MASTERARBEIT | MASTER'S THESIS

Titel | Title

SNP typing in cheetahs (Acinonyx jubatus) using Oxford Nanopore Technologies (MinION)

verfasst von | submitted by

Laura Marie Plazer BSc

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

Master of Science (MSc)

Wien | Vienna, 2025

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

Acknowledgements

I would like to sincerely thank everyone who contributed to the completion of this thesis. Their support was invaluable, and this work would not have been possible without them.

Firstly, I would like to thank my (unofficial) thesis advisor Dr. Stefan Prost of the University of Oulu who helped me a lot to gain a better understanding of the subject matter since I wasn't versed in genetical research and laboratory work. Even though he pursued his career in Finland he made sure to always be available to answer my questions in any way possible.

I also owe many thanks to Priv. Doz. Dr. Elisabeth Haring who especially helped me during the writing process for this thesis. Through her guidance I was able to learn many important things concerning thesis writing. Her questions concerning my work also helped me gain a better understanding of the material and stimulated my thought process to help me express myself more understandably.

Moreover, I would like to thank everybody at the Central Research Laboratory (ZFL) of the Natural History Museum Vienna. Without the help of Julia Schindelar and Alexandra Wanka who showed me the way around the laboratory I would still be struggling to calculate primer concentrations. Even though my visits in the laboratory, were few and far between everyone always gave me a warm welcome and lent a helping hand, either while pipetting or calculating annealing temperatures, no matter how often I came by to conduct just one more PCR.

Finally, I want to thank my friends, family and everyone that helped me through one of the many tough spots I was in while conducting this thesis. I am very grateful for every support I got while trying to juggle too many things at the same time.

Thank you.

Abstract

Large carnivores tend to be threatened by a multitude of factors including human wildlife conflict, illegal wildlife trade, habitat loss and degradation. Their loss can have significant impacts on the structure and function of ecosystems. New technologies like high-throughput DNA sequencing (HTS) platforms, however, can aid in their conservation efforts. Genomic analyses can provide us with information needed to evaluate species and subspecies to develop efficient conservation strategies. However, biodiversity rich areas often lack laboratory infrastructure therefore portable and affordable devices like the Oxford Nanopore Technologies' (ONT) MinION sequencer can make a big difference.

The main focus of this work lays in the use of the MinION's Short Fragment Mode (SFM), which enables the sequencing of very small sequences of less than 200 bp. This study used a set of SNPs previously designed for threatened cheetahs (*Acinonyx jubatus*), with amplicon lengths between 63 bp and 92 bp. Questions concerning the reliability of this SFM in light of the devices error rate of 1 % - 5 % as well as the applicability while using multiple cheetah SNP primers and potentially degenerated DNA of museums material are addressed.

Results show that the MinION sequencer and SFM can be used to reliably call SNP sites, even using potentially degenerated DNA. Despite the MinION's error rate, reliable SNP calling was possible, opening the path for the devices application for research in situ.

However, results also show that further research concerning annealing temperatures of the used SNP primers during multiplexing should be conducted for optimal amplification results.

Zusammenfassung

Große Raubtiere sind oft durch eine Vielzahl von Faktoren bedroht, darunter Konflikte zwischen Menschen und Wildtieren, illegaler Wildtierhandel, Lebensraumverlust und -degradierung. Ihr Verlust kann erhebliche Auswirkungen auf die Struktur und Funktion von Ökosystemen haben. Neue Technologien wie high-throughput DNA-Sequenzierung (HTS) können jedoch zu ihren Schutzmaßnahmen beitragen. Genetische Analysen liefern wichtige Informationen zur Bewertung von Arten und Unterarten, um effiziente Schutzstrategien entwickeln zu können.

Oft verfügen Biodiversitäts-Hotspots jedoch nicht über die erforderliche Laborinfrastruktur. Daher können tragbare und kostengünstige Geräte wie der MinION-Sequenzierer von Oxford Nanopore Technologies (ONT) einen entscheidenden Unterschied machen.

Der Schwerpunkt dieser Arbeit liegt auf der Nutzung des Short-Fragment-Mode (SFM) des MinION, der das Sequenzieren sehr kurzer Sequenzen von weniger als 200 bp ermöglicht. In dieser Studie wurde eine Reihe von zuvor entwickelten SNPs für bedrohte Geparden (*Acinonyx jubatus*) verwendet, diese besitzen Längen zwischen 63 bp und 92 bp.

Es wurden Fragen zur Zuverlässigkeit des SFM angesichts der Fehlerquote des Geräts von 1 % bis 5 % sowie zur Anwendbarkeit bei der Verwendung mehrerer Geparden-SNP-Primer und möglicherweise degradierter DNA aus Museumsmaterial untersucht.

Die Ergebnisse zeigen, dass der MinION-Sequenzierer und SFM zuverlässig SNP-Stellen identifizieren können, selbst bei potenziell degradierter DNA. Trotz der Fehlerrate des MinION war eine verlässliche SNP-Analyse möglich, was den Weg für den Einsatz des Geräts in der Feldforschung ebnet.

Dennoch weisen die Ergebnisse darauf hin, dass weitere Untersuchungen zu den Annealing-Temperaturen der verwendeten SNP-Primer bei Multiplexing erforderlich sind, um optimale Amplifikationsergebnisse zu erzielen.

Table of contents

Abbreviations

ANGSD	Analysis of next generation Sequencing Data
BC	barcode
bp	base pairs
BWA	Burrows-Wheeler Aligner
CO2	carbon dioxide
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
°C	degrees celsius
dNTPs	deoxynucleotide triphosphate
DNA	desoxyribonucleic acid
HTS	High-throughput DNA sequencing
IUCN	International Union for Conservation of Nature
kb	kilobase pairs
Mg	magnesium
µl	microliter
µM	micromolar
min	minutes
NCBI	Natural Center for Biotechnology Information
NHMW	Naturhistorisches Museum Wien (Natural History Museum Vienna)
NFW	nuclease-free water
ONT	Oxford Nanopore Technologies
PCR	polymerase chain reaction
sec	seconds
SNP	single nucleotide polymorphisms
km2	square kilometer
USB	Universal Serial Bus
V	volt

1 Introduction

1.1 Why do we care about large carnivores?

Within the class Mammalia, several representatives of large carnivores captivate through their strength and beauty, making them some of the most feared and admired animals. Their numbers, however, are rapidly declining because of threats including human wildlife conflict (Ripple et al., 2014), habitat loss and degradation (Purvis et al., 2000; Schmidt-Küntzel et al., 2024). These factors in combination with their tendency for slow life histories, low population densities and wide-ranging behaviour (Cardillo et al., 2004; Ripple et al., 2014) add to their vulnerability and extinction risk.

Within the order Carnivora we can find 245 terrestrial species that inhabit almost every major habitat in the world. The 31 largest carnivores among them (including only species with an average adult body mass of more than 15 kg) belong to the five families: Canidae, Felidae, Mustelidae, Ursidae and Hyaenidae (Ripple et al., 2014).

The loss of such large carnivores can have significant impacts on the structure and functions of ecosystems as well as the services they provide (Beschta & Ripple, 2009; Ripple et al., 2014; Díaz et al., 2019). Ripple et al. (2014) also mentioned positive economic effects that large carnivores can have by fuelling tourism through their iconic nature. While the effects on economic services can be beneficial the effect large carnivores have on ecosystems are manifold. They play crucial roles in structuring the ecosystem's food-web via limiting the number of herbivores present within an ecosystem through predation and limit the number of mesopredators through intraguild competition. Furthermore, by limiting the number of herbivores, plants are allowed to grow and store more CO_2 providing a buffering effect on climate change (Ripple et al., 2014). Therefore, the loss of species on high trophic levels can have devastating and cascading impacts on ecosystems (Ripple et al., 2014; Díaz et al., 2019).

1.2 The Cheetah *Acinonyx jubatus* (Schreber, 1775)

1.2.1 Population numbers and threats

The cheetah *Acinonyx jubatus* (Schreber, 1775) is a large carnivore and apex predator, which faces extreme challenges within the Anthropocene. As one of the most wide ranging carnivores

with home ranges that can span up to 3,000 km^2 (Marker et al., 2008; Durant et al., 2017) they face a plethora of threats including fragmentation and habitat loss, human wildlife conflict, prey loss by overhunting and wildmeat harvesting as well as illegal trade (Durant et al., 2017; Tricorache et al., 2018; Magliolo et al., 2021).

The global population of adult and adolescent cheetahs is estimated to include around 7,100 individuals that are spread across 33 populations (Durant et al., 2022). These populations only occur in about 9 % of their historical distribution range which shows the horrific effects that fragmentation can have on species distribution (Durant et al., 2017; Prost et al., 2022). The biggest cheetah population includes more than one-half of the world's cheetahs and spans across six southern African countries (Angola, Botswana, Mozambique, Namibia, South Africa, Zambia); only one more population includes over 1,000 individuals that are spread across Kenya and Tanzania, the other 31 populations include between 200 or less than 10 individuals. The Asiatic cheetah *A. j. venaticus* occupies only 2 % of its historical range and currently is found exclusively in Iran with one population containing less than 30 individuals (Durant et al., 2017).

The wide-ranging nature of cheetahs make conservation efforts difficult since most individuals will ultimately occur outside of protected areas (PA) (Durant et al., 2017). Human intolerance furthers the threat these animals face, cheetahs hunting on commercial ranches are often trapped and killed to protect livestock. The removal through such lethal actions caused by conflict control or by trophy hunting led to the annual death of 650-890 individuals from commercial ranches resulting in the loss of 9,588 cheetahs between 1978 and 1994 alone. More recent removals between 2008 and 2014 resulted in the extermination of 267 cheetahs on commercial farms in Namibia (Weise et al., 2015).

1.2.1 Taxonomy of *Acinonyx jubatus*

The International Union for Conservation of Nature (IUCN) Cat Specialist Group currently recognizes four subspecies of cheetah, namely *A. jubatus venaticus* (South-West Asia, currently only found in Iran), *A. jubatus hecki* (North-West Africa), *A. jubatus sommeringii* (North-East Africa) and *A. jubatus jubatus* (South and East Africa) (Kitchener et al., 2017). However, traditionally the cheetah was classified into five subspecies, including *A. jubatus raineyi* (East Africa) (Krausmann & Morales, 2005). This subspecies is currently treated as a synonym of *A. j. jubatus* because of their close genetic relationship based on mitochondrial DNA analyses

(O'Brien et al., 1987; Charruau et al., 2011; Prost et al., 2022). However, Prost et al. (2022) recently argued that these two subspecies should be split following the traditional subspecies classification as they show a high degree of genetic separation on the whole-genome level.

Also, the determination of intraspecific taxonomy can have major impacts on conservation efforts and should therefore be taken into consideration while mapping out conservation strategies. The cheetah is globally listed as *Vulnerable* by the IUCN (Durant et al., 2014; Weise et al., 2015; Prost et al., 2022). However, Durant and colleagues (2017) argued that the cheetah meets all IUCN criteria to be recategorized as *Endangered.* Indeed, of the four currently recognized subspecies only one is classified as *Vulnerable*, namely *A. j. jubatus*, while *A. j. soemmeringii* has recently been reclassified as *Endangered* (Durant et al., 2023) and *A. j. venaticus* and *A. j. hecki* are listed as *Critically Endangered* by the IUCN (Durant et al., 2014). Due to the synonymisation of *A. j. raineyi* with *A. j. jubatus*, a conservation status of the former is absent.

1.2.2 The importance of subspecies classification

The classification and status as a species or subspecies have major consequences concerning conservation status and efforts. Genetic analyses can help with establishing a taxonomic rank for a certain population as well as allowing species/subspecies identification. Within *A. jubatus*, a low genetic variability has been detected, which probably resulted from a bottleneck 10,000-12,000 years ago at the end of the Pleistocene (Charruau et al., 2011). This could explain the close relationship between *A. jubatus raineyi* and *A. j. jubatus* that has been reported by O'Brien et al. (1987) and Charruau et al. (2011) based on mitochondrial sequences, which had been an argument to synonymize *A. j. raineyi* (Kitchener et al., 2017). However, the question remained whether and to which extent hybridisation between subspecies occurs and whether nuclear genome data would deliver a different picture. Recent genomic analyses using genome-wide single nucleotide polymorphisms (SNPs) show that the East African cheetah (*A. j. raineyi*) is genetically quite separated form *A. j. jubatus* and indeed closer to *A. j. sommeringii,* which is also geographically the closest subspecies (Prost et al., 2022). This supports the division of *A. jubatus* into of five subspecies as postulated by Krausmann and Morales (2005), and to treat the East African cheetah (*A. j. raineyi*) as a separate subspecies.

1.2.3 Impact of illegal wildlife trade

Cheetahs are listed in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) since 1975. Species under Appendix I are not allowed to be traded commercially (with a few exceptions). Captive-bred cheetahs from commercial farms, however, are allowed to be traded by certain facilities registered by CITES (Nowell, 2014; Tricorache et al., 2018, 2021; Magliolo et al., 2021).

The breeding of cheetahs in captivity can be challenging (Nowell, 2014; Tricorache et al., 2018) and the amount of captive-bred individuals does not seem to satisfy the demand for trophy hunting, body parts for traditional medicine, as decorative ornaments or as pets (Tricorache et al., 2018). Therefore, many wild living individuals that roam outside of protected areas fall prey to illegal wildlife hunts driving the species further into extinction (Nowell, 2014; Durant et al., 2017; Tricorache et al., 2018). Even though the trade in cheetahs is monitored, the fact that cheetahs can be legally traded provides paths for illegal trade.

1.3 Genomic analyses and miniaturized laboratory equipment

As mentioned above, genomic analyses in cheetahs may provide information concerning taxonomic status of populations or subspecies as well as to evaluate their conservation status. Besides the possibility to investigate population structure in cheetah populations in general, genetic identification might also allow to identify the geographic origin of individuals confiscated in illegal trade.

