THE APPLICATION OF NEW TECHNOLOGIES IN CONSERVATION GENETICS

by

Christiaan De Jager Labuschagne

Thesis submitted for the degree Philosophiae Doctor (Ph.D.)
in Genetics at the University of the Free State

Supervisor: Prof P. Grobler

Co-Supervisors: Dr D.L. Dalton & Prof A. Kotzé

January 2016

I, Christiaan De Jager Labuschagne, the undersigned, declare that the thesis hereby submitted to the University of the Free State for the degree Philosophiae Doctor (Ph.D) in Genetics and the work contained therein is my own original work and has not previously, in its entirety, or in part, been submitted to any University for a degree.

_____

Signature

ACKNOWLEDGEMENTS

I would like to thank the following people and institutions for their contributions during my project:

Prof P. Grobler, Department of Genetics, University of the Free State; Dr D.L. Dalton, Molecular Genetic, Research Department, The National Zoological Gardens of South Africa; Prof A. Kotzé, Manager: Research and Scientific Services, The National Zoological Gardens of South Africa, for professional guidance, continued support, assistance and encouragement to complete this research.

The **National Zoological Gardens Research and Ethics Scientific Committee** for their inputs and approval of the research project

Work at the **National Zoological Gardens of South Africa (NZG)** was partially supported by National Research Foundation (NRF) grant (number 79732), of the Society, Ecosystems and Change (SEAChange) Programme. **Khomotjo J. Sekhu** and **Freddy Mabetlela** provided technical assistance with the screening of samples.

The zoo and aquarium facilities, **Two Oceans Aquarium** (Cape Town) and **uShaka Sea World** (Durban) in South Africa, for providing biological samples.

Pan African Association of Zoos and Aquaria (PAAZA) for sharing studbook information.

Songimvelo Nature Reserve in South Africa for contributing biological samples.

The former **Advanced Biomolecular Research Cluster** at the **University of the Free State** for research funding.

Dr Oliver Preisig and **inqaba biotec** for granting me the opportunity to pursue a PhD while working full-time; as well as the use of their laboratory facilities.

My colleagues and friends for their encouragement and motivation.

Lastly, I thank **my family** for their unwavering support.

TABLE OF CONTENTS PAGE

LIST OF TABLES PAGE

PART I: REVIEW OF TECHNOLOGIES IN CONSERVATION GENETICS

PART II: DEVELOPMENT AND APPLICATION OF CONSERVATION GENETICS TECHNIQUES IN AFRICAN PENGUIN

PART III: DEVELOPMENT AND APPLICATION OF CONSERVATION GENETICS TECHNIQUES IN RHINOCEROS

LIST OF FIGURES **PAGE**

PART I: REVIEW OF TECHNOLOGIES IN CONSERVATION GENETICS

PART II: APPLICATION AND DEVELOPMENT OF CONSERVATION GENETICS TECHNIQUES IN AFRICAN PENGUIN

PART III: DEVELOPMENT AND APPLICATION OF CONSERVATION GENETICS TECHNIQUES IN RHINOCEROS

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols

Δ	Delta
µ	micro: 10^{-6}
°C	degrees centigrade
µl	micro litre
µM	micro molar

Abbreviations

$(NH_4)_2SO_4$	Ammonium sulfate
A	Mean number of alleles per locus
A	Adenine (in primer sequence)
aa	Amino acid
ABI	Applied Biosystems, Inc.
ACDB	African Centre for DNA Barcoding
ADU	Avian Demography Unit
AFLP	Amplified Fragment Length polymorphism
Ala	Alanine
AMOVA	Analysis of Molecular Variance
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
ATP	Adenosine Triphosphate
ATP6	ATP synthase F0 subunit 6
ATP8	ATP synthase F0 subunit 8
biotin-dUTP	Biotin deoxyuridine triphosphate
BOLD	Barcode of Life Data Systems
bp	Base pair
C	Cytosine (in primer sequence)
CATS	Comparative Anchor Tagged Sites
CEL I	Celery endonuclease I
CF	Captive Facility
CHD	Chromo Helicase DNA-binding gene
CI	Confidence Intervals
COX	Cytochrome c oxidase
CR	Control Region
CSB	Conserved Sequence Blocks
Cys	Cysteine
Cytb	Cytochrome b
dATP	Deoxyadenosine triphosphate
dbSNP	Single Nucleotide Polymorphism database
dCTP	Deoxycytidine triphosphate

ddH_20	Double distilled water
D_{EST}	Estimate of differentiation
dGTP	Deoxyguanosine triphosphate
DHU	Dihydrouridine
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide Triphosphate
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediamine tetra-acetic acid: $C_{10}H_{16}N_2O_8$
Endo V	Endonuclease V
F	Forward primer
F(Null)	The F score for the null hypothesis that the locus is in Hardy-Weinberg Equilibrium
FAM	6-carboxyfluorescein
FIASCO	Fast Isolation by AFLP of Sequences Containing repeats
F_{IS}	Inbreeding coefficient
F_{st}	Fixation index
G	Guanine (in primer sequence)
G	Gram
G''_{ST}	Genetic distance
Gln	Glutamine
Glu	Glutamic Acid
Gly	Glycine
GS	Genome sequence
G_{ST}	Genetic Distance
H	Heavy strand
H_e / H_z / H_{exp}	Expected heterozygosity
His	Histidine
H_o / H_{obs}	Observed heterozygosity
HW / HWE	Hardy-weinberg equilibrium
IBOL	International Barcode Of Life
ISIS	International Species Information System
IUCN	International Union for Conservation of Nature
K	Most probable number of populations
K	number of alleles
Kb	kilo base pair
kcal/mol	kilocalorie per mole
KCl	Potassium chloride
L	Light Strand
LD	Linkage Disequilibrium
Leu	Leusine
Ile	Isoleucine
Ln	Posterior Probabilities

LOD	Logarithm of odds
Lys	Lysine
M	milli: 10^{-3}
M	molar: moles per litre
MBT	Temporal moment-based
Met	Methionine
Mg	Milligram
$MgCl_2$	magnesium chloride
$MgSO_4$	Magnesium sulfate
Min	Minutes
mM	milli molar
MMCMC	Markov Chain Monte Carlo
Mol	Molar
MS	Microsatellites
Mt	mitochondrial
MT	Mitochondrial DNA Marker
mtDNA	mitochondrial DNA
Mya	Million years ago
N	nano: 10^{-9}
N	Mean sample size / Number of samples
Na / N_a±SE	Number of alleles
NaCl	Sodium Chloride
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (Reduced)
ND	NADH dehydrogenase
N_e	Effective Population Size
Ne-1P	Average non-exclusion probability for one candidate parent
Ne-2P	Average non-exclusion probability for one candidate parent given the genotype of a known parent of the opposite sex
N_{eA}	Number of effective alleles
N_es	Small Effective Population Size
Ng	Nanogram
nM	nanomolar
No	Number
Ns	non-significant
Nt	nucleotide
NZG	National Zoological Gardens of South Africa
P	Probability values for exact tests of HW proportions
PAAZA	Pan African Association for Zoos and Aquaria
PCG	Protein Coding Genes
PCR	Polymerase Chain Reaction
Phe	Phenylalanine

PIC	Polymorphic Information Content
pmol	pico mol: 10^{-12}
Pro	Proline
R	Reverse primer
RAPD	Random Amplified Polymorphic DNA
RNA	ribonucleic acid
Rpm	revolutions per minute
rRNA	ribosomal RNA
s / sec	seconds
SANBI	South African National Biodiversity Institute
SANCCOB	Southern African Foundation for the Conservation of Coastal Birds
SDS	Sodium dodecyl sulphate
Ser	Serine
SNPs	Single Nucleotide Polymorphisms
SPARKS	Single Population Analysis and Record Keeping System
SSC	Single-strand conformational polymorphism
ssDNA	single stranded DNA
SSLP	Simple Sequence Length Polymorphisms
SSR	Simple Sequence Repeat
STR	Short Tandem Repeat
T	Thymine (in primer sequence)
TE	Tris-EDTA
TEN100	10 Mm Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.5
TENIKWA	Tenikwa Wildlife Awareness Centre
Thr	Threonine
T_m	melting temperature
Tris-HCl	Tris hydrochloride
tRNA	Transfer RNA
Trp	Tryptophan
Tyr	Tyrosine
U	units
U/µl	units per microliter
uH_z	Unbiased expected Heterozygosity
uShaka	Sea World Durban
V	volts
Val	Valine
VNTR	Variable Number of Tandem Repeats
ZR	Zymo Research

LIST OF PUBLICATIONS

The following published works form part of this thesis:

Published

Labuschagne C., Kotzé A., Grobler P., Dalton D.L. 2014. The complete sequence of the mitochondrial genome of the African Penguin *(Spheniscus demersus), Gene*, 534: 113-118.

Labuschagne C., Kotzé A., Van Wyk A.M., Dalton D.L. 2013. Isolation and characterization of species-specific microsatellite loci in African Penguin *(Spheniscus demersus). Conservation Genetic Resources*, 5: 169-171.

Labuschagne C., Kotzé A., Dalton D.L. 2012. Isolation and characterization of SNP markers for African Penguin *(Spheniscus demersus). Conservation genetic resources*, 4 (4): 1067-1069.

Labuschagne C., Nupen L., Kotzé A., Grobler P.J., 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(19): 4389-4399.

Labuschagne C., Kotzé A., Grobler P., Dalton D.L. 2013. A targeted gene approach to SNP discovery in the White Rhino *(Ceratotherium simum)*. *Conservation Genetic Resources*, 5: 265-266.

Labuschagne, C., Nupen, L., Kotzé, A., Grobler, J. P. and Dalton, D. L. 2015. Endonuclease V digestion for SNP discovery and marker development in White Rhino *(Ceratotherium simum). Conservation Genetics Resources*, 7 (1): 53-56.

Labuschagne C., Kotzé A., Grobler P., Dalton D.L. 2016. Genetic monitoring of *ex-situ* African Penguin (*Spheniscus demersus*) populations in South Africa, *African Zoology*, 51 (2): 83-90.

Accepted

Labuschagne C., Dalton D.L., Grobler J.P., Kotzé A. SNP discovery and characterisation in White Rhino *(Ceratotherium simum)* with application to parentage assignment, *Genetics and Molecular Biology*, accepted May 2016.

SUMMARY

Over the past decade, the development of high-throughput DNA techniques has expanded the scope of conservation genetics and molecular markers have become indispensable tools for the management of wildlife species and populations. There are several molecular markers available for biodiversity analysis, but their selection depends on the objective of the study, the molecular information sought (and reliability thereof) and the facilities and/or resources available. In order to develop and apply new genetic techniques I have decided on using one bird and one mammal species of interest in South Africa. The bird species chosen is the African Penguin (*Spheniscus demersus*) which has suffered serious population declines and is listed in the IUCN Red Data Book as an endangered species. Due to world-wide attention to rhinoceros conservation and population decline, the white rhinoceros *(Ceratotherium simum)* was selected as mammal species. Three different markers and their utility in aid of South African wildlife biodiversity conservation were investigated in these diverse species. The complete mitochondrial genome of the African Penguin was sequenced. The *Spheniscus demersus* mtDNA genome is very similar, both in composition and length, to both the *Eudyptes chrysocome* and *E. minor* genomes. This is the first report of the complete nucleotide sequence for the mitochondrial genome of the African Penguin. These results can be subsequently used to provide information for penguin phylogenetic studies and insights into the evolution of genomes. Furthermore, the study reported eight species specific microsatellite markers as well as 31 SNP markers as new molecular tools for the investigation, management and reintroduction of African penguin. Utilising these new tools, the study generated molecular genetic information to verify/complement studbook-based pedigree data from *ex-situ* populations of African Penguin. In addition, we compared the relative and combined utility of MS and SNP markers for parentage assignment. We found that a combined subset of these two types of markers attained a > 99% correct cumulative parentage assignment probability. This study further reported on 34 novel SNP markers for the white rhinoceros, identified through sequencing of CATS loci as well as SNP enriched libraries. The utility of 33 Single Nucleotide Polymorphisms and 10 microsatellites in isolation and in combination for assigning parentage in captive white rhinoceros were compared. It was found that a combined dataset of SNPs and microsatellites was most informative and showed the highest confidence level. This study thus provides a useful set of SNP and MS markers for parentage and relatedness testing in white rhinoceros. Furthermore, assessment of the utility of SNP and MS markers over multiple (> three) generations and the incorporation of a larger variety of relationships among individuals (e.g. half-siblings or cousins) is strongly recommended. Developed SNP markers could be used to define the genetic mating system of this species, for forensic applications and to determine population structure and variability when other markers prove problematic.

OPSOMMING

Oor die afgelope dekade, het die ontwikkeling van hoë-deurset DNA tegnieke die omvang van bewaringsgenetika aansienlik uitgebrei en molekulêre merkers het onontbeerlike hulpmiddels vir die bestuur van wild spesies en bevolkings geword. Daar is verskeie molekulêre merkers beskikbaar vir die ontleding van biodiversiteit, maar die keuse hang af van die doel van die studie, die vereisste molekulêre inligting (en betroubaarheid daarvan) en die fasiliteite en / of hulpbronne beskikbaar. Een voël en een soogdier spesie van belang in Suid-Afrika was gekies vir die ontwikkeling en toepassing van nuwe genetiese tegnieke. Die Afrika pikkewyn (*Spheniscus demersus*), wat ernstige bevolkings afnames toon en gelys is in die IUCN Rooidataboek as 'n bedreigde spesie, was gekies as voëlspesie. Die Witrenoster (*Ceratotherium simum*) is gekies as soogdier spesies as gevolg van die wêreld-wye aandag aan renoster bewaring en bevolkings afname. Drie verskillende klasse merkers en hul nut vir die Suid-Afrikaanse wilds-biodiversiteit bewaring was ondersoek in hierdie diverse spesies. Die volledige mitochondriale genoom volgorde van die Afrika pikkewyn was bepaal. Die *S. demersus* mtDNA genoom is baie soortgelyk, beide in samestelling en lengte, aan beide die *Eudyptes chrysocome* en *E. minor* genome. Hierdie is die eerste verslag van die volledige nukleotiedvolgorde vir die mitochondriale genoom van die Afrika pikkewyn. Hierdie resultate kan gebruik word as inligting vir pikkewyn filogenetiese studies en kan insigte in die evolusie van genome te verskaf. Verder het die studie agt spesie-spesifieke mikrosatelliet merkers asook 31 SNP-merkers geidentifiseer as nuwe molekulêre hulpmiddels vir die ondersoek, bestuur en hervestiging van Afrika-pikkewyne. Deur gebruik te maak van die merkers het die studie genetiese inligting gegenereer om stamboom data van ex-situ bevolkings van Afrika pikkewyne aan te vul en te verifieer. Daarbenewens is die relatiewe en gekombineerde nut van MS en SNP-merkers vir ouerskap-bepaling vergelyk. Daar is bevind dat 'n gekombineerde substel van hierdie twee tipe merkers 'n > 99% kumulatiewe ouerskap waarskynlikheid moontlik maak. Hierdie studie het verder 34 nuwe SNP-merkers vir die Wit renoster geïdentifiseer deur volgorde bepaling van CATS lokusse sowel as SNP verrykingsbiblioteke. Die nut van 33 SNP-merkers en 10 mikrosatelliete was afsonderlik en in kombinasie vergelyk vir toepassing in ouerskap-bepaling op Witrenosters in gevangeneskap. Daar is gevind dat 'n gekombineerde datastel van SNPs en mikrosatelliete die mees insiggewend was en die hoogste vlak van betroubaarheid bied. Hierdie studie bied dus 'n nuttige stel SNP en MS merkers vir ouerskap en verwantskap bepaling in Witrenostesr. Verder word die assessering van die nut van SNP en MS merkers oor verskeie (>3) generasies en met 'n groter verskeidenheid van moontlike verhoudings tussen individue (bv half-broers en susters of neefs) sterk aanbeveel. Die ontwikkelde SNP-merkers kan gebruik word om die genetiese paringstelsel van hierdie spesie te definieer, vir forensiese ondersoeke en om die bevolking-struktuur en verskeidenheid te bepaal wanneer ander merkers problematies is.

PART I: REVIEW OF TECHNOLOGIES IN CONSERVATION GENETICS

Introduction

African Penguin *(Spheniscus demersus)*

African Penguin, also referred to as Jackass, Cape or Black-footed Penguin, are the only penguin species breeding along the coast of Africa (Figure 1) and form an important part of the marine avifauna diversity of the Benguela upwelling ecosystem off southern Africa (Shannon and Crawford, 1999).

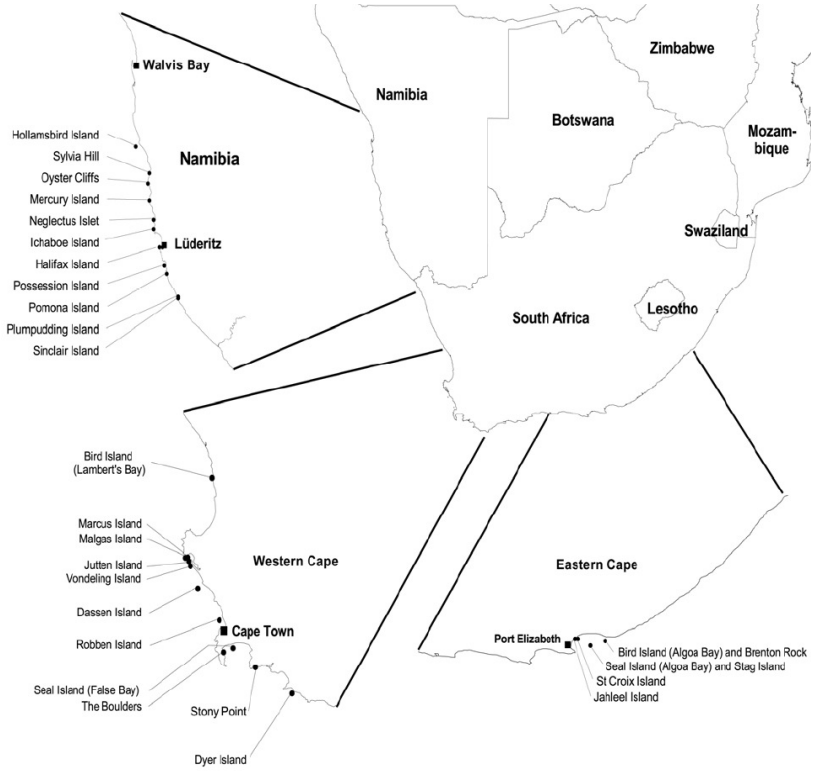


Figure 1: African Penguin colony distribution extending from Hollamsbird Island adjacent to central Namibia to Bird Island, Algoa Bay, on the south-eastern coast of South Africa (Kemper, 2006)

The genus *Spheniscus* contains three other species. The Galapagos Penguin (*S. mendiculus*) that mainly breed on Fernandina and Isabela Islands (Vargas *et al.,* 2005), Humboldt Penguin (*S. humboldti*) that breed in Peru and Chile and Magellanic Penguin (*S. magellanicus*) in Chile, Argentina and the Falkland Islands (Wilson *et al.,* 1995). The Magellanic Penguin is the only true seasonal breeders of the four species, migrating away from their breeding localities during the non-breeding season (Yorio *et al.,* 2001). African and Humboldt Penguins may travel long distances, but tend to remain near or at breeding localities when not breeding (Culik and Luna-Jorquera, 1997; Whittington *et al.,* 2005). All four species nest in burrows scraped into the ground using their beaks and clawed feet (Boswall and MacIver, 1975; Boersma, 1977; Paredes and Zavalaga, 2001).

All four *Spheniscus* species have similarities in the plumage pattern of juvenile and adult birds apart from their species-specific markings (Williams 1995). African Penguin is patterned black above and white below, with a characteristic white head stripe and black breast band. Juvenile African Penguin is grey to brown above, fading into white underparts. During the first moult, after fledging, the juveniles acquire their adult plumage and become indistinguishable from sexually mature adults. These young birds reach sexual maturity at three years or older (Crawford *et al.,* 1999).

African Penguin is classified as endangered under the IUCN threat status, since the species is undergoing a rapid population decline with no signs of reversing (BirdLife International, 2013). The Namibian population was estimated at around 5,000 breeding pairs in 2008, while the South African population was estimated at about 21,000 breeding pairs in 2009 (BirdLife International, 2013). In the early 1990s, the overall *S. demersus* population was estimated at 180 000 + penguins (Crawford *et al.,* 1995).

During the first half of the 20th century the commercial exploitation of eggs and removal of guano nesting substrata are thought to have drastically reduced numbers of penguins (Underhill *et al.,* 2006). The last authorised egg collections were made in 1967 and the practise no longer pose a significant threat (Shelton *et al.,* 1984). The loss of guano in which burrows may be constructed has reduced breeding success by exposing penguins to heat stress and making them more vulnerable to displacement from breeding sites by larger animals, such as the Cape Fur Seal (*Arctocephalus pusillus pusillus*) (Shannon and Crawford, 1999; Sherley *et al.,* 2012). Artificial nests have been implemented as a conservation tool with some success (Sherley *et al.,* 2012). In the second half of the 20th century, declines are largely attributed to food shortages caused by large-scale purse-seine commercial fisheries exploiting schooling epipelagic fish, the main prey of African Penguin (Shannon and Crawford, 1999). Within three months of the establishment of a 20 km no-take zone at the St Croix Island colony, breeding African Penguins decreased their foraging effort by 25–30% and their daily energy expenditure by approximately 43 percent, shifting their core foraging areas from outside to within the area closed to fishing. This provides some support for the affects of commercial fishing on penguins (Pichegru *et al.,* 2010). An eastward shift in sardine (*Sardinops sagax*) and anchovy (*Engraulis capensis*) populations have also been implicated, with the biomass of these species near the largest breeding islands west of Cape Town falling sharply since 2002 (BirdLife International, 2013). At present other factors that adversely influence penguin populations include mortality during oil spills, competition with Cape Fur Seals for breeding space, predation by seals, Feral cats (*Felis catus*) and Kelp gulls (*Larus dominicanus*) as well as entanglement in fishing gear and other marine debris (Crawford *et al.,* 2000, Whittington *et al.,* 2000, David *et al.,* 2003). There has been a dramatic increase in the number of birds oiled since 1990 with the Apollo Sea spill of 1994 oiling 10 000 penguins and killing 5 000 (Underhill *et al.,* 1999). The Treasure spill of 2000 oiled 19 000 penguins and killed 2 000 (Crawford *et al.,* 2000). Physiological and behavioural problems affecting breeding success have been indicated in oiled birds, further contributing to the population decline after an oil spill (Nel and Whittington, 2003; Barham *et al.,* 2007).

At the time of this study, very limited genetic information was available for the African Penguin, highlighting the need for development of markers in aid of the management and conservation of this endangered species.

The White Rhinoceros *(Ceratotherium simum)*

The white rhinoceros belongs to the perissodactyl family Rhinocerotidae, which includes four genera, *Rhinoceros* (Indian and Javan rhinoceroses), *Dicerorhonus* (Ssumatran rhinoceros), *Diceros* (black rhinoceros) and *Ceratotherium* (white rhinoceros) (Xu and Arnason, 1997). *Ceratotherium* and *Diceros* occur in Africa, whereas *Rhinoceros* and *Dicerorhonus* are found in Asia (Xu and Arnason, 1997). The white rhinoceros is the largest of the five species of rhinoceros with males averaging at 2300 kg and females at about 1700 kg (Macdonald, 2001). White rhinoceros are lpolygynandrous, solitary grazers with the females ranging over much larger areas than males. Following ritualized fighting, males set up territories in small, high-quality forage areas (Owen-Smith, 1975; White *et al.,* 2007). Females mark their territories with dung, urine and broken vegetation while visiting male territories to mate (Owen-Smith, 1975; Gyseghem, 1984). Females reach sexual maturity from six to seven years, while males reach sexual maturity from seven to ten years. The gestation period is around 16 months with birth intervals per calf at two to three years (Owen-Smith, 1973).

The white rhinoceros is one of the great success stories of modern wildlife conservation, with numbers growing from as few as 50-100 animals in the 1880s, to approximately 20,000 white rhinoceros remaining in the wild today (Emslie and Brooks, 1999; Emslie, 2011). South Africa is home to about three-quarters of world's remaining rhinoceros with between 8 000 to 9 290 white rhinoceros surviving in the Kruger National Park, roughly one-quarter on private land while national and provincial parks authorities host approximately 15 700 black and white rhinoceros (Ferreira 2013; Huebschle, 2016). The trade in conservational rhinoceros horns is however a problem in many parts of the world, especially in parts of Asia, where the horns are used traditionally as material in sculptures or as drug products for medicinal purposes (Florescu *et al.,* 2003; Hsieh *et al.,* 2003; Seror *et al.,* 2002) adding constant pressure on remaining populations.

The two predominant molecular methods used in genotyping of rhinoceros used to be random amplified polymorphic DNA (RAPD) analysis and microsatellite analysis (Seror *et al.,* 2002; Florescu *et al.,* 2003; Harley *et al.,* 2005). However, due to the low diversity observed in the overall white rhinoceros population, these techniques have limited application (Florescu *et al.,* 2003; Nielsen *et al.,* 2008). In a study conducted by Florescu *et al.* (2003) the total number of variable markers developed from 40 000 colonies (enriched microsatellite-containing clones) were limited to five, therefore emphasizing the need for a time and cost-effective approach to developing suitable markers. Guerier and colleagues (2012) successfully utilized microsatellite genotypes from 11 loci together with comprehensive historic information for parentage assignment in a managed free-ranging population of white rhinoceros. Polymorphic information dontent (PIC) of 0.357 and mean

heterozygosity across 31 individuals of 0.447 was observed, indicating low genetic diversity (Guerier *et al*., 2012).

Genetic diversity data, for both captive and wild populations, form an important tool in successful reproductive management, population viability assessments and diversity conservation with regards to translocation of animals and establishing breeding programmes (Seror *et al.,* 2002; Harley *et al.,* 2005). Genetic data may further be used to create geographic-specific allele frequency maps to determine geographic origin of seized wildlife contraband, which offer powerful law enforcement tools for a variety of forensic applications (Wasser *et al*., 2015).If the poaching continues to escalate it may destroy a critically important success story in conservation history, and eliminate a special species and beautiful beast whose prehistoric links, cultural importance, and ecological role in Africa cannot be replaced.

Aim of the study

The broad aim of this study is to expand on the the molecular tools available for African Penguin (full mitochondrial genome, species-specific microsatellite markers and single nucleotide polymorphisms) and white rhinoceros (single nucleotide polymorphisms). The study presented in this thesis is the first investigation to develop new tools for African Penguin and white rhinoceros. The development of these marker sets contribute to further genetic research studies aimed at conservation efforts for both species. Chapter three describes the first complete sequence of the mitochondrial genome of the African Penguin. Chapters four and five focus on the isolation and characterisation of species-specific microsatellite loci and single nucleotide polymorphisms in African penguin. Chapters six and seven include studies where the novel markers are employed in order to assess their utility in parentage assignment as well as to provide an assessment of the genetic diversity of captive penguin populations in South Africa. In these chapters it is hypothesised that a set of various markers could accurately be used to assign parentage and that the genetic variability in the *ex-situ* penguin populations and subsequent generations would be lower than *in-situ* populations due to a reduction in population size. Chapters eight and nine include details on the development of a SNP enrichment protocol in white rhinoceros for cost effective SNP discovery utilizing Endo V in heteroduplex reduction libraries. Chapter ten investigates the utility of these markers in parentage assignment. Conclusions drawn from these results are described in Chapter 11 where additional avenues of markers development for theses species are provided.

Specific aims

This study included two different species; African Penguin and white rhinoceros with an overall aim to develop new molecular tools for conservation of these species. The developed tools were investigated for their utility in parentage assignment and were additionally used to assess genetic diversity in the African Penguin *ex-situ* populations. The specific aims of the study were:

1. To sequence and annotate the whole mitochondrial genome of the African Penguin.
2. To isolate and characterize novel species-specific microsatellite markers for African Penguin.

