Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Forest Science

Pride and prejudice

– Extra-group paternity in lions and the effects of marker density on kinship and relatedness estimates

Lejon och lögner

– Effekterna av markördensitet för analyser av föräldraskap- och släktskapsanalyser

Julia L. Jansson

Examensarbete i ämnet biologi
Department of Wildlife, Fish, and Environmental studies
Umeå
2017

Pride and prejudice

- Extra-group paternity in lions and the effects of marker density on kinship and relatedness estimates

Lejon och lögner

- Effekterna av markördensitet för analyser av föräldraskap- och släktskapsanalyser

Julia L. Jansson

Supervisor: Göran Spong, Dept. of Wildlife, Fish, and Environmental Studies

Assistant supervisor: Anita Norman, Dept. of Life Sciences, San Diego Zoo

Examiner: Carl-Gustaf Thulin, Dept. of Wildlife, Fish, and Environmental Studies

Credits: 30 HEC
Level: A2E
Course title: Master degree thesis in Biology at the Department of Wildlife, Fish, and Environmental Studies
Course code: EX0633
Programme/education:

Place of publication: Umeå
Year of publication: 2017
Cover picture: Julia L. Jansson
Title of series: Examensarbete i ämnet biologi
Number of part of series: 2017:11
Online publication: http://stud.epsilon.slu.se

Keywords: Lion, Mating structure, Extra-group Paternity, Single Nucleotide Polymorphism (SNP), Marker density, *Panthera leo*

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Forest Science
Department of Wildlife, Fish, and Environmental Studies

Abstract

The development effective high-throughput sequencing techniques has allowed further insight to the mating systems of wild animal populations and challenged our perception that reproductive behaviour can be predicted from social structure. A recent study of lions *(Panthera leo*) in Etosha National Park, Namibia, found evidence of extra-group paternity (EGP) which has never before been confirmed among lions. Relatedness estimated from earlier research on lions in Selous Game Reserve, Tanzania, suggests that this population also deviates from a strict within-pride mating system. In this study, the same population of wild lions from Selous Game Reserve was visited two times, with twelve years in between, and with the aid of RAD-sequenced SNP markers, pedigrees were reconstructed and pairwise relatedness calculated to explore their mating system and compare with their social structure. Additionally, the effect of marker density (i.e. number of markers) on the precision of the kinship and relatedness estimates was explored. Evidence of EGP was found in the Selous population; of the paternal relationships detected 60% (n=15), were between members of different prides. This result calls for more studies of the reproductive behaviours of the different lion populations that remain to see how the behaviour is distributed. Increasing marker density positively affected the accuracy of the pairwise relatedness estimates in the range of 100-1000 SNPs. In the context of parentage analysis increasing the number of markers lead to relatedness values of the relationships detected to have a mean closer to the theoretically expected value of 0.5, however the number of relationships detected decreased considerably with marker densities above 200. Thus, future studies conducting parentage analyses should not simply use as many markers as possible, but instead consider the optimal number markers in the context of their data.

1 Introduction

Knowledge of reproductive behaviour is fundamental to our understanding of the social and genetic structure of wild populations and therefore also to the application of wildlife management, especially when it comes to the management of endangered species. Research has shown that observational data is unreliable when studying animal mating systems. Thus, the discovery of molecular methods of paternity analysis completely revolutionized the field (Clutton-Brock and Isvaran 2006, Lyke et al. 2013). The recent development of high-throughput sequencing technology has greatly increased the possibilities for studies of kinship and relatedness, enabling fast and cost-effective sequencing. High-throughput sequencing has made the development of large numbers of markers feasible even for non-model species (Grover and Sharma 2016). The optimal number of markers for these types of analyses is not clear, however, and in this study the effects of marker density (i.e. number of markers) on the precision of kinship and relatedness estimates are explored by investigating the mating structure of the lions (*Panthera leo*) of Selous Game Reserve, Tanzania.

1.1 Lion social structure

Lions are by far the most social member of the Felidae family, and their social organization can be divided into two types: residents and nomads (Schaller 1972). Resident lions form groups called prides with between 4 and, in extreme cases, 30 members with mean size of around 15. The females make up the stable core of the group and are usually directly related to each other; they normally stay with the same pride in the same area their whole life. Male lions are more transitory members of the pride. They form coalitions of about 2-4 adult males that take over a pride and stay until they are ousted by another coalition (Schaller 1972). The exception is the Tsavo population in Kenya where there is only one male per pride (Lyke et al. 2013). Prides can scatter widely within the pride area splitting into smaller groups but the pride is still a distinct unit and trespassers are not welcome. However, both female and male lions have been noted to be more aggressive towards intruders of their own sex when protecting their territory (Schaller 1972).

Nomadic lions do not hold a territory but and wander in a larger range either solitarily or in coalitions (Schaller 1972). For a male lion being nomadic is a normal part of their life history as they are pushed out of their natal pride when they become sexually mature at around 3.5 to 4 years of age (Schaller 1972). About 80% of females are philopatric and stay with their natal pride, the remaining lionesses get pushed out at around the same age as the males, often settling in the outskirts of their natal pride area (Schaller 1972, Spong et al. 2002).

1.2 Extra-group paternity (EGP)

During his extensive study of the lions of Serengeti, Schaller (1972) suggested that resident lionesses may court with nomadic males and males of neighbouring prides as a means to avoid inbreeding. However, when parentage of lions could first be tested genetically no evidence of breeding outside the confines of the pride, i.e. extra-group paternity (EGP) could be found in a sample of 78 cubs in 11 prides (Gilbert et al. 1991). This was in agreement with the general view at the time; that EGP was unimportant in mammalian mating systems. In stark contrast to the studies on birds, only low levels of EGP could be detected when paternity was first analysed in social mammals during the 1990's (Clutton-Brock and Isvaran 2006). Since then many social mammals have been found to exhibit extra-group copulatory behaviour (Clutton-Brock and Isvaran 2006) and when Lyke et al. (2013) tested the paternity of lion cubs in Etosha National Park, Namibia, EGP occurred at a frequency of 41% and in five of seven prides where paternity was analysed. Spong et al. (2002) also found, in Selous Game Reserve, Tanzania, that measures of relatedness deviated from what would be expected if prides members mated exclusively with each other.