The development of high-throughput DNA sequencing (HTS) technologies provided essential tools for the assessment and quantification of biodiversity on our planet and, consequently, for wildlife conservation. HTS platforms can produce millions of individual sequencing reads from several hundreds to thousands of samples concurrently (Vasiljevic et al., 2021). However, high costs, bulky equipment and expensive infrastructure are disadvantages of most HTS technology that result in a high dependency on external service providers. The development of "nanopore technology" by Oxford Nanopores Technologies' (ONTs; https://nanoporetech.com/) provided a recent technological innovation (Deamer et al., 2016). The MinION sequencing platform integrates the use of HTS in a portable, low-cost device capable of long read outputs and rapid real-time analysis (Jain et al., 2016).

Nanopore technology relies on protein pores, so called “nanopores”, that function as biosensors. These nanopores are embedded within an electrically resistant biological membrane that lies in an electrolyte solution. While this solution is fed by a constant voltage an ionic current is produced. Due to this current negatively charged single stranded DNA and RNA molecules are fed through the nanopores from the negatively charged *cis* pole to the positively charged *trans* pole (Wang et al., 2021). During library preparation a motor protein is added to the DNA or RNA strands. This protein unwinds the double-stranded DNA into single strands and controls the speed at which the strands are fed through. The changes in the ionic current that occur during this procedure refer to the nucleotide sequence of the strand and can be decoded with the use of computational algorithms (Wang et al., 2021).

Quick and easy to use library preparation protocols are readily available and a standard laptop computer can be used for data analysis through a single USB connection. Additionally, the MinION is able to generate ultra-long reads up to >50 kb and has a short fragment mode capable of reads with as little as 20 bp in length (*nanoporetech.com*, 2024)

Yet, apart from its many advantages, the platform is limited by its relatively higher error rate compared to other HTS platforms it currently lies between 1 % and 5 % for raw reads (Liu-Wei et al., 2024). However, regular updates in the sequencing chemistry as well as constant improvements to the nanopores cause the error rate to decline. Furthermore, most bioinformatic tools used for data analysis are specifically designed to handle the error rate of MinION, e.g., concerning struggles during sequencing of low complexity regions. Many studies (Pomerantz et al., 2018; Srivathsan et al., 2018; Krehenwinkel, Pomerantz, Henderson, et al., 2019; Seah et al., 2020; Vasiljevic et al., 2021) have shown that the data produced by the MinION sequencer is sufficiently accurate for the generation of reliable consensus sequences for species identification.

1.4 Aims and Questions

The main goal of this work was to test the use of ONTs MinION MK1B nanopore sequencer for molecular subspecies assignment in cheetahs (*Acinonyx jubatus*).

While the focus of this research question is especially concerned with the technical possibility of using miniaturized laboratory equipment, questions concerning the applicability of the data

produced by the MinION technology, for conservation and other areas will be addressed at the same time.

Questions addressed in the course of this work:

- Can the short fragment mode of the MinION reliably be used to call SNP sites from short amplicons?
- Does the MinION technology allow for reliable calling of SNP sites, despite the considerable error rate?
- Can the SNP typing with the MinION be used for old and potentially low-quality DNA of museum material?
- Can the MinION be a useful tool in the fight against illegal wildlife trade?
- Does the MinION have the potential to support local capacity building and aid the fight against wildlife crime within remote areas?

2 Material and Methods

2.1 Material

Altogether, 36 DNA samples of *Acinonyx jubatus* were analysed in the current study, provided by Stefan Prost (University of Oulu and Natural History Museum Vienna) and Pamela Burger (Vetmeduni, Vienna) and their teams. The samples comprised all subspecies of *Acinonyx jubatus* including *A. j. raineyi* (despite its subspecies status is disputed): *A. j. jubatus* (7), *A. j. hecki* (4), *A. j. raineyi* (11), *A. j. sommeringii* (5), and *A. j. venaticus* (9). The sample numbers as well as the subspecies assignments of all samples used are listed in Table 1. Volumes of DNA samples differed between samples from less than 2 µl to more than 20 µl. Sample concentration was measured before amplification with the exception of samples with less than 2 µl in volume to preserve enough DNA for amplification. Measurements were taken using the Qubit™ Fluorometer (Invitrogen) by mixing 1 µl of DNA with fluorescent dyes (that only emit signals when attached to DNA) before measurement.

Table 1:List of DNA samples used for SNP typing. too high: concentration of sample exceeded measurable concentration; too low: concentration of the sample was beneath a measurable threshold; samples without concentration values contained less than 2 µl and were not measured.

Sample number	Subspecies	DNA concentration (ng/µl)	Origin
4362	A. *j. venaticus*	too high	Iran
12070	*A. j. hecki*	0.12	Lybia
13102	*A. j. jubatus*	52.20	Democratic Republic of the Congo
38432	*A. j. soemmeringii*	12.00	East Africa (confiscated in Djibouti)
56128	*A. j. raineyi*	19.40	Tanzania
56277	*A. j. venaticus*	17.50	Iran
56291	*A. j. raineyi*	1.61	Tanzania
AJ119b	*A. j. raineyi*	5.34	Tanzania
AJ120b	*A. j. raineyi*	1.77	Tanzania
AJ139b	*A. j. venaticus*	-	Turkmenistan
AJ154b	*A. j. jubatus*	30.60	Botswana
AJ178b	*A. j. venaticus*	0.54	Iraq
AJ078b	*A. j. hecki*	-	Algeria
AJ079b	*A. j. venaticus*	-	Afghanistan
AJ095n	*A. j. jubatus*	7.20	South Africa
AJ013	*A. j. jubatus*	too high	Angola
AJ06	*A. j. jubatus*	1.54	Zimbabwe
AJ063	*A. j. venaticus*	7.58	Iran
AJ228	*A. j. raineyi*	10.30	Kenia
AJ268	*A. j. raineyi*	5.34	Kenia
AJ271	*A. j. hecki*	20.60	Nigeria
AJ010	*A. j. raineyi*	0.65	Kenia
AJ017	*A. j. jubatus*	0.44	Angola
AJ021	*A. j. hecki*	0.13	Senegal
AJ023	*A. j. venaticus*	0.14	India
AJ024	*A. j. raineyi*	too low	Tanzania
AJ045	*A. j. soemmeringii*	too low	Sudan
AJ047	*A. j. soemmeringii*	too low	Sudan

AJ068	*A. j. soemmeringii*	too low	Ethiopia
AJ081	*A. j. soemmeringii*	2.32	Sudan
AJ129	*A. j. raineyi*	too low	Tanzania
AJ141	*A. j. venaticus*	too low	Syria
AJ159	*A. j. jubatus*	1.16	Zimbabwe
AJ160	*A. j. jubatus*	21.40	Zimbabwe
AJ251	*A. j. raineyi*	too low	Tanzania
AJ343	*A. j. venaticus*	0.62	Iran

2.2 Methods

2.2.1 PCR amplification of SNP sites

As a guideline for amplification and sequencing, the protocol from Pomerantz et al. (2022) was used. This protocol presents valuable simplified guidelines for DNA amplification and sequencing using miniaturized laboratory equipment in the field.

For the amplification of target regions, SNP primers developed by Magliolo et al. (2020) specifically designed for cheetahs were used. From the originally 218 designed primer pairs (Magliolo et al., 2020), 20 were chosen at random (Table 2). Stock solutions of primers had a concentration of 100 µM which were further diluted in nuclease-free water (NFW) to create 10 µM dilutions. The product of the various primer pairs showed a maximum length of 92 base pairs (bp) and minimum length of 63 bp (see Table 2). All pre-amplification steps were conducted in the clean room of the Laboratory of Molecular Systematics at the Natural History Museum Vienna to minimize possible sources of contamination.

To determine the best strategy for short fragment amplification two approaches were tested (see 2.2.1.2 and 2.2.1.3). The first approach (two-step PCR) followed the protocol of Pomerantz et al. (2020) using a first and second step PCR with minor adjustments. For the second approach (single-step PCR) the Native Barcoding Kit (ONT) was used. In addition, two different DNA polymerases (Taq DNA Polymerase, QIAGEN, Hilden, Germany; Multiplex PCR Master Mix, QIAGEN, Hilden, Germany) were tested.

Table 2: List of SNP primer sequences (5´- 3´), including product length (bp). The colours represent the way the primers were pooled in four primer sets. -Set 1 -Set 2 -Set 3 -Set 4. Primer sets underwent further pooling into Mix A (Set 1 + Set 2) and Mix B (Set 3 + Set 4) (see 2.2.1.3 below). T_{ann} shows annealing temperature in °C

Primer	Forward	Reverse	Product length	T_{ann}
cheetah_genome_SNP3	CCCTAGAGGATTTGTATGTGATCAACT	GTCAATGAAAGTTGGCTGAGTACCT	88	60.42/61.26
cheetah_genome_SNP6	GCCGAGGCCAACTGTGT	AACCTCCACCAAATGGCTCAT	66	59.93/59.92
cheetah_genome_SNP34	GGGAGCTAGTAACAGCCTTCTTAAA	CTACAAATTGGGTCATTTTAGCATGCA	75	60.34/61.04
cheetah_genome_SNP37	CGGAGAAGGTATCGCTCTTGTG	CCTAATGCGCTCCCGATGA	63	60.8/59.63
cheetah_genome_SNP57	TCCCTAGGCAACACGTGGATA	GAGAAAAGCCACTGAAGAGACAGA	85	60.62/60.50
cheetah_genome_SNP77	ATGATGGTGCCAGTGGAGAAG	GCCGTGGTGACACTTTGG	83	60.06/58.97
cheetah_genome_SNP99	ACGGAAATTTACAGGGAAGAAAGCT	GTCTTATCACTGTGATGTCAACCTTGA	76	60.98/61.22
cheetah_genome_SNP101	CTGGAGTCAGGCCAATCAGATT	ATTTGCTGCAACTACCACTACACTA	72	60.09/60.51
cheetah_genome_SNP114	ATGTCTCTACCACTTGTTATACCTACACA	CTTCATCCCATATCACTCCCTCTTC	77	61.41/60.28
cheetah_genome_SNP117	CCATCACTAGAGCTTCTGATGCTT	CACCCGAATTACATCCTAGGCATAG	82	60.44/60.85
cheetah_genome_SNP119	CCAGAAGATGGTGTTGAAGTTTCCT	CCCATGCAGGGCAGACT	78	61.21/58.90
cheetah_genome_SNP146	GAAGCTTGAGATTGTGAAACAGCT	GGTGATAGCAAGTAATGTCCTCATGA	92	60.02/60.69
cheetah_genome_SNP153	AGAAACAGAAGCTTGAATTGCATTCC	CCCCCCATGAACTTGTTTTCG	76	61.01/59.73
cheetah_genome_SNP157	GAACTCAGGCCAAAACAAATCCAT	CTGTAACCACAAATAATGAGTTTTGGGAAA	79	60.26/61.27
cheetah_genome_SNP169	TTCTTCCACAGTCAAAGGTAAAGCT	CAAGTTCACACTGTGCCTTCATTT	74	60.63/60.44
cheetah_genome_SNP187	GGTGGCTTGGATTGTACTTGACTTT	CTCCAGGCTTCTGCACACA	76	61.73/59.93
cheetah_genome_SNP200	GGCATAGGCTTATGAACCTTTACCT	GCAACTGCTTAGTACGGTATTACCT	83	60.69/60.68
cheetah_genome_SNP217	GCAGAATATGGTCGCTCCAGAT	CAGGGACACTGAAGTGGTTGA	73	60.29/59.86
cheetah_genome_SNP225	GTGGCTGATTAATGCCCAACAT	GGAGTTCAACGCCTGAAGGAT	84	59.57/60.34
cheetah_genome_SNP231	TCTCCCTAGTACACAGCTGATTGT	CCTTGCCTTTATACTTTTCTTCTAGCTTTG	79	60.81/61.21

2.2.1.1 Optimization of annealing temperature

Online tools for the calculation of annealing temperatures estimated different temperatures for the SNP primers. Magic-BLAST (Boratyn et al., 2019) revealed very similar annealing temperatures for all primers pairs (58.63 – 61.73 °C), the Thermo Scientific T_m Calculator estimated annealing temperatures of primers at 65 °C. For the optimization of the annealing temperature a gradient PCR, was conducted using the Eppendorf Mastercycler nexus X2. For this Step three primer pairs (cheetah_genome_SNP3, cheetah_genome_SNP99, cheetah_genome_SNP157) were chosen at random, conditions for the cycling protocol can be seen in Table 3A.

2.2.1.2 Two-Step PCR

2.2.1.2.1 First step

The first step PCR was conducted using the cycling protocol as seen in Table 3B. To gain optimal results varying concentrations of SNP primers and DNA were tested.

Seven primer pairs were used with 10 samples for the two-step amplification following Pomerantz et al. (2022). The primer pairs were equipped with ONTs universal tails (universal tail forward primer: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGACTC-3', universal tail reverse primer: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3') needed for the second step PCR. For better understanding of the workflow, a schematic overview can be seen in Figure 1. In the first trials the PCR contained a volume of 25 µl consisting of 17.375 µl NFW, 2.5 µl 10X PCR Buffer, 1.5 µl Mg^{2+} (1.5 µM), 0.5 µl dNTPs (1.6 µM), 0.125 µl (0.635 units) Taq Polymerase (QIAGEN®), 2 µl primer mix (1 µM) and 1 µl DNA (1/10 dilutions).

For the verification of successful amplification, agarose gel electrophoresis runs were conducted. The gels used were prepared with 2 % agarose to provide a dense gel that would allow the small amplicons of 63 bp to 92 bp to be accurately identified. The wells within the gel were loaded with 4 µl of PCR product and 1 µl of loading dye (Midori Xtra). For length comparison of fragments within the gel 1 µl of a 50 kb ladder (NIPPON Genetics®) was used. Gel electrophoresis was conducted at 100 V for at least 15 minutes.

Table 3: PCR program used for 1st and 2nd step PCR amplifications in the first run and amplification of the second run.

