present.
```

#--make-bed to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25.bed

#+ /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25.bim +

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25.fam ... done.

#--recode ped to

#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25.ped +

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25.map ... done.

#pamela@PamelaiMac:/Applications/plink_mac$ ./plink --file /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25 --maf 0.01 --make-bed --recode --out /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001

#PLINK v1.90b5 64-bit (14 Nov 2017) www.cog-genomics.org/plink/1.9/

(C) 2005-2017 Shaun Purcell, Christopher Chang GNU General Public License v3

#Logging to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001.log.

#Options in effect:

#--file /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25

#--maf 0.01

#--recode

#--out /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001

#16384 MB RAM detected; reserving 8192 MB for main workspace.

#.ped scan complete (for binary autoconversion).

#Performing single-pass .bed write (884 variants, 31 people).

#--file:

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001-temporary.bed

#+

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001-temporary.bim

#+

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001-temporary.fam

#written.

##884 variants loaded from .bim file.

##31 people (0 males, 0 females, 31 ambiguous) loaded from .fam.

#Ambiguous sex IDs written to

#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001.nosex .

#Using 1 thread (no multithreaded calculations invoked).

#Before main variant filters, 31 founders and 0 nonfounders present.

#Calculating allele frequencies... done.

#Total genotyping rate is 0.783426.

#0 variants removed due to minor allele threshold(s)

#(--maf/--max-maf/--mac/--max-mac).

#884 variants and 31 people pass filters and QC.

#Note: No phenotypes present.

#--make-bed to

#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001.bed +

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001.bim +

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001.fam ...

#done.

#--recode ped to

#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001.ped +

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001.map ...

#done.

#./plink --file /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50 --maf 0.01 --make-bed --recode --out /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001

#PLINK v1.90b5 64-bit (14 Nov 2017) www.cog-genomics.org/plink/1.9/

(C) 2005-2017 Shaun Purcell, Christopher Chang GNU General Public License v3

#Logging to /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001.log.

#Options in effect:

#--file /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50

#--maf 0.01

#--recode

#--out /Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001

#16384 MB RAM detected; reserving 8192 MB for main workspace.

#.ped scan complete (for binary autoconversion).

#Performing single-pass .bed write (884 variants, 23 people).

#--file:

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001-temporary.bed

#+

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001-temporary.bim

#+

/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001-temporary.fam

#written.

#884 variants loaded from .bim file.

#23 people (0 males, 0 females, 23 ambiguous) loaded from .fam.

#Ambiguous sex IDs written to

#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001.nosex

#.

#Using 1 thread (no multithreaded calculations invoked).

#Before main variant filters, 23 founders and 0 nonfounders present.

#Calculating allele frequencies... done.

#Total genotyping rate is 0.972359.

#3 variants removed due to minor allele threshold(s)

#(--maf/--max-maf/--mac/--max-mac).

#881 variants and 23 people pass filters and QC.

#Note: No phenotypes present.

#--make-bed to

#/Applications/plink_mac/Cheetah-ddrad-raw-snps_mind999_geno25_maf001_mind50_maf001.bed

```
#+

#  /Applications/plink_mac/Cheetah-ddrad-raw-
snps_mind999_geno25_maf001_mind50_maf001.bim

#+

#  /Applications/plink_mac/Cheetah-ddrad-raw-
snps_mind999_geno25_maf001_mind50_maf001.fam

#... done.

#--recode ped to

#/Applications/plink_mac/Cheetah-ddrad-raw-
snps_mind999_geno25_maf001_mind50_maf001.ped

#+

#  /Applications/plink_mac/Cheetah-ddrad-raw-
snps_mind999_geno25_maf001_mind50_maf001.map

#... done.

#*******************************************************

#CHEETAH DATA SET NEW FILTERING WITHOUT PUMA AND LYNX

#*******************************************************

#plink --file Cheetah_only-ddrad-raw-snps --mind 0.999 --recode --out
Cheetah_only_mind999

PLINK v1.90b5 64-bit (14 Nov 2017)           www.cog-genomics.org/plink/1.9/

  (C) 2005-2017 Shaun Purcell, Christopher Chang   GNU General Public License v3

Logging to Cheetah_only_mind999.log.