3. To isolate and characterize novel SNP markers for African Penguin.
4. To investigate the application of the developed tools to determine the level of genetic variation, population structure and effective population size in an *ex-situ* African Penguin population.
5. To investigate the application of microsatellite and SNP markers in parentage assignment in an *ex-situ* African Penguin population.
6. To isolate and characterize novel SNPs through a targeted gene approach for white rhinoceros.
7. To develop SNP markers for white rhinoceros through Endonuclease V mediated SNP enrichment of reduced representation libraries.
8. To investigate the application of Microsatellite and SNP markers in parentage assignment in white rhinoceros.

References

Barham P.J., Underhill L.G., Crawford R.J., Leshoro T.M. 2007. Differences in breeding success between African Penguins (*Spheniscus demersus*) that were and were not oiled in the MV Treasure oil-spill in 2000. *Emu*, 107(1): 7-13.

BirdLife International. 2013. *Spheniscus demersus*. The IUCN Red List of Threatened Species. Version 2015.2. <www.iucnredlist.org>. Downloaded on 27 July 2015.

Boersma P.D. 1977. An ecological and behavioural study of the Galápagos Penguin. *Living Bird,* 15: 43-93.

Boswall J. and MacIver D. 1975. The Magellanic Penguin *Spheniscus magellanicus*. In: B. Stonehouse (ed.), The biology of penguins. MacMillan: London. pp. 271-305.

Crawford R.J.M., Williams A.J., Hofmeyr J.H., Klages N.T.W., Randall R.M., Cooper J., Dyer B.M., Chesselet Y. 1995. Trends of African Penguin *Spheniscus demersus* populations in the 20th century. *South African Journal of Marine Science,* 16: 101-118.

Crawford R.J.M., Shannon L.J., Whittington P.A. 1999. Population dynamics of the African Penguin *Spheniscus demersus* at Robben Island, South Africa. *Marine Ornithology,* 27: 139-147.

Crawford R.J.M., Davis S.A., Harding R., Jackson L.F., Leshoro T.M., Meyer M.A., Randall R.M., Underhill L.G., Upfold L., van Dalsen A.P., van der Merwe E., Whittington P.A., Williams A.J., Wolfaardt A.C. 2000. Initial impact of the Treasure oil spill on seabirds off western South Africa. *South African Journal of Marine Science,* 22: 157-176.

Culik B.M. and Luna-Jorquera G. 1997. The Humboldt Penguin *Spheniscus humboldti*: a migratory bird? *Journal für Ornithologie,* 138: 325-330.

David J.H.M., Cury P., Crawford R.J.M., Randall R.M., Underhill L.G., Meyer M.A. 2003. Assessing conservation priorities in the Benguela ecosystem, South Africa: analysing predation by seals on threatened seabirds. *Biological Conservation,* 114: 289-292.

Emslie R. 2011. *Ceratotherium simum*. In: IUCN 2012. IUCN Red List of Threatened Species. Version 2012.1. <www.iucnredlist.org>. Downloaded on 06 September 2012.

Emslie R.H. and Brooks M. 1999. African rhino: status survey and conservation action plan. Gland and Cambridge, IUCN/ SSC African Rhino Specialist Group, pp. i-x, 1-92.

Ferreira S. 2013: Kruger National Park's response to poaching: Lessons learnt. 11th IUCN SSC African Rhino Specialist Group. Laikipia County, Kenya.

Florescu A., Davila J.A., Scott C., Fernando P., Kellners K., Morales J.C., Melnick D., Boag P.T., van Coeverden de Groot P. 2003. Polymorphic microsatellites in white rhinoceros. *Molecular Ecology Notes,* 3: 344-345.

Guerier A.S., Bishop J.M., Crawford S.J., Schmidt-Kuentzel A., Stratford K.J. 2012. Parentage analysis in a managed free ranging population of southern white rhinoceros: genetic diversity, pedigrees and management. *Conservation Genetics,* 2012: 1-14.

Gyseghem R.V. 1984. Observations on the ecology and behaviour of the northern white rhinoceros *Ceratotherium simum cottoni*. *Zeitschrift füer Säugetirkunde,* 49: 348-358.

Harley E.H., Baumgarten I., Cunningham J., O'Ryan P. 2005. Genetic variation and population structure in remnant populations of Black Rhinoceros, *Diceros bicornis*, in Africa. *Molecular Ecology,* 14: 2981-90.

Hsieh H., Huang L., Tsai L., Kuo Y., Meng H., Linacre A., Lee J.C. 2003. Special identification of rhinoceros horns using the cytochrome b gene. *Forensic Science International,* 136: 1-11.

Huebschle A.M. 2016. A game of horns: transnational flows of rhino horn. Dissertation published in Cologne, IMPRS-SPCE (Studies on the Social and Political Constitution of the Economy) -, pp. 1-424.

Kemper J. 2006. Heading towards extinction? Demography of the African Penguin in Namibia (Doctoral dissertation, University of Cape Town).

Macdonald D. 2001. The New Encyclopedia of Mammals. Oxford University Press, Oxford.

Nel D.C. and Whittington P.A. 2003. The Rehabilitation of Oiled African Penguins: a Conservation Success Story. *BirdLife South Africa*, Cape Town.

Nielsen L., Meehan-Meola D., Kilbourn A., Alcivar-Warren A. 2008. Characherisation of microsatellite loci in the black rhinoceros *(Diceros bicornis)* and white rhinoceros *(Ceratotherium*

simum): their use for cross species amplification and differentiation between the two species. *Conservation Genetics*, 9: 239-242.

Owen-Smith R. N., 1973. The behavioural ecology of the white rhinoceros (Doctoral dissertation, University of Wisconsin).

Owen-Smith N. 1975. The social ethology of the white rhinoceros *Ceratotherium simum* (Burchell, 1817). *Zeitschrift für Tierpsychologie*, 38: 337-384.

Paredes R. and Zavalaga C.B. 2001. Nesting sites and nest types as important factors for the conservation of Humboldt penguins (*Spheniscus humboldti*). *Biological Conservation,* 100: 199-205.

Pichegru L., Grémillet D., Crawford R.J.M., Ryan P.G. 2010. Marine no-take zone rapidly benefits endangered penguin. *Biology Letters*, 6(4): 498-501.

Seror B., Zorovsky Y., Terkel A., Katcoff D.J. 2002. Genetic Kinship and Social Structure in a Herd of Square-Lipped Rhinoceroses *(Ceratotherium simum simum*) at the Zoological Center, Tel Aviv/Ramat-Gan, Israel. *Zoo Biology,* 21:561-571.

Shannon L.J. and Crawford R.J.M. 1999. Management of the African Penguin *Spheniscus demersus* – insights from modelling. *Marine Ornithology,* 27: 119-128.

Shelton P.A., Crawford R.J.M., Cooper J., Brooke R.K. 1984. Distribution, population size and conservation of the Jackass Penguin *Spheniscus demersus*. *South African Journal of Marine Science,* 2: 217-257.

Sherley R.B., Barham B.J., Barham P.J., Leshoro T.M., Underhill L.G. 2012. Artificial nests enhance the breeding productivity of African Penguins (*Spheniscus demersus*) on Robben Island, South Africa. *Emu*, 112(2): 97-106.

Underhill L.G., Bartlett P.A., Baumann L., Crawford R.J.M., Dyer B.M., Gildenhuys A., Nel D.C., Oatley T.B., Thornton M., Upfold L., Williams A.J., Whittington P.A., Wolfaardt A.C. 1999. Mortality and survival of African Penguins *Spheniscus* demersus involved in the Apollo Sea oil spill: an evaluation of rehabilitation efforts. *Ibis,* 141: 29-37.

Underhill L.G., Crawford R.J.M., Wolfaardt A.C., Whittington P.A., Dyer B.M., Leshoro T.M., Ruthenburg M., Upfold L., Visagie J. 2006. Regionally coherent trends in colonies of African Penguins *Spheniscus demersus* in the Western Cape, South Africa, 1987–2005. *African Journal of Marine Science,* 28: 697-704.

Vargas H., Lougheed C., Snell H. 2005. Population size and trends of the Galápagos Penguin *Spheniscus mendiculus*. *Ibis,* 147: 367-374.

Wasser S.K., Brown L., Mailand C., Mondol S., Clark W., Laurie C., Weir B.S. 2015. Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. *Science*, 349(6243): 84-87.

White A.M., Swaisgood R.R., Czekala N. 2007. Ranging patterns in white rhinoceros *Ceratotherium simum simum*: Implications for mating strategies. *Anim. Behav*. 74: 349-356.

Whittington P.A., Crawford R.J.M., Huyser O., Oschadleus H.D., Randall R.M., Ryan P.G., Shannon L.J., Wolfaardt A.C., Cooper J., Lacy R., Ellis S. 2000. African Penguin Population and Habitat Viability Assessment. Final Report. IUCN/SSC Conservation Breeding Specialist Group, Apple Valley, Minnesota, pp.54.

Whittington P.A., Randall R.M., Randall B.M., Wolfaardt A.C., Crawford R.J.M., Klages N.T.W., Bartlett P.A., Chesselet Y.J., Jones R. 2005. Patterns of movements of the African Penguin in South Africa and Namibia. *African Journal of Marine Science,* 27: 215-229.

Williams T.D. 1995. The penguins. Oxford University Press, New York.

Wilson R.P., Duffy D.C., Wilson M.P., Araya B. 1995. Aspects of the ecology of species replacement in Humboldt and Magellanic penguins in Chile. *Le Gerfaut,* 85: 49-61.

Yorio P., García Borboroglu P., Potti J., Moreno J. 2001. Breeding biology of Magellanic Penguins *Spheniscus magellanicus* at Golfo San Jorge, Patagonia, Argentina. *Marine Ornithology,* 29: 75-79.

Xu X. and Arnason U. 1997. The complete mitochondrial DNA sequence of the white rhinoceros, *Ceratotherium simum*, and comparison with the mtDNA sequence of the Indian rhinoceros, *Rhinoceros unicornis*. *Molecular phylogenetics and evolution*, 7(2): 189-194.

CHAPTER 2

Literature review

Conservation and genetics

A central issue in conservation genetics is the level of genetic variation present, a prerequisite for evolution (Pertoldi *et al.,* 2007; Väli *et al.,* 2008). The rates of adaptive evolution needs to, at least, match the rate of environmental change in order for a population to persist (Pertoldi *et al.,* 2007). Two potential consequences may be envisioned for loss of genetic variability: (i) low genetic variability can be a threat in the long-term for adapting and evolving in disturbed habitats and under changing environmental conditions; and (ii) inbreeding may occur in small, fragmented and isolated populations, i.e. increased relatedness and homozygosity between individuals as well as autozygosity, posing an immediate threat to fitness in such a population (Pertoldi *et al.,* 2007; Väli *et al.,* 2008). An important prerequisite for the design of conservation strategies is information on the speed at which populations become inbred (Pertoldi *et al.,* 2007). A common rescue-strategy adopted by conservation genetics includes the increase of gene flow among populations for the maintenance of genetic diversity and alleviating inbreeding depression (Pertoldi *et al.,* 2007). However, high levels of gene flow can reduce the capacity of populations to stay adapted to local conditions or introduce mal-adapted genes that can reduce viability of populations, known as out-breeding depression (Pertoldi *et al.,* 2007). A further important issue in conservation genetics is the current structure as well as a history of a population or species, both in a demographic and phylogenetic sense (Pertoldi *et al.,* 2007). Evaluation of levels of genetic diversity is therefore common in population genetics and is particularly important in conservation genetics (Väli *et al.,* 2008). Comprehensive management plans for any species of conservation concern should include plans for maintaining existing genetic diversity, both to ensure ability to adapt to changing environments and to preserve the possibility of future speciation (Lacy, 1997).

Nuclear DNA markers

Mitochondrial DNA Markers

Mitochondrial DNA (mtDNA) is maternally inherited, degrades slower than nuclear DNA, evolves approximately ten-fold faster than nuclear DNA and is highly conserved in taxa. Animal mtDNA contain 36 or 37 genes, 22 for tRNA, 12 or 13 for subunits of the mitochondrial membrane and two for rRNAs (Boore, 1999). There is also a noncoding control region that plays a role in the transcription and replication of mtDNA molecules (Wan *et al.,* 2004). Both 16S and 12S rDNA genes are useful for investigating genetic diversity in higher taxa such as phyla and mid-level taxa such as families. Protein-coding mtDNA genes, such as the cytochrome b (cyt*b*) gene, are preferred for genetic diversity analyses at lower level taxa (genera, species) because of their fast evolutionary rate (compared to the rRNA markers). The non-coding region of mtDNA, the control

region, exhibits higher levels of variation than protein-coding genes and is useful for detecting molecular diversity in species and can also aid in the identification of conservation units. Overall, mtDNA markers are preferably used for the study of evolutionary relationships and genetic diversity, but not recent evolutionary events (Arif and Khan, 2009, Wan *et al.,* 2004).

The D-loop region from mitochondrial DNA has been used to investigate the evolutionary history of the Gentoo and Adélie Penguins on Admiralty Bay, King George Island (Dantas *et al.,* 2014). Pena and colleagues (2014) utilised mitochondrial DNA hypervariable region I as well as intron 7 of the β-fibrinogen gene to investigate the demographic history and population structure on Gentoo Penguin populations in Antarctica. Murata and Murakami (2013) utilised control region fragments as well as cytochrome b sequences to investigate genetic diversity in captive African Penguins in Japan and found that these penguins may be derived from two distinct maternal lines.

The complete mitochondrial genome (16,829 nt) of the indian rhinoceros was sequenced and assembled from a clone library by Xu and colleagues in 1996. The following year, Xu and Arnason reported the complete mitochondrial genome (16,832 nt) of the white rhinoceros. Willerslev and colleagues (2009) reported the complete mitochondrial genomes of the extinct ice-age woolly rhinoceros (16,436 nt) (*Coelodonta antiquitatis*) as well as Javan (16,417), Sumatran (16,466 nt) and black rhinoceroses (16,411 nt) as part of a rhinoceros phylogeny study. The afore mentioned genome sequences were determined through next generation sequencing on the Roche FLX analyzer (Willerslev *et al.,* 2009).

Microsatellites

Microsatellites are the most popular marker for conservation genetic studies in animals. Microsatellites, also known as short tandem repeats (STR), simple sequence repeats (SSR) or simple sequence length polymorphisms (SSLP) (Chambers and MacAvoy, 2000) are tandemly repeated DNA sequences, generally consisting of 1-6 nucleotide repeats (Zane *et al.,* 2002). They form part of Variable Number of Tandem Repeats (VNTR) (Nakamura *et al.,* 1987), which includes satellites (Britten and Kohne, 1968), minisatellites (Jeffreys *et al.,* 1985) and microsatellites (Litt and Luty, 1989). Microsatellites are highly polymorphic in length when analysed between different individuals of the same species and are found throughout prokaryotic (Field and Wills, 1996; Gur-Arie *et al.,* 2000) and eukaryotic genomes in both non-coding and (rarely) coding regions (Sutherlands and Richards, 1995).

In conservation genetic research, microsatellite markers can be divided into two groups, namely; cross-species markers and species-specific markers. Cross-species microsatellites are markers that have been developed for one species and used in closely related species (Schlötterer *et al.,* 1991; FitzSimmons *et al.,* 1995; Rico *et al.,* 1996; Gemmel *et al.,* 1997; Primmer and Ellegren, 1998). Cross-species amplification success decreases with increasing evolutionary distance (Primmer *et al.,* 1996). Although cross-species amplification is possible (Schlötterer *et al.,* 1991; FitzSimmons *et al.,* 1995; Rico *et al.,* 1996; Gemmell *et al.,* 1997; Primmer and Ellegren, 1998), if a species is being studied for the first time, species-specific microsatellites may have to be isolated

de novo (Zane *et al.,* 2002) because of the ineffectiveness of cross-species amplification over an evolutionary time scale (Primmer *et al.,* 2005).

Microsatellite detection requires sequence information and traditionally, microsatellite isolation methods consisted of colony hybridization-screening of partial genomic libraries with repeat containing probes followed by sequencing as illustrated in Figure 3 (Rassmann *et al.,* 1991). Protocols have also been developed for the production of DNA libraries enriched for microsatellite loci based on primer extension, as illustrated in Figure 4 or selective hybridization, as depicted in Figure 5 (Karagyozov *et al.,* 1993; Ostrander *et al.,* 1992). The primer extension methods involve the construction of a primary library containing fragmented genomic DNA inserted into a phagemid or phage vector in order to obtain a single strand DNA (ssDNA) library (Ostrander *et al.,* 1992; Paetkau 1999). Utilising repeat-specific oligonucleotides, the ssDNA is used as template for a primer extension reaction, which generates a double-stranded product only from vectors containing the desired repeat (Zane *et al.,* 2002). Selective hybridization protocols involve producing small genomic fragments that are ligated to a known sequence - a vector or an adaptor. Selective hybridization is then performed with an oligonucleotide containing several tandem repeats of the motif to be enriched as a probe (Zane *et al.,* 2002).

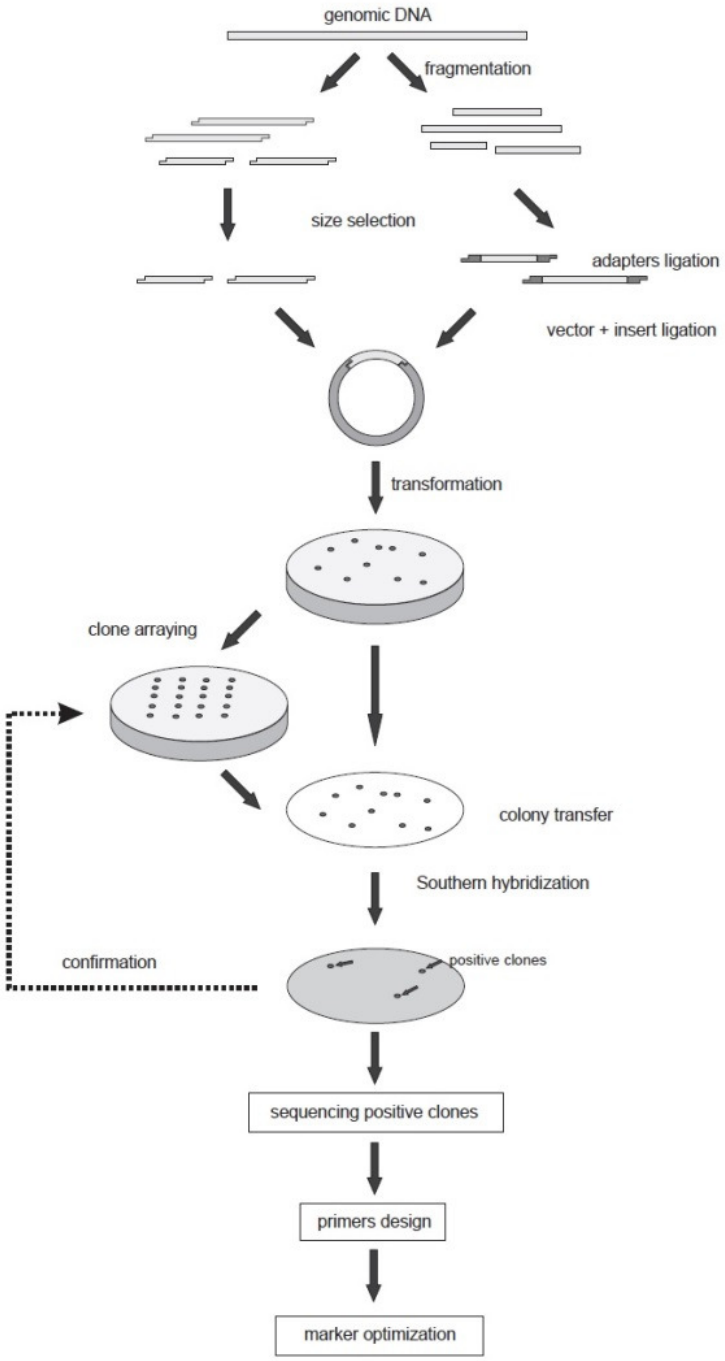


Figure 3: Schematic representation of traditional methods for microsatellite isolation (adapted from Zane *et al.* 2002).

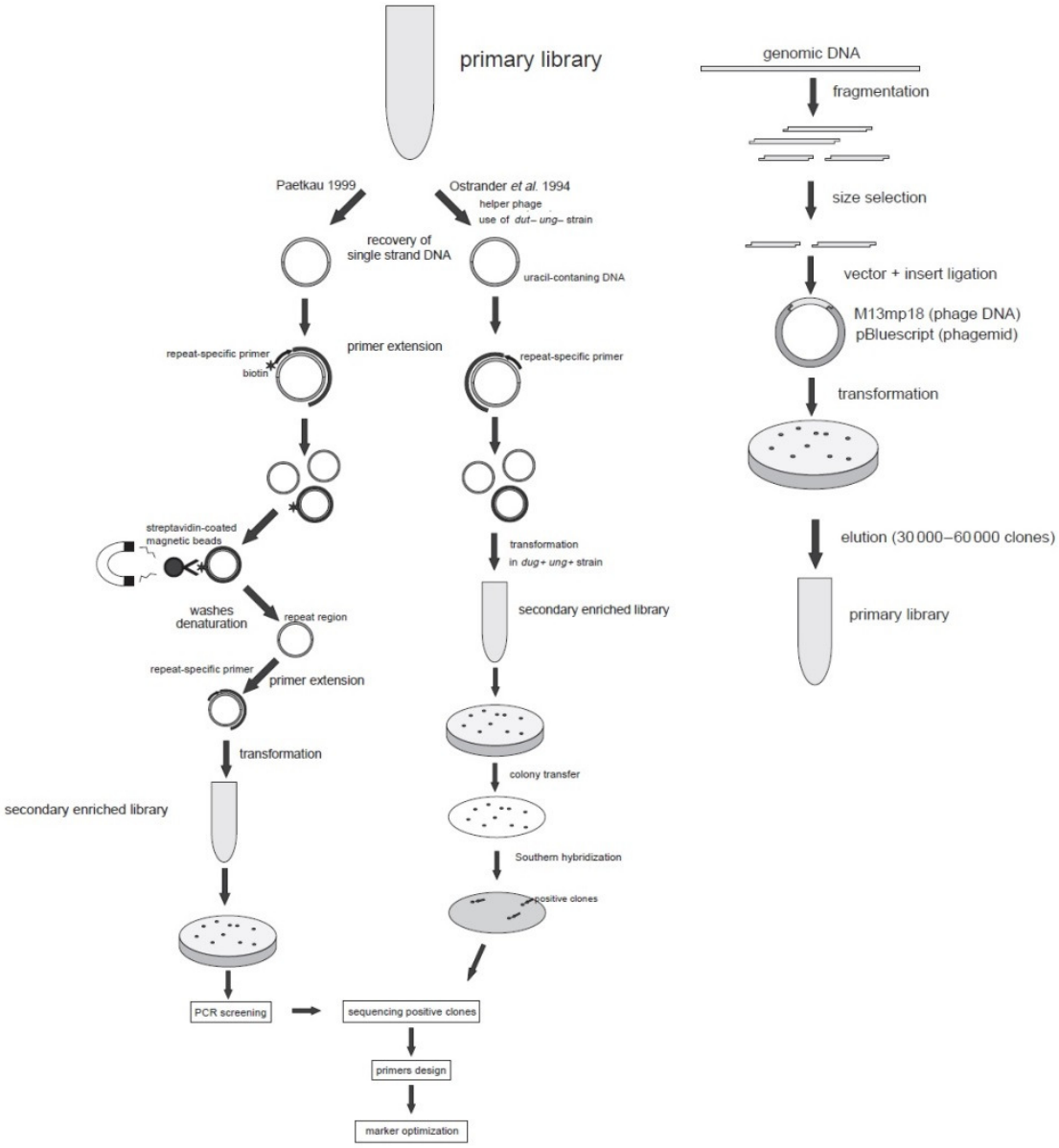


Figure 4: Schematic representation of primer extension microsatellite enrichment protocols (adapted from Zane *et al.* 2002).

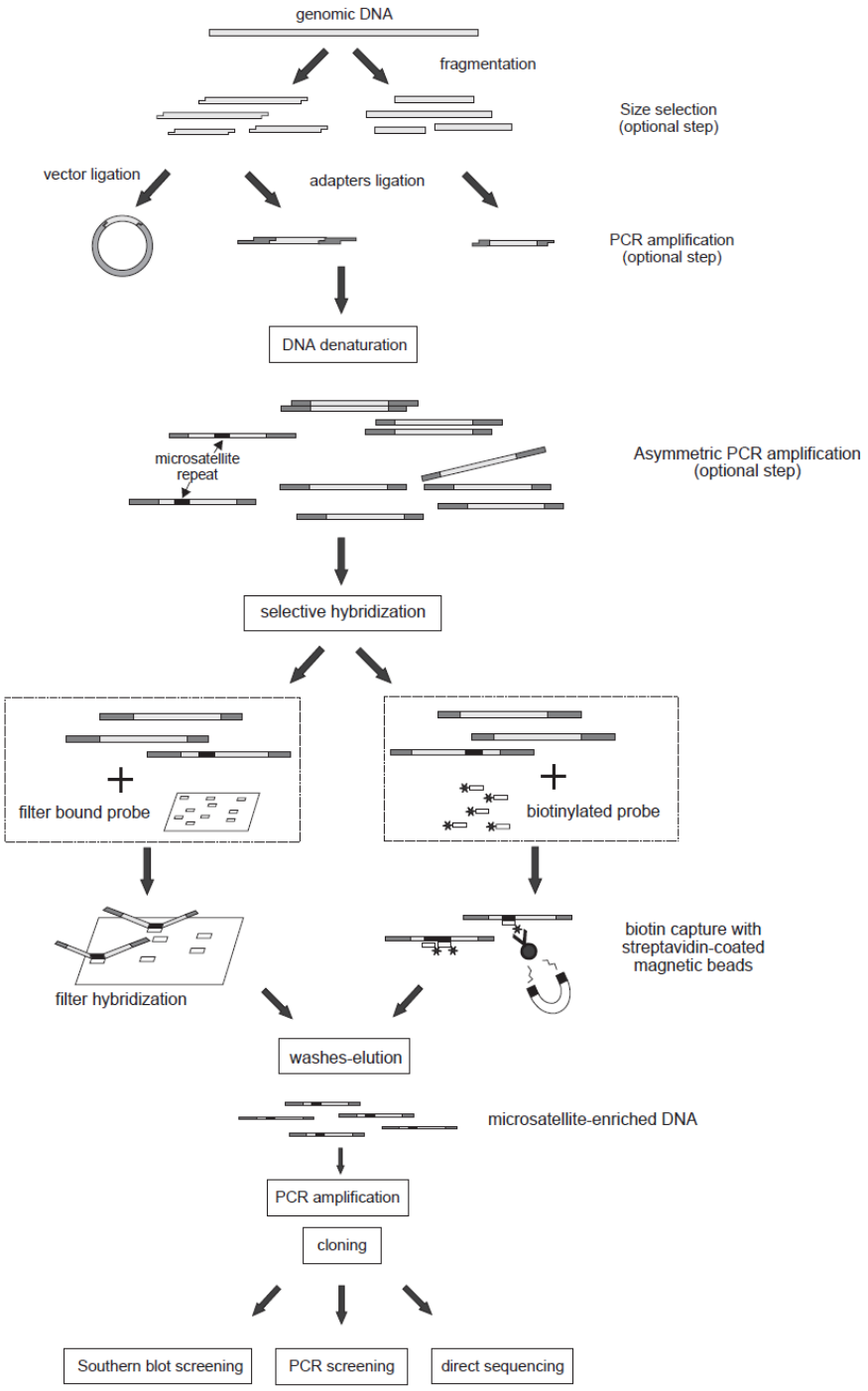


Figure 5: Schematic depiction of selective hybridization protocols for microsatellite discovery (by Zane *et al.* 2002).