A: Cycling protocol for gradient PCR

Initial denaturation	3 min	95 °C	
Denaturation	15 sek	95 °C	x34
Annealing	15 sek	68 °C 72 °C 77 °C 79 °C	
Extension	30 sek	72 °C	
Final extension	1 min	72 °C	

B: Cycling protocol 1st step

Initial denaturation	15 min	95 °C	
Denaturation	30 sek	95 °C	x34
Annealing	90 sek	68 °C	
Extension	60 sek	72 °C	
Final extension	10 min	72 °C	

C: Cycling protocol 2nd step

Initial denaturation	3 min	95 °C	
Denaturation	15 sek	95 °C	x15
Annealing	15 sek	53 °C	
Extension	30 sek	72 °C	
Final extension	1 min	72 °C	

2.2.1.2.2 Second step

In the second step PCR ONT index primers (see Table 4) were added to all successfully amplified samples. This step used a PCR volume of 25 µl containing 17.357 µl NFW, 2.5 µl 10X PCR Buffer, 1.5 µl Mg^{2+} (1.5 µM), 0.5 µl dNTPs (1.6 µM), 0.125 µl (0.635 units) Taq

DNA Polymerase, 1 µl ONT index primer forward, 1 µl ONT index primer reverse, 1 µl amplification product of the first step PCR (1/10 dilution). The DNA amplification followed the protocol seen in Table 3C and was performed using a Mastercycler ® nexus X2 (Eppendorf). Agarose gel electrophoresis was performed as described above for the verification of successful amplification.

The clean-up of the DNA amplicons, library preparation and end prep followed the protocol by Pomerantz et al. (2022) with the exception that SPRI Beads were used other than AMPure XP beads as suggested in the protocol.

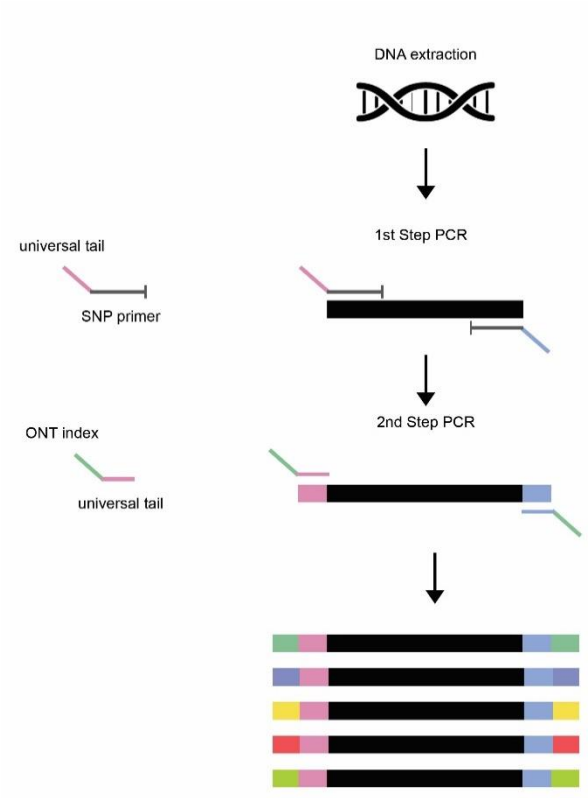


Figure 1: Visualisation of the workflow involved within the Two-Step PCR following the protocol by Pomerantz et al. (2022).

Table 4:Samples and assigned ONT index primers including universal tails (italics) used in the Two-Step PCR.

Sample	Forward	Sequence 5'-3'	Reverse	Sequence 5'-3'
63	BC01_fw	AAGAAAGTTGTCGGTGTCTTTGTG	BC01_rv	AAGAAAGTTGTCGGTGTCTTTGTG
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
53	BC02_fw	TCGATTCCGTTTGTAGTCGTCTGT	BC02_rv	TCGATTCCGTTTGTAGTCGTCTGT
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
51	BC03_fw	GAGTCTTGTGTCCCAGTTACCAGG	BC03_rv	GAGTCTTGTGTCCCAGTTACCAGG
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
121	BC04_fw	TTCGGATTCTATCGTGTTTCCCTA	BC04_rv	TTCGGATTCTATCGTGTTTCCCTA
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
256	BC05_fw	CTTGTCCAGGGTTTGTGTAACCTT	BC05_rv	CTTGTCCAGGGTTTGTGTAACCTT
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
354	BC06_fw	TTCTCGCAAAGGCAGAAAGTAGTC	BC06_rv	TTCTCGCAAAGGCAGAAAGTAGT
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
99	BC07_fw	GTGTTACCGTGGGAATGAATCCTT	BC07_rv	GTGTTACCGTGGGAATGAATCCTT
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
AJZ001	BC09_fw	AACTAGGCACAGCGAGTCTTGGTT	BC09_rv	AACTAGGCACAGCGAGTCTTGGTT
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG
AJZ003	BC10_fw	AAGCGTTGAAACCTTTGTCCTCTC	BC10_rv	AAGCGTTGAAACCTTTGTCCTCTC
		ACACTCTTTCCCTACACGA		GTGACTGGAGTTCAGACGTG

2.2.1.3 Single-Step PCR

As later runs were made with more primer pairs the concentrations of single primers used in the primer mix needed to be adjusted. The 20 primer pairs used for these trials were grouped into 4 sets (each consisting of 5 primer pairs; Table 2) with concentrations of 100 µM. These primers were not equipped with universal tails, because in this protocol index sequences are added to the amplicons for individual identification via ligation during the sequencing library preparation. The 4 primer sets were further diluted by adding 90 µl NFW to 10 µl of each set. The resulting sets had concentrations of 10 µM.

PCRs using each primer set as well as combinations of these sets were carried out to test possible interferences of primers with each other, as the original primers were not developed

for multiplex PCR. After this, tests with DNA concentrations were conducted. Many samples were excluded in these trials because of limitations concerning the available amount and/or concentration of the sample (Table 1).

Trials with these 4 sets (consisting of five primer pairs each; see above) were performed in 25 µl using two approaches: (1) using Taq DNA Polymerase: 2.5 µl 10X PCR Buffer, 1.5 µl Mg^{2+} (1.5 µM), 0.5 µl dNTPs (1.6 µM), 0.125 µl (0.625 units) Taq Polymerase (QIAGEN®), 2 µl primer mix (0.16 µM) and 1 µl DNA. (2) using the Multiplex PCR Master Mix (12.5µl Multiplex PCR Master Mix; QIAGEN®), 2 µl primer mix (0.8 µM) with DNA varying to contain concentrations of approximately 5-10 ng/µl with primer concentrations between 0.1 µM and 1 µM.

According to the tests two sets of primer pairs (mix of 10 primer pairs each) were selected (Mix A and Mix B; see Table 2), which were used for the final analysis. Subsequently, amplicons were shipped to the University of Oulu, Finland, for library preparation and sequencing using the Native Barcoding Kit (ONT).

2.2.2 Data analysis

The data analysis was conducted using a standard laptop (HP Sprectre x360 Convertible 13-aw0xxx) and Docker 4.22.1 (Merkel, 2014) for the creation of a container to ensure consistency during data analysis runs independent of the host laptops environment. Anaconda was used to install necessary programmes for data analysis such as SAMtools (Li et al., 2009), ANGSD (Analysis of next generation Sequencing Data; Korneliussen et al., 2014) and BWA (Burrows-Wheeler Aligner; Jung & Han, 2022). BWA was used for the reference mapping against the NCBI genome isolate of *Acinonyx jubatus* Ajub_Pintasa_27869175 (NCBI RefSeq assembly GCF_027475565.1). Post-processing of the files was carried out using SAMtools and ANGSD (used for SNP calling).

Commands for data preparation were as follows:

- Indexing of the reference file:

```
bwa index snp_ref_new.fasta
```

Then in a loop:

- Start of the loop:

  *for filein *.fasta;do*

- Defining the basename of the files to analyse:

  bn='basename$file.fastq'

- Mapping against a reference file using BWA and conversion of a SAM file into a BAM file:
  bwa mem snp_red_new.fasta $file | samtools view -Sb | samtools sort -o ${bn}.sorted.bam
- Indexing of created BAM file using SAMtools
  samtools index ${bn}.sorted.bam
- Call SNPs with angsd:
  angsd -I ${bn}.sorted.bam -GL 2 -doMaf 2 -SNP_pval 1e-9 -doMajorMinor 1 -out${bn}
- Unzip the SNP call file
  gzip -d ${bn}.mafs.gz
- End of loop:
  done

For verification, each called SNP was inspected by visualizing the mapping files in Tablet 1.21.02.08 (Milne et al., 2013), a graphical viewing application for next generation sequence alignments and assemblies.

Additionally, a replicate experiment was conducted using these amplicons. To assess the accuracy, seven samples were amplified twice independently and sequenced separately to create replicates for comparative analysis.

3 Results

3.1 DNA amplification

3.1.1 SNP primer performance

During work in the laboratory, the primers developed by Magliolo et al. (2021) underwent several test runs for optimization. The first part of this optimisation included the assessment of the optimal annealing temperature for all primers. The use of different online tools such as Magic-Blast and Thermo Fischer Scientific T_m Calculator (*thermofisher.com*, 2024) resulted in the calculation of different annealing temperatures. Whilst Magic-Blast suggests annealing temperatures of 58.63 – 61.73 °C, the T_m Calculator calculates about 65 °C for each primer. A gradient PCR, that incorporated these temperatures showed best amplification results for the primers at 68 °C. Seeing how these online tools and the gradient PCR greatly differed in temperature a review of annealing temperatures for each primer as well as primer pools would be recommended for further projects.

Having determined these conditions for the cycling protocol, adjusting of primer and DNA concentrations followed. The adjustment of primer concentrations took several trials since the SNP primers were not designed for a multiplex approach. Even though primer concentrations were adjusted the success rate of amplifications was rather low, this however could also be a result of the annealing temperature since it was not tested for a multiplex approach. For trials with several primers lower annealing temperatures might yield better results.

3.1.2 Two-Step PCR results

During DNA amplification using the two-step PCR method following the protocol of Pomerantz et al. (2022) issues concerning contamination of amplicons occurred. Several runs to investigate the potential source of contamination (such as purity of DNA samples, issues during agarose gel loading, contamination of PCR chemicals and possible workflow issues) revealed a contamination within primer stocks. In order to investigate whether the successful amplification was cheetah DNA, all libraries were sequenced on the MinION. Because of these contamination issues the data acquired through the Two-Step PCR was not considered during data interpretation.

3.1.3 Single-Step PCR results

For the single-step PCR using the Native Barcoding Kit (Zarate, 2022) newly acquired primers were used. The amplification of samples proved challenging due to the high number of primers within one PCR. Several trials (5 to 20 primer pairs in one reaction) were conducted to determine the maximum number of primers with acceptable results. Finally, it was concluded that 20 primer pairs per PCR were too many even when concentrations deemed tolerable according to the QIAGEN® Multiplex PCR Handbook as well as the QIAGEN® Taq PCR Handbook. Amplifications were therefore conducted using 10 primer pairs in two amplification reactions. The primer pairs were split into sets of 5 as can be seen in Table 2. Set 1 and Set 2 were pooled into Mix A (10 primer pairs), Set 3 and Set 4 were pooled into Mix B (10 primer pairs).

Not all samples amplified successfully, as each sample was amplified twice (once with Mix A and once with Mix B) some samples could be successfully amplified with only one mix. The gel electrophoresis revealed a successful amplification of 21 out of 36 samples, of which nine were amplified with Mix A and 12 with Mix B.

3.2 Bioinformatics

During bioinformatical analysis a visual comparison between SNP sites called by ANGSD and generated BAM files by BWA was conducted. Due to the fact that the chosen SNP sites are not heterozygous in all individuals not every SNP site could be found in every sample. Furthermore, some SNP sites had no or only a very small number of sequenced reads (< 7 reads). To view which SNP sites could be verified in which sample see Table 5. Overall, 280 SNP sites were investigated, of these 99 were called by ANGSD. Of these 99 calls 26 did not reach the expected 50% of heterozygous site during visual verification. When heterozygous the base pair substitution was coherent within a SNP for all samples. Within several amplicons ANGSD detected more than the expected SNP site, these SNPs could partially be verified in some amplicons. However, these sites are located at the edges of the aligned sequences often showing very little coverage. Positions of these additional SNP sites can be found the 8.1 Appendix 1 since they are not included within the further analysis.

Table 5: Overview of SNP sites within samples. In each cell, left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD. Green fields show sites called by ANGSD that did not withstand the visual verification and could only show 10 – 30% of heterozygous sites. A more detailed table can be found in 8.1 Appendix 1.