1.3 Parentage analysis

The current procedure of molecular parentage analysis can be summarized into the following steps: collection of DNA in form of for example tissues or faeces, DNA sequencing, and assigning parentage using molecular markers to identify and pair the genotypes of offspring and their parents with available software packages (Jones et al. 2010). Since their discovery in 1987, microsatellites have been the marker of choice for most parentage analyses. However, the recent advancements in the field of high-throughput sequencing has resulted in an increased use of single nucleotide polymorphisms (SNPs, Grover and Sharma 2016). SNPs are, as their name suggests, polymorphisms at single nucleotide loci in the genome (Grover and Sharma 2016). Only the four bases of DNA are possible alleles for a SNP which leads to a much lower resolving power than microsatellites, especially since a majority of SNPs are biallelic (Hauser et al. 2011). As a result, a higher number of markers are needed for a parentage analysis using SNPs compared to microsatellites.

Nevertheless, due to the efficiency of screening, easy transferability between laboratories and low genotyping error, recent studies have found that SNPs exceed the performance of microsatellites in parentage analyses. Hauser et al. (2011) found that SNPs had a higher assignment success than microsatellites in their parentage and kinship analyses in a wild sockeye salmon (*Oncorhynchus nerka*) population and similar results were found by Labuschagne et al. (2015) in a study on a captive population of African penguins (*Spheniscus demersus*). The authors of both studies note however, that even a low error rate could become a problem if a very large number of loci are screened as multilocus error scales linearly with each added locus while the explaining power for each locus decreases beyond a certain point (Hauser et al. 2011, Labuschagne et al. 2015). The number of markers (i.e. marker density) needed to reliably estimate kinship and relatedness has indeed been a topic for debate (Glaubitz et al. 2003, Städele and Vigilant 2017).

1.4 RAD-sequencing

With the constant advancement of high-throughput sequencing techniques, getting large numbers of markers sequenced poses a diminishing problem, and the relevance of data reduction and strict filtering to preserve only the data with the highest quality becomes increasingly important (Norman et al. 2013). The technique used in this study; Restriction site-associated DNA sequencing (RAD-Seq; Davey & Blaxter 2011) reduces the complexity of the genome using restriction enzymes that cuts the DNA into fragments at specific sites. By choosing only fragments with certain lengths it allows the same subsets of the DNA to be sequenced in each individual, thereby creating reduced representation libraries (RRL). This method of SNP-discovery makes population-scale sequencing of species without a reference genome feasible at a fraction of the cost of whole genome sequencing *(*Davey et al. 2013).

1.5 Aim

In this study, the same population of lions from Selous Game Reserve was visited twice, in 1997-1998 and 2011-2012, with the goal of further exploring the mating system and social structure of these big cats using cutting-edge molecular methods. The effects of marker density on the precision of the kinship and relatedness estimates were also investigated.

2 Method

2.1 Study site and sample collection

Sample collection was carried out in the northern sector of Selous Game Reserve in south-eastern Tanzania (73*5'S, 38*15' E). The study site covers ca 1000 km^2 and is dominated by wooded savannah, thorn woodland, and *Brachystegia* and *Combretum* thickets. Samples were collected from a wild population of lions in 1997-1998 and 2011-2012.

Individual lions were identified by their whisker spot patterns and with the help of a picture library in addition to other physical characteristics (Spong et al. 2002). In total 84 lions in 21 prides were identified and used in the analyses during the study. Tissue biopsies for DNA extraction were collected from the lions using a CapChur CO_2 pistol with biopsy darts developed from plans by Karech et al. (1987) and the tissue samples were stored in 100 mM EDTA 95% ethanol solution at ambient temperature in the field and at -20 °C in the laboratory. No samples were taken from cubs younger than 1 year old as the risk of injury was considered too high.

For tracking purposes, one female per pride was fitted with a radio collar (MOD-500 Telonics, Mesa, AZ, USA). In such cases, the lioness was sedated by darting her with a Telinject (Saugus, CA, USA) (Vario 3V) CO_2 rifle with 200 mg Telazol (Tiletamine-zolzapam) and 100 mg Rompun (Xylazine) from 10-15 m away from inside a vehicle. When the collar was in place and the lioness showed signs of recovery, the effect of the Rompun was reversed by injecting 17 mg of Yohimbine restoring the lion to full mobility within two hours.

2.2 DNA extraction and RAD-sequencing

In the laboratory, DNA extraction was performed on the tissue samples using the QIAsymphony SP and QIAsymphony DNA kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. A spectrophotometer (NanoDrop, Thermo Fisher, Massachusetts, USA) was used to assess nucleotide purity and quantity. Additionally, the DNA quality was visually assessed by gel electrophoresis with the Kodak Electrophoresis Documentation and Analysis system 120 (Eastman Company, Rochester, USA).

Using the restriction enzyme EcoR1 (Thermo Fisher, Massachusetts, USA), 1 to 5 μg of each DNA sample was digested, for 16 hours according the manufacturer's instructions. This cuts the DNA into fragments at specific sites. The quality of the result was assessed visually with another gel electrophoresis. The samples were sent to the Science for Life Laboratories (SciLifeLab, Stockholm, Sweden), which performed the subsequent library construction, preparation, and sequencing. As part of the RAD-seq procedure, the complexity of the genome is reduced by only sequencing fragments of a certain length (Davey and Blaxter 2010). In this case fragments between 300 to 600 base pairs long were extracted and blunt end repaired. Paired-end, multiplexed adapters, that enable the fragments to bind to primers in the sequencing process (Davey and Blaxter 2010), were ligated to the fragments. Equimolar concentrations were measured and the sequencing was carried out on one lane of Illumina HiSeq2000. The result was sorted and aligned in Stacks (version 1.44, Catchen et al. 2011, Catchen et al. 2013) and SNP detection was performed using with the settings: m 3, M 2.

2.3 Marker selection

According to Hosking et al. (2004) loci that are not in Hardy-Weinberg Equilibrium (HWE) are more likely to harbour genotyping error, therefore GENEPOP (version 4.2, Raymond and Rousset 1995, Rousset 2008) was used to perform a Hardy-Weinberg exact test and only loci that were in HWE ($p<0.1$) were used in further analyses. For similar reasons, non-biallelic SNPs were filtered out and were not included in any analyses.

Only the 1000 loci with the highest minor allele frequency (MAF) were used to calculate pairwise relatedness and the parentage analyses as markers with very low MAF contain little information. For calculating the inbreeding coefficients all loci in that were in HWE ($p<0.1$) were used. Homozygosity was plotted against MAF to locate any outliers or unexpected patterns, such as a high homozygosity coupled with a high MAF (Figure 1). Too many homozygotes is probably an indication of high allelic dropout.