These older procedures may be costly, time consuming (Squirrell *et al.,* 2003; Pashley *et al.,* 2006; Parchman *et al.,* 2010) and the yield may be lower compared to modern techniques (Santana *et al.,* 2009). Advances in cloning and next-generation sequencing have reduced the cost and time associated with microsatellite isolation (Abdelkrim *et al.,* 2009) and increased the yield (Santana *et al.,* 2009).

Seven microsatellite markers for the Humboldt Penguin were previously developed through enrichment of a size-selected genomic DNA library by magnetic bead selection with a biotin-

labelled oligonucleotide probe targeting CA repeats (Schlosser *et al.,* 2003). Twelve microsatellite markers were developed in the Yellow-eyed Penguin using a similar technique using biotinylated probes targeting GT, GA, AAC, AAG, ACT and ATC repeats (Bossenkool *et al.,* 2008). Ahmed and coworkers (2009) created an enriched library utilising GT and CT probes, as well as tetra-nucleotide repeats GTAA, CTAA, TTTC and GATA, developing 25 microsatellite markers for the Macaroni Penguin. In application of these markers, genetic diversity and population structure has been investigated using microsatellite markers in the Galapagos Penguin (Nims *et al.,* 2008). Microsatellite markers, in conjunction with mitochondrial markers, have been used to investigate population structure in the Magellanic and Yellow-eyed Penguins (Bossenkool *et al.,* 2009; Bouzat *et al.,* 2009). Paternity testing has been successfully performed in a captive population of 39 Adélie Penguins using eight microsatellite markers (Sakaoka *et al.,* 2014).

In the indian rhinoceros, mitochondrial and microsatellite markers have been used successfully to investigate genetic diversity and population structure (Zschokke *et al.,* 2011). Microsatellite markers for the indian rhinoceros were identified through enrichment of a size-selected genomic DNA library by magnetic bead selection with biotin-labelled oligonucleotide repeats (Zschokke *et al.,* 2003). In the sumatran rhinoceros, microsatellite markers have been developed through secondary screening of clone libraries with radio-labelled oligonucleotide probes (Scott *et al.,* 2004). These markers have been used to investigate genetic diversity and population structure (Goossens *et al.,* 2013). In the javan rhinoceros, mitochondrial and microsatellite markers have been used to investigate genetic diversity and population structure (Fernando *et al.,* 2004; Fernando *et al.,* 2006). For the black rhino, microsatellites have been developed through probe screening of a clone library (Rondebosch *et al.,* 1999), as well as the use of sequence data from public databases (Nielsen *et al.,* 2008). Microsatellites have been used in both population structure and parentage analysis in black rhino (Garnier *et al.,* 2001, *Kotzé et al.,* 2014).

As microsatellites can be used on a wide variety of species and because of their high degree of polymorphisms, these markers are a valuable tool in determining population structure (Bruford and Wayne, 1993), relatedness (Morin *et al.,* 1994), neonatal fitness (Coltman *et al.,* 1998), male mating success (Coltman *et al.,* 1999), bottleneck events (Luikart *et al.,* 1998a, b), evolutionary relationships, demographic history (Goldstein *et al.,* 1999) and detecting hybridization (Gottelli *et al.,* 1994; MacHugh, 1997; Goodman *et al.,* 1999; Evans *et al.,* 2001; Nijman *et al.,* 2003; Grant *et al.,* 2004; Gay *et al.,* 2008; Pastorini *et al.,* 2009).

Single nucleotide polymorphisms

SNPs (single nucleotide polymorphisms), a relatively new and promising tool in conservation genetics, allow us to survey both the neutral (non-coding) variation, as well as genes under selection (coding region) of genomic DNA and are single base pair positions at which different sequence alternatives can occur in a population, as indicated in Figure 2 (Pertoldi *et al.,* 2007; Ryynänen and Primmer, 2006). These markers are appealing for evolutionary and population genetic studies (Morin *et al.,* 2004; Namroud *et al.,* 2008; Stinchcombe & Hoekstra 2008; Slate *et*

al., 2009) as they represent the most abundant type of DNA variation in the vertebrate genome and are distributed across the entire genome providing broader genome coverage as compared to mitochondrial DNA or microsatellites (Pertoldi *et al.,* 2007; Ryynänen and Primmer, 2006). In addition, SNPs offer higher recovery of information from degraded DNA samples since the DNA target sequence in SNP-based genotyping is appreciably shorter (50 - 70 bp) than of that in microsatellite-based genotyping (80 - 300 bp) (Butler *et al.,* 2007; Morin *et al.,* 2004; Pertoldi *et al.,* 2007; Ryynänen and Primmer, 2006). An additional advantage is that individual SNPs may be associated with a phenotype. However, a whole genome association study is required to identify genomic regions with genes that influence traits or disease (Hirschhorn & Daly 2005; Karlsson *et al.,* 2007; McCarthy *et al.,* 2008). In contrast to microsatellites, SNP genotyping reveals polymorphisms directly on the DNA sequence and thus data is automatically standardized across chemistries, hardware platforms and laboratories (Glover *et al.,* 2010; Smith *et al.,* 2005). Furthermore, the development of high through-put genotyping platforms permits simultaneous genotyping of thousands of loci, enabling the identifications of highly diagnostic panels (Glover *et al.,* 2010). Since SNP genotype codes are independent of the genotyping system, the data can be standardised in public databases so that the data can be compared directly among studies (Morin *et al.,* 2004). Currently SNP loci for humans as well as other species are presented in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP), which enables information about the locus to be stored and searched, along with some information about alleles frequencies in a sample population (Morin *et al.,* 2004). The disadvantage of SNPs is that since they are bi-allelic, several SNPs are required to obtain the level of information gained from microsatellites.

Numerous approaches for SNP discovery have been described, including some also used for genotyping (Vignal *et al.,* 2002). Most methods are based on the comparison of locus-specific sequences, generated from different individuals (Vignal *et al.,* 2002). In humans, much of the SNP discovery has been done *in silico* with genomic information from multiple individuals already deposited in public databases screened for putative polymorphisms (Morin *et al.,* 2004). Such data is not available for most non-model organisms and SNPs have to be found through laboratory screening (Morin *et al.,* 2004).

SNP discovery efforts mostly use existing sequence information for closely related species or from genomic regions flanking microsatellites, which may limit the breadth of taxa and/or the genomic distribution of SNP loci used (Cramer *et al.,* 2008). Less restrictive methods of SNP discovery are needed to allow conservation geneticists to apply these powerful markers more broadly (Cramer *et al.,* 2008). One method suggested is shotgun cloning, which requires no pre-existing sequence data and may be readily applied to all taxa (Cramer *et al.,* 2008). This method is, however, completely random and time consuming involving random cloning of size selected fragments, sequencing and design of primers from successful clones and then screening of individuals with the selected loci for possible variation (Cramer *et al.,* 2008).

For most non-model organisms, SNPs have to be discovered through sequencing of segments of the genome from multiple individuals (Morin *et al.,* 2004). The targeted gene or genomic region

approach exploits the regions from multiple species from which PCR primers can be designed to amplify the orthologous gene regions in related species referred to as CATS (Comparative Anchor Tagged Sites) loci (Kwok, 2001; Morin *et al.,* 2004). The use of CATS loci allows the discovery of SNPs in genes of known function as well as known genomic location in some species, thus some genomic information is associated with the loci even without prior genomic characterization of the target species (Aitken *et al.,* 2004). Candidate genes of known function that might be predicted to influence fitness in a particular environment can be identified and sequenced (Morin *et al.,* 2004). Markers identified as closely linked with genes influencing fitness might provide a better indicator of levels of adaptive variation within populations and their potential to respond to changing environmental conditions (Morin *et al.,* 2004).

The targeted gene approach using CATS primers has been successfully used to identify 18 SNP markers for the sperm whale *(Physeter macrocephalus)* (Morin *et al.,* 2007). Recent SNP discovery strategies have resulted in characterization of SNPs in many natural populations of vertebrates to address several evolutionary, ecological and conservation issues. For example, SNPs have been applied for the identification of cryptic vole species, to investigate the level of genome introgression in a passerine bird hybrid zone and to study the population genetics of wolves (Ryynänen and Primmer, 2006). SNPs have been used in a wide variety of fish studies to assign country of origin (Smith *et al.,* 2005), determine genetic structure of populations (Wirgin *et al.,* 2007) and identify hybrids (He *et al.,* 2003). In domestic species SNPs have been used for individual verification and parentage (Werner *et al.,* 2004), population structure and origin and identification of desirable breeding traits (Beuzen *et al.,* 2000). However, only a few wildlife species have been investigated, which may be due to the relatively expensive genome-wide, species-specific research projects required to identify SNPs. Affordable SNP panels can thus be developed using cross-species SNPs. Ogden *et al.,* 2012 has identified polymorphic markers in the Scimitar-horned Oryx and Arabian Oryx by screening an array of SNPs previously developed for cattle. However, this method will only be applicable to species closely related to domestic species with previously identified SNPs.

Next generation sequencing

Next generation sequencing (NGS) consists of several very different technologies, each with its own set of characteristics, which rapidly generates huge amounts of sequence data in a cost effective way (Ekblom and Galindo 2011). A genomic or a transcriptomic route can be followed for NGS, with the latter utilising complementary DNA (produced from the mRNA of a specific tissue or life stage) as starting template. Data obtained may be used for studying nucleotide variation as well as transcriptome and gene expression analysis (Ekblom and Galindo 2011). Advances in NGS technology have been driven predominantly by major commercial entities since the completion of the Human Genome Project in 2003, including 454 Life Sciences Inc (Roche Applied Science, Branford, Connecticut), Illumina Inc (San Diego, California), and ThermoFisher Scientific Inc (Waltham, Massachusetts) (Roy et al., 2016). These NGS platforms are predominantly distinguished by different sequencing chemistries (synthesis versus ligation), methods for clonal

PCR of DNA fragments (bead-based emulsion PCR versus flow cell bridge PCR) and targeted approach (hybrid capture versus PCR amplification) (Roy et al., 2016). NGS enables the gathering of genomic information across multiple individuals at a genome-wide scale, both in mapping crosses and natural populations, opening new avenues in population genetics, quantitative trait mapping, comparative genomics and phylogeography (Etter *et al*., 2011). Restriction site associated DNA genotyping (RAD-seq) uses Illumina sequencing technology for the simultaneous discovery and typing of tens to hundreds of thousands of SNPs in hundreds of individuals for minimal investment of recourse (Etter *et al*., 2011). Trucchi and colleagues (2014) employed RAD-Seq in inferring part demography in the king penguin (*Aptenodytes patagonicus*). The development of genotyping-by-sequencing (GBS) approaches allow a targeted fraction of the genome (reduced representation library created by restriction enzymes, capture probes or transcriptome sequencing) to be sequenced with NGS technology rather than the entire genome (Narum *et al*., 2013). A challenge of NGS sequencing data analysis is the introduction of sequencing error, although low at 0.1-1% per nucleotide, it still becomes a significant source when millions of reads are considered simultaneously further compounded when assembling paralogous regions (Etter *et al*., 2011).

Utility of genetic markers

Currently, neutral genetic markers are used to assess inbreeding levels, genetic variation, population structure and phylogenetic or conservation units (Pertoldi *et al.,* 2007; Väli *et al.,* 2008). Neutral markers, however, do not provide insight into adaptive variation unless a large fraction of these markers are tightly linked to the relevant quantitative-trait loci (Pertoldi *et al.,* 2007). An additional problem associated with neutral markers lies in the nature of the two most commonly used markers, microsatellites and mitochondrial DNA sequences (Pertoldi *et al.,* 2007). Both these markers are fast evolving with high mutational rates providing high information content, but at a price of homoplasy, which together with other problems such as the presence of null alleles and complex mutation patterns, may introduce ambiguity to data analysis (Glover *et al.,* 2010; Morin *et al.,* 2004; Pertoldi *et al.,* 2007). Furthermore, scoring is platform-dependent and standardization often has difficulties, making inter-laboratory collaboration a challenge (Glover *et al.,* 2010; Smith *et al.,* 2005). Publication bias is another concern in microsatellite markers (Väli *et al.,* 2008). Markers with limited variability in initial screenings may not be considered worthwhile to use in more extensive population surveys or reports thus, regardless of the overall level of genomic variability, only the most polymorphic loci are sought resulting in a bias in estimates of the overall levels of genomic diversity (Väli *et al.,* 2008).

Genetic tools assist in understanding the nature of life and offer conservation applications in assessments of genetic variation within populations, biological parentage, kinship, gender identification, population structure, phylogeography, wildlife forensics, speciation, hybridization, introgression and phylogenetics. Currently, there are three genetic tools that can be used for wildlife conservation research, namely; single nucleotide polymorphisms (SNPs), mitochondrial DNA and microsatellites (Figure 2). Each tool has several advantages and disadvantages and is used to answer various conservation questions (as discussed below). However, the current availability of validated markers for South African wildlife species is limited (as shown in Table 1).

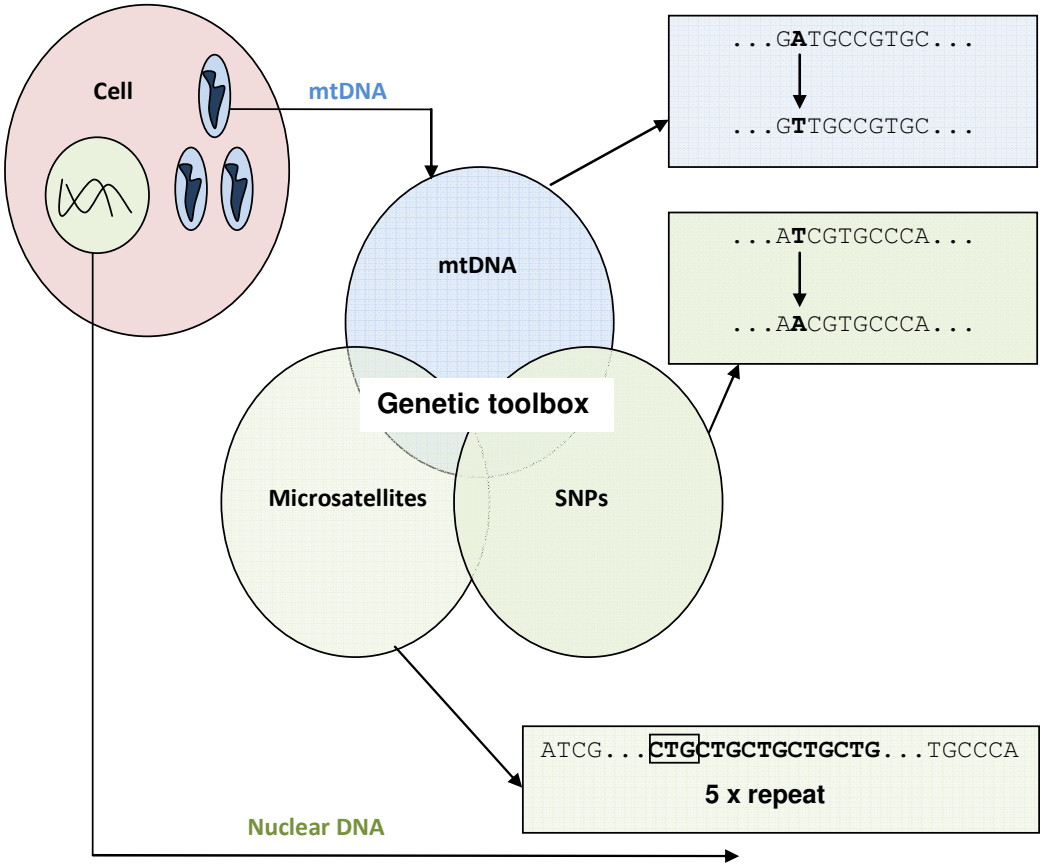


Figure 2: Diagrammatic representation of the genetic toolbox indicating mitochondrial DNA (mtDNA) and nuclear DNA markers (microsatellites and SNPs). Adapted from Morin *et al.,* 2004.

Table 1: List of genetic markers available for genetic conservation research in South African wildlife species. ✓ indicates informative; ✓✓ indicates highly informative and x indicated not informative. MS = microsatellite marker; MT = mitochondrial DNA marker and SNP = single nucleotide polymorphism.

Species	Genetic tool	Variation	Regional ID	Species ID	Population ID	Individual ID	Parentage	Hybridisation	Reference/s
African wild cat *(Felis lybica)* and domestic cat *(F. Catus)*	MS					✓		✓ ✓	Wiseman *et al.,* 2000; Le Roux *et al.,* 2014.
African wild dog *(Lycaon pictus)*	MS		x			✓	✓		Girman *et al.,* 1997; Girman *et al.,* 2001
	MT		x				✓		Girman *et al.,* 1997; Girman *et al.,* 2001
African Penguin (*Spheniscus demersus*)	MT	✓			✓				Murata and Murakami, 2014.
Black-backed jackal *(Canis mesomelas)*	MS	✓			✓ ✓				James R, 2014
Black faced impala (*Aepyceros melampus petersi*) and common impala (*A. m. Melampus*)	MS	✓ ✓			✓	✓		✓	Lorenzen and Siegismund, 2004; Eblate *et al.,* 2011; Schwab *et al*., 2012.
Black rhinoceros (*Diceros bicornis*)	MS	✓	✓	✓		✓			Cunningham *et al.,* 1999; Brown and Houlden, 1999; Nielsen *et al.,* 2008; Cain *et al.,* 2014; *Kotzé et al.,* 2014.
	MT		✓						*Kotzé et al.,* 2014
Black wildebeest (*Connochaetes gnou*) and blue wildebeest (*C. taurinus*)	MS	✓			✓			✓	Grobler *et al.,* 2005; Arctander *et al.,* 1999; Roed *et al.,* 2011.
	MT				✓				Arctander *et al.,* 1999
Bontebok *(Damaliscus pygargus pygargus*) and blesbok *(D. pygargus philllipsi)*	MS	✓		✓				✓ ✓	Dalton *et al.,* 2011; van Wyk *et al.,* 2013.
Brown hyena *(Hyaena brunnea)*	MS	✓ ✓	x				✓		Knowles *et al.,* 2009
Buffalo (*Syncerus caffer*)	MS	✓ ✓	✓			✓ ✓	✓ ✓		Van Hooft *et al.,* 1999; Simonsen *et al.,* 1998; O'Ryan *et al.,* 1998; Greyling *et al.,* 2008.
	MT	✓	✓		x				Van Hooft *et al.,* 2002;
	SNP	✓							Wenink *et al.,* 1998
Cape Clawless otter *(Aonyx capensis)* and the spotted necked otter *(Lutra maculicollis)*	MT			✓ ✓					Madisha *et al.,* 2015
Cape Parrot	MS			✓ ✓					Coetzer *et al.,* 2015

(Poicephalus robustus)	MT			✓ ✓					
Chacma baboon *(Papio ursinus sensu lato)*	MT		✓ ✓						Sithaldeen *et al.,* 2015
Cheetah *(Acinonyx jubatus*)	MS	✓ ✓				✓ ✓	✓ ✓		Marker *et al.,* 2008; Dalton *et al.,* 2013
	MT	✓							Freeman *et al.,* 2001
Common reedbuck *(Redunca arundinum)*	MT			✓					Dalton and Kotzé, 2011
Eland *(Taurotragus oryx)*	MS	✓							Eblate *et al.,* 2011
Elephant (*Loxodonta africana*)	MS	✓ ✓	✓ ✓		✓	✓ ✓			Whitehouse and Harley, 2001; Wasser *et al.,* 2004; Nyakaana *et al.,* 2002; Eggert *et al.,* 2002
	MT	✓	✓ ✓		✓				Nyakaana *et al.,* 2002; Eggert *et al.,* 2002
Giraffe *(Giraffa camelopardalis)*	MS	✓ ✓	✓ ✓		✓ ✓				Brown *et al.,* 2007; Huebinger *et al.,* 2002
	MT	✓ ✓	✓ ✓		✓ ✓				Brown *et al.,* 2007; Bock *et al.,* 2014
Grants gazelle (*Nanger granti*)	MS	✓							Eblate *et al.,* 2011
Hartebeest (*Alcelaphus buselaphus*)	MS	✓			✓				Arctander *et al.,* 1999; Eblate *et al.,* 2011
	MT	✓			✓				Arctander *et al.,* 1999
Hippopotamus (*Hippopotamus amphibious*)	MT	✓	✓		✓	✓			Okello *et al.,* 2005
Leopard *(Panthera pardus):*	MS	✓			□ □				McManus *et al.,* 2015
	MT	✓			✓ ✓				
Lion *(Panthera leo)*	MS	✓	✓		✓ ✓	✓ ✓	✓ ✓		Spong *et al.,* 2002; Bertola *et al.,* 2014; Miller *et al.,* 2014
	MT		✓						Bertola *et al.,* 2014
Montain Zebra *(Equus zebra*)	MS	✓ ✓			✓ ✓	✓			Moodley and Harley, 2005
	MT	✓			✓ ✓				Moodley and Harley, 2005
Nile crocodile *(Crocodylus niloticus)*	MS	✓	✓ ✓	✓					Hekkala *et al.,* 2010
Nyala *(Tragelaphus angasii)* and greater kudu *(T. strepsiceros)*	MT			✓ ✓					Dalton *et al*., 2014; Grobler *et al*., 2005b.
	MS							✓ ✓	
Plains Zebra (*Equus quagga*)	MS	✓			x	✓			Lorenzen *et al.,* 2008
	MT	✓			x				Lorenzen *et al.,* 2008
Red-billed oxpeckers *(Buphagus erythrorhynchus)*	MS	✓ ✓							Spies *et al.,* 2012
Roan antelope (*Hippotragus equines*)	MS	✓	✓		✓	✓			Alpers *et al.,* 2004; Eblate *et al.,* 2011
	MT	✓	✓		✓				Alpers *et al.,* 2004

Samango monkey *(Cercopithecus albogularis)*	MS			✓					Dalton *et al.,* 2015
	MT			✓					
Spotted hyena *(Crocuta crocuta)*	MS	✓					✓		Wilhelm *et al.,* 2003; Libants *et al.,* 2000
South African Climbing Mice (*Dendromus sp.*)	MT			✓					Solano *et al.,* 2014
Southern Ground-Hornbills *(Bucorvus leadbeateri)*	MS	✓ ✓			✓				Theron *et al.,* 2013
Temminck's ground pangolin *(Smutsia temminckii)*	MT			✓ ✓					Du Toit *et al.,* 2014
Topi (*Damaliscus korrigum*)	MS	✓							Eblate *et al.,* 2011
Vervet monkey *(Chlorocebus sensu lato)*	MT				✓ ✓				Turner *et al.,* 2015
Warthog (*Phacochoerus africanus*)	MS	✓ ✓	✓ ✓		✓ ✓				Muwanika *et al.,* 2003; Muwanika *et al.,* 2007;
	MT	✓	✓ ✓		✓ ✓				Muwanika *et al.,* 2003; Muwanika *et al.,* 2007
White rhinoceros (*Cerathotherium simum simum*)	MS	✓		✓		✓			Florescu *et al.,* 2003; Nielsen *et al.,* 2008.

Individual identification and parentage

Individual identification and parentage using molecular methods is emerging as a critical tool in conservation genetics and molecular ecology. In addition, non-invasive sampling (faeces, shed hair, or shed feathers) in many conservation studies is of prime importance. Individual identification is required for monitoring of animal movements and abundance, in forensic applications and in behavioural studies. Molecular markers can also be used to determine the relationship and relatedness between individuals when the ancestry is unknown. Parentage and individual identification analysis assumes that loci are error-free, independent (unlinked) and that population allele frequencies are accurately estimated (Boehnke and Cox, 1997).

Several tools exist for individual identification and parentage, including SNPs and microsatellite markers. However, unlike human identification kits, there are currently no commercial kits available for wildlife DNA testing for forensic applications and for the conservation community. Currently, there are several studies that have identified markers for individual identification and parentage (Table 1); however few of these studies have validated their marker sets. In order to validate molecular tools, a population study on the species of interest would be required using a number of markers to develop a database. Several markers with high heterozygosity (approaching 0.50) should be included in the database (Miller *et al.,* 2002) to provide sufficient statistical power for parentage analysis and individual identification (Chakraborty *et al.,* 1999). Markers should be selected only if they display: absence of linkage, have a high probability of parentage exclusion and

individual genetic identity and have a high level of genetic diversity (Kanthaswamy *et al.,* 2006). Numerous studies would then have to be conducted to validate the markers including species-specificity, sensitivity testing, peak height ratio and intra- and inter-locus colour balance as well as sequencing of markers to evaluate genotype data (Shibuya *et al.,* 1994). In addition, the presence of null alleles may affect the interpretation of data and the presence of null alleles would have to be investigated. Null alleles occur when one of the alleles does not amplify during the polymerase chain reaction (PCR) due to a mutation (Blouin, 2003), thus a heterozygous individual (an) will be typed as a homozygous individual (aa). The presence of null alleles may eliminate possible candidates as parents in parentage analysis. Thus analysis of data should include statistical programmes that correct for the possibility of null alleles (Wagner *et al.,* 2006). In regards to non-invasive sampling, it is of critical importance that pilot studies also be conducted to assess the probability of identity as well as technical errors that may arise (Taberlet and Luikart, 1999). The database would require geographically representative populations of the species and would include information on locus informativeness, allele frequencies, distribution of genetic variation, match probability estimates as well as inbreeding coefficients (Kanthaswamy, *et al.,* 2006).

Hybridization

Hybridization is the inbreeding of individuals from different genetic populations, despite their taxonomic status (Rhymer and Simberloff, 1996). The mating of heterospecific individuals are more commonly referred to as hybridization but hybridization has also been applied to the mating of individuals of different sub-species and also mating of individuals that are taxonomically the same, but differ genetically (Rhymer and Simberloff, 1996). Hybridization has led to the extinction of many populations of plant and animal taxa (Rhymer and Simberloff, 1996) and is therefore extremely problematic for rare taxa that come into contact with more abundant taxa. To date, the most powerful molecular methods used to the detect hybridization include SNPs and microsatellites. Mitochondrial DNA is less likely to assist in hybridizing testing as it is maternally inherited. Thus it has been proposed that any hybrid or subsequent generations would possess only maternal mitochondrial DNA (Ward *et al.,* 2005). In Southern Africa there are several species that can hybridize, these include: Black-faced impala x common impala (Green and Rothstein, 1998); Mallard ducks x indigenous waterfowl; African wild cat x domestic cat (Wiseman *et al.,* 2000); donkeys and Burchell's zebra (Gray, 1972); red Hartebeest and Blesbok (Robinson *et al.,* 1991); gemsbok and scimitar-horned Oryx (Brooke *et al.,* 1986). In addition, many populations of springbok and impala that were genetically distinct have been mixed via translocation (Brooke *et al.,* 1986). Thus far, markers that can accurately discriminate between pure and hybrid animals in South Africa have only been reported in Bontebok and Blesbok (van Wyk *et al.,* 2013), as well as African wildcat and domestic cat (Wiseman *et al.,* 2000) as shown in Table 1.

Genetic variation and diversity

Genetic diversity is the degree of genetic variation in a population or among species and is expressed in different characteristics, including; phenotypes, behaviour, proteins and enzymes (Frankham *et al.,* 2010). Genetic diversity allows populations to adapt to environmental changes.

These environmental changes include climate change, pests, parasites, food sources, predators, competitors, pollution and infectious disease. Large populations with high levels of genetic diversity are better able to adapt to these changes since they have a large genetic reservoir (Frankham *et al.,* 2010). Due to genetic drift and inbreeding small populations generally have lower levels of genetic diversity and are therefore not able to adapt to environmental changes as well compared to larger populations (Frankham *et al.,* 2010). Bijlsma and Loeschcke (2005) indicated that genetic diversity enables populations to adapt to environmental changes such as heavy metal pollutants, pesticides, herbicides. Inbreeding is associated with loss in genetic diversity which causes a decrease in population fitness and ultimately species survival. Reed and Frankham (2003) and Leimu and colleagues (2006) indicated a positive correlation between population fitness and genetic diversity. Inbreeding increases homozygosity, consequently exposing more deleterious alleles and causing reduced fitness in populations which can lead to inbreeding depression (Frankham *et al.,* 2010). Currently both SNPs and microsatellites can be used to determine levels of genetic diversity.