Options in effect:

  --file Cheetah_only-ddrad-raw-snps

--recode

--mind 0.999

--out Cheetah_only_mind999

16384 MB RAM detected; reserving 8192 MB for main workspace.

.ped scan complete (for binary autoconversion).

Performing single-pass .bed write (50316 variants, 88 people).

--file: Cheetah_only_mind999-temporary.bed + Cheetah_only_mind999-temporary.bim

+ Cheetah_only_mind999-temporary.fam written.
```

#50316 variants loaded from .bim file.

#88 people (0 males, 0 females, 88 ambiguous) loaded from .fam.

Ambiguous sex IDs written to Cheetah_only_mind999.nosex .

#59 people removed due to missing genotype data (--mind).

IDs written to Cheetah_only_mind999.irem .

Using 1 thread (no multithreaded calculations invoked).

Before main variant filters, 29 founders and 0 nonfounders present.

Calculating allele frequencies... done.

#Total genotyping rate in remaining samples is 0.505764.

#50316 variants and 29 people pass filters and QC.

Note: No phenotypes present.

--make-bed to Cheetah_only_mind999.bed + Cheetah_only_mind999.bim +

Cheetah_only_mind999.fam ... done.

--recode ped to Cheetah_only_mind999.ped + Cheetah_only_mind999.map ... done.

#---

#plink --file Cheetah_only_mind999 --geno 0.25 --recode --out
Cheetah_only_mind999_geno25

PLINK v1.90b5 64-bit (14 Nov 2017) www.cog-genomics.org/plink/1.9/

(C) 2005-2017 Shaun Purcell, Christopher Chang GNU General Public License v3

Logging to Cheetah_only_mind999_geno25.log.

Options in effect:

--file Cheetah_only_mind999

--geno 0.25

--recode

--out Cheetah_only_mind999_geno25

16384 MB RAM detected; reserving 8192 MB for main workspace.

.ped scan complete (for binary autoconversion).

Performing single-pass .bed write (50316 variants, 29 people).

--file: Cheetah_only_mind999_geno25-temporary.bed +

Cheetah_only_mind999_geno25-temporary.bim +

Cheetah_only_mind999_geno25-temporary.fam written.

#50316 variants loaded from .bim file.

#29 people (0 males, 0 females, 29 ambiguous) loaded from .fam.

Ambiguous sex IDs written to Cheetah_only_mind999_geno25.nosex .

Using 1 thread (no multithreaded calculations invoked).

Before main variant filters, 29 founders and 0 nonfounders present.

Calculating allele frequencies... done.

#Total genotyping rate is 0.505764.

#48939 variants removed due to missing genotype data (--geno).

#1377 variants and 29 people pass filters and QC.

Note: No phenotypes present.

--make-bed to Cheetah_only_mind999_geno25.bed + Cheetah_only_mind999_geno25.bim

+ Cheetah_only_mind999_geno25.fam ... done.

--recode ped to Cheetah_only_mind999_geno25.ped +

Cheetah_only_mind999_geno25.map ... done.

#---

#plink --file Cheetah_only_mind999_geno25 --maf 0.01 --recode --out
Cheetah_only_mind999_geno25_maf001

PLINK v1.90b5 64-bit (14 Nov 2017) www.cog-genomics.org/plink/1.9/

(C) 2005-2017 Shaun Purcell, Christopher Chang GNU General Public License v3

Logging to Cheetah_only_mind999_geno25_maf001.log.

Options in effect:

--file Cheetah_only_mind999_geno25

--maf 0.01

--recode

--out Cheetah_only_mind999_geno25_maf001

16384 MB RAM detected; reserving 8192 MB for main workspace.

.ped scan complete (for binary autoconversion).

Performing single-pass .bed write (1377 variants, 29 people).

--file: Cheetah_only_mind999_geno25_maf001-temporary.bed +

Cheetah_only_mind999_geno25_maf001-temporary.bim +

Cheetah_only_mind999_geno25_maf001-temporary.fam written.

#1377 variants loaded from .bim file.

#29 people (0 males, 0 females, 29 ambiguous) loaded from .fam.

Ambiguous sex IDs written to Cheetah_only_mind999_geno25_maf001.nosex .

Using 1 thread (no multithreaded calculations invoked).

Before main variant filters, 29 founders and 0 nonfounders present.

Calculating allele frequencies... done.

#Total genotyping rate is 0.767586.

#1170 variants removed due to minor allele threshold(s)

(--maf/--max-maf/--mac/--max-mac).

#207 variants and 29 people pass filters and QC.

Note: No phenotypes present.

--make-bed to Cheetah_only_mind999_geno25_maf001.bed +

Cheetah_only_mind999_geno25_maf001.bim +
Cheetah_only_mind999_geno25_maf001.fam

... done.

--recode ped to Cheetah_only_mind999_geno25_maf001.ped +

Cheetah_only_mind999_geno25_maf001.map ... done.

plink --file Cheetah_only_mind999_geno25_maf001 --mind 0.5 --recode --out
Cheetah_only_mind999_geno25_maf001_mind50

PLINK v1.90b5 64-bit (14 Nov 2017) www.cog-genomics.org/plink/1.9/

(C) 2005-2017 Shaun Purcell, Christopher Chang GNU General Public License v3

Logging to Cheetah_only_mind999_geno25_maf001_mind50.log.

Options in effect:

--file Cheetah_only_mind999_geno25_maf001

--recode

--mind 0.5

--out Cheetah_only_mind999_geno25_maf001_mind50

16384 MB RAM detected; reserving 8192 MB for main workspace.

.ped scan complete (for binary autoconversion).

Performing single-pass .bed write (207 variants, 29 people).

--file: Cheetah_only_mind999_geno25_maf001_mind50-temporary.bed +