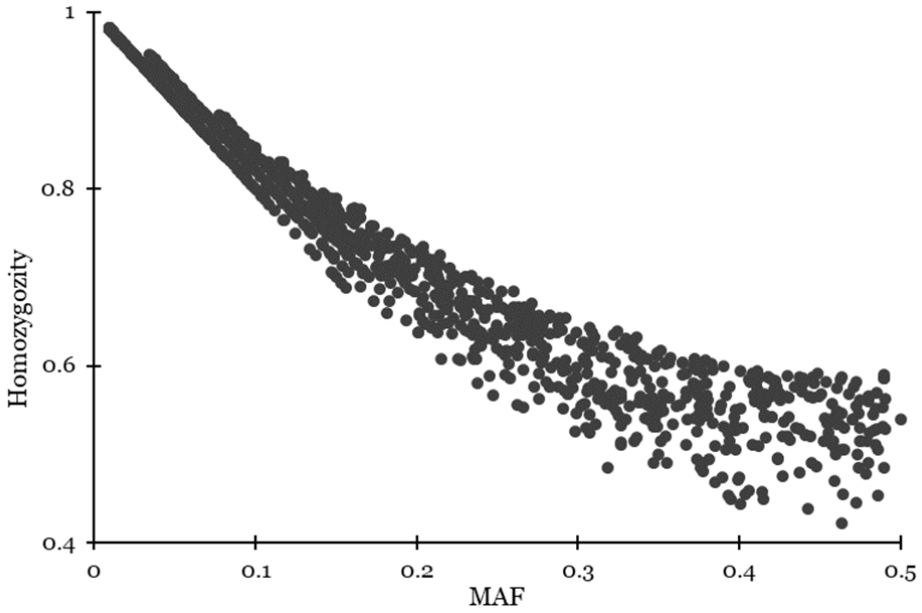

Figure 1. Homozygosity against minor allele frequency (MAF) of the loci.

2.4 Relatedness

In the software R (version 3.3.2) the package related (version 1.0) was used to calculate pairwise relatedness (R) between all individual lions (Lynch and Ritland 1999, Pew et al. 2015). To explore how the number of markers affected the result, ten different levels of markers were used: starting at 100 and adding 100 each iteration up to 1000 markers. The markers were sorted from highest to lowest MAF so that all iterations contained the markers with the highest MAF to retain the most amount of information. Four parent-offspring relationships from the parentage analysis that were fairly certain were chosen based on what was plausible in relation to pride structure and based on field observations, and their R-values as well as the variance in R between them were plotted against the level markers.

Pride structure was investigated using the pairwise R-values of the analysis with 1000 markers. Female to female R-values within prides were compared to female to female R-values between prides with a Welch Two Sample t-test carried out in R. Additionally, the mean relatedness of males to their pride-females was compared to that of the females within prides with a Welch Two Sample t-test in R. The analyses included only adult lions, 4 years and older, as including younger lions would result in parents being paired with their cubs, with which they are obviously related, resulting biased estimates.

2.5 Parentage analysis

The parentage analysis was conducted in the software FRANz (version 2.0.0, Riester et al. 2009), which is a pedigree reconstruction program designed to handle codominant genomic markers, such as SNPs and microsatellites, from wild populations with multiple, overlapping generations (Riester et al 2009). The algorithm can take into account metadata such as: sex, year of birth and death, and known relationships. FRANz uses Markov Chain Monte Carlo (MCMC) sampling to estimate the statistical confidence and Simulated Annealing to find the Maximum Likelihood Pedigree as described in Almudevar (2003, Riester et al. 2009).

In this study, the sex and year of birth were known for most individuals. For the remaining lions, it was estimated from the picture library. The age of a lion can be determined by examining a combination of traits such as: body size, nose pigmentation, wear of teeth, accumulation of scars and mane development in male lions (Miller et al. 2016). To enhance the accuracy of the parentage analysis the span of years in which males and females are able to reproduce can be set in FRANz.

According to Schaller et al. (1972), female lions generally begin reproducing earliest at the age of 3 and male lions at the age of around 3.5 to 4 years. Thus, 3 years was used for females and 4 for males. The oldest lion in this study was 14 years old, and according to Schaller et al. (1972), breeding seems to decline at around 10-15 years. Thus, 15 years was set as the upper limit for reproduction.

Genotyping errors are one of the main factors affecting the power of parentage analyses. In FRANz the error rate is used to determine the number of mismatches that will be allowed between a putative parent and an offspring before exclusion (Riester et al. 2009). Using three animals that were sampled at two separate occasions, the error rate could be estimated by comparing the two samples of each individual and counting each time an allele differed between them. To get the mean error rate for all used loci, the number of mismatching alleles was divided by the total number of alleles, and then divided by two as both samples are as likely to contribute with error.

In case of multiple samples of the same individual, only the sample of with the least amount of missing data was kept in the analysis. Due to errors in identification during sampling in the field or mistakes in the laboratory, a few individuals that were marked as different animals displayed high R-values in the relatedness analysis indicating that they were the same lion. As one could not be certain as to which one of the individuals was the one with correct metadata, most of these animals had to be removed. However, in a couple of cases, due to prior knowledge of for example a sibship-relationship, the right individual could be discerned by means of their relatedness to other lions. One individual was missing over 90% of the loci due to bad genotyping and was therefore excluded from the analysis.

To find the optimum number of markers to use in the parentage analysis, the results from the analysis of relatedness were used. The parentage analysis was run with 11 different levels of marker densities: 50, 100, 200 and then 100 was added each iteration up to 1000 markers. As when conducting the analysis of relatedness, the loci were sorted by MAF, so that the SNPs with the highest MAF were always included in the analysis. From each of the iterations the maximum likelihood pedigree was examined and the number of parents assigned as well as the mean R-value from the analysis of relatedness assigned to these parent-offspring pairs were extracted. The R-values used were from the iteration with 1000 markers as these seemed most precise (more on this in the Relatedness section of the results). Both the number of parents and the R-values were plotted against the number of markers used.

2.6 Inbreeding

The inbreeding coefficient, F_{is}, was calculated for the samples collected in 1997/98 and 2011/12 separately using GENEPOP (version 4.2, Raymond and Rousset 1995, Rousset 2008) and the results were compared with a Welch Two Sample t-test in R (version 3.3.2).

3 Results

3.1 Marker selection

A total of 2757 SNPs were sequenced with a minimum average read depth of 10 per SNP. Out of those, 2454 were biallelic. The Hardy-Weinberg exact test showed that 1800 loci were in HWE ($p < 0.1$). The homozygosity plotted against MAF was deemed to look reasonable enough to continue with the data. After sorting the markers for highest MAF, the first 1000 had a frequency between 0.5 and 0.06.