Geographic genetic differences and origin

Molecular genetic markers can also be used to determine genetic partitions among geographically isolated populations as well as to identify units for conservation management purposes (Avise and Ball 1990; Moritz 1994; Fraser and Bernatchez 2001). Due to fragmentation of our landscape and the establishment of borders between farming ventures and conservation systems, populations of species are continually being separated into smaller sub-populations. The resulting reduced gene flow due to isolation can result in a higher probability of local extinction due to human impact or by demographic or environmental stochasticity (Roelke-Parker *et al.,* 1996). The developed metapopulation structures may thus vary genetically (McCauley, 1991) and will have to be managed through the use of different conservation strategies. Thus the understanding of the genetic structure of wildlife populations is essential. In addition, from a forensic standpoint the ability to trace individuals within a species to a particular geographic area (Alacs *et al.,* 2010) is of particular importance. Geographic origins of individuals can only be determined if genetic structure occurs within the regions of interest. In order to determine population structure, samples need to be collected from each region and analysed using microsatellite markers, SNPs and/or mtDNA, as indicated in Table 1. Statistical programmes such as STRUCTURE (Pritchard *et al.,* 2000; Falush *et al.,* 2003) can then be used for either phylogeographic assessments or for population assignment.

DNA Barcoding

DNA barcoding has been identified as a rapid and practical molecular tool that can be used to identify species due to species-specific variation in short DNA sequences from one or a few selected genomic regions (Spooner, 2009). In wildlife law enforcement cases, a common problem is identifying the species of origin of carcasses, meat or blood; in the absence of other morphological characteristics. Amplification and sequencing using universal primers for cyt*b* and cytochrome c oxidase I (COI) genes provide a reliable test for the identification of species in wildlife

forensic cases (Branicki *et al.,* 2003; Hebert *et al.,* 2003). However, these genes enable accurate animal species identification only in the case where adequate reference sequence data exists (Dawnay *et al.,* 2007). Discrimination between species is then determined via sequence differences (Alacs *et al.,* 2003; Hsieh *et al.,* 2001; Lo *et al.,* 2006; Wong *et al.,* 2004; Dalton and Kotzé, 2011). Although only a few studies have published barcodes for South African wildlife species, several freely accessible databases exist, such as the Barcode of Life Data Systems (BOLD) which includes COI-barcode libraries of voucher specimens. The National Zoological Gardens of South Africa (NZG) is undertaking the barcoding of South African terrestrial vertebrate species (birds, mammals, reptiles and amphibians) as part of a collaborative project with South African Biodiversity Institute (SANBI) and the African Centre for DNA Barcoding (ACDB) at the University of Johannesburg. This project forms part of a larger initiative, the Consortium for Barcode of Life (CBOL). CBOL is hosted by the Shithsonian Institute in the USA. The goal of this initiative is to create a reference library verified for all species; a quick way of identifying organisms in nature simply by comparing a DNA barcode to an online reference library linked to relevant information for each species.

References

Abdelkrim J., Robertson B.C., Stanton J.A.L., Gemmell N.J. 2009. Fast, cost effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*. 46: 185-192.

Ahmed S., Hart T., Dawson D.A., Horsburgh G.J., Trathan P.N., Rogers A.D. 2009. Isolation and characterization of macaroni penguin (*Eudyptes chrysolophus*) microsatellite loci and their utility in other penguin species (Spheniscidae, AVES). *Molecular ecology resources*, *9*(6): 1530-1535.

Aitken N., Smith S., Schwarz C., Morin P.A. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology* 13: 1423-1431.

Alacs E., Alpers D., De Tores P., Dillon M., Spencer P.B.S. 2003. Identifying the presence of quokkas (*Setonix brachyurus*) and other macropods using cytochrome b analyses from faeces. *Wildlife Res*. 30: 41-47.

Alacs E.A., Georges A., Fitzsimmons N.N., Robertson J. 2010. DNA detective: a review of molecular approaches to wildlife forensics. *Forensic Sci. Med. Pathol.* 6(3): 180-194.

Alpers D.L., Van Vuuren B.J., Arctander P., Robinson T.J. 2004. Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Mol Ecol*. 13: 1771-1784.

Arctander P., Johansen C., Coutellec-Vreto M.A. 1999. Phylogeography of three closely related African bovids (tribe Alcelaphini). *Mol. Biol. Evol.* 16(12): 1724-1739.

Arif I.A. and Khan H.A. 2009. Molecular markers for biodiversity analysis of wildlife animals: a brief review. *Biodiversity Conser*. 32.1: 9-17.

Avise J.C. and Ball R.M. 1990 Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surv Evol Biol*. 7: 45-67.

Bertola L.D., Tensen L., van Hooft P., White P.A., Driscoll C.A., Henschel P., Caragiulo A., Freedman I.D., Sogbohossou E.A., Tumenta P.N., Jirmo T.H., de Snoo G.R., de Iongh H.H., Vrieling K. 2015. Autosomal and mtDNA Markers Affirm the Distinctiveness of Lions in West and Central Africa. *PloS one*, *10*(10): e0137975.

Beuzen N.D., Stear M.J., CHANG K.C. 2000. Molecular markers and their use in animal breeding. *Vet. J*. 160: 42-52.

Bijlsma R. and Loeschcke V. 2005. Environmental stress, adaptation and evolution: an overview. *J. Evol. Biol.* 18: 744-749.

Blouin M.S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends Ecol. Evolut.* 18: 503-511.

Bock F., Fennessy J., Bidon T., Tutchings A., Marais A., Deacon F., Janke A. 2014. Mitochondrial sequences reveal a clear separation between Angolan and South African giraffe along a cryptic rift valley. *BMC evolutionary biology*, *14*(1): 219.

Boehnke M. and Cox N.J. 1997. Accurate Inference of Relationships in Sib-Pair Linkage Studies. *Am. J. Hum. Genet*. 61: 423-429.

Boessenkool S., King T.M., Seddon P.J., Waters J.M. 2008. Isolation and characterization of microsatellite loci from the yellow-eyed penguin (*Megadyptes antipodes*). *Molecular ecology resources*, *8*(5): 1043-1045.

Boessenkool S., Star B., Waters J.M., Seddon P.J. 2009. Multilocus assignment analyses reveal multiple units and rare migration events in the recently expanded yellow-eyed penguin (*Megadyptes antipodes*). *Molecular ecology*, *18*(11): 2390-2400.

Boore J.L. 1999. Animal mitochondrial genomes. *Nucl. Acids Res.* 27 (8): 1767-1780.

Bouzat J.L., Walker B.G., Boersma P.D. 2009. Regional genetic structure in the Magellanic penguin (*Spheniscus magellanicus*) suggests metapopulation dynamics. *The Auk*, *126*(2): 326-334.

Branicki W., Kupiec T., Pawlowski R. 2003. Validation of cytochrome b sequence analysis as a method of species identification. *J. Forensic Sci.* 48: 83-87.

Britten R.I. and Kohne D.E. 1968. Repeated sequences in DNA.*Science*. 161: 529-540.

Brooke R.K., Lloyd P.H., De Villiers A.L. 1986. Alien and translocated terrestrial vertebrates in South Africa. Macdonald I.A.W., Kruger, F.J., Ferrar A.A. Editors, The Ecology and Management of Biological Invasions in Southern Africa, Oxford University Press, Cape Town, pp. 63-74.

Brown D.M., Brenneman R.A., Koepfli K.P., Pollinger J.P., Milà B., Georgiadis N.J., Louis E.E., Grether G.F., Jacobs D.K., Wayne R.K. 2007. Extensive population genetic structure in the giraffe. *BMC Biol*. 5: 57.

Brown S.M. and Houlden B.A. 1999. Isolation and characterization of microsatellite markers in the black rhinoceros (*Diceros bicornis*). *Mol. Ecol.* 8: 1559-1561.

Bruford M.W. and Wayne R.K. 1993. Microsatellites and their application to population genetic studies. *Curr. Opin. Genet. Dev.* 3: 939-943.

Butler J.M., Coble M.D., Vollone P.M. 2007. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol* (3):200-205.

Cain B., Wandera A.B., Shawcross S.G., Edwin-Harris W., Stevens-Wood B.A.R.R.Y., Kemp S.J., Watts P.C. 2014. Sex-Biased Inbreeding Effects on Reproductive Success and Home Range Size of the Critically Endangered Black Rhinoceros. *Conservation biology*, *28*(2): 594-603.

Chakraborty R., Stivers D.N., Su B., Zhong Y., Budowle B. 1999. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis*. 20(8): 1682-96.

Chambers G.K. and Macavoy E.S. 2000. Microsatellites: Consensus and controvery. *Comp Biochem Physiol B Biochem Mol Biol*. 126 (4): 455-76.

Coetzer W.G., Downs C.T., Perrin M.R., Willows-Munro S. 2015. Molecular systematics of the Cape Parrot (*Poicephalus robustus*): implications for taxonomy and conservation. *PloS one*, *10*(8): e0133376.

Coltman D.W., Bowen W.D., Wright J.M. 1998. Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites. *Proc. Roy. Soc. B* 265: 803-809.

Coltman D.W., Pilkington J.G., Smith J.A., Pemberton J.M. 1999. Parasite-mediated selection against inbred soay sheep in a free-living, island population. *Evolution*. 53: 1259–1267.

Cramer E.R.A., Stenzler L., Talaba A.L., Makarewich C.A., Vehrencamp S.L., Lovette I.J. 2008. Isolation and characterization of SNP variation at 90 anonymous loci in the banded wren (*Thryothorus pleurostictuc*). *Conserv Genet*. 9(6):1657-1660.

Cunningham J., Harley E.H., O'ryan C. 1999. Isolation and characterization of microsatellite loci in black rhinoceros (*Diceros bicornis*). *Electrophoresis*. 20: 1778-1780.

Dalton D.L. and Kotzé A. 2011. DNA Barcoding as a tool for species identification in three forensic wildlife cases in South Africa. *Forensic Sci. Int.* 207: e51-54.

Dalton D.L., van Wyk A.M., Kotzé A. 2011. Isolation and characterization of microsatellite loci in blesbok (*Damaliscus pygargus phillipsi*). *Conservation Genetics Resources*, *3*(4): 677-679.

Dalton D.L., Charruau P., Boast L., Kotzé A. 2013. Social and genetic population structure of free-ranging cheetah in Botswana: implications for conservation. *European journal of wildlife research*, *59*(2): 281-285.

Dalton D.L., Tordiffe A., Luther I., Duran A., van Wyk A.M., Brettschneider H., Ooshuizen A., Modiba C., Kotzé A. 2014. Interspecific hybridization between greater kudu and nyala. *Genetica*, *142*(3): 265-271.

Dalton D.L., Linden B., Wimberger K., Nupen L.J., Tordiffe A.S., Taylor P.J., Madisha M.T., Kotzé A. 2015. New Insights into Samango Monkey Speciation in South Africa. *PloS one*, *10*(3): e0117003.

de Mendonça Dantas G.P., de Oliveira L.R., Marasco A.C.M., de Araujo J., Hurtado R., Durigon E.L., Morgante J.S. 2014. Demographic History of the Gentoo Penguin (*Pygoscelis papua*) and the Adélie Penguin (*Pygoscelis adeliae*) on Admiralty Bay, King George Island, Antarctica. *Waterbirds*, 37(4): 410-418.

Dawnay N., Ogden R., McEwing R., Carvalho G.R., Thorpe R.S. 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. *J. Forensic Sci.* 173: 1-6.

Du Toit Z., Grobler J.P., Kotzé A., Jansen R., Brettschneider H., Dalton D.L. 2014. The complete mitochondrial genome of Temminck's ground pangolin (*Smutsia temminckii*; Smuts, 1832) and phylogenetic position of the Pholidota (Weber, 1904). *Gene*, *551*(1): 49-54.

Eblate E.M., Lughano K.J., Sebastian C.D., Peter M.L., Knut R.H. 2011. Polymorphic microsatellite markers for genetic studies of African antelope species. *Afr. J. Biotechnol*. 10 (56): 11817-11820.

Eggert L.S., Rasner C.A., Woodruff D.S. 2002 The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers. *Proc. R. Soc. Lond. B.* 269: 1993-2006.

Ekblom R. and Galindo J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity, 107(1): 1-15.

Etter P.D., Bassham S., Hohenlohe P.A., Johnson E.A., Cresko W.A. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Molecular methods for evolutionary genetics*, 157-178.

Evans B.J., Supriatna J., Melnick D.J. 2001. Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution*. 55: 1686-1702.

Falush D., Stephens M., Pritchard J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 164: 1567-1587.

Fernando P. and Melnick D.J. 2004. Genetic Analysis of the Javan Rhinoceros in relation to its management. *Unpublished Report submitted to the United States Fish and Wildlife Service, Rhinoceros and Tiger Conservation Fund*.

Fernando P., Polet G., Foead N., Ng L.S., Pastorini J., Melnick D.J. 2006. Genetic diversity, phylogeny and conservation of the Javan rhinoceros (*Rhinoceros sondaicus*). *Conservation Genetics*, *7*(3): 439-448.

Field D. and Wills C. 1996. Long, polymorphic microsatellites in simple organisms. *Proc Biol Sci.* 22; 263 (1367): 209-15.

Fitzsimmons N.N., Moritz C., Moore S.S. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.* 12: 432-440.

Florescu A., Davila J.A., Scott C., Fernando P., Kellner K., Morales J.C., Melnick D., Boag P.T., Van Coeverden De Groot P. 2003. Polymorphic microsatellites in white rhinoceros. *Mol. Ecol. Resour.* 3: 344-345.

Frankham R. 2010. Inbreeding in the wild really does matter. *Heredity*. 104(2):124.

Fraser D.J. and Bernatchez L. 2001. Adaptive evolutionary conservation: Towards a unified concept for defining conservation units. *Mol. Ecol.* 10: 2741–2752.

Freeman A.R., Machugh D.E., Mckeown S., Walzer C., Mcconnell D.J., Bradley D.G. 2001. Sequence variation in the mitochondrial DNA control region of wild African cheetahs (*Acinonyx jubatus*). *Heredity*. 86: 355-362.

Garnier J.N., Bruford M.W., Goossens B. 2001. Mating system and reproductive skew in the black rhinoceros. *Molecular Ecology*, *10*(8): 2031-2041.

Gay L., Crochet P.A., Bell D.A., Lenormand T. 2008. Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*. 62: 2789-2806.

Gemmell N.J., Allen P.J., Goodman S.J., Reed J.Z. 1997. Interspecific microsatellite markers for the study of pinniped populations. *Mol. Ecol.* 6: 661-666.

Girman D.J., Mills M.G.L., Geffen E., Wayne R.K. 1997. A molecular genetic analysis of social structure, dispersal, and interpack relationships of the African wild dog (*Lycaon pictus*). *Behav Ecol Socialbiol*. 40: 187-198.

Girman D.J., Vilà C., Geffen E., Creel S., Mills M.G.L., Mcnutt J.W., Ginsberg J., Kat P.W., Mamiya K.H., Wayne R.K. 2001. Patterns of population subdivision, gene flow and genetic variability in the African wild dog (*Lycaon pictus*). *Mol Ecol*. 10: 1703-1723.

Glover K.A., Hansen M.M., Lien S., Als T.D., Høyheim B., Skaala Øystein. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics* 11: 2.

Goldstein D.B., Roemer G.W., Smith D.A., Reich D.E., Bergman A., Wayne R.K. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics*. 151: 797-801.

Goodman S.J., Barton N.H., Swanson G., Abernethy K., Pemberton J.M. 1999. Introgression through rare hybridization: A genetic study of a hybrid zone between red and sika deer (genus Cervus) in Argyll, Scotland. *Genetics* 152: 355-371.

Goossens B., Salgado-Lynn M., Rovie-Ryan J.J., Ahmad A.H., Payne J., Zainuddin Z.Z., Ambu L.N. 2013. Genetics and the last stand of the Sumatran rhinoceros *Dicerorhinus sumatrensis*. *Oryx*, *47*(03): 340-344.

Gottelli D., Sillero-Zubiri C., Applebaum G.D. 1994. Molecular genetics of the most endangered canid: the European wolf *Canis simensis*. *Mol. Ecol.* 3: 301-312.

Grant P.R., Grant B.R., Markert J.A., Keller L.F., Petren K. 2004. Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*. 58: 1588-1599.

Gray A.P. 1972. Mammalian hybrids. A check list with bibliography, 2nd ed. Commonwealth Agricultural Bureaux, Farnham Royal, 262 pp.

Green W.C.H. and Rothstein A. 1998. Translocation, Hybridization, and the Endangered Black-Faced Impala. *Conserv. Biol.* 12: 475-480.

Greyling B.J., Kryger P., Du Plessis S., Van Hooft W.F., Van Helden P., Getz W.M., Bastos A.D.S. 2008. Development of a high-throughput microsatellite typing approach for forensic and population genetic analysis of wild and domestic African Bovini. *Afr. J. Biotechnol*. 7 (5): 655-660.

Grobler J.P., Hart G.B., Grobler N., Kotzé A., Botha K., Tiedemann R. 2005. The genetic status of an isolated black wildebeest (*Connochaetes gnou*) population from the Abe Bailey nature reserve, South Africa: Microsatellite data on a putative past hybridization with blue wildebeest (*C. taurinus*). *Mamm. Biol*. 70: 35-45.

Grobler J.P., Pretorius D.M., Botha K., Kotze A., Hallerman E.M., Jansen van Vuuren B. 2005b. An exploratory analysis of geographic genetic variation in southern African nyala (*Tragelaphus angasii*). *Mammalian Biology*, 70: 291-299.

Gur-Arie R., Cohen C.J., Eitan Y., Shelef L., Hallerman E.M., Kashi Y. 2000. Simple sequence repeats in Escherichia coli: abundance, distribution, composition, and polymorphism. *Genome Res*. 10: 62-71.

He C., Chen L., Simmons M., Li P., Kim S., Liu Z.J. 2003. Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Anim Genet*. 34: 445-448.

Hebert P.D.N., Ratnasingham S., Dewaard J.R. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lon.* B 270 (Suppl. 1): 96-99.

Hekkala E.R., Amato G., Desalle R., Blum M.J. 2010. Molecular assessment of population differentiation and individual assignment potential of Nile crocodile (*Crocodylus niloticus*) population. *Conserv Genet*. 11: 1435-1443.

Hirschhorn J.N. and Daly M.J. 2005. Genome-wide association studies for common diseases and complex traits. *Nature reviews genetics*, 6(2): 96-108.

Hsieh H.M., Chiang H.L., Tsai L.C., Lai S.Y., Huang N.E., Linacre A., Lee J.C. 2001. Cytochrome b gene for species identification of the conservation animals. *Forensic Sci. Int.* 122: 7-18.

Huebinger R.M., Pierson D.J., De Maar T.W., Brown D.M., Brenneman R.A., Louis E.E. 2002. Characterization of 16 microsatellite marker loci in the Maasai giraffe (*Giraffa camelopardalis tippelskirchi*). *Mol. Ecol. Notes.* 2: 531-533.

James R. 2014. The population dynamics of the black-backed jackal *(Canis mesomelas*) in game farm ecosystems of South Africa. Diss. University of Brighton.

Jeffreys A.J., Wilson V., Thein S.L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature*. 314: 67-73.

Kanthaswamy S., Von Dollen A., Kurushima J.D., Alminas O., Rogers J., Ferguson B., Lerche N.W., Allen P.C., Smith D.G. 2006. Microsatellite Markers for Standardized Genetic management of Captive Colonies of Rhesus Macaques (*Macaca mulatta*). *Amer. J. Primatol*. 68: 73-95.

Karagyozov L., Kalcheva I.D., Chapman V.M. 1993 Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Research*, 21: 3911-3912.

Karlsson E.K., Baranowska I., Wade C.M., Hillbertz N.H.S., Zody M.C., Anderson N., Lindblad-Toh K. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature genetics*, 39(11): 1321-1328.

Knowles J.C., Van Coeverden De Groota P.J., Wieselb I., Boaga P.T. 2009. Microsatellite variation in Namibian brown hyenas (*Hyaena brunnea*): Population structure and mating system implications. *J Mammal*. 90: 1381-1391.

Kotzé A., Dalton D.L., du Toit R., Anderson N., Moodley Y. 2014. Genetic structure of the black rhinoceros (*Diceros bicornis*) in south-eastern Africa. *Conservation Genetics*, *15*(6): 1479-1489.

Kwok P. 2001. Methods for genotyping single nucleotide polymorphisms. *Ann. Rev. Genomics Hum. Genet.* 2: 235-58.

Lacy R.C. 1997. Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy*, 78(2): 320-335.

Le Roux J.J., Foxcroft L.C., Herbst M., MacFadyen S. 2015. Genetic analysis shows low levels of hybridization between African wildcats (*Felis silvestris lybica*) and domestic cats (*F. s. catus*) in South Africa. *Ecology and evolution*, *5*(2): 288-299.

Leimu R., Mutikainen P., Koricheva J., Fischer M. 2006. How general are positive relationships between plant population size, fitness and genetic variation? *J. Ecol*. 94: 942-952.

Libants S., Olle E., Oswald K., Scribner K.T. 2000. Microsatellite loci in the spotted hyena *Crocuta crocuta*. *Mol Ecol*. 9: 1443-1445.

Litt M. and Luty J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet.* 44 (3): 397-401.

Lo C.F., Lin Y.R., Chang H.C., Lin J.H. 2006. Identification of turtle shell and its preparations by PCR-DNA sequencing method. *J. Food Drug Anal.* 14: 153-158.

Lorenzen E.D. and Siegismund H.R. 2004. No suggestion of hybridization between the vulnerable black-faced impala (*Aepyceros melampus petersi*) and the common impala (*A. m. melampus*) in Etosha National Park, Namibia. *Mol Ecol*. 13: 3007-3019.

Lorenzen E.D., Arctander P., Siegismund H.R. 2008. High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. *Mol Ecol*. 17: 2812-2824.

Luikart G.L., Sherwin W.B., Steele B.M., Allendorf F.W. 1998a. Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Mol. Ecol.* 7: 963-974.

Luikart G.L., Sherwin W.B., Steele B.M., Allendorf F.W. 1998b. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *J. Hered*. 89: 238-247.

Machugh D.E., Shriver M.D., Loftus R.T., Cunningham P., Bradley D.G. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and Zebu cattle (*Bos Taurus and Bos indicus*). *Genetics*. 146: 1071-1086.

Madisha M.T., Ponsonby D., Schwaibold U., Kotzé A., Jansen R., Brettschneider H., Dalton D.L. 2015. Differentiation of two South African otter species (*Aonyx capensis* and *Lutra maculicollis*) from spraint based on partial CytB primer sets. *Global Ecology and Conservation*, *4*: 8-13.

Marker L.L., Wilkerson A.J.P., Sarno R.J., Martenson J., Breitenmoser-Würsten C., O'Brien S.J., Johnson W.E. 2008. Molecular genetic insights on Cheetah (*Acinonyx jubatus*) ecology and conservation in Namibia. *J. Hered*. 99: 2-13.

McCarthy M.I., Abecasis G.R., Cardon L.R., Goldstein D.B., Little J., Ioannidis J.P., Hirschhorn J.N. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, *9*(5): 356-369.

Mccauley D.E. 1991. Genetic consequences of local population extinction and recolonization. *Trends Ecol. Evol.* 6: 5-8.

McManus J.S., Dalton D.L., Kotzé A., Smuts B., Dickman A., Marshal J.P., Keith M. 2015. Gene flow and population structure of a solitary top carnivore in a human-dominated landscape. *Ecology and evolution*, *5*(2): 335-344.

Miller C.R., Joyce, P., Waits L.P. 2002. Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, 160: 357-366.

Miller S.M., Harper C.K., Bloomer P., Hofmeyr J., Funston P.J. 2014. Evaluation of Microsatellite Markers for Populations Studies and Forensic Identification of African Lions (*Panthera leo*). *Journal of Heredity*, *105*(6): 762-772.

Moodley Y. and Harley E.H. 2005. Population structuring in mountain zebras (*Equus zebra*): The molecular consequences of divergent demographic histories. *Conservation Genetics*. 6: 953-968.

Morin P.A., Luikart G., Wayne R.K. SNP Workshop GRP. 2004. SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19: 208-216.

Morin P.A., Moore J.J., Chakraborty R., Jin L., Goodall J., Woodruff D.S. 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science*. 265: 1193-1201.

Morin P.A., Aitken N.C., Rubio-Cisneros N., Dizon A.E., Mesnick S. 2007. Characterization of 18 SNP markers for sperm whale *(Physeter macrocephalus). Molecular Ecology Notes,* 7: 626-630.

Moritz C. 1994. Defining evolutionarily significant units for conservation. *Trends Ecol. Evol*. 9: 373-375.

Murata M. and Murakami M. 2014. Two Distinct mtDNA Lineages among Captive African Penguins in Japan. *The Journal of Veterinary Medical Science*, *76*(4): 559.

Muwanika V.B., Nyakaana S., Siegismund H.R., Arctander P. 2003. Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*. 91: 361-372.

Muwanika V.B., Nyakaana S., Siegismund H.R., Arctander P. 2007. Population genetic structure of the common warthog (*Phacochoerus africanus*) in Uganda: evidence for a strong philopatry among warthogs and social structure breakdown in a disturbed population. *Afr. J. Ecol*. 45: 22-30.

Nakamura Y., Leppert M., O'Connell P., Wolff R., Holm T., Culver M., Martin C., Fujimoto E., Hoff M., Kumlin E., White R. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science,* 235:1616-1622.

Namroud M.C., Beaulieu J., Juge N., Laroche J., Bousquet J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* 17: 3599-3613.

Narum S.R., Buerkle C.A., Davey J.W., Miller M.R., Hohenlohe P.A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22(11): 2841-2847.

Nielsen L., Meehan-Meola D., Kilbourn A., Alcivar-Warren A. 2008. Characterization of microsatellite loci in the black rhinoceros (*Diceros bicornis*) and white rhinoceros (*Ceratotherium simum*): their use for cross-species amplification and differentiation between the two species. *Conserv. Genet.* 9(1): 239-242.

Nijman I.J., Otsen M., Verkaar E.L.C., De Ruijter C., Hanekamp E., Ochieng J.W., Shamshad S., Rege J.E.O., Hanotte O., Barwegen M.W., Sulawati T., Lenstra J.A. 2003. Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. *Heredity*. 90:10-16.

Nims B.D., Vargas F.H., Merkel J., Parker P.G. 2008. Low genetic diversity and lack of population structure in the endangered Galápagos penguin (*Spheniscus* mendiculus). *Conservation Genetics*, *9*(6): 1413-1420.

Nyakaana S., Arctander P., Siegismund H.R. 2002. Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. *Heredity*. 82: 90-98.

Ogden R., Baird J., Senn H., McEwing R. 2012. The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: a case study from Arabian and scimitar-horned oryx. *Conservation Genetics Resources*, 4(2): 471-473.

Okello J.B.A., Nyakaana S., Masembe C., Siegismund H.R., Arctander P. 2005. Mitochondrial DNA variation of the common hippopotamus: evidence for a recent population expansion. *Heredity.* 95: 206-215.

O'Ryan C., Harley E.H., Bruford M.W., Beaumont M., Wayne R.K., Cherry M.I. 1998. Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Anim. Conserv.* 1: 85-94.

Ostrander E.A., Jong P.M., Rine J., Duyk G. 1992. Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences of the USA,* 89: 3419-3423.

Paetkau D. 1999. Microsatellites obtained using strand extension: An enrichment protocol. *Biotechniques*, 26: 690-697.