```
  Cheetah_only_mind999_geno25_maf001_mind50-temporary.bim +
  Cheetah_only_mind999_geno25_maf001_mind50-temporary.fam written.
#207 variants loaded from .bim file.
29 people (0 males, 0 females, 29 ambiguous) loaded from .fam.
Ambiguous sex IDs written to Cheetah_only_mind999_geno25_maf001_mind50.nosex .
7 people removed due to missing genotype data (--mind).
IDs written to Cheetah_only_mind999_geno25_maf001_mind50.irem .
Using 1 thread (no multithreaded calculations invoked).
Before main variant filters, 22 founders and 0 nonfounders present.
Calculating allele frequencies... done.
#Total genotyping rate in remaining samples is 0.970575.
#207 variants and 22 people pass filters and QC.
Note: No phenotypes present.
--make-bed to Cheetah_only_mind999_geno25_maf001_mind50.bed +
  Cheetah_only_mind999_geno25_maf001_mind50.bim +
  Cheetah_only_mind999_geno25_maf001_mind50.fam ... done.
--recode ped to Cheetah_only_mind999_geno25_maf001_mind50.ped +
  Cheetah_only_mind999_geno25_maf001_mind50.map ... done.
#=========================================================================
#MAKE ADEGENET INPUT FILE
#/plink --file Cheetah_only_mind999_geno25_maf001 --mind 0.5 --recodeA --out
Cheetah27_207SNP #gives a .raw file
library("devtools")
install_github("thibautjombart/adegenet")
library("adegenet")
install.packages("pegas")
library(pegas)
install.packages("hierfstat")
library(hierfstat)
#import the file into R ADEGENET
```

```
CH <- read.PLINK("/Applications/plink_mac/Cheetah27_207SNP.raw") # imports the file and converts it into a genlight object

CH

#check the file

CH

#Make a PCA

pca_CH <- glPca(CH)

#Select the number of axes:

#  6

pca_CH

#Plot the PCA

scatter(pca_CH, posi="topleft")

pca_CH_colorplot <-colorplot(pca_CH$scores,pca_CH$scores, transp=TRUE, cex=4)

abline(h=0,v=0, col="grey")

add.scatter.eig(pca_CH$eig[1:10],2,1,2, posi="topright", inset=.05, ratio=.3)

s.label(pca_CH$scores)

s.class(pca_CH$scores, fac=pop(CH), col=funky(15))#plots the samples with lines and circles

s.class(pca_CH$scores, fac=pop(CH), col=transp(funky(2),.6), axesel=FALSE, cstar=0, cpoint=3)#plots the samples as dots

add.scatter.eig(pca_CH$eig[1:10],3,2,3, posi="topright", ratio=.3) #adds the Eigenvalues to the plot

#Discriminant analysis of prinical components

grpCH <- find.clusters(CH)

Choose the number PCs to retain (>=1): 20

dapc_CH <- dapc(grpCH) # n.pca=5, n.da=1)

==========================================================================

further analyses

==========================================================================

#Check for relatedness in the samples using PLINK IBS matrix

/plink --file  Cheetah_only_21_mind999_geno25_maf001_mind50 --cluster --matrix

#CREATE A GENETIC DISTANCE MATRIX / Identity by state IBS

#for modifying nexus file to open in splitstree
```

plink --file Cheetah_only_21_mind999_geno25_maf001_mind50 --distance-matrix --out Cheetah_only_21_mind999_geno25_maf001_mind50

#for plink 1.9! doesn´t work in 1.7

#input is ped file

#RUN SPLITSTREE FOR A NEIGHBOR NETWORK ANALYSIS

#1) Create a 1-IBS distance matrix in PLINK

cd Plink folder

/plink --file Cheetah_only_21_mind999_geno25_maf001_mind50 --cluster --distance-matrix --out Cheetah_only_21_mind999_geno25_maf001_mind50

#use IBS.mibs file

#2) Use the PED file to convert it into a NEXUS file in PGDspider

#launch the GUI with java -Xmx10g -Xms512m -jar PGDSpider2.jar

#3) Copy-paste the distance matrix and the correct headers into the nexus file; the headers are found in the example folder of splitstree for an input file using a distance matrix

#BEGIN DISTANCES;

#DIMENSIONS ntax=21;

#FORMAT labels=left diagonal triangle=both;

#MATRIX

#Admixture

#Make bed/bim/fam in plink

plink --file Cheetah_only_21_mind999_geno25_maf001_mind50 --make-bed --out Cheetah_only_21_mind999_geno25_maf001_mind50

#Convert into admixture input file, use bed as input

plink --bfile Cheetah_only_21_mind999_geno25_maf001_mind50 --recode12 --out Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in

#run:

Admixture /genetics/Burger/cheetah-ddrad-all/Rundeck_Feb/Admix_In/Cheetah21_207SNP/Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in.ped 2 | tee /genetics/Burger/cheetah-ddrad-all/Rundeck_Feb/Admix_In/Cheetah21_207SNP/Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in

#GET THE CV errors

burgerp@PamelaBurger:/Applications/admixture_macosx-1.3.0$ grep -h CV Cheetah_only_21_Admix_log*.out > Cheetah_only_21_Admix_CVerrors.txt

#plot in R

#import the Q tables

tblK2 <- read.table("/Users/burgerp/Dropbox/MHC_CHEETAH/hier_pamela_schau/Admixture_results/Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in.2.Q")

tblK3 <- read.table("/Users/burgerp/Dropbox/MHC_CHEETAH/hier_pamela_schau/Admixture_results/Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in.3.Q")

tblK4 <- read.table("/Users/burgerp/Dropbox/MHC_CHEETAH/hier_pamela_schau/Admixture_results/Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in.4.Q")

tblK5 <- read.table("/Users/burgerp/Dropbox/MHC_CHEETAH/hier_pamela_schau/Admixture_results/Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in.5.Q")

#import the ped file for the names list

names <- read.table("/Users/burgerp/Dropbox/MHC_CHEETAH/hier_pamela_schau/Admixture_results/Cheetah_only_21_mind999_geno25_maf001_mind50_Admix_in.ped", sep=" ") [,2]

names2 <- substr(as.character(names), 1,2)

#plot and add the axis to each barplot

par(mfrow=c(4,1))

barplot(t(as.matrix(tblK2)), col=rainbow(2), border=NA)

axis(1, at = seq(0.6,25.2, 1.2), labels = names, las=2, tck=0, col="white", cex=0.6) # calculate #individuals times wanted text width 1.2

barplot(t(as.matrix(tblK3)), col=rainbow(3), border=NA)

axis(1, at = seq(0.6,25.2, 1.2), labels = names, las=2, tck=0, col="white", cex=0.6)

barplot(t(as.matrix(tblK4)), col=rainbow(4), border=NA)

axis(1, at = seq(0.6,25.2, 1.2), labels = names, las=2, tck=0, col="white", cex=0.6)

barplot(t(as.matrix(tblK5)), col=rainbow(5), border=NA)

axis(1, at = seq(0.6,25.2, 1.2), labels = names, las=2, tck=0, col="white", cex=0.6)

==========

R Studio

==========