3.2 Relatedness

The four parent-offspring relationships chosen had R-values that varied between 0.42 to 0.64 at 100 loci and 0.45 to 0.54 at 1000 loci (Figure 2). The variation in R-value decreased with an increasing

number of loci (Figure 3), and they also seem to be getting closer to 0.5 (Figure 2), which is the expected R-value between a parent and their offspring (Städele and Vigilant 2017). Thus, a higher number of markers should yield an R-value closer to reality and the values from the iteration with 1000 markers were used in further analyses.

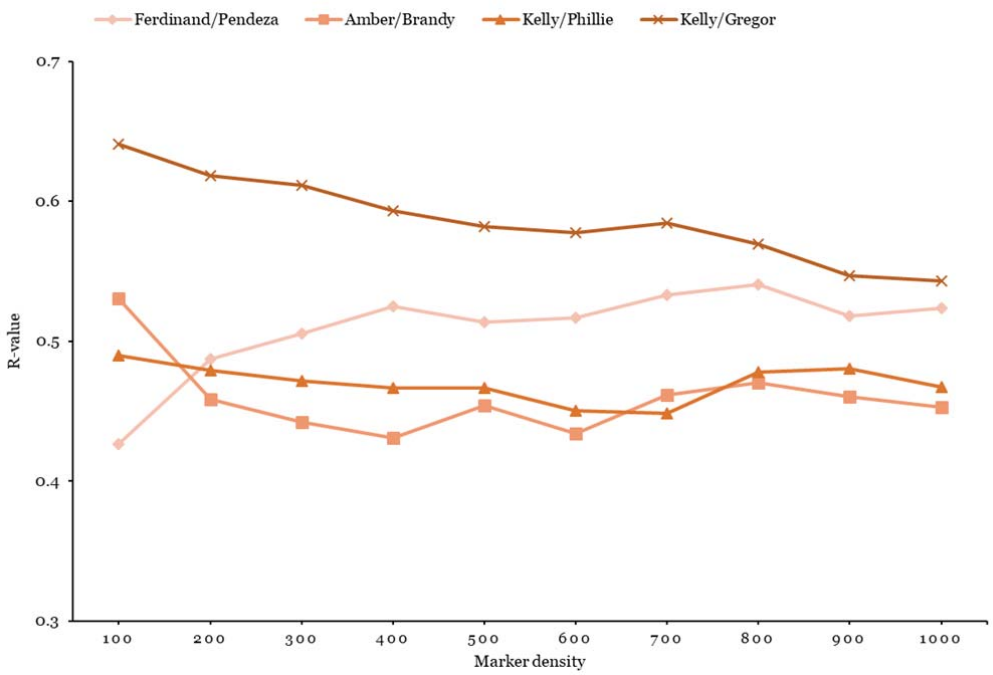

Figure 2. The R-values of four parent-offspring relationships (Parent/Offspring) against the number of markers used in the analyses.

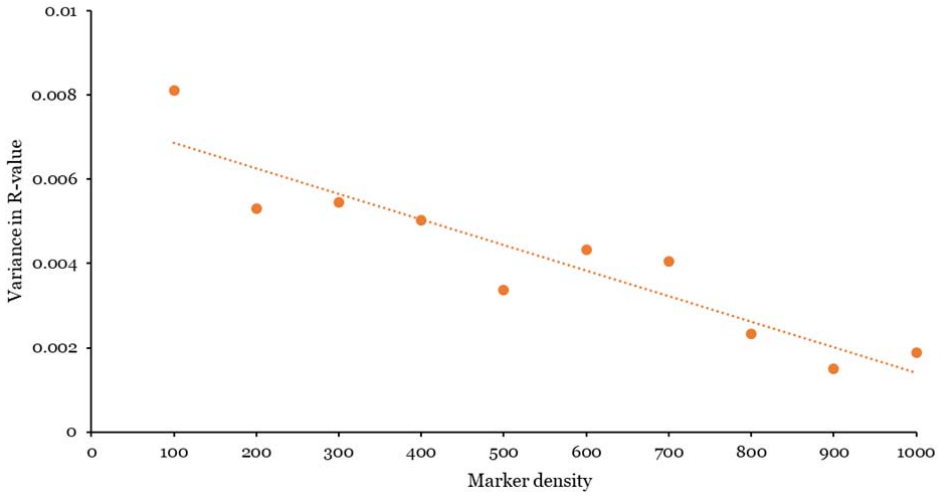

Figure 3. The variance in R between the four parent-offspring relationships (Figure 2) against the number of markers used in the analysis of relatedness. Trendline is a linear regression described by the equation: $y=-6^{-6}+0.0075$ ($R^2 = 0.8536$).

The adult female lions within a pride were significantly more related to each other compared to females between prides, with the mean R-value within a pride being 0.051 and between prides -0.019 ($p = 0.001$, Welch Two Sample t-test, Figure 4). The difference between the mean R-value of the adult males to the adult females within prides, 0.070, and between the adult females within prides, 0.051, was not significant ($p = 0.639$, Welch Two Sample t-test).

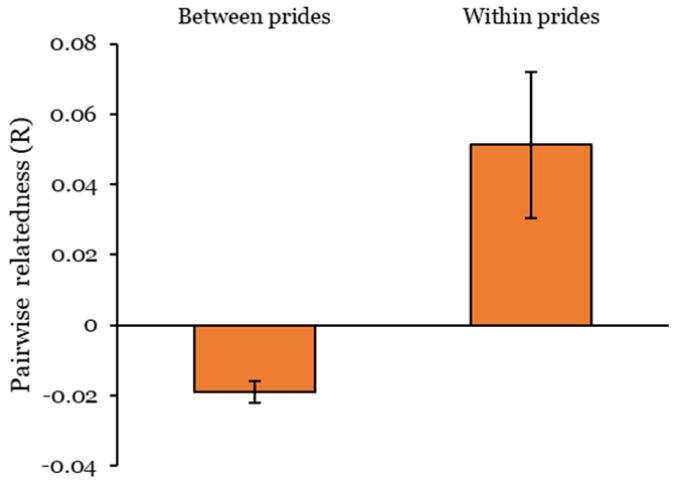

Figure 4. Mean pairwise relatedness of adult females between prides and between adult females within prides. Error bars represent standard ±1 error of the mean.