Parchman T., Geist K., Grahnen J., Benkman C., Buerkle C.A. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics.* 11: 180.

Pashley C.H., Ellis J.R., Mccauley D.E., Burke J.M. 2006. EST databases as a source for molecular markers: lessons from Helianthus. *J. Hered*. 97: 381-388.

Pastorini J., Zaramody A., Curtis D.J., Nievergelt C.M., Mundy N.I. 2009. Genetic analysis of hybridization and introgression between wild mongoose and brown lemurs. *BMC Evol. Biol.* 9: 32.

Pena F., Poulin E., Dantas G.P., Gonzalez-Acuna D., Petry M.V., Vianna J.A. 2014. Have historical climate changes affected Gentoo Penguin *(Pygoscelis papua)* populations in Antarctica?. *PloS one*, *9*(4).

Pertoldi C., Bijlsma R., Loeschcke V. 2007. Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodivers Conserv* 16: 4147-4163.

Primmer C.R., Ellegren H., Saino N., Møller A.P. 1996. Directional evolution in germline microsatellite mutations. *Nat. Genet.* 13: 391-393.

Primmer C.R. and Ellegren H. 1998. Patterns of molecular evolution in avian microsatellites. *Mol. Biol. Evol*. 15 (8): 997-1008.

Primmer C.R., Painter N.J., Koskinen M.T., Palo J.U., Merilä J. 2005. Factors affecting avian cross-species microsatellite amplification. *J Avian Biol*. 36: 348-360.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155: 945-959.

Rassmann K., Schlötterer C., Tautz D. 1991. Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis,* 12: 113-118.

Reed D.H. and Frankham R. 2003. The correlation between population fitness and genetic diversity. *Conservation Biology,* 17: 230-237.

Rhymer J.M. and Simberloff D. 1996. Extinction by Hybridization and Introgression. *Ann. Rev. Ecol. Syst*. 27: 83-109.

Rico C., Rico I., Hewitt G. 1996. 470 million years of conservation of microsatellite loci among fish species. *Proc. R. Soc. Lon.* B. 263: 549-557.

Robinson T.J., Morris D.J., Fairall N. 1991. Interspecific hybridization in the Bovidae: Sterility of *Alcelaphus buselaphus* x *Damaliscus dorcas* F1 progeny. *Biol. Conserv.* 58: 345-356.

Røed K.H., Ernest E.M., Midthjell L., Msoffe P.L.M. 2011. Identification and characterization of 17 microsatellite loci in the blue wildebeest, *Connochaetes taurinus*. *Conservation Genet. Resour.* 3(1): 181-183.

Roelke-Parker M.E., Munson L., Packer C., Kock R., Cleaveland S., Carpenter M., O'Brien S.J., Pospischil A., Hofmann-Lehmann R., Lutz H., Mwamengele G.L.M., Mgasa M.N., Machange G.A., Summer B.A., Appel M.J.G. 1996. A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature*. 379: 441-445.

Rondebosch C.T. 1999. Isolation and characterization of microsatellite loci in black rhinoceros (*Diceros bicornis*). *Electrophoresis*, *20*: 1778-1780.

Roy S., LaFramboise W.A., Nikiforov Y.E., Nikiforova M.N., Routbort M.J., Pfeifer J., Pantanowitz L. 2016. Next-Generation Sequencing Informatics: Challenges and Strategies for Implementation in a Clinical Environment. Archives of pathology & laboratory medicine.

Ryynänen H.J. and Primmer C.R. 2006. Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primers exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo sala*r) and other salmonid fishes. *BMC Genomics*, 7: 192.

Sakaoka K., Suzuki I., Kasugai N., Fukumoto Y. 2014. Paternity testing using microsatellite DNA markers in captive Adélie penguins (*Pygoscelis adeliae*). *Zoo biology*, *33*(5): 463-470.

Santana Q., Coetzee M., Steenkamp E., Mlonyeni O., Hammond G., Wingfield M., Wingfield B. 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques*. 46(3): 217-23.

Schlosser J.A., Garner T.W.J., Dubach J.M., McElligott A.G. 2003. Characterization of microsatellite loci in Humboldt penguin (*Spheniscus humboldti*) and cross-amplification in other penguin species. *Molecular Ecology Notes*, *3*(1): 62-64.

Schlötterer C., Amos B., Tautz D. 1991. Conservation of polymorphic simple sequence loci in cetacean species. *Nature,* 354: 63-65.

Scott C., Foose T., Morales J.C., Fernando P., Melnick D.J., Boag P.T., Van Coeverden De Groot P.J. 2004. Optimization of novel polymorphic microsatellites in the endangered Sumatran rhinoceros (*Dicerorhinus sumatrensis*). *Molecular Ecology Notes*, *4*(2): 194-196.

Schwab P., Debes P.V., Witt T., Hartl G,B., Hmwe S.S., Zachos F.E., Grobler, J.P. 2012. Genetic structure of the common impala (*Aepyceros melampus melampus*) in South Africa: phylogeography and implications for conservation. *Journal of Zoological Systematics and Evolutionary Research,* 50: 76-84.

Shibuya H., Collins B.K., Huang T.H-M., Johnson G.S. 1994. A polymorphic (AGGAAT)n tandem repeat in an intron of the canine von Willebrand factor gene. *Anim. Genet*. 25: 122.

Simonsen B.T., Siegismund H.R., Arctander P. 1998. Population structure of African buffalo inferred from mtDNA sequences and microsatellite loci: high variation but low differentiation. Mol. Ecol. 7 (2): 225-237.

Sithaldeen R., Ackermann R.R., Bishop J.M. 2015. Pleistocene Aridification Cycles Shaped the Contemporary Genetic Architecture of Southern African Baboons. *PLoS ONE* 10(5): e0123207

Slate J., Gratten J., Beraldi D., Stapley J., Hale M., Pemberton J.M. 2009. Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica*. 136: 97-107.

Smith C.T., Baker J., Park L., Seeb L.W., Elfstrom C., Abe S., Seeb J.E. 2005. Characterization of 13 single nucleotide polymorphism markers for chum salmon. *Mol. Ecol. Notes*. 5: 259-262.

Solano E., Taylor P.J., Rautenbach A., Ropiquet A., Castiglia R. 2014. Cryptic Speciation and Chromosomal Repatterning in the South African Climbing Mice Dendromus (*Rodentia, Nesomyidae*). *PloS one*, *9*(2): e88799.

Spies R., Dalton D.L., Labuschagne C., Kotzé A. 2012. Isolation and characterization of species-specific microsatellite loci in red-billed oxpeckers (*Buphagus erythrorhynchus*). *Conservation Genetics Resources*, *4*(4): 869-871.

Spong G., Stone J., Creel S., Björklund M. 2002. Genetic structure of lions (*Panthera leo L.)* in the Selous Game Reserve: implications for the evolutionary sociality. *J Evolution Biol*. 15: 945-953.

Spooner D.M. 2009. DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am. J. Bot.* 96 (6): 1177-89.

Squirrel J., Hollingsworth P.M., Woodhead M. 2003. How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* 12: 1339-1348.

Stinchcombe J.R. and Hoekstra H.E. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*. 100: 158-170.

Sutherland G.R. and Richards R.I. 1995. Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci. USA* 92: 3636-3641.

Taberlet P. and Luikart G. 1999. Non-invasive genetic sampling and individual identification. *Biol. J. Linn. Soc.* 68: 41-55.

Theron N., Dalton D., Grobler J.P., Jansen R., Kotzé A. 2013. Molecular insights on the re-colonization of the Limpopo Valley, South Africa, by Southern Ground-Hornbills." *Journal of Ornithology* 154(3): 727-737.

Trucchi E., Gratton P., Whittington J.D., Cristofari R., Le Maho Y., Stenseth N.C., Le Bohec C. 2014. King penguin demography since the last glaciation inferred from genome-wide data. In Proc. R. Soc. B 281: 20140528.

Turner T.R., Coetzer W.G., Schmitt C.A., Lorenz J.G., Freimer N.B., Grobler J.P. 2015. Localized population divergence of vervet monkeys (*Chlorocebus* spp.) in South Africa: Evidence from mtDNA. *American journal of physical anthropology*.

Väli U., Einarsson A., Waits L., Ellegren H. 2008. To what extent do microsatellites markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology,* 17: 3808-3817.

Van Hooft W.F., Groen A.F., Prins H.H.T. 2002. Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Mol. Ecol*., 11: 267-279.

Van Hooft W.F., Hanotte O., Wenink P.W., Groen A.F., Sugimoto Y., Prins H.H.T., Teale A. 1999. Applicability of bovine microsatellite markers for population genetic studies on African buffalo (*Syncerus caffer*). *Anim Genet*. 30: 214-220.

van Wyk A.M., Kotzé A., Randi E., Dalton D.L. 2013. A hybrid dilemma: a molecular investigation of South African bontebok (*Damaliscus pygargus pygargus*) and blesbok (*Damaliscus pygargus phillipsi*). *Conservation Genetics*, *14*(3): 589-599.

Vignal A., Milan D., SanCristobal M., Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics.*Genetics Selection Evolution*, *34*(3): 275-306.

Wagner A.P., Creel S., Kalinowski S.T. 2006. Estimating relatedness and relationships using microsatellite loci with null alleles. *Heredity*. 97(5): 336-45.

Wan Q.H., Wu H., Fujihara T., Fang S.G. 2004. Which genetic marker for which conservation genetics issue? *Electrophoresis*. 25: 2165-2176.

Ward R.D., Zemlak T.S., Innes D.H., Last P.R., Hebert P.R. 2005. DNA barcoding Australia's fish species. *Phil. Trans. R. Soc. B.* 360: 1847-1857.

Wasser S.K., Shedlock A.M., Comstock K., Ostrander E.A., Mutayoba B., Stephens M. 2004. Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade. *Proc Natl Acad Sci USA*. 10: 14847-14852.

Wenink P.W., Groen A.F., Roelke-Parker M.E., Prins H.H.T. 1998. African buffalo maintain high genetic diversity in the major histocompatibility complex in spite of historically known population bottlenecks. 1998. *Mol Ecol*. 7: 1315-1322.

Werner F.A.O., Durstewitz G., Habermann F.A., Thaller G., Krämer W., Kollers S., Buitkamp J., Georges M., Brem G., Mosner J., Fries R. 2004. Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. *Anim Genet*. 35: 44-49.

Whitehouse A.M. and Harley E.H. 2001. Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. *Mol Ecol*. 9: 2139-2149.

Wilhelm K., Dawson D.A., Gentle L.K., Horsfield G.F., Schlötterer C., Greig C., East M., Hofer H., Tautz D., Burke T. 2003. Characterization of spotted hyena, *Crocuta crocuta* microsatellite loci. *Mol Ecol Notes*. 3: 360-362.

Willerslev E., Gilbert M.T.P., Binladen J., Ho S.Y., Campos P.F., Ratan A., Schuster S.C. 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution.*BMC evolutionary biology*, *9*(1): 95.

Wirgin I., Kovach A.I., Maceda L., Roy N.K., Waldman J., Berlinsky D.L. 2007. Stock identification of Atlantic cod in U. S. waters using mircosatellite and single nucleotide polymorphism DNA analyses*. T. Am. Fish. Soc*. 136: 375-391.

Wiseman R., O'Ryan C., Harley E.H. 2000. Microsatellite analysis reveals that domestic cat (*Felis catus*) and southern African wild cat (*F. lybica*) are genetically distinct. *Anim. Conserv.* 3: 221-228.

Wong K.L., Wang J., But P.P.H., Shaw P.C. 2004. Application of cytochrome b DNA sequences for the authentication of endangered snake species. *Forensic Sci. Int.* 139: 49-55.

Xu X. and Arnason U. 1997. The complete mitochondrial DNA sequence of the white rhinoceros, *Ceratotherium simum*, and comparison with the mtDNA sequence of the Indian rhinoceros, *Rhinoceros unicornis*. *Molecular phylogenetics and evolution*, *7*(2): 189-194.

Xu X., Janke A., Arnason U. 1996. The complete mitochondrial DNA sequence of the greater Indian rhinoceros, *Rhinoceros unicornis*, and the Phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+ Cetacea). *Molecular biology and evolution*, *13*(9): 1167-1173.

Zane L., Baegelloni L., Patarnello T. 2002. Strategies for microsatellite isolation: a review. *Mol Ecol*. 11: 1-16.

Zschokke S., Armbruster G.F., Ursenbacher S., Baur B. 2011. Genetic differences between the two remaining wild populations of the endangered Indian rhinoceros (*Rhinoceros unicornis*). *Biological conservation*, *144*(11): 2702-2709.

Zschokke S., Gautschi B., Baur B. 2003. Polymorphic microsatellite loci in the endangered Indian rhinoceros, *Rhinoceros unicornis*. *Molecular Ecology Notes*, *3*(2): 233-235.

PART II: DEVELOPMENT AND APPLICATION OF CONSERVATION GENETICS TECHNIQUES IN AFRICAN PENGUIN

CHAPTER 3

The Complete Sequence of the Mitochondrial Genome of the African Penguin (*Spheniscus demersus*)

Abstract

The complete mitochondrial genome of the African Penguin (*Spheniscus demersus*) was sequenced. The molecule was sequenced via next generation sequencing and primer walking. The size of the genome is 17,346 bp in length. Comparison with the mitochondrial DNA of two other penguin genomes that have so far been reported was conducted namely; Little Blue Penguin (*Eudyptula minor*) and the Rockhopper penguin (*Eudyptes chrysocome*). This analysis made it possible to identify common penguin mitochondrial DNA characteristics. The *Spheniscus demersus* mtDNA genome is very similar, both in composition and length to both the *E. chrysocome* and *E. minor* genomes. The gene content of the African Penguin mitochondrial genome is typical of vertebrates and all three penguin species have the standard gene order originally identified in the chicken. The control region for *Spheniscus demersus* is located between tRNA-Glu and tRNA-Phe and all three species of penguins contain two sets of similar repeats with varying copy numbers towards the 3' end of the control region, accounting for the size variance. This is the first report of the complete nucleotide sequence for the mitochondrial genome of the African Penguin, *Spheniscus demersus.* These results can be subsequently used to provide information for penguin phylogenetic studies and insights into the evolution of genomes.

Introduction

Mitochondrial DNA (mtDNA) is generally a 15-23 kb double-strand circular genome in animals and plays an important role in the process of metabolism and programmed cell death (Cao *et al.,* 2006). This genome generally contains 13 protein-coding genes, two ribosomal RNAs, 22 transfer RNAs and a non-coding control region (D-loop) of variable length that contains signals required for replication and transcription (Wolstenholme, 1992; Ruokonen and Kvist, 2002). Being separate from the nucleus, the mitochondrial genome has several characteristics that make it unique; including maternal inheritance, its small size, fast evolutionary rate, limited recombination and relatively conserved gene content and organization (Brown, 1983; Wu *et al.,* 2003; Cao *et al.,* 2006). Due to these traits mtDNA have been used extensively for testing hypotheses of microevolution, studying population structure, phylogeography and phylogenetic relationships at various taxonomic levels (Cao *et al.,* 2006; Zhou *et al.,* 2009).

Since useful information can be identified from many of the mitochondrial genes and due to primers being functional for a wide range of taxa, the number of complete mitochondrial genomes is steadily increasing (Sammler *et al.,* 2011). Complete mitochondrial genomes provide sets of genome-level characteristics, which are useful for modelling genome evolution and phylogenetic inference (Gibb *et al.,* 2007; Lei *et al.,* 2010). These characteristics include base composition,

genetic codon variation, gene content and gene arrangement, tRNA and rRNA gene secondary structures and modes of replication and transcription (Lei *et al.,* 2010). To date, complete mitochondrial genomes have been reported for only two penguin species, the Little Blue Penguin (*Eudyptula minor*) and the Rockhopper penguin (*Eudyptes chrysocome*) (Slack *et al.,* 2003; Watanabe *et al.,* 2006). This study reports the complete mitochondrial genome of the African Penguin (*Spheniscus demersus)* along with a comparative analysis of the complete mtDNA genome with the two other penguin species.

Methods and Materials:

DNA sample

A DNA sample prepared for a previous study on microsatellite development in *Spheniscus demersus* (Labuschagne *et al.,* 2013) was used. The blood sample was from a captive breeding adult African Penguin in a colony located in the KwaZulu-Natal Province of South Africa. Total genomic DNA was isolated using the Qiagen DNeasy® Blood and Tissue Kit.

Primer design, PCR amplification and DNA sequencing

Initial primers were designed based on 60 reads identified as mitochondrial from a next generation sequencing dataset (7,706 reads) generated in a previous study (Labuschagne *et al.,* 2013) on the GS FLX (Roche). The majority of the reads mapped to NADH dehydrogenase subunit 2 (38 reads) followed by NADH dehydrogenase subunit 6 (seven reads) and NADH dehydrogenase subunit 4 (six reads). Gaps were then filled using the primer walking method. Assemblies, mapping and primer design were performed in CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark). Primer design parameters were set to a minimum melting temperature (Tm) of 53 °C, maximum Tm of 62 °C, primer length 16-21 bp and remaining settings on default. Assembly setting were set to auto-trim, minimum aligned read length of 30 bp, alignment stringency medium, ambiguity nucleotides and all other settings as default. All PCRs were performed utilising an ABI 9700 thermal cycler (Applied Biosystems, Foster City, CA). Amplification reactions were done in a final volume of 25 µl containing 30 ng DNA, 25 pM of each primer and 2X DreamTaq® Green Master Mix (Thermo Fisher Scientific, Vilnius, Lithuania). Thermal cycling consisted of initial denaturation at 95 °C for 5 min, 45 cycles of denaturation at 95 °C for 30 s, annealing at 55-59 °C for 30 s, extension at 72 °C for 5 min, followed by final extension at 72 °C for 10 min.

Resulting amplicons were inspected on 1 % agarose gels followed by purification utilizing the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Orange, CA, USA). Purified templates were sequenced utilising a Big Dye V3.1 Terminator Kit (Applied Biosystems, Foster City, CA) and the ABI 3500XL genetic analyser (Applied Biosystems, Foster City, CA) following manufacturer's instructions.

Table 2: Primers used in amplification and sequencing of the *S. demersus* mitochondrial genome.

No. of primer pair	Name	Sequence (5'-3')	Binding Site
1	PenMIT-1	GACCCACACCATATACCACAACAGG	15236-15260
	PenMIT-2	GGCACCGCCAAGTCCTTAGAG	604-624
	PenMIT-INT-1	CACGAGATAAGTCATAGC	15501-15518
	PenMIT-INT-2	CATTCTTTCCCCCTACAC	16110-16127
2	PenMIT-3	GCATGGCACTGAAGATGCCAAG	23-44
	PenMIT-4	GGTGACGGGCGGTATGTACG	926-945
3	PenMIT-5	GAGAACTACGAGCACAAACGC	577-597
	PenMIT-6	GGCTTTTCACCTCTACTAACAAGTC	1452-1476
4	PenMIT-7	GGTAAGTCGTAACAAGGTAAG	1003-1023
	PenMIT-8	GCCATTCATACAAGTCTCG	2092-2110
5	PenMIT-9	GAGTAATTTAAAGGAGGTACAG	1581-1602
	PenMIT-10	CTCTGCCACGCTAGCGGT	2724-2741
6	PenMIT-11	CGAGACTTGTATGAATGGC	2092-2110
	PenMIT-12	GGCTAGGATTATTGGGAATAAA	3585-3606
7	PenMIT-13	CCATATCAAGCCTAGCAGTTTAC	3146-3168
	PenMIT-14	GTGATGGTTGTTCCTAGAAGG	4040-4060
8	PenMIT-15	ATTATTCCCAATAATCCTAGCC	3585-3606
	PenMIT-16	CTTTGAAGGCCTTCGGTTTG	5058-5077
9	PenMIT-17	CTCATCTATCTCCCATCTAGGC	4544-4565
	PenMIT-INT-3	GATAGTTTTTCTATGAGTATGAG	7236-7258
	PenMIT-INT-4	CTCTTAGCACACATCAATGAGC	5282-5303
	PenMIT-INT-5	GGGCTCATAGTATTGGAGG	6420-6438
10	PenMIT-19	GTACAAGAAAGGAAGGAATCG	6951-6971
	PenMIT-20	GATGAGTATGAACGTGATTATG	9559-9580
	PenMIT-INT-6	CCTCAAAGCCATCGGACACCA	7391-7411
	PenMIT-INT-7	GGTTTGATTCCTGTTGG	9052-9068
11	PenMIT-21	CAGAACTAGGTGGACAATGACC	9029-9050
	PenMIT-22	CCTGCACCTGCTTCACAG	10183-10200
12	PenMIT-23	CATAATCACGTTCATACTCATC	9559-9580
	PenMIT-24	GGTTAGGATGATTGTTAGGG	11438-11457
	PenMIT-INT-8	CCTTCGCCCTCATACCAGTAC	10143-10163
	PenMIT-INT-9	CCTATGTGGCTTACGGAGGAG	11133-11153
13	PenMIT-25	GTACACTACACCTAACAATACTAG	10784-10807
	PenMIT-26	GTACTAGGCTAATTAAGAAGGCAG	11993-12016
14	PenMIT-27	CCCTAACAATCATCCTAACC	11438-11457
	PenMIT-28	GTCCTAGTTGGCTGGATGTG	12780-12799
	PenMIT-INT-10	GCCTTCTTAATTAGCCTA	11995-12012
15	PenMIT-29	GGAGACATCGGCCTCATCCTAAG	12400-12422
	PenMIT-30	GGGTGGAATGGGATTTTGTCG	14334-14354
	PenMIT-INT-11	TACTTCTACTAACAATC	8607-8623
16	PenMIT-31	CGCAAACGGAGCCTCATTCTTC	13938-13959
	PenMIT-32	GCAGGAACTGCTATGACTTATCTCG	15503-15527
	PenMIT-INT-12	CCAACCTCATTATCCTAAC	14678-14696

Sequence assembly and sequence analysis

Sequences were checked, assembled and annotated in CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark). The boundaries of the protein-coding genes and rRNA genes were inferred by comparisons with the amino acid sequence of proteins and the nucleotide sequence of other birds including Flamingo (*Phoenicopterus roseus*; EF532932), Pacific Loon (*Gavia pacifica*; AP009190), White Stork (*Ciconia ciconia*; AB026818), Red-throated Loon (*Gavia stellate*; AY293618), Little Blue Penguin (*Eudyptula minor;* AF362763) and Rockhopper Penguin (*Eudyptes chrysocome*; NC_008138). The tRNA genes were identified by their cloverleaf secondary structure

using tRNA-scan SE 1.21 (Lowe and Eddy, 1997) as well as ARWEN (online version) (Laslett and Canbäck, 2008) and verified by comparison with homologous sequences of other birds (mentioned above). Comparisons were made by forming assemblies between homologous sequences in CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark) using min aligned read length of 20 bp, alignment stringency low, ambiguity nucleotides and all other settings as default. The complete mtDNA sequence of *Spheniscus demersus* reported in this article was deposited in GenBank under accession number KC914350. CLC Bio Genomics work bench 5.0 (CLC Bio, Aarhus, Denmark) was used to draw a maximum likelihood phylogeny between AB026818, AF362763, NC_008138 and KC914350 utilizing Neighbor Joining as starting tree algorithm, General Time Reversible as substitution model and bootstrapping of 1000 replicates.

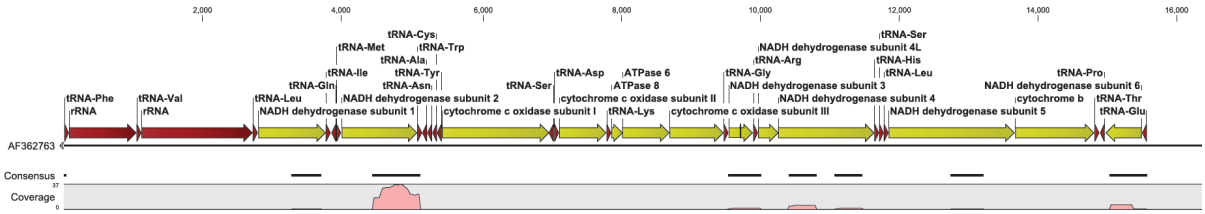


Figure 6: Illustration of GS reads identified as mtDNA mapped to a previously sequenced Little Blue Penguin mtDNA genome (AF362763) to estimate gap sizes to be covered.

Results and Discussion:

Mitochondrial genome organization

The complete mitochondrial genome of *Spheniscus demersus* as determined in this study, is 17,346 bp in length (Figure 7), which is comparable to *Eudyptes chrysocome* (16,930 bp) and *Eudyptula minor* (17,611 bp).

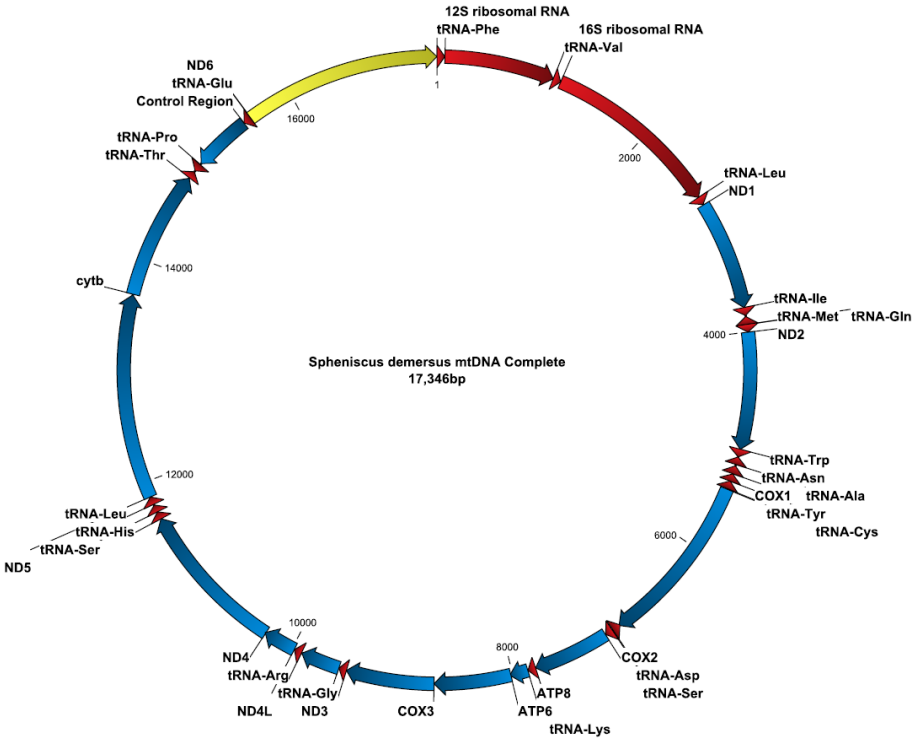


Figure 7: Genetic map of the *S. demersus* mitochondrial genome. Annotation arrows indicate orientation of genes; COX1-3 indicates cytochrome oxidase subunits 1-3; ATP6/8, ATPase subunits 6-8; ND1-6/4L, NADH dehydrogenase 1-6/4L and CytB, cytochrome b. For designation of transfer RNAs, tRNAs and the three-letter code for amino acids is used.

This length is not absolute, however, due to heteroplasmy caused by differences in the number of repeated motifs, ACAACAAACAACAA, at the 3' end of the control region (CR). Heteroplasmy has also been reported in *E. minor* and *E. chrysocome* (Slack *et al.,* 2003; Watanabe *et al.,* 2006). *Spheniscus demersus* mtDNA genome shows 88.94 % (91.77 % excluding CR) similarity to *E. chrysocome* and 89.66 % (91.75 % excluding CR) similarity with *E. minor*, while *E. minor* and *E. chrysocome* have 87.26 % (91.25 % excluding CR) similarity. *Spheniscus demersus* and *E. minor* share a more recent common ancestor and group together but are both partitioned on a separate branch from *E. chrysocome*, as illustrated in Figure 8.