MIX A	SNP3		SNP37		SNP99		SNP114		SNP101		SNP57		SNP34		SNP6		SNP117		SNP77	
4362_A.19	191	HE	2411	HE	45	HO	9	HO	10489	HO	1655	HE	3010	HE	957	HO	5100	HO	458	HO
4362_A.65	233	HE	4013	HE	178	HO	16	HO	16263	HO	2680	HE	6758	HE	318	HO	5100	HO	344	HO
13102_A.66	2	LC	121	HE	2	LC	2	LC	1090	HE	54	HE	99	HE	20	HE	143	HE	4	LC
56128_A.67	4	LC	58	HE	2	LC	3	LC	2477	HE	63	HO	79	HE	30	HE	124	HE	2	LC
56177_A.21	35	HO	709	HO	15	HO	5	LC	10913	HE	947	HO	798	HO	1108	HE	820	HO	25	HO
AJ13_A.05	20	HO	177	HE	9	HE	4	LC	4962	HO	327	HO	455	HO	449	HO	459	HO	20	HO
AJ13_A.25	190	HO	2062	HE	142	HE	5	LC	16323	HO	1867	HO	5299	HO	268	HO	7248	HO	61	HE
AJ24_A.27	1	LC	8	HE	0	LC	0	LC	65	HE	14	HE	16	HE	5	LC	18	HE	0	LC
AJ64_A.08	55	HO	984	HO	25	HO	1	LC	8411	HO	1423	HO	1227	HE	429	HO	1824	HO	126	HO
AJ64_A.29	32	HO	588	HO	26	HE	4	LC	9854	HO	831	HO	898	HE	115	HE	1616	HO	38	HO
AJ160_A.15	69	HO	1341	HO	22	HE	6	LC	7416	HO	860	HO	1350	HO	468	HO	1960	HO	197	HE
AJ178_A.32	5	LC	121	HE	4	LC	2	LC	3856	HE	136	HE	210	HE	46	HE	193	HE	13	HO
MIX B	**SNP157**		**SNP225**		**SNP231**		**SNP169**		**SNP187**		**SNP153**		**SNP119**		**SNP217**		**SNP200**		**SNP146**	
4362_B.20	3143	HO	12368	HE	873	HE	158	HO	16572	HE	670	HE	109	HO	5273	HE	224	HO	1776	HO
4362_B.68	1045	HO	1590	HE	1524	HE	91	HO	1335	HE	1335	HE	109	HO	1514	HE	30	HO	849	HO
13102_B.69	60	HE	1566	HE	3787	HO	2	LC	29	HE	29	HE	4	LC	419	HO	18	HO	13	HO
56128_B.70	626	HE	17748	HO	1581	HE	7	HE	233	HO	233	HO	28	HE	6288	HO	22	HE	220	HO
56177_B.22	421	HE	7280	HO	245	HO	28	HE	113	HO	113	HO	14	HE	1958	HO	20	HE	193	HO
56291_B.71	7	HO	105	HE	86	HO	0	LC	140	HE	4	LC	2	LC	18	HO	1	LC	5	LC
AJ13_B.07	17	HO	188	HE	1	LC	0	LC	1	LC	1	LC	0	LC	23	HE	0	LC	1	LC
AJ13_B.26	203	HO	9017	HE	34	HE	4	LC	12915	HO	25	HE	12	HO	1925	HE	15	HO	111	HE
AJ24_B.28	31	HE	357	HE	2036	HE	0	LC	356	HE	15	HO	1	LC	93	HE	9	HO	14	HE
AJ64_B.13	203	HE	8527	HO	409	HO	38	HO	11893	HO	178	HO	26	HO	2955	HO	50	HO	416	HO
AJ64_B.30	4	LC	145	HE	39	HO	1	LC	170	HO	0	LC	0	LC	93	HO	0	LC	9	HO
AJ154_B.23	27	HE	5753	HO	275	HO	2	LC	6356	HO	9	HO	1	LC	493	HO	6	LC	26	HE
AJ160_B.16	159	HO	5462	HO	325	HO	11	HE	8658	HO	123	HO	20	HE	1936	HO	29	HE	321	HO
AJ178_B.24	29	HE	29	HE	2481	HE	7	HO	430	HE	17	HE	1	LC	80	HE	4	LC	10	HO
AJ271_B.18	15	HE	1396	HO	5	LC	0	LC	1416	HO	2	LC	0	LC	130	HO	0	LC	15	HO
AJ271_B.31	38	HE	5461	HO	585	HE	0	LC	5375	HO	15	HE	3	LC	578	HO	1	LC	20	HO

Each primer set used for analysis contained 10 SNP primer pairs, within the analysis we could however detect 20 SNP primer pairs per sample (an example of this can be seen in Table 6). Therefore, samples also included calls with ANGSD in SNP sites that were not expected to be seen due to a probable cross-contamination of indices (these are not handling errors). For data analysis only the primer pairs expected, Set A primers for all sample IDs including A and Set B primers for all sample IDs including B, were considered.

Due to very low read numbers in multiple samples (Table 5) some possible SNP sites may not have been called in this experiment. To verify this possibility several samples were analysed twice producing replicates (Table 7-13) to verify SNP typing accuracy.

Table 6: Example of cross-contamination of samples. First analyses results showed amplifications of 20 instead of 10 primer pairs. Sample 4362_A.19 amplified with Mix A ☐ showed amplifications of SNP regions of Mix B ☐ that were not used with this sample. Left column beneath the sample indicates read numbers, middle column shows if SNP sites were heterozygous (HE), or homozygous (HO) and right column shows if and what base pair (bp) SNPs were called by ANGSD. NC indicates no call through ANGSD.

	4362_A.19		
cheetah_genome_SNP3	191	HE	50
cheetah_genome_SNP37	2.411	HE	32
cheetah_genome_SNP99	45	HO	NC
cheetah_genome_SNP114	9	HO	NC
cheetah_genome_SNP101	10.489	HO	NC
cheetah_genome_SNP57	1.655	HE	55
cheetah_genome_SNP34	3.010	HE	35
cheetah_genome_SNP6	957	HO	NC
cheetah_genome_SNP117	5.100	HO	NC
cheetah_genome_SNP77	458	HO	NC
cheetah_genome_SNP157	33	HE	27
cheetah_genome_SNP225	318	HE	44
cheetah_genome_SNP231	99	HO	NC
cheetah_genome_SNP169	3	HO	NC
cheetah_genome_SNP187	370	HE	45
cheetah_genome_SNP153	6	HO	NC
cheetah_genome_SNP119	0	HO	NC
cheetah_genome_SNP217	84	HE	30+45
cheetah_genome_SNP200	4	HO	NC
cheetah_genome_SNP146	11	HE	42

3.2.1 Replicates

Analysis of the replicates could be evaluated in 73 % of SNP calls. 27% were not considered because of low coverage. The consensus between evaluated SNP sites lies at 92% for hetero- and homozygous sites. Of the 7 replicates three pairs were amplified using primer Mix A, 4 pairs were amplified using primer Mix B.

3.2.1.1 Replicates of 4362_A

Between the two replicates, four heterozygous SNP sites were called by ANGSD. All the called SNP sites could be verified visually. A comparison of the replicates showed significant differences in amount of reads per amplicon. Both replicates show very low as well as very high coverage.

Table 7: Replicates of sample 4362_A. A in the sample name signifies that Mix A was used for amplification. Left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD.

MIX A	SNP3		SNP37		SNP99		SNP114		SNP101		SNP57		SNP34		SNP6		SNP117		SNP77	
4362_A.19	191	HE	2411	HE	45	HO	9	HO	10489	HO	1655	HE	3010	HE	957	HO	5100	HO	458	HO
4362_A.65	233	HE	4013	HE	178	HO	16	HO	16263	HO	2680	HE	6758	HE	318	HO	5100	HO	344	HO

3.2.1.2 Replicates of 4362_B

Differences in read amounts can also be observed for these replicates as seen in Table 8. There is no pattern detectable as to what replicate had a better read coverage. However, all SNPs called by ANGSD (SNP225, SNP231, SNP187, SNP153, SNP217) could be validated in both replicates. During visual verification SNP157 appears to have many heterozygous sites making it appear as valid SNP site, closer inspection however, shows an error during sequence alignment. This can be observed in both replicates. Although about 50% of reads included these false heterozygous sites ANGSD did not call a SNP site.

Table 8: Replicates of sample 4362_B. B in the sample name signifies that Mix B was used for amplification. Left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD.

MIX B	SNP157		SNP225		SNP231		SNP169		SNP187		SNP153		SNP119		SNP217		SNP200		SNP146	
4362_B.20	3143	HO	12368	HE	873	HE	158	HO	16572	HE	670	HE	109	HO	5273	HE	224	HO	1776	HO
4362_B.68	1045	HO	1590	HE	1524	HE	91	HO	1335	HE	1335	HE	109	HO	1514	HE	30	HO	849	HO

3.2.1.3 Replicates of AJ13_A

Replicates of AJ13 overlap in two verifiable SNP sites (SNP37 and SNP 99). SNP77 however was only called in replicate *25* which had an overall higher coverage for this SNP site (Table 9). Visual verification showed a heterozygous site in replicate *05* that did not meet the expected 50% the heterozygous site was only present in 20% and was therefore not called by ANGSD.

Table 9: Replicates of sample AJ13_A. A in the sample name signifies that Mix A was used for amplification. Left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD.

MIX A	SNP3		SNP37		SNP99		SNP114		SNP101		SNP57		SNP34		SNP6		SNP117		SNP77	
AJ13_A.05	20	HO	177	HE	9	HE	4	LC	4962	HO	327	HO	455	HO	449	HO	459	HO	20	HO
AJ13_A.25	190	HO	2062	HE	142	HE	5	LC	16323	HO	1867	HO	5299	HO	268	HO	7248	HO	61	HE

3.2.1.4 Replicates of AJ13_B

Comparison of replicates of sample AJ13 could only show one overlapping verifiable SNP site in SNP 217 (Table 10).

Replicate *26* showed additional verifiable SNP sites for SNP153 and SNP146. SNP231 only shows 35% of heterozygous sites but was still called by ANGSD. Visual verification also revealed a heterozygous site within SNP157 that was not called by ANGSD. Looking into the sorted.bam. file of this replicate, the same mismatched alignment as in 4362_B can be observed, showing a false heterozygous site. SNPs 231, 169, 187, 153, 119, 200 and 146 show too little coverage in either one or both of the replicates.

Table 10: Replicates of sample AJ13_B. B in the sample name signifies that Mix B was used for amplification. Left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD. Green fields show sites called by ANGSD that did not withstand the visual verification and could only show 10 – 30% of heterozygous sites.

MIX B	SNP157		SNP225		SNP231		SNP169		SNP187		SNP153		SNP119		SNP217		SNP200		SNP146	
AJ13_B.07	17	HO	188	HE	1	LC	0	LC	1	LC	1	LC	0	LC	23	HE	0	LC	1	LC
AJ13_B.26	203	HO	9017	HE	34	HE	4	LC	12915	HO	25	HE	12	HO	1925	HE	15	HO	111	HE

3.2.1.5 Replicates of AJ64_A

In these replicates there was only one SNP site that overlapped in both replicates namely SNP34 as seen in Table 11.

The additionally called SNP sites within replicate *29*, showing only about 20% heterozygosity during visual verification. ANGSD however wrongly called both sites heterozygous.

Table 11: Replicates of sample AJ64_A. A in the sample name signifies that Mix A was used for amplification. Left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD. Green fields show sites called by ANGSD that did not withstand the visual verification and could only show 10 – 30% of heterozygous sites.

MIX A	SNP3		SNP37		SNP99		SNP114		SNP101		SNP57		SNP34		SNP6		SNP117		SNP77	
AJ64_A.08	55	HO	984	HO	25	HO	1	LC	8411	HO	1423	HO	1227	HE	429	HO	1824	HO	126	HO
AJ64_A.29	32	HO	588	HO	26	HE	4	LC	9854	HO	831	HO	898	HE	115	HE	1616	HO	38	HO

3.2.1.6 Replicates of AJ64_B

SNP calling with ANGSD yielded different results for the two replicates of AJ64 for one SNP position (SNP225). SNP225 was only called in replicate *30*, this however could not be validated visually, as the alternative nucleotide was only observed in 10 % of reads.

Furthermore, SNP157 was called by ANGSD at 27 bp and 32 bp within replicate *13*, these calls could be verified visually. There were no called sites for SNP157 in replicate *30* due to low read numbers of four reads.

SNP225 was only called in replicate *30*, this however could not be validated visually, heterozygosity can only be observed in about 10% of reads.

Table 12: Replicates of sample AJ64_B. B in the sample name signifies that Mix B was used for amplification. Left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD. Green fields show sites called by ANGSD that did not withstand the visual verification and could only show 10 – 30% of heterozygous sites.

MIX B	SNP157		SNP225		SNP231		SNP169		SNP187		SNP153		SNP119		SNP217		SNP200		SNP146	
AJ64_B.13	203	HE	8527	HO	409	HO	38	HO	11893	HO	178	HO	26	HO	2955	HO	50	HO	416	HO
AJ64_B.30	4	LC	145	HE	39	HO	1	LC	170	HO	0	LC	0	LC	93	HO	0	LC	9	HO

3.2.1.7 Replicates of AJ271_B

Within SNP157 ANGSD called SNP sites at 27 bp and 32 bp which could be validated visually in both replicates. There were no other SNP sites called in replicate *18*. Replicate *31* showed two more call sites in SNP231 and SNP153 which could be visually verified. However, these showed low coverage in replicate *18*.

Table 13: Replicates of sample AJ271_B. B in the sample name signifies that Mix B was used for amplification. Left column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD.

MIX B	SNP157		SNP225		SNP231		SNP169		SNP187		SNP153		SNP119		SNP217		SNP200		SNP146	
AJ271_B.18	15	HE	1396	HO	5	LC	0	LC	1416	HO	2	LC	0	LC	130	HO	0	LC	15	HO
AJ271_B.31	38	HE	5461	HO	585	HE	0	LC	5375	HO	15	HE	3	LC	578	HO	1	LC	20	HO

3.2.2 Nanoplot report

The Nanoplot report showed a mean Phred Q-score of 11.6 for the sequenced reads (Table 14) indicating a base calling accuracy of more than 90 %. This is a common metric to assess sequencing platform accuracy, that implies the probability of incorrect calls by the sequencer (*Quality Scores for Next-Generation Sequencing*, o. J.).

Figure 2 shows the read lengths of sequenced amplicons mapped against the average read quality. Multiple aggregations can be observed. The most compact aggregation can be seen at about 100 bp length, other less compact aggregations follow at 200 bp, 300 bp and 400 bp, elevating the Median read length to 373 bp (Table 14). Average read quality for all read lengths is concentrated at an Average read quality between 10 and 15. More information concerning Nanoplot statistics can be found in 8.2 Appendix 2.

Table 14: Shows an excerpt of the Nanoplot report given after sequencing with the MinION, only including successfully indexed amplicons of the Single-Step PCR. "Mean read quality" shows the Phred score, "Mean read length" shows mean read length of all amplified sequences, "Number of reads" shows the total number of reads of indexed samples, "Total bases" shows total number of bases of indexed samples.