```
#CONVERT PED AND MAP FILE INTO A .RAW FILE IN PLINK

plink --file

#open the Rpackage ADEGENET

install.packages('adegenet')

library("adegenet")

install.packages("pegas")

library("pegas")

#LOAD RAW FILW INTO R

CHEETAH <-
read.PLINK("Cheetah_only_21_mind999_geno25_maf001_mind50_Cluster.raw",
parallel=FALSE)

#MAKE PCA

pca_CHEETAH <- glPca(CHEETAH)

#Select number of axes:

# min2

# see loadings (X,Y coordinates)

pca_CHEETAH$scores

pca_CHEETAH

#MAKE SCATTERPLOT

scatter(pca_CHEETAH, posi="topleft")

#Colorplot

colorplot(pca_CHEETAH$scores, pca_CHEETAH$scores, axes=NULL, transp=TRUE, cex=3,
xlab="PC 1", ylab="PC 2")

#add title

title("PCA of CHEETAH\naxes 1-2")

abline(v=0,h=0,col="grey", lty=2)

add.scatter.eig(pca_CD_NG$eig[1:40],2,1,2, posi="topleft", inset=.05, ratio=.3)

textxy(pca_CHEETAH$scores, pca_CHEETAH$scores, pca_CHEETAH$IID)

# use the text function to add labels to the positions given by the coordinates you used in
plot

s.label(dfxy = df, xax=1, yax=2, label=noms, clabel=0.7, boxes=FALSE, grid=FALSE,
addaxes=FALSE)
```

```
#safe colorplot as object to input for NJ colouring

pca_COL <-colorplot(pca_CHEETAH$scores, pca_CHEETAH$scores, transp=TRUE, cex=3, xlab="PC 1", ylab="PC 2")

#MAKE NJ TREE

tre <- nj(dist(as.matrix(CHEETAH)))

plot(tre, typ="fan", cex=0.8)

#use "MYcol" to assign colours one by one, then make "col=MYcol" in next command line

MYcol <- c("green", "green", "green", "green", "yellow", "red", "red","red", "red", "red", "red", "red", "red", "red", "red", "red", "red", "red", "red", "red", "red", "yellow")

#or assign colours from eg colorplot, save, and put name in "col=name"

tiplabels(pch=20, col=pca_COL, cex=4)

#"col="filename colorplot"" imports colours from saved colorplot

#"typ" changes tree style, possible: phylogram, cladogram, fan, unrooted

# "CHEETAH$ind.names" to see list of ID+name, to assign colours etc

# "pch" colour symbol format

#assign selected colours to match colorplot

MYcol <- c("darkturquoise", "orchid1", "slateblue1")

#s.class plot

s.class(pca_CHEETAH$scores, fac=pop(CHEETAH), xax=2,yax=3, col=MYcol)

#MAKE DAPC

#First, find clusters in your data set

DCheetah <- tab(CHEETAH, freq=TRUE, NA.method="mean")

grp <- find.clusters(DCheetah)

dapc1 <- dapc(DCheetah, grp$grp)

dapc1

scatter(dapc1)

scatter(dapc1, posi.da="topleft")