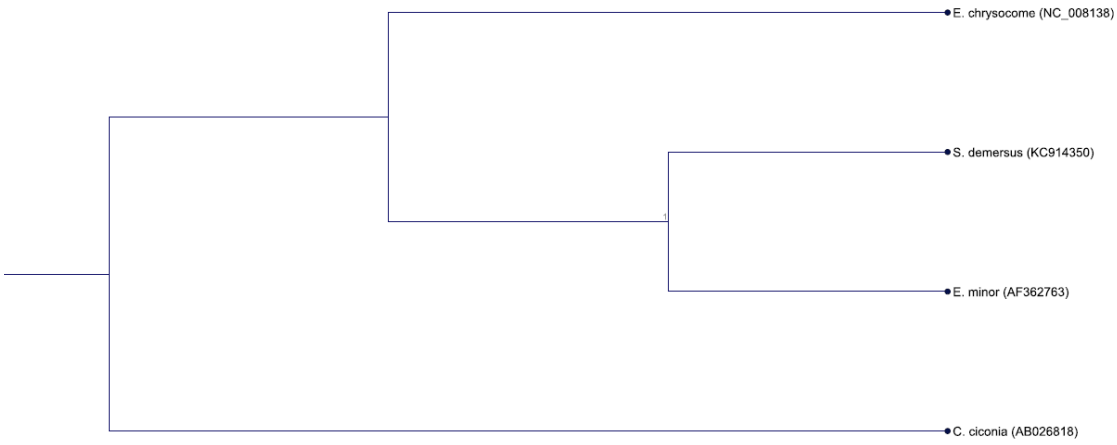


Figure 8: Maximum Likelihood Phylogeny based on the mitochondrial genomes, excluding the control region showing the relationships among three penguin taxa and the white stork outgroup.

This observation is in agreement with our current understanding of their relationships. Baker *et al.,* (2006) indicated that based on 2802 bp of nuclear and 2889 bp of mtDNA; *Spheniscus demersus*, *E. minor* and *E. chrysocome* diverged from the older Antarctic genera approximately 34-25 mya. The authors further indicated that *Spheniscus demersus*, *E. minor* grouped together, but were both partitioned from *E. chrysocome.* The gene content of the African Penguin mt genome is typical of vertebrates, consisting of 13 protein coding genes (PCGs), 22 tRNAs and two rRNAs. As seen in other birds, NADH dehydrogenase subunit 6 and 8 tRNAs are transcribed from the light strand, while the other 12 protein coding genes, 14 tRNAs and two rRNAs are located on the heavy strand. Although variation in gene order has been described among avian mt genomes, all three penguin species have the standard gene order originally identified in the chicken (Desjardins and Morais, 1990). The nucleotide composition of the *S. demersus* mt genome (H strand) (A=30.77 %; C=32.56 %; G=13.58 %; T=23.08 %) is similar to that of *E. chrysocome* (A=30.67 %; C=32.88 %; G=13.85 %; T=22.60 %), *E. minor* (A=30.96 %; C=31.84 %; G=13.53 %; T=23.67 %) and other avian species. The A+T content of 53.85 % is within range for avian mt genomes (51.6 %-55.7 %) and very similar to the other two penguin genomes (*E. chrysocome*=53.27 %; *E. minor*=54.63 %). One extra cytosine is present in NAD3 in all three penguin species. The extra nucleotide has been described in several other bird species as well as some turtles and is thought not to be translated (Mindell *et al.,* 1998). Russell and Beckenbach (2008) suggested that certain mitochondrial translation systems have the ability to tolerate frameshift insertions using programmed translational frameshifting, but the function of the extra nucleotide in NAD3 and its phylogenetic implications are still unclear (Kan *et al.,* 2010).

Codon usage and sequence features of protein-coding genes

The usage of initial and termination signals as well gene length in comparison with two other penguin species is given in Table 3.

Table 3: Length and start/stop codons of mitochondrial protein-encoding genes of three penguin species.

Gene	Species	***S. demersus*** (current study)	***E. chrysocome*** (Watanabe *et al.,* 2006)	***E. minor*** (Slack *et al.,* 2003)
ND1	Length (bases/amino acid)	978/325	978/325	978/325
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	AGG	AGG	AGG
ND2	Length (bases/amino acid)	1041/346	1041/346	1041/346
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAG	TAG	TAG
COX1	Length (bases/amino acid)	1551/516	1551/516	1551/516
	Start Codon	GTG(Val)	GTG(Val)	GTG(Val)
	Stop Codon	AGG	AGG	AGG
COX2	Length (bases/amino acid)	684/227	684/227	684/227
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAA	TAA	TAA
ATP8	Length (bases/amino acid)	165/54	165/54	165/54
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAA	TAA	TAA
ATP6	Length (bases/amino acid)	684/227	684/227	684/227
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAA	TAA	TAA
COX3	Length (bases/amino acid)	784/261	784/261	784/261
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	T--	T--	T--
ND3	Length (bases/amino acid)	352/116	352/116	352/116
	Start Codon	ATC(Ile)	ATT(Ile)	ATC(Ile)
	Stop Codon	TAA	TAA	TAA
ND4L	Length (bases/amino acid)	297/98	297/98	297/98
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAA	TAA	TAA
ND4	Length (bases/amino acid)	1380/459	1380/459	1380/459
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAG	TAG	TAG
ND5	Length (bases/amino acid)	1818/605	1818/605	1821/606
	Start Codon	ATG(Met)	ATG(Met)	GTG(Val)
	Stop Codon	TAA	TAA	TAA
Cytb	Length (bases/amino acid)	1143/380	1146/381	1143/380
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAA	TAA	TAA
NAD6 (L)	Length (bases/amino acid)	519/172	519/172	519/172
	Start Codon	ATG(Met)	ATG(Met)	ATG(Met)
	Stop Codon	TAG	TAG	TAG

The most common start codon is ATG. In COX1, all three penguin species as with most other birds (Slack *et al.,* 2003) use the nonstandard start codon GTG. *Eudyptula minor* uses the same start codon for ND5, whereas the other two species use the standard ATG. The use of GTG in ND5 has also been described in ducks (Readhead duck, *Aythya americana*) and goose (Greater White-fronted Goose, *Anser albifrons*) (Slack *et al.,* 2003). Furthermore, ATC (*S. demersus* and *E. minor)* and ATT (*E. chrysocome*) are used as start codons in ND3. This unusual start codon (isoleucine) has thus far only been found in ND3 in passerines (Watanabe *et al.,* 2006). Stop codon usage is consistent for all three penguins across all 13 PCGs. As in the mtDNA genome of other birds, TAA is the most frequent stop codon. TAG is used for ND2, ND4 and ND6, while AGG was used for ND1 and COX1. Among neognath birds, ND4 is usually terminated through TAA or incomplete stop codons, TA- and T-- (Slack *et al.,* 2003). All three penguin species use the incomplete stop codon T-- in COX3 as described in other birds. The terminal T serves as the stop signal after it is completed to UAA by post-transcriptional polyadenylation (Ojala *et al.,* 1981). Identical gene length was observed among the three penguin species for 11 PCGs. *Eudyptula minor* contains one extra amino acid (aa) in ND5 (606 aa), while *E. chrysocome* contains one extra aa in Cytb (381 aa) when compared to the other two species. Varying sizes for ND5 have been reported previously with sizes ranging from 603 aa for the tinamou to 607 aa for a duck (Slack *et al.,* 2003). Cytb size reports for birds are mostly 379 aa and 380 aa (Slack *et al.,* 2003; Watanabe *et al.,* 2006; Kan *et al.,* 2010). The longest mtDNA PCG for all three species is NAD5, while the shortest is ATP8, as described in other birds. All three penguin species have one less aa than most other birds in both ATP8 and NAD6 (Slack *et al.,* 2003).

Spacers and overlaps

A total of 19 intergenic spacers ranging from 1 bp to 1758 bp, are found in the mtDNA genome of *S. demersus* (Table 4). Among these, the longest non-coding region (1758 bp) is found between tRNA-Glu and tRNA-Phe and this will be discussed further under the control region section. Nineteen intergenic spacers are also found in *E. minor, while E. chrysocome* had only 18. Excluding the CR, the intergenic spacers amount to 80 bp in *S. demersus*, 64 bp in *E. chrysocome* and 60 bp in *E. minor.* The *S. demersus* mtDNA genome seems less compact when compared to the other two penguin species. In general, although length may vary, spacer and overlap positions are mostly conserved across the three penguin species. However, *S. demersus* contains a 8 bp spacer instead of an overlap observed in the other two species between tRNA-Ser(AGY) and tRNA-Leu(CUN). Furthermore, *E. minor* contains an 8 bp spacer between tRNA–Met and NAD2 while the other two species have no spacer. The overlaps can be divided into at least four classes. The first class are those overlaps between H and L strand-encoded elements:1 bp between tRNA-Gln(L)/tRNA–Met and 9 bp between COX1/tRNA-Ser(UCN)(L). Since different RNA transcripts are involved, these do not comprise genuine overlaps. The second class involve those overlaps on TAR stop codons: 2 bp overlap between a TAG stop codon in NAD2 and the start of tRNA-Trp; 1 bp overlap between ATP6 TAA stop codon and COX3. It may be that these are not true overlaps, but rather represent endonucleolytic cleavage sites producing incomplete stop codons (Ojala *et al.,* 1981). The third class consists of overlaps between the coding sequences of PCGs: 10 bp between

ATP8/ATP6 and 7 bp between NAD4L/NAD4 (in all birds). These overlaps are always associated with different reading frames, but more information is required regarding the generation and processing of mt protein-coding transcripts (Slack *et al.,* 2003). Finally, the fourth class is made up of the remaining overlaps and involve unknown mechanisms: a 2 bp overlap between an AGG stop codon in NAD1 and the start of tRNA-Ile; a 1 bp overlap between tRNA-Cys(L) andtRNA-Tyr(L); a 1 bp overlap between tRNA-Ser(AGY) and tRNA-Leu(CUN) (not present in *S. demersus*).

Table 4: Length indicated in base pairs (bp) of penguin control regions, intergenic spacers and overlaps

Region	***S. demersus*** (current study)	***E. chrysocome*** (Watanabe *et al.,* 2006)	***E. minor*** (Slack *et al.,* 2003)
Control region	1758	1376	2040
tRNA-Phe/12 S rRNA	-	-	-
12 S Rrna/tRNA-Val	-	-	-
tRNA-Val/16S rRNA	-	-	-
16S Rrna/tRNA-Leu(UUR)	-	-	-
tRNA-Leu(UUR)/NAD1	5	4	5
NAD1/tRNA-Ile	2 overlap	2 overlap	2 overlap
tRNA-Ile/tRNA-Gln(L)	9	9	9
tRNA-Gln(L)/tRNA-Met	1 overlap	1 overlap	1 overlap
tRNA–Met/NAD2	-	-	8
NAD2/tRNA-Trp	2 overlap	2 overlap	2 overlap
tRNA-Trp/tRNA-Ala(L)	1	1	1
tRNA-Ala(L)/tRNA-Asn(L)	13	2	2
tRNA-Asn(L)/tRNA-Cys(L)	2	2	2
tRNA-Cys(L)/tRNA-Tyr(L)	1 overlap	1 overlap	1 overlap
tRNA-Tyr(L)/COX1	6	1	1
COX1/tRNA-Ser(UCN)(L)	9 overlap	9 overlap	9 overlap
tRNA-Ser(UCN)(L)/tRNA-Asp	6	5	4
tRNA-Asp/COX2	2	2	2
COX2/tRNA-Lys	1	1	1
tRNA-Lys/ATP8	1	1	1
ATP8/ATP6	10 overlap	10 overlap	10 overlap
ATP6/COX3	1 overlap	1 overlap	1 overlap
COX3/tRNA-Gly	-	-	-
tRNA-Gly/NAD3	-	-	-
NAD3/tRNA-Arg	4	4	4
tRNA-Arg/NAD4L	1	1	1
NAD4L/NAD4	7 overlap	7 overlap	7 overlap
NAD4/tRNA-His	1 overlap	1 overlap	1 overlap
tRNA-His/tRNA-Ser(AGY)	-	-	-
tRNA-Ser(AGY)/tRNA-Leu(CUN)	8	1 overlap	1 overlap
tRNA-Leu(CUN)/NAD5	-	-	-
NAD5/Cytb	6	7	7
Cytb/tRNA-Thr	4	4	3
tRNA-Thr/tRNA-Pro(L)	11	13	9
tRNA-Pro(L)/NAD6(L)	13	12	13
NAD6(L)/tRNA-Glu(L)	2	2	3
tRNA-Glu(L)/CR	-	-	-
CR/ tRNA-Phe	-	-	-

Transfer RNA and rRNA genes

A total of 22 tRNAs are found interspersed in the mtDNA genome of *S. demersus* and range in size from 66 bp (tRNA-Ser(AGY) to 76 bp (tRNA-Trp and tRNA-Ser(UCN). The tRNAs include two tRNA-Leu and two tRNA-Ser. These tRNAs correspond to the standard set found in other metazoan mtDNAs. Most of the tRNAs could be folded into the canonical cloverleaf secondary structure with examples in Figure 9A.

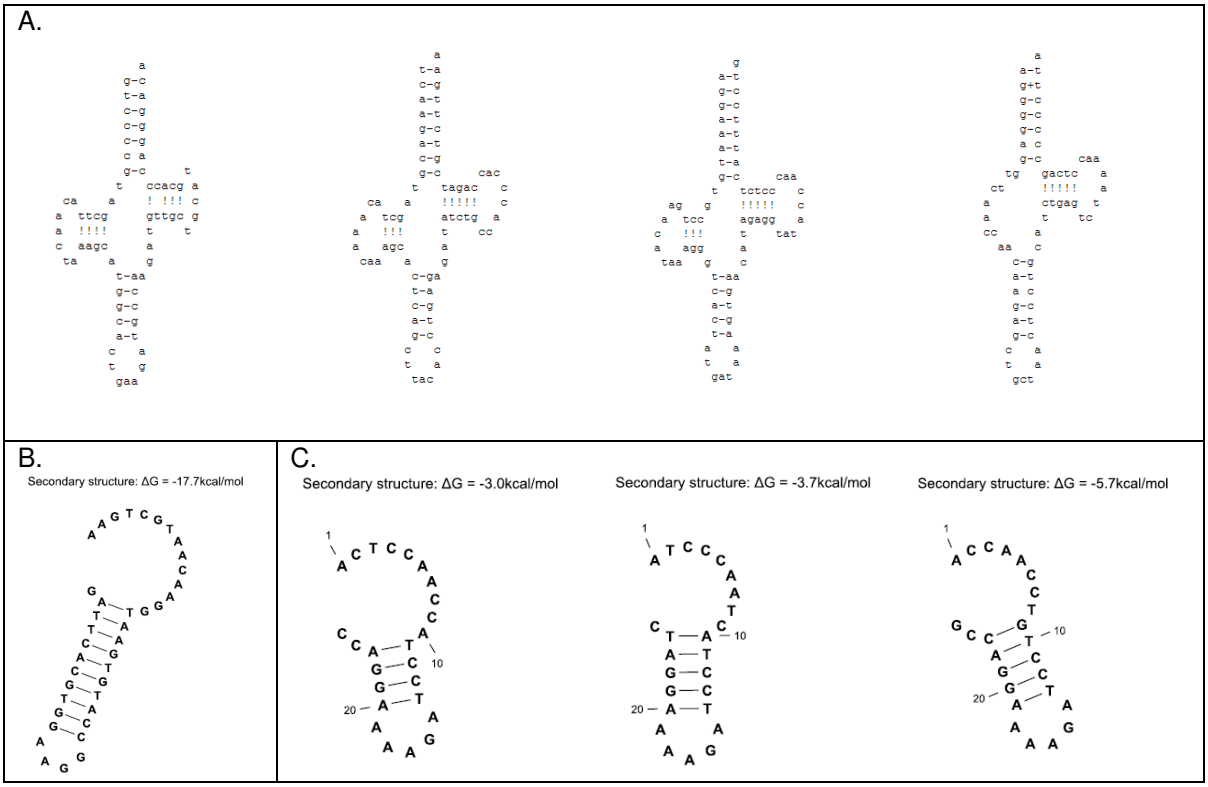


Figure 9: Inferred secondary structures (A) Secondary structures of 4 tRNAs found in *S. demersus* namely; tRNA-Phe, tRNA-Val, tRNA-Ile, tRNA-Ser(AGY). (B) Secondary structure formed at the 3' end of 12 S rRNA found in *S. demersus*. (C) Secondary structures formed at the 3' end of 16 S rRNA in *S. demersus*; *E. minor* and *E. chrysocome* respectively

As in vertebrates in general, the secondary structure of tRNA-Ser(AGY) lacks the DHU arm. Located between tRNA-Phe and tRNA-Val, the 12 S rRNA gene of *S. demersus* was 980 bp (Figure 9B), 4 and 5 bp longer than those described in *E. chrysocome* and *E. minor* respectively. The 16 S rRNA gene (Figure 9C), located between tRNA-Val and tRNA-Leu, was 1606 bp and was 2 bp shorter than *E. minor*, but 11 bp longer than *E. chrysocome*. Asakawa *et al.,* (1995) suggested that a stem and loop structure around the 3' end of 12 S rRNA and 16 S rRNA could play an essential role in the protein synthesis and transcriptional regulation in mitochondria, respectively. A conserved 39 bp at the 3' end of the 12 S rRNA gene in *S. demersus* was also inferred to have a stable stem and loop structure (Figure 9B) showing free energy of -17.7 kcal/mol. Although the 3' end of 16 S rRNA of the penguins showed some sequence variation, they could still be folded into stem and loop structures (Figure 9C). In eutherians, the L-strand origin of replication is usually

located between tRNA-Asn and tRNA-Cys, but is missing in *S. demersus* with the two tRNAs separated by only 2 bp. The absence of an origin of replication at this position is consistent with other birds described (Mindell *et al.,* 1998b). Desjardins and Morais (1989) proposed that it is possible for origin of L-strand replication to be initiated within the CR.

Control region

The mtDNA Control Region (CR) is responsible for transcription and replication of the mitochondrial genome (Taanman, 1999). As in the majority of birds for which data is available the CR for *S. demersus* is located between tRNA-Glu and tRNA-Phe. The CR of *S. demersus* is 1758 bp in length, which is longer than *E. chrysocome* (1376 bp), but shorter than what has been reported for *E. minor* (2040 bp). All three species contain two sets of similar repeats (TCGATACAYWTTACAYTTYWWYTTTCTCTAAAATTTCATTAABRYAYRATARCAACYCTTYGTT GCYATCDYCTTTACTGTA and ACAACAAACAACAA) with varying copy numbers towards the 3' end of the CR, accounting for the size variance. Conserved sequence blocks (CSB-1,-2, and-3) have been identified in the CR of several vertebrates and may be involved in the origin of H-strand replication (Walberg and Clayton, 1981). Only CSB-1 (TATTTGTTGAATGCTTGTTAGACATAA) could be identified in *S. demersus*. A cytosine string (CCCCCCCCTACCCCCC) located close to the 5' end of the CR is similar to the motif observed in other avian species such as Struthioniformes, Galliformes and Falconiformes. The motif consists' of a G/C stem and a loop containing a TCCC motif that may be involved in H–strand termination (Ruokonen and Kvist, 2002). This motif in the CR has also been reported in African side-necked turtle (*Pelomedusa subsrufa*) (Zardoya and Meyer, 1998). In chickens and lesser snow geese (*Ansercaerulescens caerulescens*) the motifs have the potential to form a stable hairpin structure (Quinn and Wilson, 1992). However, in the three penguins discussed here, the C-stretch is not followed by a G-stretch. Thus the repeat sequence is unable to form a hairpin secondary structure. Reasons behind conservation of the C-stretch is still unknown and the role of this sequence is currently unknown (Ruokonen and Kvist, 2002). The termination-associated sequence motif TATAT was identified 33 bp downstream from the C-stretch in *S. demersus*, but was not present in the other two penguins. The termination-associated sequence motif TACAT, immediately preceding the TATA motif in *S. demersus* is present in all three species. The highly conserved bird similarity box (CACTGATGCACTTTG) was identified approximately 821 bp downstream from the C-stretch in all three penguins. The high level of sequence conservation suggests that the bird similarity box may play a key role in the replication and transcription of the mitochondrial genome in Aves (Bing *et al.,* 2006).

In summary, this is the first report of the complete nucleotide sequence for the mitochondrial genome of the African Penguin, *Spheniscus demersus.* The *Spheniscus demersus* mtDNA genome is very similar, both in composition and length to both the *E. chrysocome* and *E. minor* genomes. These results can be subsequently used to provide information for penguin phylogenetic studies and insights into the evolution of genomes.

References

Asakawa S., Himeno H., Miuraand K., Watanabe K. 1995. Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. *Genetics,* 140: 1047-1060.

Baker A.J., Pereira S.J., Haddrath O.P., Edge K. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. *Proc. R. Soc*. B 273: 11-17.

Brown W.M. 1983. Evolution of animal mitochondrial DNA. pp. 62-88. In: Evolution of Genes and Proteins, Edited by Nei, M. and Koehn, R. K. Sinauer, Sunderland, Massachusetts.

Bing X., Fei M.A., Yi S., Qing-Wei L. 2006. Comparative Analysis of Complete Mitochondrial DNA Control Region of Four Species of Strigiformes. *Acta Genetica Sinica* 33 (11): 965-974.

Cao S., Wu B., Yan P., Hu Y., Su X., Jiang Z. 2006. Complete nucleotide sequences and gene organization of mitochondrial genome of *Bufo gargarizans*. *Mitochondrion,* 6: 186-193.

Desjardins P. and Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome. *J. Mol. Biol*. 212: 599-634.

Gibb G.C., Kardailsky O., Kimball R.T., Braun E.L., Penny D. 2007. Mitochondrial genomes and avian phylogeny: Complex characters and resolvability without explosive radiations. *Mol. Biol. Evol.* 24(1): 269-280.

Kan X.Z., Li X.F., Lei Z.P., Wang M., Chen L., Gao H., Yang Z.Y. 2010. Phylogeny of major lineages of galliform birds (*Aves: Galliformes*) based on complete mitochondrial genomes. *Genetics and Molecular Research,* 9 (2): 1204-1216.

Labuschagne C., van Wyk A.M., Kotzé A., Grobler P., Dalton D.L. 2013. Isolation and characterization of species-specific microsatellites loci in African Penguin (*Spheniscus demersus*). *Conservation Generics Resources* 5: 169-171.

Laslett D. and Canbäck B. 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics,* 24**:** 172-175.

Lei R., Shore D.G., Brenneman R.A., Engberg S.E., Sitzmann B.D., Bailey C.A., Kimmel L.M., Randriamampoinona R., Ranaivoarisoa J.F., Louis Jr. E.E. 2010. Complete sequence and genome organization of the mitochondrial genome for Hubbard's sportive lemur (*Lepilemur hubbardorum*). *Gene,* 464: 44-49.

Lowe T.M. and Eddy S.R. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res*. 25: 955-964.

Mindell D.P., Sorenson M.D., Dimcheff D.E., 1998a. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol. Biol. Evol*. 15: 1568-1571.

Mindell D.P., Sorenson M.D., Dimcheff D.E. 1998b. Mutiple independent origins of mitochondrial gene order in birds. *Proc. Natl. Acad. Sci. USA,* 95: 10693-10697.

Ojala D., Montoya J., Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria *Nature,* 290: 470-474.

Ruokonen M. and Kvist L. 2002. Structure and evolution of the avian mitochondrial control region.*Molecular Phylogenetics and Evolution,* 23: 422-432.

Sammler S., Bleidorn C., Tiedemann R. 2011. Full mitochondrial genome sequence of two endemic Philippine hornbill species (*Aves: Bucerotidae*) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics,* 12: 35.

Slack K.E., Janke A., Penny D., Arnason U. 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. *Gene,* 302 (1-2): 43-52.

Taanman J.W. 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta* 1410: 103-123.

Walberg M.W. and Clayton D.A. 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucleic Acids Res*. 9: 5411-5421.

Watanabe M., Nikaido M., Tsuda T.T., Kobayashi T., Mindell D., Cao Y., Okada N., Hasegawa M. 2006. New candidate species most closely related to penguins. *Gene,* 378: 65-73.

Wolstenholme D.R. 1992. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* 141:173-216.

Wu X.B., Wang Y.Q., Zhou K.Y., Zhu W.Q., Nie J.S., Wang Z.L. 2003. Complete mitochondrial DNA sequence of Chinese alligator *Alligator sinensis* and phylogeny of crocodiles. *Chin. Sci. Bull.* 48 (19): 2050-2054.

Zhou Y., Zhang Jia-Y., Zheng Rong-Q., Yu Bao-G., Yang G. 2009. Complete nucleotide sequence and gene organization of the mitochondrial genome of *Paa spinosa* (*Anura: Ranoidae*). *Gene,* 447: 86-96.

CHAPTER 4

Isolation and characterization of species-specific microsatellite loci in African Penguin (*Spheniscus demersus*)

Abstract

Eight microsatellite markers were developed via pyrosequencing of a microsatellite-enriched library for the African Penguin (*Spheniscus demersus*). These microsatellite loci displayed 2 to 6 alleles with expected heterozygosity values ranging between 0.316 to 0.782 and observed heterozygosity between 0.381 to 0.84. These loci may be suitable for assessing patterns of genetic variability in African Penguin. This is the first development of species-specific markers for the African Penguin.

Introduction

African Penguin (*Spheniscus demersus*) populations have decreased substantially in South Africa and Namibia in the 20th century. The species' population trend is noted as still declining in the IUCN's Red List (IUCN, 2011) and currently there are estimated to be fewer than 31,000 breeding pairs left (ADU, 2007). The decrease is likely due to a number of factors including; competition for food with seals (Crawford *et al.,* 1992) and commercial fisheries (Frost *et al.,* 1976), predation by seals, oil spills (Morant *et al.,* 1981; Adams, 1994; Underhill *et al.,* 1999), and loss of habitat. We report the isolation and characterization of eight novel species specific markers suitable for investigating population genetic structure, gene flow and levels of genetic diversity in the African Penguin.

Materials and methods

Blood samples were collected from 25 breeding captive adult African Penguins in a colony located in the KwaZulu Natal Province of South Africa. For each individual, 30 microlitres (µL) blood was collected on filter paper. DNA extraction was conducted using the Qiagen DNeasy® Blood and Tissue Kit. The extraction protocol as outlined in the manufacturer protocol was followed. Microsatellite enrichment was performed using Fast Isolation by AFLP of Sequences Containing repeats (FIASCO) (Zane *et al.,* 2002*;* Cortinas *et al.,* 2006)..The microsatellite-enriched library was sequenced on the Roche 454 GS-FLX platform at Inqaba biotec (Pretoria, Gauteng, South Africa) generating 7706 reads. Sample preparation and analytical processing was performed at Inqaba biotec using the manufacturer's protocol. The program MSATCOMMANDER version 0.8.1 (Faircloth, 2008) was used to search the resulting reads for microsatellite motifs between 2 and 6 bp and with ≥8 repeats in length. A total of 1791 reads were identified containing microsatellite repeats. Primers flanking repeat regions were designed using PRIMER 3 software for 12 loci (Rozen and Skaletsky, 1997).