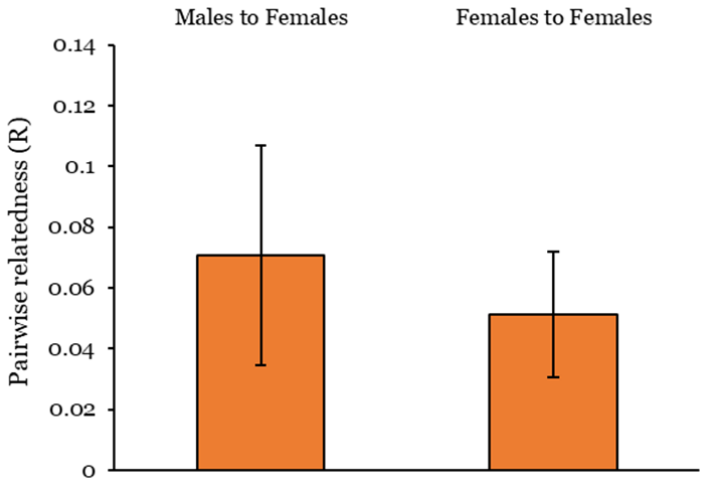

Figure 5. Mean pairwise relatedness of adult females to males within prides and adult females to females within prides. Error bars represent standard ±1 error of the mean.

3.3 Parentage analysis

The genotyping error rate was estimated to be 0.017, this was used in all parentage analyses and no certain loci were found to more error-prone. The number of parent-offspring relationships detected in addition to the mean R-value of these relationships in each of the 11 analyses are presented in Figure 6. To achieve a pedigree with the highest number of parents assigned, with the R-value closest to 0.5, the pedigree generated with 300 markers was chosen. It is to be noted that the individuals that were assigned as parent and offspring did not tend to change with the changing marker density, except for in a couple of cases.

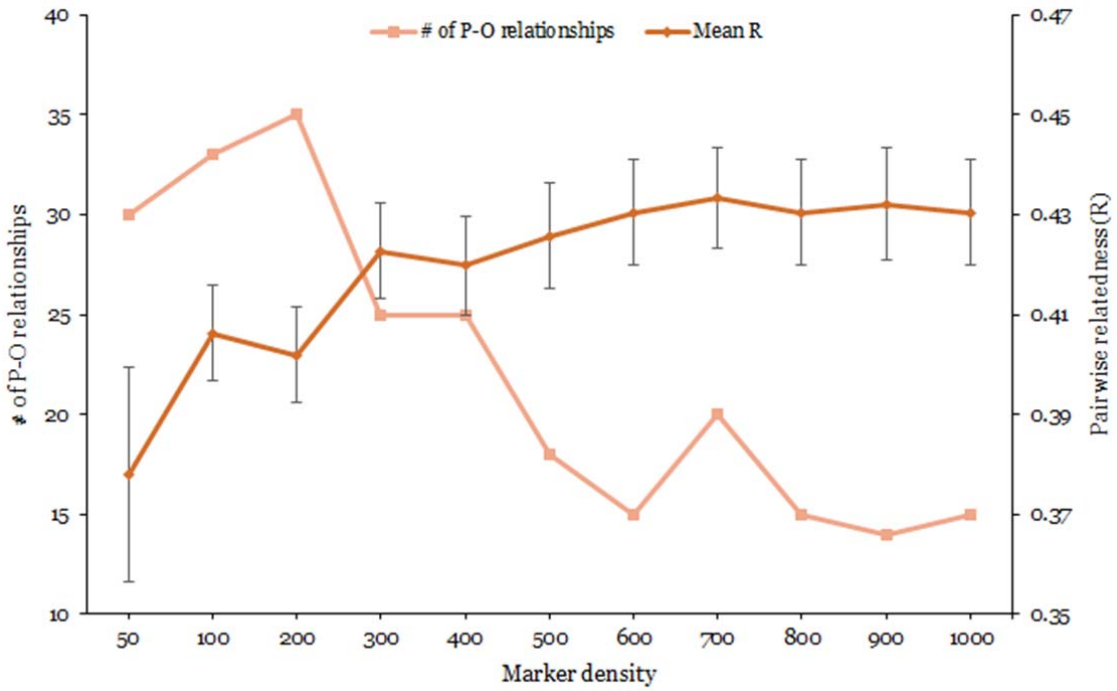

Figure 6. The light line with squares depicts the number of parent-offspring relationships and the dark line with the diamond markers depicts the mean R-value of the parent-offspring relationships against the number of markers used in each parentage analysis. Error bars represent standard ±1 error of the mean.

About one third of the study population was successfully genotyped, and among those 25 parent-offspring relationships were detected in the analysis using 300 markers (Figure 5). Out of those 15 relationships were paternal and 10 were maternal. The mean pairwise relatedness of all relationships, calculated with 1000 markers, was 0.423. Out of the paternal relationships 9 (60%) were detected between individuals from different prides and in 6 (40%) cases they were from the same pride. Among the maternal relationships, 1(10%) was found between prides, in 1 (10%) case the female was part of a pride and the offspring was observed as nomadic and in 8 (80%) cases offspring and mother were observed in the same pride.