	Amplicons
Mean read quality	11.6
Median read length	373.0
Number of reads	478,828.0
Total bases	203,205,193.011.6

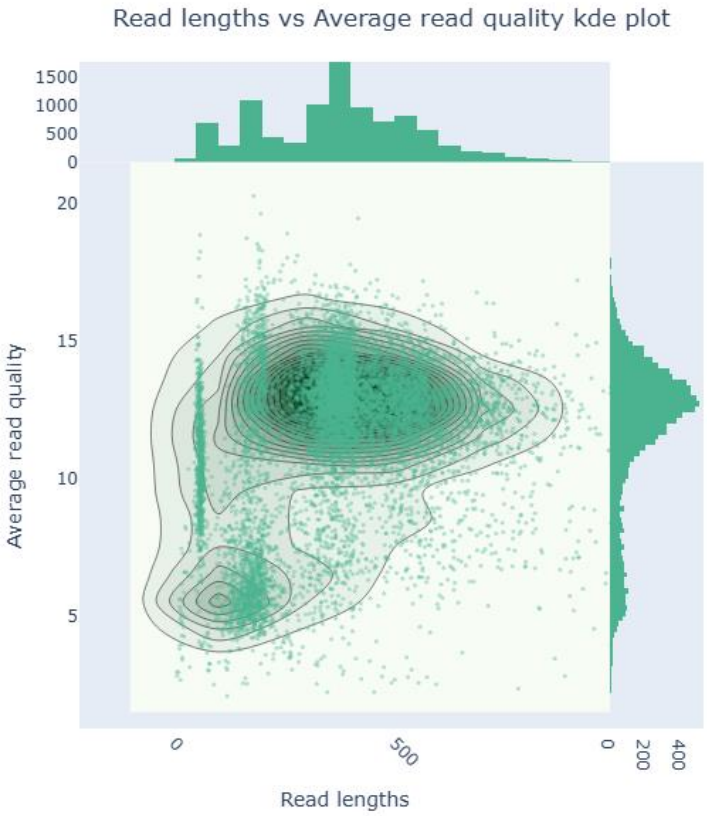


Figure 2: Nanoplot showing "Read lengths" mapped against "Average read quality".

4 Discussion

4.1 SNP primer performance

Here the performance of SNP primers developed by Magliolo et al. (2021) will be discussed under the aspect that the primers were not designed to be used for multiplexing. Furthermore, the primers were used with different annealing temperatures and DNA concentrations as recommended by Magliolo et al. (2021). By accident the conducted gradient PCR for annealing temperature adjustment were wrongly programmed into the thermocycler (Table 3), hence a suboptimal temperature range was applied resulting in too high annealing temperatures. However, these temperatures still resulted in successful amplifications while using only one primer pair.

Several trials for multiplexing could determine working primer mixes that showed an up to 100 % success rate during amplification for the Two-Step PCR. Success rate during trials using more primer mixes and samples during the Single-Step PCR were at 31 %. The different combination of primers, especially larger pools of primers, during the Single-Step PCR could be part of the lower success rates. Since not every primer was tested with all other primers, interference between primers cannot be excluded. Additionally, of the 36 provided DNA samples 67 % had starting concentration of less than 10 ng/µl. Even though the amount of DNA within the samples were adjusted to include at least 5 ng/µl within the PCR, some samples could not reach this concentration because of low volumes provided. Magliolo et al. (2021) determined the optimal DNA concentrations for the SNP primers to be between 10 – 30 ng/µl, this range could only be provided within 45 % samples. Two of the samples that showed successful amplification had lower concentrations within the PCR of 8.05 ng/µl (sample: 56291) and a too low to measure starting concentration (sample: AJZ024). A higher success for SNP amplification could be reached by further adjusting the annealing temperatures when dealing with primer mixes as well as using DNA samples with higher concentrations.

4.2 Utility of ONTs MinION sequencer

The use of miniaturized laboratory equipment in this work was to establish its usefulness for sequencing programs in the field as well as verifying the accuracy of the short fragment mode released in 2022. Advantages of the MinION sequencer include its small size, the capability of

short as well as long read sequencing and its relatively low price. One could argue that the MinIONs error rate compares badly with sequencing methods such as Sanger or Illumina (Delahaye & Nicolas, 2021). Each update for the MinION however, increased and keeps increasing the devices accuracy. While the error rate lay at 5 to 15 % in 2021 (Wang et al., 2021) we can see that it decreased drastically and has an error rate of 2-3 % (Liu-Wei et al., 2024) in 2024. The errors commonly seen while working with Nanopore technology include deletions, mismatches and insertions especially within low-complex regions such as stretches of homopolymers of three or more bases. For these the Nanopore technology can struggle to sequence the number accurately. However, multiple studies were able to show that the MinION is capable of producing highly accurate consensus sequences with coverages of 1,000 reads or lower (Krehenwinkel, Pomerantz, & Prost, 2019; Vasiljevic et al., 2021). Krehenwinkel et al. (2019) further showed that even as little as 10 reads per amplicon can reach accuracies of 98 % to 99 %. Studies using the MinION for forensic SNP typing and wildlife forensic DNA barcoding already stated the high possibility of this technology becoming a routine choice for performing forensic profiling (Cornelis et al., 2019; Delahaye & Nicolas, 2021; Vasiljevic et al., 2021). Given that the SNPs selected in Magliolo et al. (2021) were not within long homopolymers, the standard errors of the MinION device should not have a huge effect on the SNP typing accuracy. In the next section, I will discuss the sequencing and SNP typing errors encountered more in detail.

4.3 Bioinformatic analyses

The results acquired through bioinformatical analyses will be discussed in the following. It is noteworthy that ANGSD was used for SNP calling, because of its advantage for handling low coverage data. However, several cases occurred where ANGSD made wrong calls identifying homozygous sites as heterozygous. This occurred four times during the replicate experiment, 26 times overall out of 280 SNPs, showing an error rate of 9 % for SNP sites called by ANGSD with less than 50 % heterozygous sites. Therefore, it is advisable to reach higher coverage for better results, as well as experimenting with other SNP calling tools, e.g. ones that were especially made to handle nanopore data or include nanopore error models, for optimal outcomes.

4.3.1 Nanoplot reports

Nanoplot reports showed an overall agreeable Phred Q-score of 11.6. The median read length however lies at 373 bp which is longer than the expected short fragments of 63 bp to 92 bp. The plot seen in Figure 2 shows read lengths of approximately 100 bp as expected as well as aggregations of reads at 200 bp, 300 bp and 400 bp. These aggregation at double, triple, etc. read lengths are probably caused by chimeras that could have originated during PCR. However, it is more likely that these chimeras probably consist of multiple amplicons connected to each other during the sequencing. Even with the short fragment mode too fast analyses of the amplicons could have caused multiple amplicons to be read in quick succession creating double and triple length sequences.

4.3.2 Obstacles

As the reads were mapped against the reference sequences for all pools (A and B), reads mapping to the reference sequence for pool B could be detected in cases where only pool A primers were used and *vice versa* (Table 6). There are several causes that could explain this finding.

Firstly, miniature amounts of contamination within the test tubes could have caused the additional data. However, since working conditions in the laboratory were very clean this seems very unlikely. Furthermore, similar instances of contaminations have been reported by several other users in the Oxford Nanopore Technologies user forum (*nanoporetech.com*, 2024).

Secondly, this problem could have occurred during indexing using the Native Barcoding Kit (Zarate, 2022). While the protocol aims for the ligase to be deactivated before pooling of samples, partially active ligase within the pooled samples could have caused barcodes to be ligated to samples other than the initially barcoded ones.

Thirdly, though, very unlikely, small amounts of cross-contamination during the production of the indices by the manufacturer could cause this issue.

Fourthly, faulty assignment of barcodes could have occurred during demultiplexing using ONT's Guppy barcoder (Wick et al., 2019). This problem appears to be more likely to have occurred than the other possibilities since it is repeatedly reported within the Oxford Nanopore Technologies user forum (*nanoporetech.com*, 2024).

This cross-contamination could potentially impact SNP calls with lower coverage (< 20x). Potential cases were this could have had impacts were found in 9 % of all data. For example, in individual AJ64_B ANGSD called a heterozygous position for SNP225 in one but not the other replicate. Even though, the ratios are strongly deviating from 50 % (e.g. 12 % in replicate *30*) and thus both replicates should be scored as homozygous. The presence of a small number of reads with a T instead of a C in both replicates could be explained, either by sequencing errors (as this site is in a low complexity region) or by the cross-contamination issue. Either way, verification by eye found that the position was wrongly determined as heterozygous by ANGSD. Similar cases could be observed within the replicate experiment discussed below.

4.3.3 Replicates

To review SNP typing accuracy replicates of random samples were sequenced (see samples in Table 7-13). While most heterozygous sites called by ANGSD could be verified and a consensus can be seen between many sites within replicates, there are some cases in which ANGSD called SNP sites with far less than 50 % heterozygous sites. Within the replicates this error occurred in 6 % of all SNP sites investigated. These and other noteworthy cases will be discussed below.

4.3.3.1 4362_B

Overall, these replicates show high coverage and ANGSD called 5 SNP sites that overlap perfectly in both. Visual verification however revealed an interesting sight within SNP157 for both replicates. Looking into the sorted.bam. files it firstly appears that there is a heterozygous site for SNP157, a closer look however reveals insertions that result into a wrong alignment, making it appear as false heterozygous site as can be seen in Figure 3. Even though these sites appear to have 50 % of heterozygous sites ANGSD correctly called them as heterozygous. These are likely typical nanopore errors in short homopolymers.

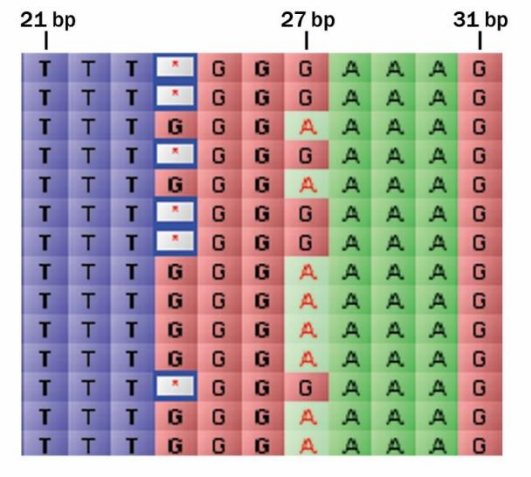


*Figure 3: Excerpt from Tablet. View of base pair position 20-31 revealing a shifted alignment making it appear position 27 is heterozygous. T: Thymine, G: Guanine, A: Adenine, *: insertion because of false alignment.*

4.3.3.2 AJ13_A

Here we can observe a discrepancy between called SNP sites for SNP77. SNP77 was called as a heterozygous site within replicate *25* but not in replicate *05*. The coverage with 20 reads should be sufficient to call the heterozygous site, according to literature (Krehenwinkel, Pomerantz, & Prost, 2019). ANGSD did not call the heterozygous site within replicate *05* and indeed visual verification revealed a homozygous site. The comparison with its replicate however clearly indicates that this site should contain a SNP position, showing that coverage of about 20 reads might not always be reliable. Although, in contrast to this, in this sample we can also see a different case, Table 9 reveals a correctly called SNP site in replicate *05* for SNP99 with only 9 reads.

4.3.3.3 AJ13_B

Sample AJ13_B illustrates the problem of low coverage very well (Table 10). Especially in replicate *07* 70 % of SNP sites could not be analysed. Within SNP157 we can again observe a mismatched alignment giving the impression of a wrong SNP site that was correctly not called by ANGSD.

4.3.3.4 AJ64_A

The replicates of sample AJ64_A match only in one called heterozygous site (Table 11).

Additionally, ANGSD called two more SNP sites within replicate *29* which could not be verified visually. The verification process revealed that heterozygous sites only occur in 10 - 20 % of reads. These calls by ANGSD are wrong and might be connected to the cross-contaminations (see 4.3.2 Obstacles). Since the heterozygous sites in SNP6 for example showed the exchange of G to A in the heterozygous sites, these could stem from contamination since in homozygous individuals this site tends to show only G.

4.3.3.5 AJ64_B

In replicate *30* we can observe again a wrong called SNP site by ANGSD for SNP 225 as in AJ64_A.

4.4 Fight against illegal wildlife trade

Protected areas (PA) provide safe spaces for many species at risk of extinction. The cheetah however mostly occurs outside of such areas (67 % of cheetah populations) due to its wide-ranging nature. Like other large terrestrial mammals, cheetahs face many threats within the Anthropocene that not only include habitat loss and fragmentation but also human-wildlife conflict and illegal trade (Durant et al., 2017, 2022; Magliolo et al., 2021; Tricorache et al., 2018). Within PAs the pressure on cheetahs remains high as well due to poor law enforcement and other anthropogenic pressures. The habitats cheetahs frequent also include some of the most marginalized and vulnerable human populations in the world (Durant et al., 2022). The illegal trade in live cheetahs does not get the same attention as illegal trade on other matters such as ivory, pangolin scales or rhinoceros horns, but puts immense pressure on cheetah populations. Confiscations in Somaliland revealed that 100 % of the illegally traded individuals belong to *Acinonyx jubatus soemmeringii* suggesting high pressure on this subspecies. This pressure recently led to the uplisting of *A. j. soemmeringii* from “vulnerable” to “endangered” (Schmidt-Künzel et al. 2023). To aid in the fight against the illegal trade of cheetahs several mechanisms have already been applied. These include, for example, microchips to individually mark animals registered for trade under Appendix I of the CITES convention. This method, however, can easily be criticised because of the easy way microchips can be switched between individuals,

e.g. from captive-bred into wild-caught individuals. Therefore, efforts to supply parental DNA within cheetahs' export papers have been implemented (Tricorache et al., 2018).

Recently, Magliolo et al. (2021) developed a SNP-based array to monitor the legal trade of CITES-registered captive-bred cheetah individuals from South Africa. While this SNP array proved to be highly efficient, it has two limitations for its use in monitoring other cheetah subspecies. Firstly, it was developed for a single subspecies and thus will likely show ascertainment bias for other subspecies. This selection bias occurs since SNP sites are more likely to be heterozygous in samples of the subspecies the SNP array was designed for than in the other subspecies investigated. It is imperative to be aware of the existence of SNP ascertainment biases as such since they could ultimately influence results.

Secondly it has to be run on the ThermoFisher Quant Studio platform therefore, it is not easily available for other cheetah range countries, especially those with lower conservation, research funding and infrastructure.

Furthermore, it would require many countries to transport the samples or DNA abroad for analysis. International conventions such as CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) and the Convention on Biological Diversity's Nagoya Protocol on Access and Benefit Sharing (https://www.cbd.int/abs) increase the difficulty for quick and easy analysis of samples if these need to be processed outside the country of origin. Therefore, ONT's MinION sequencer is assessed here to investigate if this portable and inexpensive device can be used for SNP typing. This would open up an efficient alternative to shipping samples abroad or the need to invest in more expensive technologies such as the Thermo Fisher Quant Studio. Furthermore, the use of portable and inexpensive laboratory equipment, such as the MinION, within the country of sample origin aids one of the key aspects within the Nagoya Protocol and the United Nations Sustainable Development Goals (https://sdgs.un.org/goals), namely the development of local scientific capacities and communities (Pomerantz et al., 2022).