All loci were amplified individually prior to multiplexing. Promega GoTaq® Flexi DNA polymerase (Promega Corporation) was used for amplification in 12.5 µl reactions. The final reaction conditions

were as follows: 1 X PCR buffer, 1 mM $MgCl_2$, 200 micro molar (µM) of each 2'-deoxynucleotide triphosphate (dNTP), 10 pico mol (pmol) of each of the forward and reverse primer, 1 unit (U) *Taq* DNA polymerase and 50 nano gram (ng) genomic DNA template. The PCR reaction was carried out in the BOECO TC-PRO Thermal Cycler. The conditions for PCR amplification were as follows; 5 minutes at 95°C denaturation, 30 cycles for 30 seconds at 95°C, 30 sec at 50-60°C and 30 sec at 72°C, followed by extension at 72°C for 40 min. PCR products were pooled together and run against Genescan™ 500 LIZ™ internal size standard on an ABI 3130 Genetic Analyzer (Applied Biosystems, Inc.). Samples were genotyped using GeneMapper v. 4.0 (Applied Biosystems, Inc.). All 25 individuals were genotyped with the resulting MS markers. The number of alleles per locus, observed heterozygosity (Ho) and expected heterozygosity (He) were calculated with MS toolkit (Park, 2001). GENEPOP version 4.0.10 (Raymond and Rousset, 1995) was used to test for deviation from Hardy-Weinberg (HW) proportions and to evaluate loci for gametic disequilibrium.

Results and discussion

Eight of the loci were polymorphic, with the number of alleles ranging from 2 to 6 (Table 5). Mean Ho values ranged from 0.381 to 0.84 and He varied from 0.316 to 0.782. Deviations from HW and gametic disequilibrium were not observed for any of the markers. In conclusion, the eight microsatellite loci presented here will be useful for estimations of genetic diversity, population structure and for developing a conservation management strategy for this endangered species.

Table 5: Characterization of eight microsatellite loci in *Spheniscus demersus*: F = forward primer; R = reverse primer; bp = base pairs; No = number; Na = number of alleles; He = expected heterozygosity; Ho = observed heterozygosity and *P* = probability values for exact tests of HW proportions. GenBank accession numbers are JX494403 - JX494410.

No.	Name	Fluorescent dye label	Sequence (5'-3')	Repeat unit	Allele size range in bp (Na)	He	Ho	*P*
1	PNN01F	VIC	ATGATGAGAGGGATGAATGGAC	$(GGAT)_7$	309-321 (4)	0.7086	0.84	0.0759
	PNN01R		GAGTACACCTGCGCCAGAC					
2	PNN03F	NED	ACAAACTTCCCACCTGACTGTT	$(TAC)_{12}$	362-374 (4)	0.5656	0.5	0.1260
	PNN03R		GCTCCTATTTCACGACTCATCC					
3	PNN05F	PET	CAGTGACAGGCAAGGGTCTTAT	$(TG)_{10}$	245-247 (2)	0.316	0.8333	0.1202
	PNN05R		TGAGTAAGCAATGAGTTGGCAC					
4	PNN06F	NED	TCAGAAAGGAACTGTGTAGAGGC	$(TCTA)_{10}$	127-139 (4)	0.679	0.6087	0.6699
	PNN06R		TCCTGAGTAACACTTGTGGGTG					
5	PNN07F	VIC	GAGAGATGTTCATAGCACGCAG	$(CT)_{11}$	355-363 (2)	0.579	0.7	0.3805
	PNN07R		CTACCTTCTTCTTGGTTCTGGC					
6	PNN08F	FAM	GGAAATGCCACTGAAAACCTAA	$(ATAG)_9$	127-139 (4)	0.348	0.381	1.000
	PNN08R		GATAGATGGGGAACTGGAAACA					
7	PNN09F	FAM	CTGAGCAGACAAACTGGTAAAA	$(GATA)_{12}$	356-376 (6)	0.74	0.4348	0.5401
	PNN09R		TCAACTCGTCTTTGCTTACAAC					
8	PNN12F	FAM	TGGAGGTGTTATGTTTAGCAT	$(GT)_{10}$	244-256 (6)	0.782	0.810	0.0891
	PNN12R		TTCAGTGGCTGTATTTGCTG					

References

Adams N.J. 1994. Patterns and impacts of oiling on African Penguins *Spheniscus* demersus: 1981-1991. *Biological Conservation,* 68: 35-41.

ADU (Avian Demography Unit), 2007. ADU Projects and Research. News from the colonies 1, Nov. 2007. University of Cape Town: ADU.

Crawford R.J.M., Underhill L.G., Raubenheimer C.M., Dyer B.M., Martin J. 1992. Top predators in the Benguela ecosystem: implications of their trophic position. *South African Journal of Marine Science,* 12: 675-687.

Cortinas M.N., Barnes I., Wingfield B.D., Wingfield M.J. 2006. Polymorphic microsatellite markers for the Eucalyptus fungal pathogen *Colletogloeopsis zuluensis. Mol. Ecol. Notes,* 6: 780-783.

Faircloth B.C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources,* 8: 92-94.

Frost P.G.H., Siegfried W.R., Cooper J. 1976. Conservation of the Jackass Penguin (*Spheniscus* demersus (L.)). *Biological Conservation,* 9: 79-99.

IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2. <www.iucnredlist.org>. Downloaded on 03 May 2012.

Morant P.D., Cooper J., Randall R.M. 1981. The rehabilitation of oiled Jackass Penguins *Spheniscus* demersus 1970-1980. In: Cooper, J. (Ed.). Proceedings of the Symposium of Birds of the Sea and Shore. Cape Town; African Seabird Group: 267-30.

Park S.D.E. 2001. Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection [Ph.D. thesis] University of Dublin.

Raymond M. and Rousset F. 1995. Genepop (version 1.2), population genetics software for exact tests and ecumenicism. *J Hered*. 86: 248-249.

Rozen S. and Skaletsky H.J. 1997. Primer3. Primer3 Input (primer3.cgi v 0.2c). http://www.genome.wi.mit.edu/genome_software/other/primer3.html. Accessed 10 February 2011.

Underhill L.G., Bartlett P.A., Baumann L., Crawford R.J.M., Dyer B.M., Gildenhuys A., Nel D.C., Oatley T.B., Thornton M., Upfold L., Williams A.J., Whittington P.A., Wolfaardt A.C. 1999. Mortality and survival of African Penguins *Spheniscus demersus* involved in the Apollo Sea oil spill: an evaluation of rehabilitation efforts. *Ibis.* 141: 29-37.

Zane L., Bargelloni L., Patarnello T. 2002. Strategies for microsatellites isolation: a review. *Mol. Ecol.*11**:**1-16.

CHAPTER 5

Isolation and characterization of SNP markers for African Penguin (*Spheniscus demersus*)

Abstract

We report the characterization of 30 new single nucleotide polymorphism (SNP) markers for an endangered species, the African Penguin (*Spheniscus demersus*), based on screening of a random genomic library. The polymorphisms of these SNP loci were assessed using a captive population comprising 34 individuals. The minor allele frequency ranged from 2.17 to 42.65 and the observed and expected heterozygosity ranged from 0.1 to 0.6897 and from 0.0435 to 0.4965, respectively. None of the loci deviated significantly from Hardy–Weinberg equilibrium. These SNP markers will provide a necessary addition to the genetic tools employed for understanding population structure and for developing a conservation management strategy for this endangered species.

Introduction

The African Penguin (*Spheniscus demersus*) is currently listed in the Red Data Book as an endangered species (IUCN, 2011) with populations on 25 islands around Southern Africa, from Namibia to St. Croix island (Algoa Bay). Primary threats to African Penguins include competition for food with seals (Crawford *et al.,* 1992) and commercial fisheries (Frost *et al.,* 1976), predation by seals, oil spills (Morant *et al.,* 1981; Adams, 1994; Underhill *et al.,* 1999), and loss of habitat. Molecular genetic data is increasingly important for effective conservation and management of threatened species. In addition to allowing the identification of populations sufficiently divergent to warrant independent conservation programmes, genetic data is critical in evaluating which sources of animals are most appropriate for reintroductions in areas where a threatened species has suffered local extinctions. Currently, little is known about penguin dispersal behaviour and population structure. The use of microsatellites and single nucleotide polymorphisms (SNPs) provides an alternative to behavioural and tracking studies in a species that spends much of its life off shore. To complement existing mtDNA and microsatellite markers for this species, we have detected 30 SNPs across 11 loci identified by sequencing of a random genomic library (Olsen *et al.,* 2011).

Materials and methods

A total of 34 blood samples were collected from a captive facility in the KwaZulu Natal Province of South Africa as part of an intensive metapopulation management programme. Approximately 30 microlitres (µL) blood was collected on filter paper. DNA extraction was conducted using the Qiagen DNeasy® Blood and Tissue Kit. The extraction protocol as outlined in the manufacturer protocol was followed. Random genomic libraries were constructed by digesting genomic DNA of a single isolate with the restrictions enzymes HpaII or TruI (Thermo Scientific), followed by cloning of the resulting

fragments into the CloneJet (Thermo Scientific) cloning system. Twenty two loci from the genomic library was successfully sequenced for a subset of five randomly chosen individuals resulting in approximately 8896 base pairs (bp) of sequence data per isolate. Comparison of the sequence data between the isolates utilising CLC Bio's Main workbench (Denmark), revealed 30 SNPs distributed across 11 loci. This equates to an average of one SNP every 296 bp, though SNPs were not evenly distributed and was clustered in 9 out of 11 loci. Subsequently the 11 loci that contained SNPs were amplified and sequenced in the remaining 29 isolates. Amplification was done using Dream Taq™ Green PCR Master Mix (2x) supplied by Thermo Scientific, Lithuania. The PCR mix for each locus contained 12.5 µl of 2x Dream Taq™ PCR master mix (10x Dream Taq™ buffer, dATP, dCTP, dGTP and dTTP, 0.4 mM each, 4 mM $MgCl_2$ and 1.25 U Dream Taq™ polymerase), 1 µl [10uM] of each primer (synthesized by Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa), 50 ng of template DNA and nuclease free water to reach a final volume of 25 µl. Sequencing of resulting amplicons was conducted by Inqaba Biotechnical Industries (Pty) Ltd utilising the ABI Big Dye V3.1 kit and the ABI 3500XL genetic analyser. Sequence data was screened and aligned using the Main workbench from CLC bio (Denmark). GENEPOP version 4.0.10 (Raymond and Rousset 1995) was used to calculate observed (Ho) and expected heterozygosity (He) and to test for genotypic linkage disequilibrium (LD) and departures from Hardy-Weinberg equilibrium (HWE).

Results and discussion

The minor allele frequency ranged from 2.17 to 42.65 and the observed and expected heterozygosity ranged from 0.1 to 0.6897 and from 0.0435 to 0.4965, respectively. None of the loci deviated significantly from Hardy–Weinberg equilibrium. Significant LD was observed for SNPs within Loci (P110-I-1 and P110-I-2; P110 EVE 5-1, P110 EVE 5-2 and P110 EVE 5-3; P110 EVE 10-1 and P110 EVE 10-2; P110 EVE 10-3 and P110 EVE 10-4; P110-L-4 and P110-L-5). For future studies only one of each linked set should be typed rather than the entire set presented here. In conclusion the SNPs described here should be further tested on larger populations to evaluate their usefulness in delineating population structure, individual genetic assignment and parentage determination. The utility of these markers in related species should be investigated.

Table 6: Characterization of 30 SNPs in African Penguin (*Spheniscus demersus*): F = forward primer; R = reverse primer; bp = base pairs; He = expected heterozygosity and Ho = observed heterozygosity. GenBank accession numbers are NCBI_SS529944710-529944737.

Locus	SNP Name	Fragment Size (bp)	Sequence (5'-3')	Minor allele frequency	Heterozygosity	
					He	Ho
PG NE 11	P110 NE 11-1	525	F-GCCACAGGCATTAACGCTCTG R-GTTAAACCTTGACAAACCTGCAG	16.18	0.2752	0.2647
	P110 NE 11-2			26.47	0.3951	0.4118
	P110 NE 11-3			8.82	0.1633	0.1765
PG NE 12	P110 NE 12-1	295	F-CCAGGTATTGAAATCAC R-GCTCTTAGTGTTTCAGG	3.13	0.0615	0
	P110 NE 12-2			4.69	0.0908	0.0313
PG NE 15	P110 NE 15-1	624	F-CTGCCAGAGATGCTGGCTAGC R-CTGTGGATGCCGTTTGATCCC	5.88	0.1124	0.1176
	P110 NE 15-2			27.94	0.4087	0.5
PG L	P110-L-1	579	F-CCACTCTTGGCTCTGATTATTC R-CTCTACTCTTCCTACGCAGC	20.59	0.33319	0.3529
	P110-L-2			19.12	0.3139	0.3235
	P110-L-3			25	0.3806	0.4412
	P110-L-4			41.38	0.4936	0.6897
	P110-L-5			41.07	0.4929	0.6071
	P110-L-6			10.87	0.1981	0.2174
	P110-L-7			2.17	0.0435	0.0435
	P110-L-8			2.17	0.0435	0.0435
	P110-L-9			5	0.0974	0.1
	P110-L-10			15	0.2615	0.3
PG EVE 5	P110 EVE 5-1	361	F-GGTAAAGAGCTACCTGAAG R-CAGAAACCGTTAGATTGCC	41.18	0.4917	0.5882
	P110 EVE 5-2			41.18	0.4917	0.5882
	P110 EVE 5-3			41.18	0.4917	0.5882
PG I	P110-I-1	555	F-CCAAGGGAAGGATCACAGGC R-CCTCCCAACTCCTTGTGCC	33.33	0.4513	0.4242
	P110-I-2			28.79	0.4163	0.4545

Table 6 Continued: Characterization of 30 SNPs in African Penguin (*Spheniscus demersus*): F = forward primer; R = reverse primer; bp = base pairs; He = expected heterozygosity and Ho = observed heterozygosity. GenBank accession numbers are NCBI_SS529944710-529944737.

Locus	SNP Name	Fragment Size (bp)	Sequence (5'-3')	Minor allele frequency	Heterozygosity	
					He	Ho
PG EVE 10	P110 EVE 10-1	305	F-GGGAAGGGATTCCTTTGG R-CATGAAACCCACAAGAAGG	36.76	0.4719	0.5
	P110 EVE 10-2			36.76	0.4719	0.5
	P110 EVE 10-3			27.94	0.4087	0.3235
	P110 EVE 10-4			16.18	0.2752	0.3235
PG A	P110-A1	44	F-GAGGAAGTGTTCGGTGGG R-GATCATTTGTCATGCAATGCTAGG	7.35	0.1383	0.1471
PG NE 1	P110 NE 1-1	513	F-CCAACATTTCAGTTGCCAACC R-CTTGCCAAGTGCCTGGTGTT	42.65	0.4965	0.5588
C6-306	P110 C6-306-1	255	F-TCACACCATTCAGCAACA R-CAGTGTCTTCCAAATGCAA	2.94	0.0579	0.0588
	P110 C6-306-2			26.47	0.3951	0.4118
B1-534	P110 B1-534-1	450	F-ACAAGTCGTTTCTCAGTTC R-CAACAACAGGGAGTCACA	30.88	0.4333	0.4412

References

Adams N.J. 1994. Patterns and impacts of oiling on African Penguins *Spheniscus* demersus: 1981-1991. *Biological Conservation,* 68: 35-41.

Crawford R.J.M., Underhill L.G., Raubenheimer C.M., Dyer B.M., Martin J. 1992. Top predators in the Benguela ecosystem - implications of their trophic position. *South African Journal of Marine Science,* 12: 675-687.

Frost P.G.H., Siegfried W.R., Cooper J. 1976. Conservation of the Jackass Penguin (*Spheniscus demersus*). *Biological Conservation,* 9: 79-99.

IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2. <www.iucnredlist.org>. Downloaded on 03 May 2012.

Morant P.D., Cooper J., Randall R.M. 1981. The rehabilitation of oiled Jackass Penguins *Spheniscus demersus* 1970-1980. In: Cooper, J. (Ed.). *Proceedings of the Symposium of Birds of the Sea and Shore. Cape Town; African Seabird Group*: 267-30.

Olsen M.T., Volny V.H., Bérubé M., Dietz R., Lydersen C., Kovacs K.M., Dodd R.S., Palsbøll P.J. 2011. A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Artic ringed seal (*Pusa hispida hispida*). *Mol Ecol Resources*., 11: 9-19

Raymond M. and Rousset F. 1995. Genepop (version 1.2), population genetics software for exact tests and ecumenicism. *J Hered*. 86: 248-249.

Underhill L.G., Bartlett P.A., Baumann L., Crawford R.J.M., Dyer B.M., Gildenhuys A., Nel D.C., Oatley T.B., Thornton M., Upfold L., Williams A.J., Whittington P.A., Wolfaardt A.C. 1999. Mortality and survival of African Penguins *Spheniscus* demersus involved in the Apollo Sea oil spill: an evaluation of rehabilitation efforts. *Ibis* 141: 29-37.

CHAPTER 6

Genetic monitoring of *ex-situ* African Penguin (*Spheniscus demersus*) populations in South Africa

Abstract

The African Penguin (*Spheniscus demersus*) has suffered population declines and is listed in the IUCN Red Data Book as an endangered species. The species is endemic to the coast of southern Africa, and breeding colonies are distributed on the south-western coast of Africa. Currently, African Penguins are being kept in zoo and aquarium facilities throughout South Africa. In this study, molecular genetic data based on 12 microsatellite markers from 119 African penguin samples from four facilities was generated in order to determine the level of genetic variation, population structure and differentiation, and effective population size to assist in the development of an effective captive management plan. Expected heterozygosity ranged from 0.57 to 0.62, and allelic richness from 4.2 to 5.1. However, based on differences between first and second generation captive birds, we conclude that the *ex-situ* population is at risk of losing genetic variability in the future and management programmes should include exchange of birds between captive facilities in order to induce gene flow and increase effective population size. Adding individuals from *in-situ* populations should also be considered in the future in cases where these birds cannot be rehabilitated. Molecular genetic analyses of wild penguin populations should be carried out for comparison, and to ascertain to what degree '*in-situ* genetic diversity' is represented among *ex-situ* populations. With regular re-sampling and analyses, the extent of the effect of processes such as genetic drift on diversity in the *ex-situ* penguin populations will become evident.

Introduction

The African Penguin (*Spheniscus demersus*) is endemic to southern Africa. Their populations have decreased substantially in South Africa and Namibia in the 20th century. The species' population trend is noted as still declining in the IUCN's Red List (IUCN, 2011) and currently there are estimated to be fewer than 26,000 breeding pairs left (Crawford *et al*. 2011). Initially, populations at some breeding localities declined due to excessive harvesting of their eggs, which occurred until 1967 (Shelton *et al*., 1984). In addition, guano harvesting has influenced the quality of nesting habitat. African penguins at certain islands have to nest in the open rather than in the guano harvest areas, making them more susceptible to heat stress and displacement from breeding sites by larger animals (Crawford *et al*., 1989). More recent declines are due to a number of factors including competition for food with seals (Crawford *et al*., 1992) and commercial fisheries (Frost *et al*., 1976), oil spills (Morant *et al*., 1981; Adams, 1994; Underhill *et al*., 1999), loss of habitat and lastly climate change. Penguins are currently swimming farther from their nests during incubation due to shifts in prey distribution as an effect of climate change (Boersma, 2008).

Ex-situ populations can serve a number of different roles in conservation efforts including public education, being invaluable assets for scientific discovery and as resources for supplementation or restoration of *in-situ* populations (Lacy, 2009). Currently, penguin populations are being kept in zoo and aquarium facilities throughout South Africa. As part of the management plan for this species, an African Association of Zoos and Aquaria (PAAZAB) regional studbook is maintained by the National Zoological Gardens of South Africa. The African regional studbook for the African Penguin uses the Single Population Analysis and Record Keeping System (SPARKS) developed by the International Species Information System (ISIS) and the PM2000 database programme.

There are several genetic concerns that should be taken into account for the management of *ex-situ* penguin populations. Since *ex-situ* populations are generally derived from a small number of individuals, these populations face the same threats as small and isolated natural populations. Adverse genetic changes due to founder effects, loss of genetic diversity, inbreeding depression, genetic adaptations to captivity (through selection) that are deleterious in the *in-situ* and the possible occurrence of new deleterious mutations may jeopardise the ability of *ex-situ* populations to reproduce and survive when returned to the wild (Woodworth *et al*., 2002). In addition, research has demonstrated that inbred individuals have lower resistance to disease and environmental stress (Keller *et al*., 1994). Thus genetic factors need to be considered in management plans for small and isolated *ex-situ* populations. Our research was designed to determine the level of genetic variability in the *ex-situ* populations and to elucidate captive African Penguin population structure. In addition, the effective population size of the *ex-situ* populations based on genetic data was studied. Lastly, the potential impact of genetic drift over the generation time was determined.

Study area

Blood samples were collected from 119 African Penguins from four facilities in South Africa, namely: SANCCOB =Southern African Foundation for the Conservation of Coastal Birds, Bloubergrandt, Western Cape, TOA = Two Oceans Aquarium, NZG = National Zoological Gardens of South Africa, Pretoria, Gauteng and uShaka= uShaka Marine World, KwaZulu-Natal). All birds are kept in the facilities as breeding populations. All necessary research and ethics permits were approved for the collection of samples; National Zoological Gardens of South Africa Research and Ethics Scientific Committee and the South African Department of Environmental Affairs permit number: RES2010/66.

Materials and methods

Sampling and laboratory procedures

Blood samples were collected from 119 African Penguins. For each individual, 30 µL of blood was collected on FTA® filter paper cards (Whatman, NJ, USA). DNA extraction was conducted using the Qiagen DNeasy® Blood and Tissue Kit. The extraction protocol as outlined in the manufacturer's protocol was followed. A total of 12 microsatellite markers were developed as described in Schlosser *et al*., 2003 and Labuschagne *et al*., 2013. Promega GoTaq® Flexi DNA polymerase (Promega Corporation) was used for amplification in 12.5 µl reactions. The final reaction conditions were as

follows: 1 X PCR buffer, 1 mM $MgCl_2$, 200 µM of each dNTP, 10 pmol of each of the forward and reverse primer, 1 U *Taq* DNA polymerase and 50 ng genomic DNA template. The PCR reaction was carried out in the BOECO TC-PRO Thermal Cycler. The conditions for PCR amplification were as follows: 5 minutes at 95°C denaturation, 30 cycles for 30 sec at 95°C, 30 sec at 50-60°C and 30 sec at 72°C, followed by extension at 72°C for 40 min. PCR products were pooled and run against an Genescan™ 500 LIZ™ internal size standard on an ABI 3130 Genetic Analyzer (Applied Biosystems, Inc.). Samples were genotyped using GeneMapper v.4.0 (Applied Biosystems, Inc.).

Genetic variability

MICRO-CHECKER (Van Oosterhout *et al*. 2004) was used to detect possible genotyping errors, allele dropout and non-amplified alleles (null alleles). The approaches implemented in this software package can estimate the frequency of null alleles and adjust the dataset to correct for genotyping errors. Differences in mean observed heterozygosity (H_o), mean expected heterozygosity (H_e), unbiased expected heterozygosity ($H_{z)}$, mean number of alleles and F_{IS} were determined between four penguin populations using MS Toolkit (Park 2001) and GenAlEx (Peakall & Smouse 2006). GENEPOP version 4.0.10 (Raymond and Rousset, 1995) was used to test for deviations from expected Hardy-Weinberg (HW) proportions and to evaluate loci for gametic disequilibrium. Allelic richness was calculated using the program Fstat v. 2.9.3.2 (Goudet, 1995, 2001). In addition, a sub-set of samples were analysed to determine loss of variation over time. Two generations based on studbook information could be distinguished in two populations (NZG and uShaka), namely founders (*in-situ* birds placed in captive facilities, n = 16) and their descendants (F1, F = 32). Both generations of the *ex-situ* population were treated as distinct subpopulations. For each generation the allele frequencies, mean number of alleles, H_o $H_{e,}$ H_z and F_{IS} was calculated using MS Toolkit (Park 2001) and GenAlEx (Peakall & Smouse 2006).

Effective population size

The effective population size (N_e), defined as the number of breeding individuals in an ideal population that exhibit the same dispersion of allele frequencies as the population of interest (Frankham *et al*., 2002), is a key parameter in studies of genetic diversity. We used a single-sample method to estimate current N_e based on linkage disequilibrium as implemented in the program LDNe (Waples and Do 2008). This included a biased correction (Waples 2006), shown to improve performance even with non-ideal populations (e.g. skewed sex ratios or non-random variance in reproductive success). A jack-knife method was used to obtain 95% confidence intervals (CI) on loci and estimates were calculated assuming random mating, with all alleles <0.01 excluded, following Waples and Do (2008). Effective population size was also investigated utilising two temporal methods requiring at least two samples in time; a temporal moment-based (MBT) (Waples 1989) and a Bayesian coalescent-based approach (Berthier *et al*., 2002) implemented in the program NeEstimator (Ovenden *et al*., 2007). Analysis of molecular variance (AMOVA) was performed for hierarchical partitioning of genetic variation among the four populations of penguins. AMOVA was calculated in GenAlEx (Peakall & Smouse 2006, 2012).

Population differentiation and structure

The genetic relationship between populations and individual assignments of captive penguins from four different facilities was inferred via a Bayesian clustering analysis using the programme STRUCTURE version 2.3.3 (Pritchard *et al*., 2000). Twenty runs were performed for K=1 to 20. The parameter settings specified the admixture model, with allele frequencies correlated and location information not given a priori. Each run consisted of 1 000 000 generations, with an additional 100 000 generations discarded as burn-in. All other settings were left as default. The values for the estimated ln(Pr(X|K)) were averaged, from which the posterior probabilities were calculated. The K with the greatest increase in posterior probability (ΔK, Evanno *et al*., 2005) was identified as the optimum number of sub-populations using STRUCTURE HARVESTER (Earl and vonHoldt, 2012). Importantly, ΔK cannot find the true number of populations if there is a panmictic population i.e. one that has no detectable discrete, genetically distinct groups based on the multi-locus genotypes. After the best value of K was determined, CLUMPP v.1.1.2 (Jakobsson & Rosenberg, 2003) was used to combine the results of each of the 20 replicates generated during each of the analyses, into a final result for each. The "Full Search" option in CLUMPP (M=1) was employed, with all other settings left as the default. The program DISTRUCT v.1.1 (Rosenberg, 2003) was used to visualise results from the CLUMPP analysis. In addition, GenAlEx (Peakall & Smouse 2006) was used for computing other genetic diversity parameters such as pair-wise F_{ST} (Wright, 1965) in order to assess population differentiation among the four captive populations.

Results

Genetic variability

Deviations from gametic equilibrium were not observed in any population for any of the markers when populations were analysed separately. Deviations from HWE, following Bonferroni correction was observed for one marker (B3-2) in the NZG population one marker (PNN03) in the TOA population and four markers (G2-2, SH1CA9, G3-6, PNN12) in the uShaka population. HW disequilibrium may be attributed to small population size, selection, non-random mating, inbreeding or genetic drift. Null alleles were not observed by using the approaches implemented in MICRO-CHECKER. The overall captive population of African Penguin was polymorphic at all 12 microsatellite loci investigated, with a mean number of five alleles per locus. Mean H_o values ranged from 0.52 to 0.623 and H_e varied from 0.54 to 0.62 (Table 7). Captive populations were similar in terms of genetic diversity and number of alleles, with NZG displaying the lowest values. F_{IS} values ranged from 0.001 (SANCCOB) to 0.054 (uShaka). All four populations showed higher than zero F_{IS} values indicating heterozygote shortage (Table 7) which can suggest inbreeding. A mode-shift indicator test was therefore conducted in which the alleles were organised into frequency classes. The distribution of alleles followed the normal L-shaped curve as shown in Figure 10. A comparison of number of alleles and heterozygosity in founder and offspring populations were similar (Table 8). However, F_{IS} values increased from -0.044 in the founder group to to 0.163 in the F1 generation.

Table 7: Heterozygosity values and average number of alleles per locus for four Penguin populations from South Africa over all loci. Genotyping included 12 microsatellite loci. N = mean sample size, N_a = number of alleles, N_{eA} = number of effective alleles, H_O = observed heterozygosities, H_e = expected heterozygosities, H_z = unbiased expected heterozygosity, F_{IS} = inbreeding coefficient and R = allelic richness. SANCCOB = Southern African Foundation for the Conservation of Coastal Birds, TOA= Two Oceans Aquarium, NZG = National Zoological Gardens of South Africa, uShaka = uShaka Marine World.