#MAKE A COMPOPLOT

?compoplot

compoplot(dapc1, col.pal = funky, posi="topleft", txt.leg=paste("Cluster", 1:2), lab="", ncol=1, xlab="individuals") #col=funky(2)

names <- read.table("Cheetah_only_21_mind999_geno25_maf001_mind50.ped", sep=" ") [,2]
```

```
#plot and add the axis to each barplot

par(mfrow=c(1,1))

barplot(t(as.matrix(tblK2)), col=rainbow(1), border=NA)

axis(1, at = seq(0.6,25.2, 1.2), labels = names, las=2, tck=0, col="white", cex=0.6)

#USE GENEPOP FILE

CHEETAH.gen <-
read.genepop("C:/Users/zumbroichj/Dropbox/Thesis/Analyses/CHEETAHs_only_75_missi
ngness_genpop.gen", ncode = 3L)

#CALCULATE OBSERVED VS EXPECTED HETEROZYGOSITY

#with adegenet

Cheetah21_GenInd <- read.genepop("Cheetah21_Genepop_In.gen", ncode = 3L)

sumCheetah21 <- summary(Cheetah21_GenInd)

sumCheetah21

names(sumCheetah21)

#[1] "n"          "n.by.pop"   "loc.n.all" "pop.n.all" "NA.perc"    "Hobs"       "Hexp"

calculate mean Hobs/Hexp and standard deviation over all loci in adegenet

Hobs.mean <- mean(sumCheetah21$Hobs, na.rm = TRUE)

Hobs.mean

Hexp.mean <- mean(sumCheetah21$Hexp, na.rm = TRUE)

Hexp.mean

sd(sumCheetah21$Hobs)

sd(sumCheetah21$Hexp)

Cheetah21_Asom_GenInd <- read.genepop("Cheetah21_Asom_Genepop_In.gen", ncode = 3L)

sumCheetah21_Asom <- summary(Cheetah21_Asom_GenInd)

sumCheetah21_Asom

names(sumCheetah21_Asom)

#[1] "n"          "n.by.pop"   "loc.n.all" "pop.n.all" "NA.perc"    "Hobs"       "Hexp"

Hobs.mean <- mean(sumCheetah21_Asom$Hobs, na.rm = TRUE)

Hobs.mean

Hexp.mean <- mean(sumCheetah21_Asom$Hexp, na.rm = TRUE)

Hexp.mean
```

```
sd(sumCheetah21_Asom$Hobs)

sd(sumCheetah21_Asom$Hexp)

#test, if Ho is significantly different from He

t.test(sumCheetah21$Hexp,sumCheetah21$Hobs,pair=T,var.equal=TRUE,alter="greater")

#CALCULATE INBREEDING COEFFICIENT

plink --file Cheetah21 --het – outfile Cheetah21_IC.het

Cheetah21_IC.het <- read.table("Cheetah21_IC.het", sep = "", header=TRUE)

names(Cheetah21_IC.het)

FisCheetah21_IC_mean <- mean(Cheetah21_IC.het$FIS)

FisCheetah21_IC_mean

#[1] 0.2463667

Cheetah21_IC_Asom.het <- read.table("Cheetah21_IC_Asom.het", sep = "", header=TRUE)

names(Cheetah21_IC_Asom.het)

FisCheetah21_IC_Asom_mean <- mean(Cheetah21_IC_Asom.het$FIS)

FisCheetah21_IC_Asom_mean

#[1] 0.2225662

#TEST SIGNIFICANCE for the FIS values using t.test

t.test(Cheetah21_IC.het$FIS, y = NULL, alternative = "two.sided", paired = FALSE, var.equal
= FALSE, conf.level = 0.95)

#p-value = 1.23e-05

t.test(Cheetah21_IC_Asom.het$FIS, y = NULL, alternative = "two.sided", paired = FALSE,
var.equal = FALSE, conf.level = 0.95)

#p.value=9.389e-05 - significant!
```