4.5 Conclusuion

In the following the questions posted in 1.4 will be addressed directly.

- **Can the short fragment mode of the MinION reliably be used to call SNP sites from short amplicons?**

Since the short fragment mode was already introduced in 2022 when this thesis was conducted DNA fragments of as small as 20 bp could be sequenced. Before this new mode the limit for short fragments lay at 200 bp. As the results show the MinION is capable of producing read outputs of short amplicons that can be used to detect SNP sites. The coverage of SNP sites, however, also depends on sample quality and quantity. Even with low quantities of samples, if high enough concentrations can be reached through PCR amplification, the resulting coverage can be enough for reliable analyses of SNP sites.

- **Does the MinION technology allow for reliable calling of SNP sites, despite the considerable error rate?**

During this work the MinIONs error rate did not pose any difficulties. The consensus between replicates lay at 92 % showing high agreement between the two. Out of the 10 cases where the replicates differed from each other, 6 could be explained due to low coverage. The other 4 cases are due to faulty calls from ANGSD as discussed above. One of these cases however showed too low coverage in one replicate for comparison (AJ13_B_07, SNP231). MinION sequencing errors, however, should not pose issues for the used SNP dataset as errors mostly occur in homopolymer stretches and SNPs were mostly located in heteropolymer regions. However, sequencing chemistry of the MinION is constantly evolving lowering the error rate, especially concerning homopolymer regions, thus, in the future reliability for SNP calling should be significantly increased.

- **Can the SNP typing with the MinION be used for old and potentially degenerated museums material?**

Materials used during this work consisted only of museums specimens. It is therefore concluded that the MinIONs sequencing chemistry does work on degenerated DNA.

- **Can the MinION be a useful tool in the fight against illegal wildlife trade?**

Besides the affordability and portability of the MinION, the readily available and easy to use protocols and videos make it a very useful tool to aid in local capacity sharing. It can not only be used to help teach practical hands-on courses in genomics in schools or field schools (Blanco et al., 2019; Prost et al., 2020; Salazar et al., 2020), but can potentially be performed by assistants and students aiding in research capacity building (Pomerantz et al., 2022; Watsa et al., 2020). Within countries with little research funding and infrastructure (as reviewed in Ogden et al., 2021) the MinION, with its high reliability for DNA barcoding and SNP typing, poses a great alternative to standard and more expensive sequencing platforms such as Sanger or Illumina sequencing. However, in the light of the observed unknown cross-contaminations further testing is required as it can potentially affect its application for forensic case work. Nevertheless, this contamination should be low and not negatively impact runs with high enough coverage.

- **Does the MinION have the potential to support local capacity building and aid the fight against wildlife crime within remote areas?**

The MinION has the potential to open doors to aid local law enforcement and rangers in identifying threatened species. The applicability of the MinION in remote areas was already tested by Pomerantz et al. (2018), where they concluded that this miniaturized sequencer in combination with other field-deployable equipment can be an asset for rapidly obtaining genetic information. As in the case of the cheetah *Acinonyx jubatus* where trade is partially legal under Appendix I of the CITES convention, individual identification of specimens can be crucial in the fight against illegal trade. The protocol by Pomerantz et al. (2022) starting from DNA extraction, PCR, nanopore sequencing and barcode assembly can produce results within less than 24 hours which can be of great advantage in remote areas. Furthermore, the tool can be used in education to train scientists and students (as discussed above).

5 List of Figures

6 List of Tables

7 References

Beschta, R. L., & Ripple, W. J. (2009). Large predators and trophic cascades in terrestrial ecosystems of the western United States. *Biological Conservation*, *142*(11), 2401–2414. https://doi.org/10.1016/j.biocon.2009.06.015

Blanco, M.B., Greene, L.K., Rasambainarivo, F. *et al.* Next-generation technologies applied to age-old challenges in Madagascar. *Conserv Genet* 21, 785–793 (2020). https://doi.org/10.1007/s10592-020-01296-0

Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B., & Madden, T. L. (2019). Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics*, *20*(1), 405. https://doi.org/10.1186/s12859-019-2996-x

Cardillo, M., Purvis, A., Sechrest, W., Gittleman, J. L., Bielby, J., & Mace, G. M. (2004). Human Population Density and Extinction Risk in the World's Carnivores. *PLoS Biology*, *2*(7), e197. https://doi.org/10.1371/journal.pbio.0020197

Charruau, P., Fernandes, C., Orozco-Terwengel, P., Peters, J., Hunter, L., Ziaie, H., Jourabchian, A., Jowkar, H., Schaller, G., Ostrowski, S., Vercammen, P., Grange, T., Schlötterer, C., Kotze, A., Geigl, E. M., Walzer, C., & Burger, P. A. (2011). Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: Evidence for long-term geographic isolates. *Molecular Ecology*, *20*(4), 706–724. https://doi.org/10.1111/j.1365-294X.2010.04986.x

Cornelis, S., Gansemans, Y., Vander Plaetsen, A.-S., Weymaere, J., Willems, S., Deforce, D., & Van Nieuwerburgh, F. (2019). Forensic tri-allelic SNP genotyping using nanopore sequencing. *Forensic Science International: Genetics*, *38*, 204–210. https://doi.org/10.1016/j.fsigen.2018.11.012

Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, *34*(5), 518–524. https://doi.org/10.1038/nbt.3423

Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLOS ONE*, *16*(10), e0257521. https://doi.org/10.1371/journal.pone.0257521

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.J., Visseren-Hamakers, I., Willis, K.J., Zayas, C.N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. Science. 2019 Dec 13;366(6471):eaax3100. doi: 10.1126/science.aax3100. PMID: 31831642.

Durant, S. M., Broekhuis, F., Evangelista, P., Tricorache, P., Asfaw, T., Gedow, O., & Marker, L. (2023). *Acinonyx jubatus ssp. Soemmeringii. The IUCN Red List of Threatened Species*. https://doi.org/2023: e.T231957008A232671735

Durant, S. M., Groom, R., Ipavec, A., Mitchell, N., & Khalatbari, L. (2022). *Acinonyx jubatus.The IUCN Red List of Threatened Species 2022: E.T219A124366642.* https://doi.org/10.2305/IUCN.UK.2022-1.RLTS.T219A124366642.en

Durant, S. M., Mitchell, N., Groom, R., Pettorelli, N., Ipavec, A., Jacobson, A. P., Woodroffe, R., Böhm, M., Hunter, L. T. B., Becker, M. S., Broekhuis, F., Bashir, S., Andresen, L., Aschenborn, O., Beddiaf, M., Belbachir, F., Belbachir-Bazi, A., Berbash, A., De Matos

Machado, I. B., Breitenmoser, C., Chege, M., Cillers, D., Davies-Mostert, H., Dickman, A.J., Ezekiel, F., M.S. Farhadinia, Funston, P., Henschel, P., Horgan, J., de Iongh, H.H., Jowkar, H., Klein, R., Lindsey, P.A., Marker, L., Marnewick, K., Melzheimer, J., Merkle, J., M'soka, J., Msuha, M., O'Neill, H., Parker, M., Purchase, G., Sahailou, S., Saidu, Y., Samna, A., Schmidt-Küntzel, A., Selebatso, E., Sogbohossou, E. A., Soultan, A., Stone, E., van der Meer, E., van Vuuren, R., Wykstra, M., Young-Overton, K. (2017). The global decline of cheetah Acinonyx jubatus and what it means for conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(3), 528–533. https://doi.org/10.1073/pnas.1611122114

Durant, S. M., Wacher, T., Bashir, S., Woodroffe, R., De Ornellas, P., Ransom, C., Newby, J., Abáigar, T., Abdelgadir, M., El Alqamy, H., Baillie, J., Beddiaf, M., Belbachir, F., Belbachir-Bazi, A., Berbash, A. A., Bemadjim, N. E., Beudels-Jamar, R., Boitani, L., Breitenmoser, C., Cano, M., Chardonnet, P., Collen, B., Cornforth, W. A., Cuzin, F., Gerngross, P., Haddane, B., Hadjeloum, M., Jacobson, A., Jebali, A., Lamarque, F., Mallon, K. Minkowski, S. Monfort, B. Ndoassal, B. Niagate, G. Purchase, S. Samaïla, D., Samna, A. K., Sillero-Zubiri, C., Soultan, A. E., Stanley Price, M. R., Pettorelli, N. (2014). Fiddling in biodiversity hotspots while deserts burn? Collapse of the Sahara's megafauna. *Diversity and Distributions*, *20*(1), 114–122. https://doi.org/10.1111/ddi.12157

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*(1), 239. https://doi.org/10.1186/s13059-016-1103-0

Jung, Y., & Han, D. (2022). BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics*, *38*(9), 2404–2413. https://doi.org/10.1093/bioinformatics/btac137

Kitchener, A., Breitenmoser-Würsten, C., Eizirik, E., Gentry, A., Werdelin, L., Wilting, A., & Yamaguchi, N. (2017). *A revised Felidae Taxonomy of the Felidae. The final report of the Cat Classification Task Force of the IUCN/ SSC Cat Specialist Group*. http://hdl.handle.net/10923/15093

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Krausmann, P. R., & Morales, S. M. (2005). Acinonyx jubatus. *Mammalian Species*, *771*, 1–6. https://doi.org/10.1644/1545-1410(2005)771[0001:AJ]2.0.CO;2

Krehenwinkel, H., Pomerantz, A., Henderson, J. B., Kennedy, S. R., Lim, J. Y., Swamy, V., Shoobridge, J. D., Graham, N., Patel, N. H., Gillespie, R. G., & Prost, S. (2019). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience*, *8*(5), giz006. https://doi.org/10.1093/gigascience/giz006

Krehenwinkel, Pomerantz, & Prost. (2019). Genetic Biomonitoring and Biodiversity Assessment Using Portable Sequencing Technologies: Current Uses and Future Directions. *Genes*, *10*(11), 858. https://doi.org/10.3390/genes10110858

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence

Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu-Wei, W., van der Toorn, W., Bohn, P., Hölzer, M., Smyth, R. P., & Von Kleist, M. (2024). Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. *BMC Genomics*, *25*(1), 528. https://doi.org/10.1186/s12864-024-10440-w

Magliolo, M., Prost, S., Orozco-terWengel, P., Burger, P., Kropff, A. S., Kotze, A., Grobler, J. P., & Dalton, D. L. (2021). Unlocking the potential of a validated single nucleotide polymorphism array for genomic monitoring of trade in cheetahs (*Acinonyx jubatus*). *Molecular Biology Reports*, *48*(1), 171–181. https://doi.org/10.1007/s11033-020-06030-0

Marker, L. L., Dickman, A. J., Mills, M. G. L., Jeo, R. M., & Macdonald, D. W. (2008). Spatial ecology of cheetahs on north-central Namibian farmlands. *Journal of Zoology*, *274*(3), 226–238. https://doi.org/10.1111/j.1469-7998.2007.00375.x

Merkel, D. (2014). *Docker: Lightweight Linux containers for consistent development and deployment* (239(2), 2) [Linux j]. 239(2), 2.

Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., Shaw, P. D., & Marshall, D. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, *14*(2), 193–202. https://doi.org/10.1093/bib/bbs012

Nanoporetech.com. (2024). https://nanoporetech.com/products/specifications

Nowell, K. (2014). An assessment of the conservation impacts of legal and illegal trade in cheetahs *Acinonyx jubatus*. *IUCN SSC Cat Specialist Group report prepared for the CITES Secretariat, 65th meeting of the CITES Standing Committee, Geneva, 7–11 July. CITES SC65 Doc. 39.* http://cites.org/sites/default/files/eng/ com/sc/65/E-SC65-39.pdf.

O'Brien, S. J., Wildt, D. E., Bush, M., Caro, T. M., FitzGibbon, C., Aggundey, I., & Leakey, R. E. (1987). East African cheetahs: Evidence for two population bottlenecks? *Proceedings of the National Academy of Sciences*, *84*(2), 508–511. https://doi.org/10.1073/pnas.84.2.508

Ogden, R., Vasiljevic, N., & Prost, S. (2021). Nanopore sequencing in non-human forensic genetics. *Emerging Topics in Life Sciences*, *5*(3), 465–473. https://doi.org/10.1042/ETLS20200287

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., Barrio-Amorós, C. L., Salazar-Valenzuela, D., & Prost, S. (2018). Real-time DNA barcoding in a rainforest using nanopore sequencing: Opportunities for rapid biodiversity assessments and local capacity building. *GigaScience*, *7*(4). https://doi.org/10.1093/gigascience/giy033

Pomerantz, A., Sahlin, K., Vasiljevic, N., Seah, A., Lim, M., Humble, E., Kennedy, S., Krehenwinkel, H., Winter, S., Ogden, R., & Prost, S. (2022). Rapid in situ identification of biological specimens via DNA amplicon sequencing using miniaturized laboratory equipment. *Nature Protocols*, *17*(6), 1415–1443. https://doi.org/10.1038/s41596-022-00682-x

Prost, S., Machado, A. P., Zumbroich, J., Preier, L., Mahtani-Williams, S., Meissner, R., Guschanski, K., Brealey, J. C., Fernandes, C. R., Vercammen, P., Hunter, L. T. B., Abramov, A. V., Plasil, M., Horin, P., Godsall-Bottriell, L., Bottriell, P., Dalton, D. L.,

Kotze, A., & Burger, P. A. (2022). Genomic analyses show extremely perilous conservation status of African and Asiatic cheetahs (*Acinonyx jubatus*). *Molecular Ecology*, *31*(16), 4208–4223. https://doi.org/10.1111/mec.16577

Prost, S., Winter, S., De Raad, J., Coimbra, R. T. F., Wolf, M., Nilsson, M. A., Petersen, M., Gupta, D. K., Schell, T., Lammers, F., & Janke, A. (2020). Education in the genomics era: Generating high-quality genome assemblies in university courses. *GigaScience*, *9*(6), giaa058. https://doi.org/10.1093/gigascience/giaa058