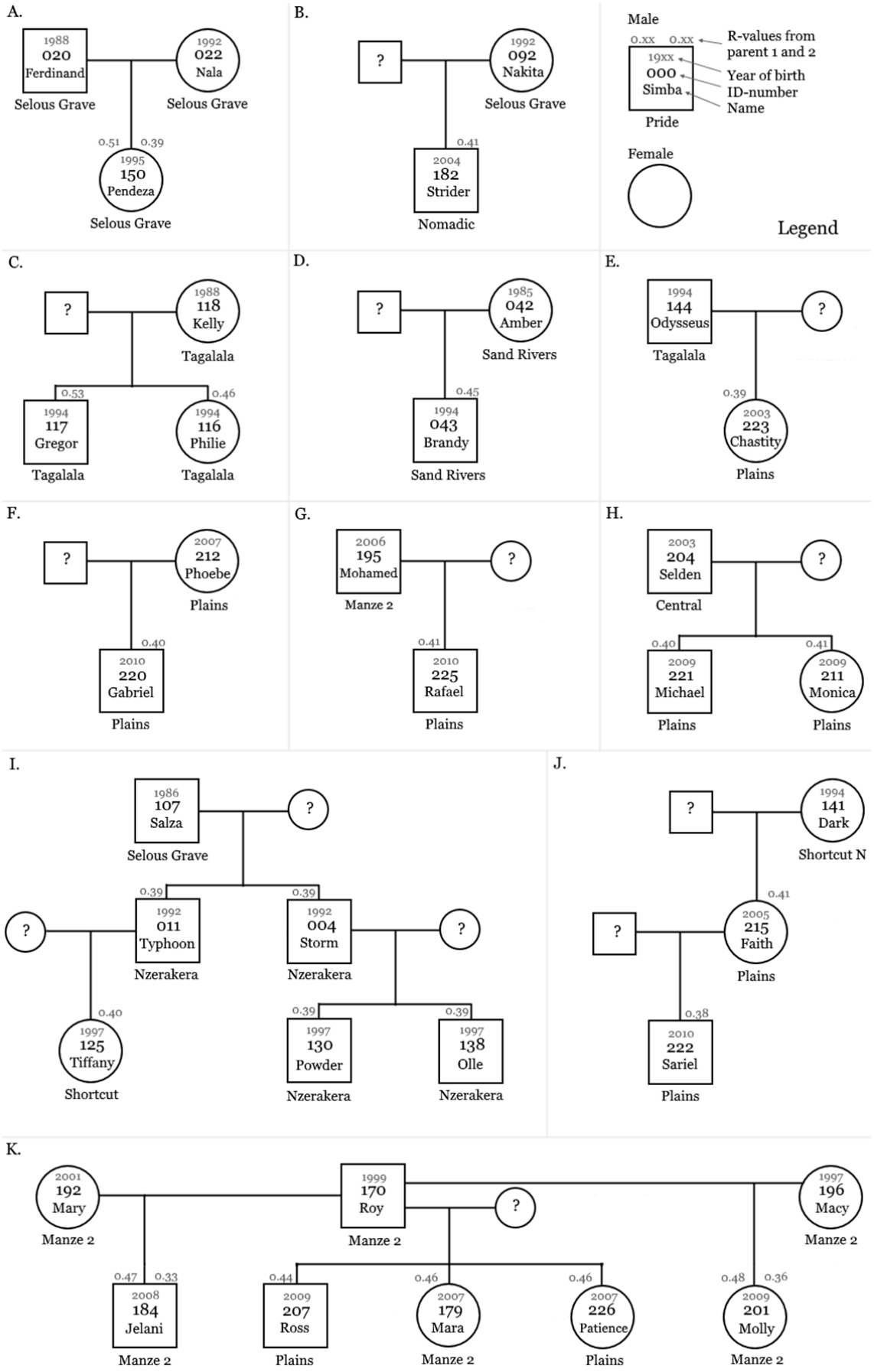
A.
1988
020
Ferdinand
Selous Grave
1992
022
Nala
Selous Grave
0.51
0.39
1995
150
Pendeza
Selous Grave
B.
?
1992
092
Nakita
Selous Grave
0.41
2004
182
Strider
Nomadic
Male
0.xx 0.xx
R-values from parent 1 and 2
19xx
000
Simba
Year of birth
ID-number
Name
Pride
Female
Legend
C.
?
1988
118
Kelly
Tagalala
0.53
0.46
1994
117
Gregor
Tagalala
1994
116
Philie
Tagalala
D.
?
1985
042
Amber
Sand Rivers
0.45
1994
043
Brandy
Sand Rivers
E.
1994
144
Odysseus
Tagalala
?
0.39
2003
223
Chastity
Plains
F.
?
2007
212
Phoebe
Plains
0.40
2010
220
Gabriel
Plains
G.
2006
195
Mohamed
Manze 2
?
0.41
2010
225
Rafael
Plains
H.
2003
204
Selden
Central
?
0.40
0.41
2009
221
Michael
Plains
2009
211
Monica
Plains
I.
1986
107
Salza
Selous Grave
?
0.39
0.39
?
1992
011
Typhoon
Nzerakera
1992
004
Storm
Nzerakera
?
0.40
0.39
0.39
1997
125
Tiffany
Shortcut
1997
130
Powder
Nzerakera
1997
138
Olle
Nzerakera
J.
?
1994
141
Dark
Shortcut N
0.41
?
2005
215
Faith
Plains
0.38
2010
222
Sariel
Plains
K.
2001
192
Mary
Manze 2
1999
170
Roy
Manze 2
?
1997
196
Macy
Manze 2
0.47
0.33
0.44
0.46
0.46
0.48
0.36
2008
184
Jelani
Manze 2
2009
207
Ross
Plains
2007
179
Mara
Manze 2
2007
226
Patience
Plains
2009
201
Molly
Manze 2

Figure 7. Pedigrees of the parent-offspring relationships detected in the parentage analysis. Sibling-relationships are not analysed.

3.4 Inbreeding

The mean inbreeding coefficients for the populations in 1997/98 and 2011/12 were 0.03 and 0.02 respectively and the difference between them was not significant (p=0.13, Welch Two Sample t-test).

4 Discussion

4.1 Mating system and pride structure

The high rates of males fathering cubs in other prides than their own supports the novel findings of Lyke et al. (2013) that lions indeed exhibit EGP, and breed outside the bounds of their apparent mating system. For example, the male lions Roy and Mohamed of pride Manze 2 and Selden of pride Central have all fathered cubs in the Plains pride even though, according to observations from other scientists at the site, the Plains pride had another resident male called Packer when these cubs were born. As all the offspring but one female were subadults when observed it is unlikely that any of them have switched pride since being born. The Manze 2 pride and the Central pride are both adjacent to the Plains pride probably providing mating opportunities when the resident male lion is less attentive.

Spong et al. (2002) found, in a study of the same population of lions in Selous Game reserve, support for so called "superprides" where neighbouring prides grouped together genetically, indicating genetic structure above the pride level. The suggested explanations were that either there were several dispersal events by males over a short distance or that past prides have split creating multiple prides with common ancestry. However, this phenomenon would also be explained by the behaviour exhibited among the male lions of this study; with resident males frequently producing offspring in neighbouring prides increasing gene flow between the groups.

The presence of these extra-group offspring raises an interesting question: is it the males the females or both sexes that seek out mating opportunities beyond the pride? Further studies with closer monitoring would be needed to be certain. The tendency to seek out extra-group copulations that lead to EGP is a trait that has been observed in both male and female mammals (Forstmeier et al. 2014). For a male, the benefit of extra-group copulatory behaviour is relatively clear; enhanced reproductive success without increased parental investment. The advantage for females, on the other hand, is heavily debated (Forstmeier et al. 2014).

According to observations of Serengeti lions by Schaller (1972), resident lionesses tend to be more aggressive towards trespassing females than outside males and Schaller further suggested that females may court with males of neighbouring prides and nomadic males as a means to avoid inbreeding. Indeed, inbreeding avoidance is one of the many hypotheses on why social animals engage in extra-group copulatory behaviour (Formeister et al. 2014). If the male lions that take over a pride have dispersed sufficiently far from their natal pride inbreeding would not pose a problem as relatedness decreases with distance. However, Spong et al. (2002) found that male dispersal distances were short in this population. The adult males in this study were not less related to their pride females than their females were to each other and thus females might find a partner with which she is less related to in another pride or mating with a nomadic male. This is a bit dubious however, as she could also risk mating with a male that has dispersed from her own pride and the females being as related to each other as with the males within a pride could also be attributed to males staying with their natal pride for longer than four years.