Purvis, A., Gittleman, J. L., Cowlishaw, G., & Mace, G. M. (2000). Predicting extinction risk in declining species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1456), 1947–1952. https://doi.org/10.1098/rspb.2000.1234

Quality Scores for Next-Generation Sequencing. (o. J.). *(2025) https://knowledge.illumina.com/instrumentation/general/instrumentation-general-reference_material-list/000006557*

Ripple, W. J., Estes, J. A., Beschta, R. L., Wilmers, C. C., Ritchie, E. G., Hebblewhite, M., Berger, J., Elmhagen, B., Letnic, M., Nelson, M. P., Schmitz, O. J., Smith, D. W., Wallach, A. D., & Wirsing, A. J. (2014). Status and Ecological Effects of the World's Largest Carnivores. *Science*, *343*(6167), 1241484. https://doi.org/10.1126/science.1241484

Salazar, A. N., Nobrega, F. L., Anyansi, C., Aparicio-Maldonado, C., Costa, A. R., Haagsma, A. C., Hiralal, A., Mahfouz, A., McKenzie, R. E., Van Rossum, T., Brouns, S. J. J., & Abeel, T. (2020). An educational guide for nanopore sequencing in the classroom. *PLOS Computational Biology*, *16*(1), e1007314. https://doi.org/10.1371/journal.pcbi.1007314

Schmidt-Küntzel, A., Yashphe, S., Hamalwa, H., Ismail, S. H., Tricorache, P., Brewer, B., O'Brien, S. J., & Marker, L. (2024). Genetic support to uplist an African cheetah subspecies, *Acinonyx jubatus soemmeringii*, imperiled by illegal trade. *Conservation Science and Practice*, *6 (1)*. https://doi.org/10.1111/csp2.13052

Seah, A., Lim, M. C. W., McAloose, D., Prost, S., & Seimon, T. A. (2020). MinION-Based DNA Barcoding of Preserved and Non-Invasively Collected Wildlife Samples. *Genes*, *11*(4), 445. https://doi.org/10.3390/genes11040445

Srivathsan, A., Baloglu, B., Wang, W., Tan, W. X., Bertrand, D., Ng, A. H. Q., Boey, E. J. H., Koh, J. J. Y., Nagarajan, N., & Meier, R. (2018). *A MinIONTM-based pipeline for fast and cost-effective DNA barcoding*. https://doi.org/10.1111/1755-0998.12890

Thermofisher.com. (2024). https://www.thermofisher.com/ch/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html

Tricorache, P., Nowell, K., Wirth, G., Mitchell, N., Boast, L. K., & Marker, L. (2018). Pets and Pelts: Understanding and Combating Poaching and Trafficking in Cheetahs. In *Cheetahs: Biology and Conservation* (S. 191–205). Elsevier. https://doi.org/10.1016/B978-0-12-804088-1.00014-9

Tricorache, P., Yashphe, S., & Marker, L. (2021). Global dataset for seized and non-intercepted illegal cheetah trade (*Acinonyx jubatus*) 2010–2019. *Data in Brief*, *35*, 106848. https://doi.org/10.1016/j.dib.2021.106848

Vasiljevic, N., Lim, M., Humble, E., Seah, A., Kratzer, A., Morf, N. V., Prost, S., & Ogden, R. (2021). Developmental validation of Oxford Nanopore Technology MinION sequence

data and the NGSpeciesID bioinformatic pipeline for forensic genetic species identification. *Forensic Science International: Genetics*, *53*. https://doi.org/10.1016/j.fsigen.2021.102493

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *39*(11), 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Watsa, M., Erkenswick, G. A., Pomerantz, A., & Prost, S. (2020). Portable sequencing as a teaching tool in conservation and biodiversity research. *PLOS Biology*, *18*(4), e3000667. https://doi.org/10.1371/journal.pbio.3000667

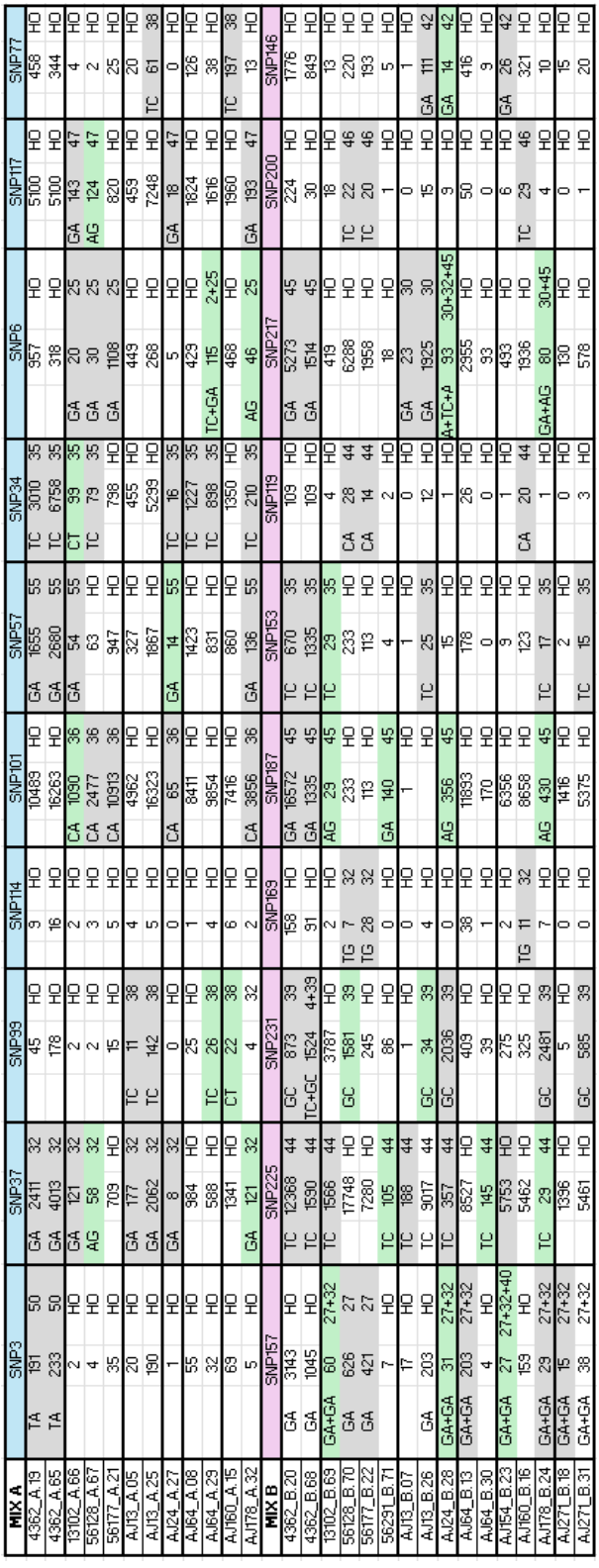
Weise, F. J., Lemeris, J. R., Munro, S. J., Bowden, A., Venter, C., Van Vuuren, M., & Van Vuuren, R. J. (2015). Cheetahs running the gauntlet: An evaluation of translocations into free-range environments in Namibia. *PeerJ*, *3*, e1346. https://doi.org/10.7717/peerj.1346

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, *20*(1), 129. https://doi.org/10.1186/s13059-019-1727-y

Zarate, M. (2022). Native Barcoding Kit 24 with amplicons (SQK-NBD112.24)-minion copy. *Oxford Nanopore Technologies*. dx.doi.org/10.17504/protocols.io.kxygx3qx4g8j/v1

8 Appendices

8.1 Appendix 1



MIX A	SNP3			SNP37		SNP99		SNP114		SNP101		SNP57		SNP34		SNP6		SNP117		SNP77	
4362_A.19	TA	191	50	GA	2411	32		45	HO	3	HO	10483	HO	GA	1655	55	TC	3010	35	357	HO	5100	HO	458	HO
4362_A.65	TA	233	50	GA	4013	32		178	HO	16	HO	16263	HO	GA	2680	55	TC	6758	35	318	HO	5100	HO	344	HO
13102_A.66		2	HO	GA	121	32		2	HO	2	HO	CA	1090	36	GA	54	55	CT	99	35	GA	20	25	GA	143	47	4	HO
56128_A.67		4	HO	AG	58	32		2	HO	3	HO	CA	2477	36		63	HO	TC	79	35	GA	30	25	AG	124	47	2	HO
56177_A.21		35	HO		709	HO		15	HO	5	HO	CA	10913	36		947	HO		798	HO	GA	1108	25		820	HO	25	HO
AJ13_A.05		20	HO	GA	177	32	TC	11	38	4	HO		4962	HO		327	HO		455	HO		449	HO		459	HO	20	HO
AJ13_A.25		190	HO	GA	2062	32	TC	142	38	5	HO		16323	HO		1867	HO		5299	HO		268	HO		7248	HO	TC 61	38
AJ24_A.27		1	HO	GA	8	32		0	HO	0	HO	CA	65	36	GA	14	55	TC	16	35		5	HO	GA	18	47	0	HO
AJ64_A.08		55	HO		984	HO		25	HO	1	HO		8411	HO		1423	HO	TC	1227	35		429	HO		1824	HO	126	HO
AJ64_A.29		32	HO		588	HO	TC	26	38	4	HO		9854	HO		831	HO	TC	898	35	TC+GA	115	2+25		1616	HO	38	HO
AJ160_A.15		69	HO		1341	HO	CT	22	38	6	HO		7416	HO		860	HO		1350	HO		468	HO		1960	HO	TC 197	38
AJ178_A.32		5	HO	GA	121	32		4	32	2	HO	CA	3856	36	GA	136	55	TC	210	35	AG	46	25	GA	193	47	13	HO
MIX B	SNP157			SNP225		SNP231		SNP163		SNP187		SNP153		SNP119		SNP217		SNP200		SNP146	
4362_B.20	GA	3143	HO	TC	12368	44	GC	873	39	158	HO	GA	16572	45	TC	670	35		109	HO	GA	5273	45		224	HO	1776	HO
4362_B.68	GA	1045	HO	TC	1590	44	TC+GC	1524	4+39	91	HO	GA	1335	45	TC	1335	35		109	HO	GA	1514	45		30	HO	849	HO
13102_B.63	GA+GA	60	27+32	TC	1566	44		3787	HO	2	HO	AG	29	45	TC	29	35		4	HO		419	HO		18	HO	13	HO
56128_B.70	GA	626	27		17748	HO	GC	1581	39	TG	32		233	HO		233	HO	CA	28	44		6288	HO	TC	22	46	220	HO
56177_B.22	GA	421	27		7280	HO		245	HO	TG	32		113	HO		113	HO	CA	14	44		1958	HO	TC	20	46	193	HO
56291_B.71		7	HO	TC	105	44		86	HO	0	HO	GA	140	45		4	HO		2	HO		18	HO		1	HO	5	HO
AJ13_B.07		17	HO	TC	188	44		1	HO	0	HO		1	HO		1	HO		0	HO	GA	23	30		0	HO	1	HO
AJ13_B.26	GA	203	HO	TC	9017	44	GC	34	39	4	HO			HO	TC	25	35		12	HO	GA	1925	30		15	HO	GA 111	42
AJ24_B.28	GA+GA	31	27+32	TC	357	44	GC	2036	39	0	HO	AG	356	45		15	HO		1	HO	A+TC+A	93	30+32+45		9	HO	GA 14	42
AJ64_B.13	GA+GA	203	27+32		8527	HO		409	HO	38	HO		11893	HO		178	HO		26	HO		2955	HO		50	HO	416	HO
AJ64_B.30		4	HO	TC	145	44		39	HO	1	HO		170	HO		0	HO		0	HO		93	HO		0	HO	9	HO
AJ154_B.23	GA+GA	27	27+32+40		5753	HO		275	HO	2	HO		6356	HO		9	HO		1	HO		493	HO		6	HO	GA 26	42
AJ160_B.16		159	HO		5462	HO		325	HO	TG	11	32	8658	HO		123	HO	CA	20	44		1936	HO	TC	29	46	321	HO
AJ178_B.24	GA+GA	29	27+32	TC	29	44	GC	2481	39	7	HO	AG	430	45	TC	17	35		1	HO	GA+AG	80	30+45		4	HO	10	HO
AJ271_B.18	GA+GA	15	27+32		1396	HO		5	HO	0	HO		1416	HO		2	HO		0	HO		130	HO		0	HO	15	HO
AJ271_B.31	GA+GA	38	27+32		5461	HO	GC	585	39	0	HO		5375	HO	TC	15	35		3	HO		578	HO		1	HO	20	HO

The table in Appendix 1 shows a more detailed Table 5 (3.2 Bioinformatics). Here we see an overview of SNP sites within samples. In each cell, left column indicates the present nucleotides

in heterozygous sites, middle column indicates read numbers, right column shows if SNP sites were heterozygous (HE) or homozygous (HO). Read numbers lower than 7 were not interpreted because of the low coverage (LC). Grey fields show a successful visual verification of heterozygous sites called by ANGSD. Green fields show sites called by ANGSD that did not withstand the visual verification and could only show 10 – 30% of heterozygous sites.

8.2 Appendix 2

8.2.1 NanoPlot statistics report: Summary statistics

General summary:

Mean read length:	424.4
Mean read quality:	11.6
Median read length:	373.0
Median read quality:	12.4
Number of reads:	478,828.0
Read length N50:	471.0
STDEV read length:	3,198.4
Total bases:	203,205,193.0

Number, percentage and megabases of reads above quality cutoffs

>Q5: 463837 (96.9%) 183.2Mb

>Q7: 423208 (88.4%) 166.6Mb

>Q10: 372203 (77.7%) 150.3Mb

>Q12: 274550 (57.3%) 112.8Mb

>Q15: 24168 (5.0%) 8.4Mb

Top 5 highest mean basecall quality scores and their read lengths

1: 36.6 (43)

2: 36.3 (24)

3: 35.7 (10)

4: 35.4 (43)

5: 28.4 (22)

Top 5 longest reads and their mean basecall quality score

1: 968629 (5.5)

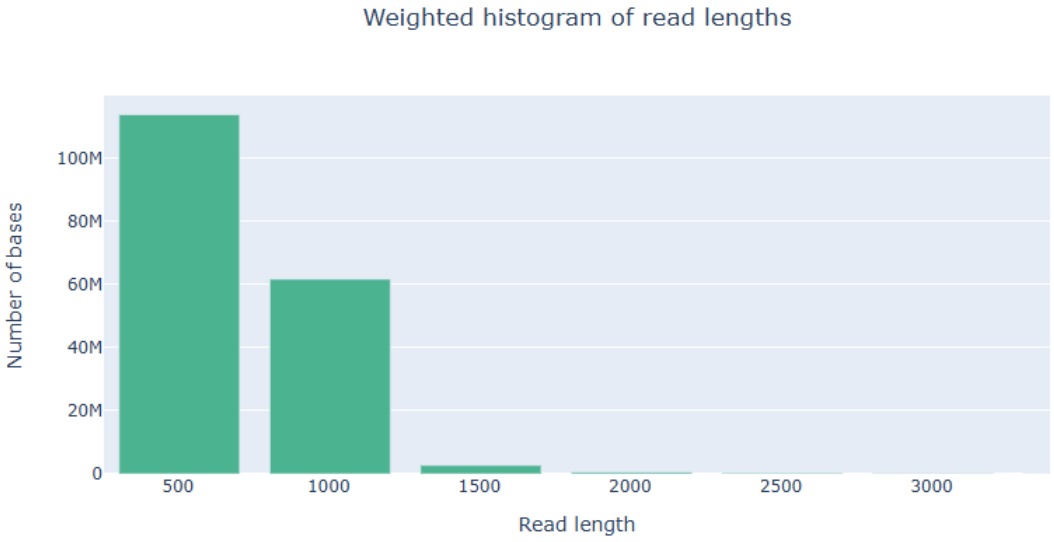
2: 833865 (3.4)

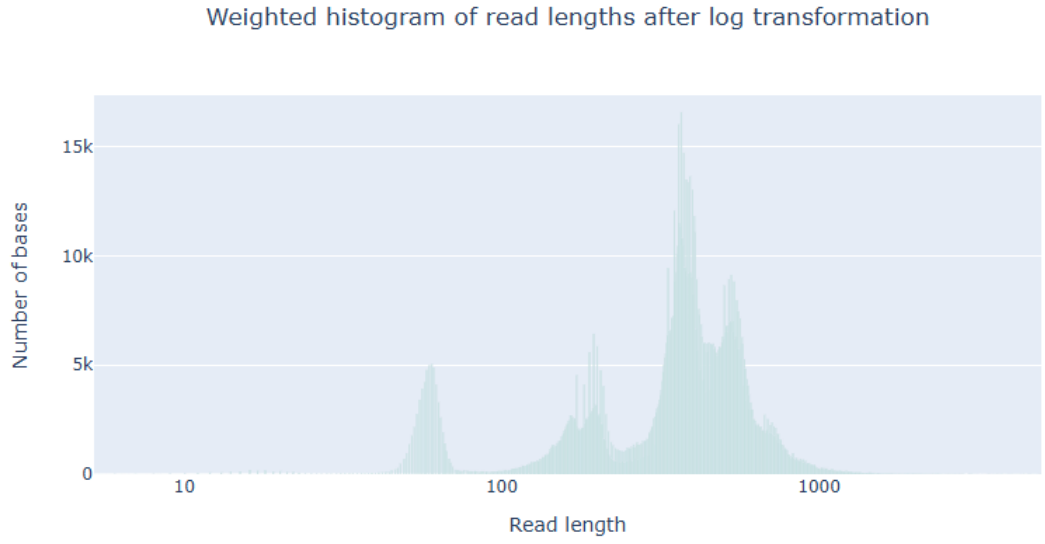
3: 473014 (6.1)

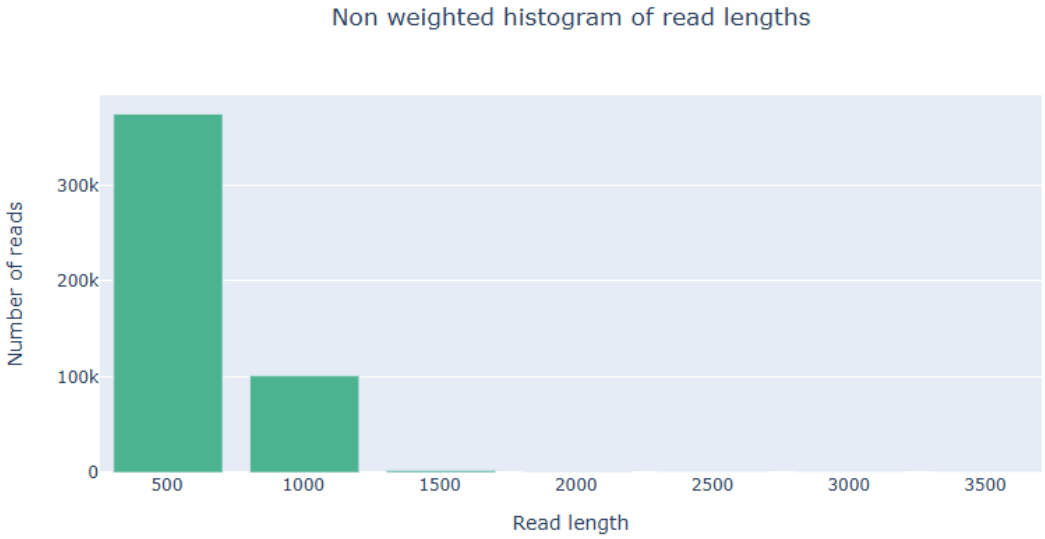
4: 449320 (5.1)

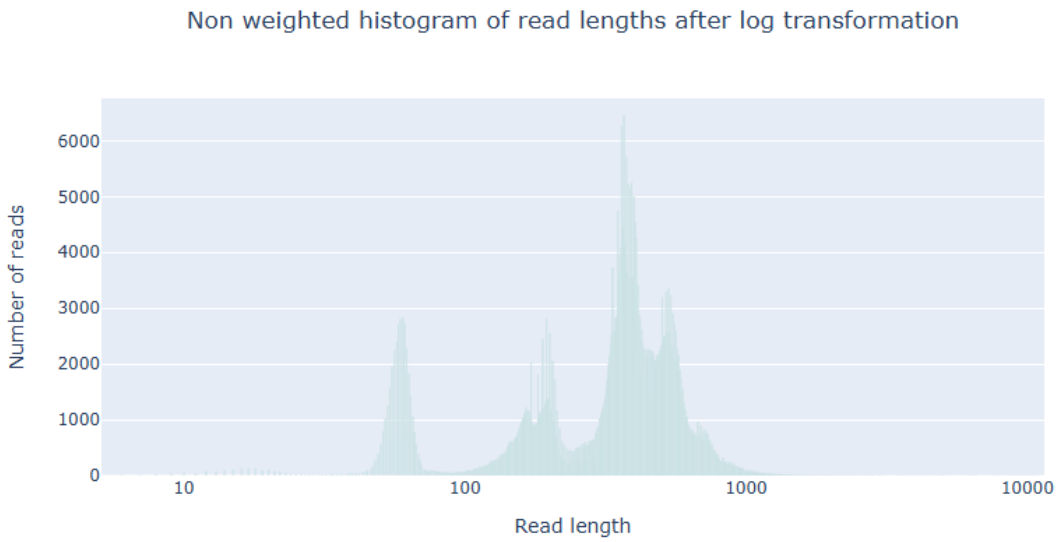
5: 440637 (3.6)

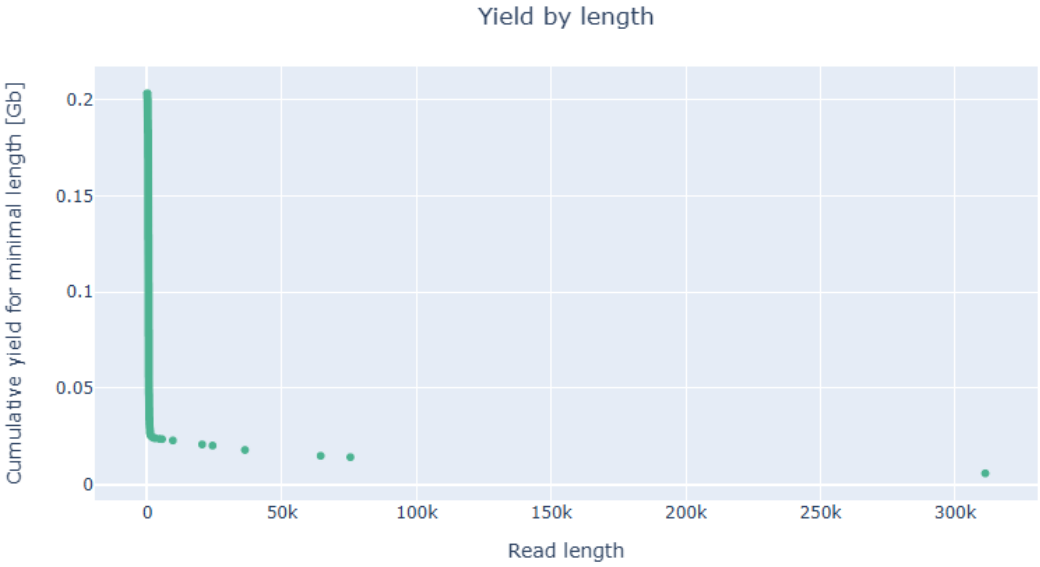
8.2.2 NanoPlot statistics report: Plots

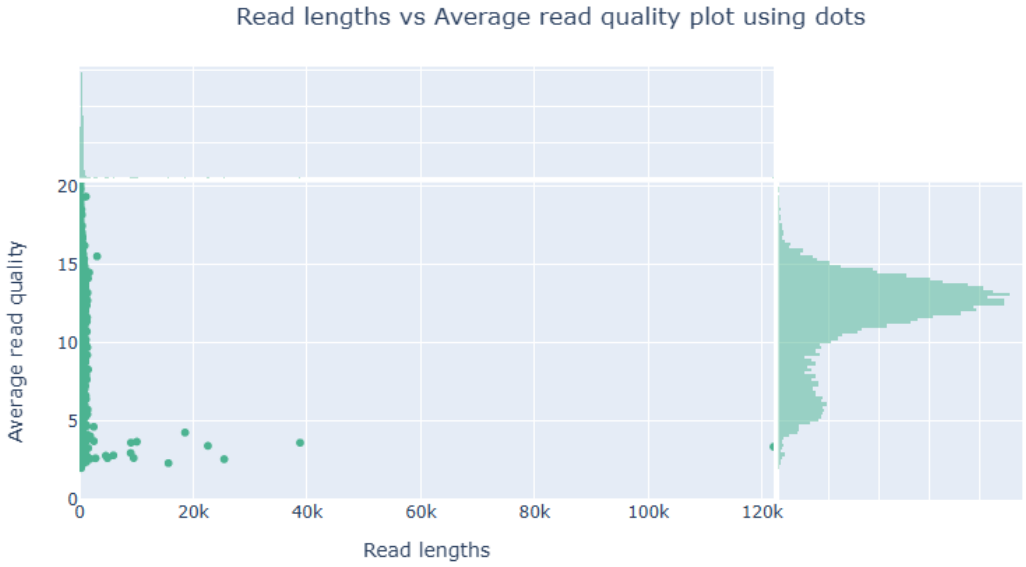


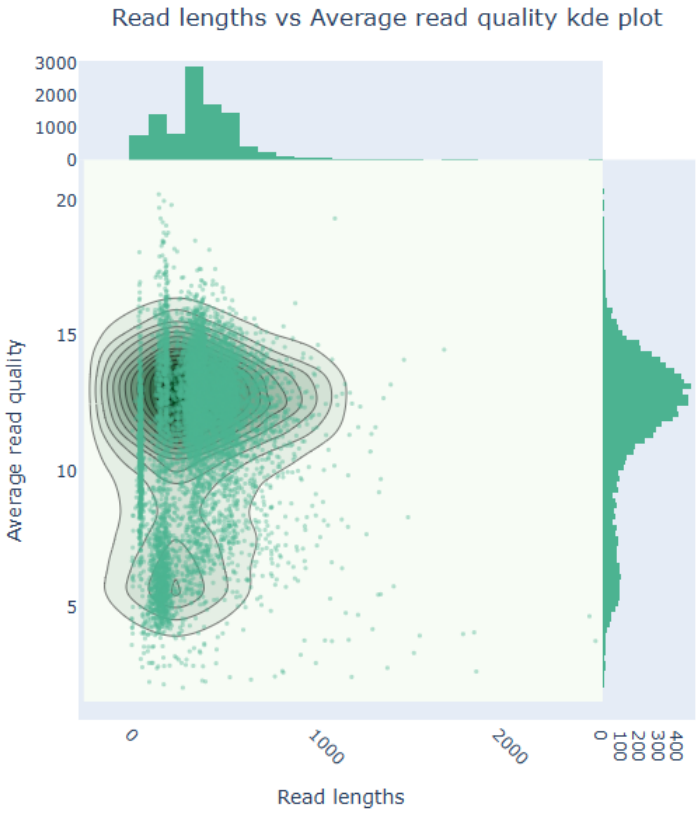



Non weighted histogram of read lengths
Number of reads
300k
200k
100k
0
500
1000
1500
2000
2500
3000
3500
Read length


Non weighted histogram of read lengths after log transformation
Number of reads
6000
5000
4000
3000
2000
1000
0
10
100
1000
10000
Read length


Yield by length
Cumulative yield for minimal length [Gb]
0.2
0.15
0.1
0.05
0
0
50k
100k
150k
200k
250k
300k
Read length


Read lengths vs Average read quality plot using dots
Average read quality
20
15
10
5
0
0
20k
40k
60k
80k
100k
120k
Read lengths


Read lengths vs Average read quality kde plot
3000
2000
1000
0
20
15
10
5
Average read quality
0
1000
2000
0
100
200
300
400
Read lengths