The risk of infanticide has also been suggested as a reason for why females would tolerate extra-group copulations, and male lions are quite infamous for this behaviour. Male lions frequently kill cubs that

are not their own to stop the female from further investing in those offspring and enable her to come into heat again (Packer and Pusey 1983). Hypothetically, if a male can remember that he has mated with a specific female when he encounters her with cubs he might not be as determined to terminate them. Considering the high frequency of EGP, without any kin recognition from the father, infanticide might even be a maladaptive behaviour. In a meta-analysis on mammal promiscuity, Wolff and Macdonald (2004) concluded that multiple-male mating by females might have evolved to confuse parentage, as female promiscuity was significantly more common in species where infanticide occurs. To my knowledge this behaviour has not been directly observed however, and one would have to be rather lucky or be very persistent while monitoring to do so. Nevertheless, the benefits for of this mechanism would be great for a female and merits further research.

However not all of the lions fathered by a lion in a different pride are necessarily a result of extra-group copulations. Male lions normally leave their natal pride (Schaller 1972), thus adult male lions are not expected to reside with same pride as their father. Salza of Selous Grave fathered two male lions in 1992 and these lions were observed in Nzerakera as adults in 1996/1997, which is probably a result of them leaving their natal pride and taking over the Nzerakera pride. Typhoon's daughter, Tiffany, is a possible case of EGP, although it could also be that Typhoon has switched pride. His brother Storm, with whom he was holding the Nzerakera pride, was later seen with the Shortcut pride where Tiffany was born. For one to be absolutely certain in these cases, closer monitoring of the different prides and their members would be needed. If the subject of interest is limited to EGP it would be safer to only sample the cubs of the year as you could be sure which the resident pride males were at the time of conception.

There was also a case of a female lion, Dark of the pride Shortcut N (Figure 7, J), who conceived an offspring (Faith) in the Plains pride. Faith was observed in the Plains pride in 2011 as an adult so she could have left her natal pride and joined or even founded the Plains pride, but as 12 years passed between the different sampling occasions it might also be that Shortcut N is the same as the Plains pride.

In general, parentage assignments can only make a "best guess" of reality, and especially in wild populations such as this one there are a number of things that complicate the process. For example, unsampled individuals, overlapping generations, age estimated with physiological attributes and the fact that none of the parents can be conclusively determined in advance all contribute to uncertainty. Therefore, even if it might be the hardest part of a study of parentage, great effort should be made to ensure that the sampling covers as large part of the study population as possible and that the metadata is as complete as possible, since the credibility of the result will only ever be as good as the quality of the data. Nevertheless, the fact that there are many uncertainties is one of the reasons for studying a wild population in the first place (Städele and Vigilant 2017).

4.2 Marker density

Given the results of the pairwise relatedness from different levels of marker density, a higher number of markers seemed to improve the accuracy of the relatedness estimates. Three out of the four chosen parent-offspring relationships remained fairly stable around 0.5, while one relationship (Pendeza/Gregor) decreased with marker density, seemingly towards 0.5, and there is a risk that the Pendeza/Gregor-relationship had a higher influence on the decreasing variation than the others. Nevertheless, seeing that the other three did not seem to decrease in accuracy with increased marker density, using a larger number of markers should not weaken the results. However, this analysis would obviously be much more certain if done on captive animals were parent-offspring relationships were already established and more relationships could be used. Previous studies, as reviewed by Städele and Vigilant (2017), have found that accuracy indeed improves with a higher number of markers. However, the increasing accuracy will certainly level out at a certain level of marker density and then using a higher number of markers will be superfluous. Even with thousands of SNPs, genetic

relatedness is a continuous parameter that will not perfectly correspond to theoretical expectations: on average full siblings share 50% of their genome but they may share considerably more or less. The coefficient is dependent on the proportion of the genome that is shared by the individuals, which could be the result of chance or because of they share a common ancestor, i.e. identical by descent (Städele and Vigilant 2017).

In the parentage analysis increasing the number of markers did not seem to have the same effect as in the pairwise relatedness analyses. Although the mean R-value of the dyads detected increased and seemed to level out at around 300-500 markers, the number of parents that were assigned decreased substantially after 200 markers after an initial increase. This suggests that after a certain point putative parents are being excluded even if they share a high R-value with the offspring. The culprit here is almost certainly genotyping error. As each new marker added contributes with less explaining power but carries the same risk of being an error (Hauser et al. 2011, Labuschagne et al. 2015), too many markers can be detrimental to the parentage assignment rather than helpful. How many markers that is optimal will most probably vary between studies as genotyping error rate varies. Consequently, studies that include parentage analyses should make an effort of finding the most optimal marker density for their own study. Moreover, finding ways to decrease the error rate will greatly improve the results of the parentage analysis. Improvements in the way parentage inference software handle genotyping error would also aid the in future studies of kinship using SNP markers.

4.3 Conclusions

The mating system of lions is not as simple as it may seem at first glance. The lions of Selous Game Reserve do not to strictly mate within their prides but rather males probably gain as much paternity as they can in their own and in neighbouring prides. Lionesses may permit trespassing males or even seek out neighbouring males to mate because of the risk of infanticide or to avoid inbreeding but further research is needed to confirm if these mechanisms are viable. EGP has also been confirmed in Etosha National Park in Namibia (Lyke et al. 2013), while older studies of Serengeti lions found no evidence of such behaviour (Gilbert et al. 1991). This calls for more studies of the reproductive behaviours of the different lion populations to determine how the behaviour is distributed and to gain a more robust understanding of lion mating structure and overall socioecology.

The development of high-throughput sequencing techniques makes genealogical studies of wild animal populations much more accessible along with the opportunity of obtaining large amounts of markers. However, seeing as marker density highly influenced the result of the parentage analyses here, future studies using SNPs should take great care in choosing the most informative SNPs with the least amount of errors and then evaluate the number of markers which produces the most accurate pedigree. The application of these techniques will allow us to further our understanding of kin relationships between wild animals which in turn can be used to study a variety of ecological phenomena, such as mating systems, kin selection and recognition, dispersal and inbreeding avoidance and ultimately guide us in the management of endangered species.

5 Acknowledgements

I would like to give a big thank you to my supervisor for all the support and for letting me realise my childhood dream of doing research on lions. I would also like to express my appreciation to my assistant supervisor Anita Norman for all the help and work with the data even from across the other side of the globe. I would also like to thank Robert Spitzer for showing me the ropes of FRANz and Helena Königsson for the laboratory work and help. A special thanks to Rebecka Strömbom, Victor Olofsson and all of my family and friends for their support and understanding. Finally, I would like to

thank the lions of Selous Game Reserve who involuntarily contributed to this study with their tissue and thereby granted us a glimpse into their lives.

References

Almudevar, A. 2003. A simulated annealing algorithm for maximum likelihood pedigree reconstruction. Theoretical Population Biology, 63, pp 63-75.

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

Catchen, J., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J., 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics, 1, pp 171-182.

Clutton-Brock, T.H. and Isvaran, K., 2006. Paternity loss in contrasting mammalian societies. Biology Letters, 2, pp 513-516.

Davey, J., & Blaxter, M. L. 2011. RADSeq: next-generation population genetics. Briefings in Functional Genomics, 9, pp 416-423.

Davey, J. W., Cezard, T., Fuentes□Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD Sequencing data: implications for genotyping. Molecular ecology, 22, pp 3151-3164.

Glaubitz, J. C., Rhodes, O. E., & DeWoody, J. A. 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Molecular Ecology, 12, pp 1039-1047.

Gilbert, D. A., Packer, C., Pusey, A. E., Stephens, J. C., & O'Brien, S. J. 1991. Analytical DNA fingerprinting in lions: parentage, genetic diversity, and kinship. Journal of Heredity, 82, 378-386.

Grover, A. and Sharma, P.C., 2016. Development and use of molecular markers: past and present. Critical Reviews in Biotechnology, 36, pp 290-302.

Forstmeier, W., Nakagawa, S., Griffith, S.C. and Kempenaers, B., 2014. Female extra-pair mating: adaptation or genetic constraint?. Trends in Ecology & Evolution, 29, pp 456-464.

Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., et al. 2004. Detection of genotyping errors by Hardy–Weinberg equilibrium testing. European Journal of Human Genetics, 12, pp 395-399.

Hauser, L., Baird, M., Hilborn, R.A.Y., Seeb, L.W. and Seeb, J.E., 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. Molecular Ecology Resources, 11, pp 150-161.

Karesh, W. B., Smith, F., & Frazier□Taylor, H. 1987. A remote method for obtaining skin biopsy samples. Conservation Biology, 1, pp 261-262.

Labuschagne, C., Nupen, L., Kotzé, A., Grobler, P.J. and Dalton, D.L., 2015. Assessment of microsatellite and SNP markers for parentage assignment in ex situ African Penguin (Spheniscus demersus) populations. Ecology and Evolution, 5, pp 4389-4399.

Lyke, M.M., Dubach, J. and Briggs, M.B., 2013. A molecular analysis of African lion (Panthera leo) mating structure and extra□group paternity in Etosha National Park. Molecular Ecology, 22, pp 2787-2796.

Lynch M, Ritland K 1999. Estimation of pairwise relatedness with molecular markers. Genetics 152, pp 1753-1766.

Miller, J. R., Balme, G., Lindsey, P. A., Loveridge, A. J., Becker, M. S., Begg, C., et al. 2016. Aging traits and sustainable trophy hunting of African lions. Biological Conservation, 201, pp 160-168.

Norman, A.J., Street, N.R. and Spong, G., 2013. De novo SNP discovery in the Scandinavian brown bear (Ursus arctos). PloS one, 8, pp 81012.

Packer, C., & Pusey, A. E. 1983. Adaptations of female lions to infanticide by incoming males. The American Naturalist, 121, pp 716-728.

Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). related: an R package for analysing pairwise relatedness from codominant molecular markers. Molecular Ecology Resources, 15, pp 557-561.

Raymond M. & Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity, 86, pp 248-249

Riester, M., Stadler, P. F., & Klemm, K. 2009. FRANz: reconstruction of wild multi-generation pedigrees. Bioinformatics, 25, pp 2134-2139.

Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources 8, pp 103-106.

Schaller, G. B. 1972. The Serengeti lion: a study of predator-prey relations. University of Chicago Press.

Spong, G., Stone, J., Creel, S., & Björklund, M. (2002). Genetic structure of lions (Panthera leo L.) in the Selous Game Reserve: implications for the evolution of sociality. Journal of Evolutionary Biology, 15, pp 945-953.

Städele, V., & Vigilant, L. 2016. Strategies for determining kinship in wild populations using genetic data. Ecology and Evolution, 6, pp 6107-6120.

Wolff, J. O., & Macdonald, D. W. 2004. Promiscuous females protect their offspring. Trends in Ecology & Evolution, 19, pp 127-134.

SENASTE UTGIVNA NUMMER

2016:16 Increased public participation as a potential human – large carnivore conflict mitigation measure
Författare: Ruben Bloemsma

2016:17 Influence of temperature and predation risk on herbivore micro habitat choice in a South African savanna
Författare: Hanna Rogers

2016:18 Elephant utilisation of and impact on a locally threatened habitat, coastal scarp forest, in South Africa
Författare: Matilda Apelqvist

2017:1 Breeding dynamics of a Golden Eagle *(Aquila chrysaetos)* population in the boreal forest of Sweden
Författare: Eirini-Lamprini Daouti

2017:2 Reforestation in the far north – Comparing effects of the native tree species *Betula pubescens* and the non-native *Pinus contorta* in Iceland
Författare: Elin Fries

2017:3 Grazing increases albedo of savanna grasslands
Författare: Linda Vedin

2017:4 Timing of ungulate browsing and its effect on sapling height and the field layer vegetation – experimental study using seasonel exclosures during one year
Författare: Filip Ånöstam

2017:5 Land use changes and its consequences on moose habitat
Författare: Ida Olofsson

2017:6 Micro habitat selection of herbivores in response to perceived predation risk and forage quality in Hluhluwe-iMfolozi game reserve
Författare: Edvin Rapp

2017:7 Risky places and risky times: Vegetation cover and carnivore olfactory cues influence patch selection and antipredator behavior of African ungulates

2017:8 Tall trees survival in relation to bottom-up and top-down drivers in Hluhluwe-iMfolozi Park, South Africa
Författare: Petter Madsen

2017:9 Prevalence of *Borrelia burgdorferi* sensu latu in rodents from two areas with varying wild ungulate densities in Southern Sweden
Författare: Jimmy Nyman

2017:10 Remotely monitoring heart-rate and feeding behaviour of fish by using electronic sensor-tags
Författare: Therese Arvén Norling

Hela förteckningen på utgivna nummer hittar du på www.slu.se/viltfiskmiljo