Population	**No. samples**	N_a	R	N_{eA}	H_O	H_e	H_z	F_{IS}
SANCCOB	16	5.2	5.11	3.1	0.63	0.62	0.64	0.010
TOA	12	4.4	5.02	3.0	0.57	0.57	0.59	0.038
NZG	37	5.00	4.50	2.6	0.52	0.54	0.55	0.053
uShaka	54	5.4	4.19	3.0	0.59	0.62	0.63	0.054
Total	119	5.0	4.6	2.9	0.58	0.59	0.60	0.039

Table 8: Heterozygosity values and average number of alleles per locus for founder and F1 generations. Genotyping included 12 microsatellite loci. N = mean sample size, N_a = number of alleles, N_{eA} = number of effective alleles, H_O = observed heterozygosities, H_e = expected heterozygosities, H_z = unbiased expected heterozygosity, F_{IS} = inbreeding coefficient. Significant values indicated in bold. NZG = National Zoological Gardens of South Africa, uShaka = uShaka Marine World.

Population	**No. samples**	N_a	N_{eA}	H_O	H_e	H_z	F_{IS}
Founders	16	4.5	2.9	0.62	0.60	0.62	-0.044
F1	32	5.3	2.9	0.51	0.61	0.61	**0.163**
Total	48	4.9	2.9	0.57	0.60	0.62	**0.059**

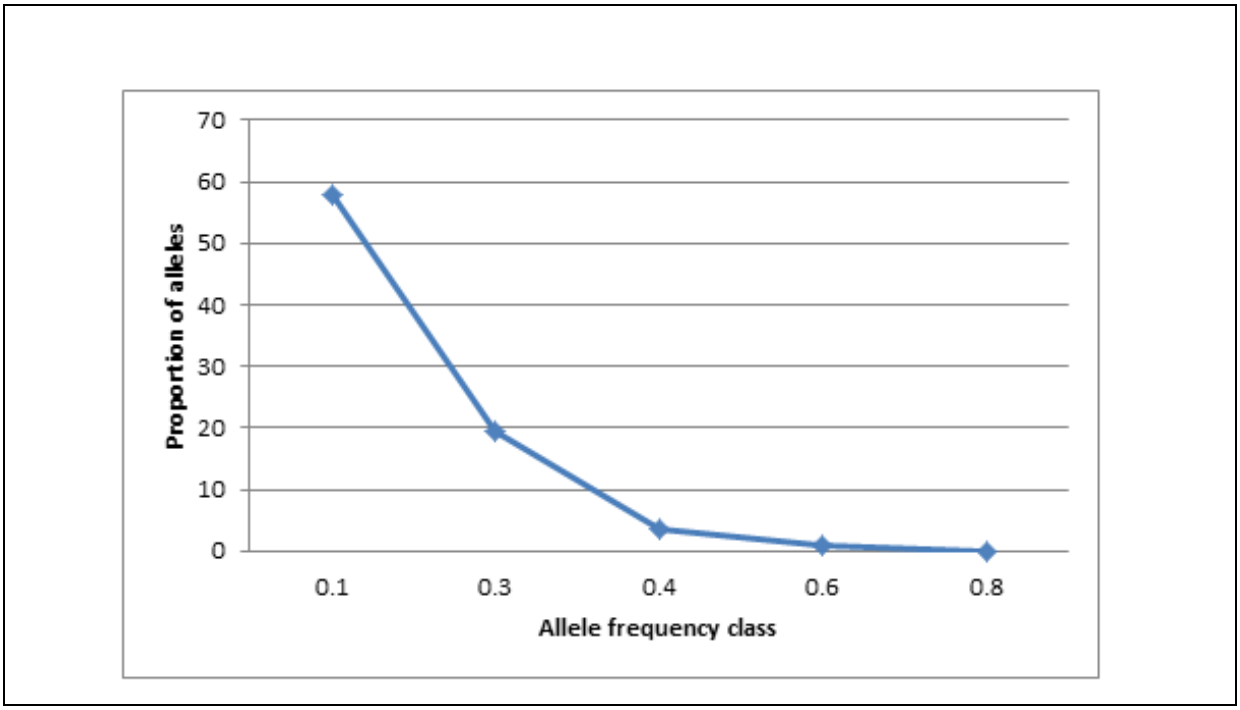


Figure 10: L-shaped curve of distribution of proportion of alleles in different allelic frequency classes in the African Penguin

Effective population size

Table 9 provides the estimates for effective population size based on a one-sample method and two temporal methods. Effective size calculated for the total captive population varied between 62.8 and 110.7 for the different methods applied. Effective size based on linkage disequilibrium (LD) for the offspring were lower than those computed for the founders. Notably, there was a 50.2% decrease in effective size between founders and offspring. Although the two temporal method estimates were very similar (62.8 and 66.8) they were smaller than that estimated by the LD method (110.7).

Table 9: Effective population size (N_e) estimates from a one-sample and two two-sample methodologies.

Population	**Estimated N_e (95% confidence interval)**		
	One-sample method	**Two-sample methods**	
	LDNe	**NeEstimator (Moment based)**	**TM3**
Founders	109.6 (59.1-370.7)		
Offspring	54.5 (41.1-75.9)		
Total Captive	110.7 (83.3-155.3)	66.8 (27.3-688.5)	62.8 (30.3-147.1)

Population structure and differentiation and Analysis of molecular variance (AMOVA)

Posterior probabilities (Ln) using Bayesian admixture analysis were calculated for K= 1-20 with K=1 being identified as the most likely true K value as K = 1 displayed the greatest posterior probability. Results indicated that there are no distinct clusters for captive penguin populations (Figure 11).

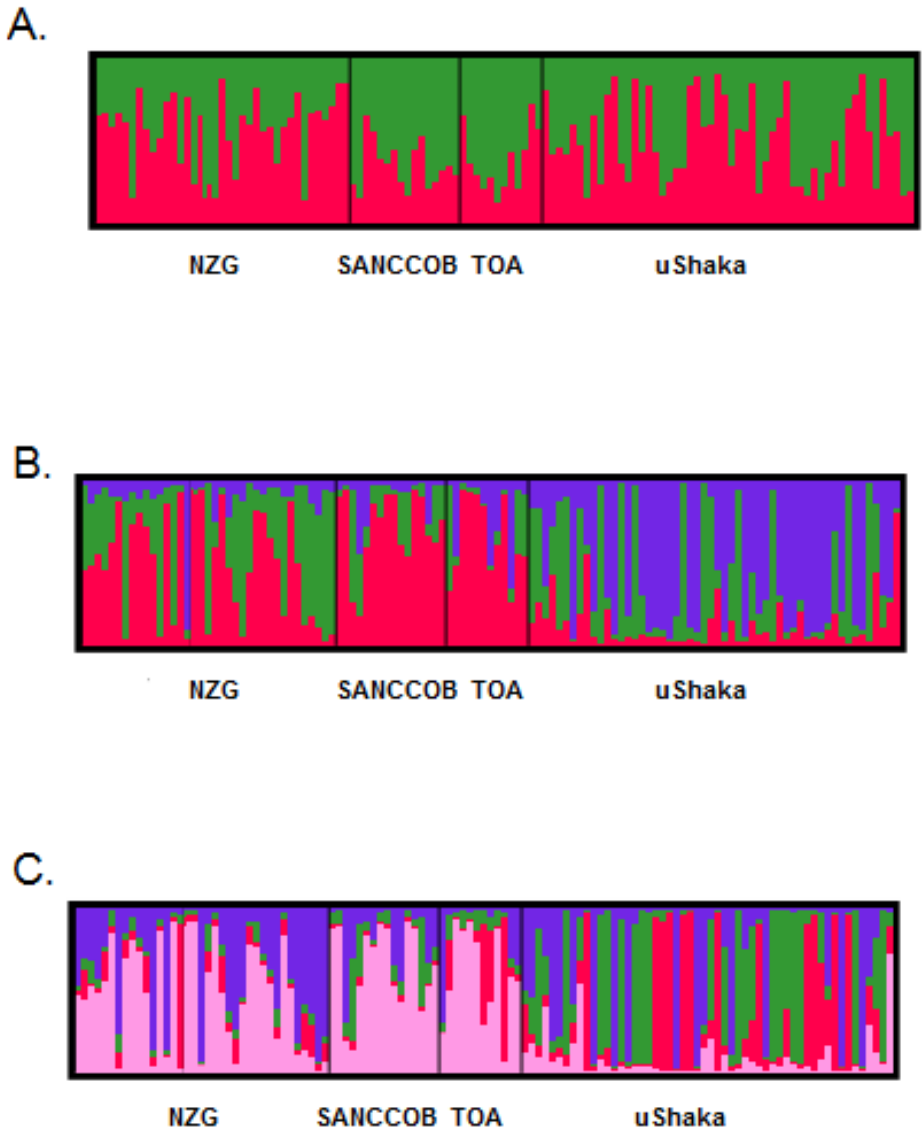


Figure 11: Cluster identity plots for the STRUCTURE analysis of captive birds (n=119). Results are shown for (A) K=2 (B) K=3 and (C) K=4. Results are based on the average of 20 runs of 1 million iterations each, for each value of K. Each vertical bar represents one individual.

Fixation indices, and their derivatives, were calculated with populations defined as the four captive institutions. All GST and corrected GST estimates, as well as DEST, were highly significant (P<0.001). GSTmax was 0.3±0.06 after correction, and overall GST was 0.03. FST was estimated as

0.05±0.01 (P<0.001) across all loci and all populations, and G”ST was estimated at 0.105±0.045. Pairwise population differentiation estimates were all highly significant (P<0.01), and consistently showed that the TOA and uShaka populations were the most different from each other (Tables 10 to 13). AMOVA-based FST was estimated based on the four captive populations to investigate the hierarchical partitioning of genetic diversity. FSTmax was 0.4, and overall FST was 0.04 (F'ST=0.1) indicating significant moderate population structure (P<0.001). AMOVA revealed 4% of the variation in allele frequencies is among captive populations. Pairwise population FST-values were > 0.1 for all comparisons involving uShaka (corrected values; Table 13), and TOA and NZG are the least differentiated. In contrast, the RST-based AMOVA for the same dataset estimated overall RST to be -0.01, with none of the variance in allele sizes explained among populations. None of the pairwise population comparisons of RST were significant, and all were less than zero.

Table 10: Pairwise population G_{ST} estimates (below the diagonal) and the associated probabilities (above the diagonal) based on 9 999 permutations of 9 999 pairwise population permutation and 10 000 bootstrap replicates.

	SANCCOB	Two Oceans	NZG	uShaka
SANCCOB		0.003	<0.001	<0.001
Two Oceans	0.019		0.009	<0.001
NZG	0.019	0.013		<0.001
uShaka	0.022	0.032	0.022	

Table 11: Pairwise population G''_{ST} estimates (below the diagonal) and the associated probabilities (above the diagonal) based on 9 999 permutations of 9 999 pairwise population permutation and 10 000 bootstrap replicates.

	SANCCOB	Two Oceans	NZG	uShaka
SANCCOB		0.004	<0.001	<0.001
Two Oceans	0.096		0.007	<0.001
NZG	0.091	0.062		<0.001
uShaka	0.117	0.156	0.107	

Table 12: Pairwise population estimates of Jost’s D (below the diagonal) between the four captive populations of African Penguins, and the associated probabilities (above the diagonal).

	SANCCOB	Two Oceans	NZG	uShaka
SANCCOB		0.004	<0.001	<0.001
Two Oceans	0.061		0.007	<0.001
NZG	0.056	0.036		<0.001
uShaka	0.077	0.101	0.065	

Table 13: Pairwise population estimates of corrected F_{ST} (F'_{ST}, from AMOVA) for all captive African Penguin populations

	SANCCOB	Two Oceans	NZG
Two Oceans	0.09		
NZG	0.09	0.06	
uShaka	0.11	0.15	0.10

Discussion

Maintaining genetic diversity is a primary goal of any captive management plan as it allows populations to adapt in response to environmental change (Frankham, 1996). All populations included in this study showed comparable and presumably adequate levels of genetic variability in terms of He and Hz, with comparable values in the *in-situ* populations (Nupen *et al*., unpublished) and captive populations. The observed genetic diversity found in this study was similar to levels reported for other species of penguin. Bouzat *et al*. (2009) reported a He value of 0.598 for *in-situ* Magellanic Penguin (*Spheniscus Magellanicus*) and Billing *et al*. (2006) indicated that He was 0.613 in Little Penguin (*Eudyptula minor*). Genetic variability in terms of number of alleles between founder populations and their offspring were found to be similar (Table 8). The offspring generation displayed a higher inbreeding coefficient in comparison to the founder generation, most likely due to an increase in relatedness of the offspring generation. Loss of alleles or heterozygosity may occur due to founder effects or reduced population size and associated non-random mating, and this loss is then maintained or exacerbated through genetic drift, the most powerful evolutionary force acting in small populations, resulting from random sampling of alleles in each generation (Allendorf, 1986). Genetic drift is also dependent on the effective population size (Hartl and Clark, 1989). Genetic drift can lead to changes in allele frequencies and fixation or loss of particular alleles. Managers should therefore aim to represent the genetic composition of the *in-situ* populations in captive populations since captive birds may in future be reintroduced into the *in-situ* as part of conservation efforts. The risk of genetic drift in captive populations is potentially high since captive populations are smaller in size, at higher risk to suffer from founder effects, genetic drift, inbreeding depression and selection for the captive environment (Nei *et al*., 1975; Leberg, 1992; Frankham *et al*., 2001).

The results of the current study indicate that the *ex-situ* populations are currently not at risk of showing the deleterious effects of inbreeding. This hypothesis was confirmed by the Mode-shift indicator test, indicating that the African penguin populations have not experienced a recent genetic bottleneck (Luikart *et al*., 1998). However, management of these populations should be directed to maintain low inbreeding levels in these populations. To avoid an excessive increase in inbreeding and minimize differentiation between subpopulations, a minimum migration rate of one migrant per generation, and subpopulation, has been suggested by some authors (Mills and Allendorf, 1996). Adding individuals from the wild will induce gene flow which will aid in maintaining genetic diversity and minimizing genetic adaptation to captivity, thereby increasing the fitness of subsequent

generations if and when they are released into their natural habitat. Re-sampling and analyses should be a continuous process, to measure the extent and effect of processes such as genetic drift on diversity in the *ex-situ* penguin populations.

Contemporary effective size (roughly, Ne that applies to the time period encompassed by the sampling effort) can be estimated on either a single sample (Tallmon *et al*., 2008; Waples and Do, 2008) or two samples (temporal) (Waples 1989; Berthier *et al*., 2002). Single sample methods estimate Ne from genetic patterns (heterozygosity, linkage disequilibrium, individual relatedness) within a single population, whereas temporal methods depend on random changes in allele frequency over time (Hare *et al*., 2011). Effective population size determines the strength of genetic drift in a population and is a crucial estimate relevant to management since it integrates genetic effects with the life history of the species, allowing for predictions of a population's current and future viability (Hare *et al*., 2011). Effective population size estimates have been determined for *in-situ* Yellow-eyed Penguins (*Megadyptes antipodes*) in South Island, New Zealand (Boessenkool *et al*., 2010) and in Southern Ground hornbills (Theron *et al*., 2013). As in the study presented here, Boessenkool *et al*. (2010) indicated that Ne estimates varied greatly depending on the type of method used. There are a number of factors that may contribute to low precision in Ne estimates: overlapping generations, violations of assumptions associated with closed populations, the number of loci used, allelic diversity as well as sample size (Waples, 2006 and Waples and Do, 2008). The LD method estimates inbreeding Ne in the generation preceding the sample by measuring the genetic result of processes acting in the parental generation and is concerned with the loss in heterozygosity (Luikart *et al*., 2010). In contrast, the temporal methods estimate variance Ne on the basis of allele frequency differences over time (Hare *et al*., 2011). The results from the temporal approach may be lower since variance Ne is more sensitive for early detection of population declines. This is due to variance Ne generally declining rapidly during bottlenecks whereas inbreeding Ne does not change until inbreeding accumulates following increased mating between relatives. Populations with small Nes have an increased risk of inbreeding depression (Frankham, 2005). Genetic drift erodes genetic variation as Ne decreases, elevating the probability of fixation of deleterious alleles and reducing the effectiveness of selection, all of which reduce overall fitness and limit adaptive responses (Hare *et al*., 2011). Captive-bred animal models suggests that Ne > 50 is needed to limit the chance of inbreeding depression, while Ne > 100 is needed to decrease the chance of mutational meltdown (Frankham *et al*., 2002; Traill *et al*., 2010). To maintain long-term evolutionary potential in the form of additive genetic variance, a Ne of at least 500-5000 has been proposed (Franklin and Frankham, 1998). Since it is unlikely that captive populations would ever approach 500, genetic variability can rather be maintained over long time spans through actions that induce gene flow. Apart from gene flow, minimization of coancestry in a subdivided population is equivalent to maximization effective population size (Caballero and Toro, 2002).

Based on microsatellite data using 12 markers, we found no significant population structure using the Bayesian approach implemented in STRUCTURE. Values obtained for pair-wise FST also suggest

low differentiation between the four populations of penguins. This is further supported by AMOVA values, where a very low level of overall variation (10%) was observed between populations compared to 90% of variation observed within the populations. This is in contrast to results reported on captive African penguins in Japan. The authors identified two distinct mitochondrial DNA lineages (Murata and Murakami, 2013). However, limited and low levels of genetic differentiation have also been reported in Adélie Penguin (*Pygoscelis adeliae*; Roeder *et al*., 2001), Magellanic Penguin (*Spheniscus Magellanicus*; Bouzat *et al*., 2009), Humboldt Penguin (*Spheniscus humboldti*; Schlosser *et al*., 2009) and *in-situ* African penguin populations (Nupen *et al*., unpublished). These authors indicated that lack of genetic differentiation may be due to gene flow between colonies and large effective population sizes. Lack of genetic differentiation using neutral markers in African Penguins is most likely due to historically unimpeded gene flow between the *in-situ* colonies, taking in to consideration that the populations in facilities have been only recently established. These results do not justify the treatment of the *ex-situ* penguin populations as separate management units but rather as a single metapopulation. However, although there is an absence of genetic structure based on neutral genetic markers there may be differences at adaptive loci and corresponding phenotypic adaptive differences between the populations. Funk *et al*. (2012) stressed the importance of such differences in defining units for conservation, and this category of variation should thus be investigated.

The baseline assessment of genetic diversity and population structure is an important first step for the establishment of a genetic monitoring program for the African Penguin. As the African Penguin population continues to decline, *ex-situ* breeding programs are increasingly being considered as potentially valuable management tools and these programs are being implemented on a small scale by a number of rehabilitation centres, zoos, aquaria and conservation authorities in South Africa. The *ex-situ* population may be at risk of losing genetic variability over time due to genetic drift, as seen from the increase in the inbreeding coefficient in the offspring generation. In order to buffer the potential negative effects of genetic drift, management strategies should focus on increasing the number of individuals in the *ex-situ* populations. Natural or artificial gene flow between the captive and natural populations could enhance Ne and could potentially restore genetic variation within this small and vulnerable population. However, the release of captive animals should be conducted with caution. Factors such as variation at markers with adaptive significance, e.g. markers that are related to stress response, such as the Major Histocompatibility Complex, should be taken into consideration, to prevent outbreeding depression (Edmands, 2006).

References

Adams N.J. 1994. Patterns and impacts of oiling on African Penguins *Spheniscus demersus*: 1981-1991. *Biological Conservation,* 68: 35-41.

ADU (Avian Demography Unit). 2007. ADU Projects and Research. News from the colonies 1, Nov. 2007. University of Cape Town: ADU.

Allendorf F.W. 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology,* 5: 181-190.

Berthier P., Beaumont M.A., Cornuet J.M., Luikart G. 2002. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics,* 160: 741-751.

Boersma P.D. 2008. Penguins as marine sentinels. *BioScience,* 58: 597-607.

Caballero A. and Toro M.A. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics,* 3: 289-299.

Crawford R.J.M., David J.H.M., Williams A.J., Dyer B.M. 1989. Competition for space: recolonizing seals displace endangered, endemic seabirds of Namibia. *Biological Conservation,* 48: 59-72.

Crawford R.J.M., Underhill L.G., Raubenheimer C.M., Dyer B.M., Martin J. 1992. Top predators in the Benguela ecosystem: implications of their trophic position. *South African Journal of Marine Science,* 12: 675-687.

Crawford R.J.M., Altwegg R., Barham B.J., Barham P.J., Durant J.M., Dyer B.M., Geldenhuys D., Makhado A.B., Pichegru L., Ryan P.G., Underhill L.G., Upfold L., Visagie J., Waller L.J., Whittington P.A. 2011. Collapse of South Africa's penguins in the early 21st century. *South African Journal of Marine Science,* 33: 139-156.

Earl D.A. and von Holdt B.M. 2012. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Research,* 4: 359-361.

Edmands S. 2006. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology,* 16: 463-475.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology,* 14: 2611-2620.

Frankham R. 1996. Relationship of genetic variation to population size in Wildlife. *Conservation Biology,* 10: 1500-1508.

Frankham R. 2005. "Genetics and extinction." *Biological Conservation*, 126: 131-140.

Frankham R., Briscoe D.A., Ballou J.D. 2002. Introduction to conservation genetics. Cambridge University Press, New York, New York, USA.

Frankham R., Gilligan D.M., Morris D.R., Briscoe D.A. 2001. Inbreeding and extinction: effects of purging. *Conservation Genetics,* 2: 279-284.

Franklin I.R. and Frankham R. 1998. How large must populations be to retain evolutionary potential? *Animal Conservation,* 1: 69-73.

Frost P.G.H., Siegfried W.R., Cooper J. 1976. Conservation of the Jackass Penguin (*Spheniscus demersus* (*L.*). *Biological Conservation,* 9: 79-99.

Funk W.C., McKay J.K., Hohenlohe P.A., Allendorf F.W. 2012. Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution,* 27: 1-8.

Goudet J. 1995. FSTAT, a computer program to calculate F-statistics. *Journal of Heredity,* 86: 485-486.

Goudet J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from: http://www2.unil.ch/popgen/softwares/fstat.htm.

Hare M.P., Nunney L., Schwartz M.K., Ruzzante D.E., Burford M., Waples R.S., Ruegg K., Palstra F. 2011. Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology,* 25: 438-449.

Hartl D.L. and Clark S.G. 1989. Principles of population genetics. Sunderland, Massachussetts: Sinauer Press. Chapter 6. An elementary introduction to the theory.

IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2. <www.iucnredlist.org>. Downloaded on 03 May 2012.

Keller L.F., Arcese P., Smith J.M., Hochachka W.M., Stearns S. 1994. Selection against inbred song sparrows during a natural population bottleneck. *Nature,* 372: 356-357.

Labuschagne C., van Wyk A.M., Kotze A., Grobler P., Dalton D.L. 2013. Isolation and characterization of species-specific microsatellites loci in African Penguin (*Spheniscus demersus*). *Conservation Genetics Resources,* 5: 169-171.

Lacy R.C. 2009. Stopping evolution: Genetic management of captive populations. Pages 58-81 in: G. Amato, R. DeSalle, O.A. Ryder, and H.C. Rosenbaum. Conservation genetics in the age of genomics. Columbia University Press, New York.

Leberg P.L. 1992. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution*, 46: 477-494.

Leberg P. 2002. Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology,* 11: 2445-2449.

Luikart G., Allendorf F.W., Cornuet J.M., Sherwin W.B. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity,* 89: 238-247.

Luikart G., Ryman N., Tallmon D.A., Schwartz M.K., Allendorf F.W. 2010. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics,* 11: 355-373.

Mills L.S. and Allendorf F.W. 1996. The one-migrant-per-generation rule in conservation and management. *Conservation Biology,* 10: 1509-1518.

Morant P.D., Cooper J., Randall R.M. 1981. The rehabilitation of oiled Jackass Penguins *Spheniscus demersus* 1970-1980. In: Cooper, J. (Ed.). Proceedings of the Symposium of Birds of the Sea and Shore. Cape Town; *African Seabird Group*, 267-30.

Murata M. and Murakami M. 2014. Two distinct mtDNA lineages among captive African Penguins in Japan. *The Journal of Veterinary Medical Science*, 76(4): 559.

Nei M., Marayuma T., Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. *Evolution,* 29: 1-10.

Ovenden J., Peel D., Street R., Courtney A.J., Hoyle S.D., Peel S.L., Podlich H. 2007. The genetic effective and adult census size of an Australian population of tiger prawns (*Penaeus esculentus*). *Molecular Ecology,* 16: 127-38.

Park S.D.E. 2001. Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection [Ph.D. thesis] University of Dublin.

Peakall R. and Smouse P.E. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes,* 6: 288-295.

Petit R.J., El Mousadik A., Pons O. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology,* 12: 844-855.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotypes. *Genetics,* 155: 945-959.

Raymond M. and Rousset F. 1995. Genepop (version 1.2), population genetics software for exact tests and ecumenicism. *Journal of Heredity,* 86: 248-249.

Schlosser J.A., Garner T.W.J., Dubach J.M., McElligott A.G. 2003. Characterization of microsatellite loci in humboldt penguin (*Spheniscus humboldti*) and cross-amplification in other penguin species. *Molecular Ecology Notes,* 3: 62-64.

Shelton P.A., Crawford R.J.M., Cooper J., Brooke R.K. 1984. Distribution, population size and conservation of the Jackass Penguin *Spheniscus demersus*. *South African Journal of Marine Science,* 2: 217-257.

Spencer C.C., Neigel J.E., Leberg P.L. 2000. Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology,* 9: 1517-1528.

Tallmon D.A., Koyuk A., Luikart G., Beaumont M.A. 2008. ONeSAMP: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources,* 8: 299-301.

Theron N., Dalton D.L., Grobler P., Jansen R., Kotzé A. 2013. Molecular insights on re-colonization of the Limpopo Valley, South Africa, by southern ground hornbills. *Journal of Ornithology,* 154: 727-737.

Traill L.W., Brook B.W., Frankham R.R., Bradshaw C.J.A. 2010. Pragmatic population viability targets in a rapidly changing world. *Biological Conservation,* 143: 28-34.

Underhill L.G., Bartlett P.A., Baumann L., Crawford R.J.M., Dyer B.M., Gildenhuys A., Nel D.C., Oatley T.B., Thornton M., Upfold I., Williams A.J., Whittington P.A., Wolfaardt A.C. 1999. Mortality and survival of African Penguins *Spheniscus demersus* involved in the Apollo Sea oil spill: an evaluation of rehabilitation efforts. *Ibis,* 141: 29-37.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M., Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Resources,* 4: 535-538.

Waples R.S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics,* 121: 379-91.

Waples R.S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics,* 7: 167-184.

Waples R.S and Do C. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources,* 8: 753-756.

Woodworth L.M., Montgomery M.E., Briscoe D.A., Frankham R. 2002. Rapid genetic deterioration in captive populations: Causes and conservation implications. *Conservation Genetics,* 3: 277-288.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution,* 19: 395-420.

CHAPTER 7

Assessment of microsatellite and SNP markers for parentage assignment in *ex-situ* African Penguin (*Spheniscus* demersus) populations

Abstract

Captive management of *ex-situ* populations of endangered species is traditionally based on pedigree information derived from studbook data. However, molecular methods provide a powerful set of complementary tools to verify studbook records and to contribute to an improved understanding of the genetic status of captive populations. Here we compare the utility of Single Nucleotide Polymorphisms (SNPs) and microsatellites (MS) and two analytical methods for assigning parentage in ten families of captive African Penguins held in South African facilities. We found that SNPs performed better than microsatellites under both analytical frameworks, but a combination of all markers was most informative. A subset of combined SNP (n=14) and MS loci (n=10) provided robust assessments of parentage. Captive or supportive breeding programmes will play an important role in future African Penguin conservation efforts as a source of individuals for reintroduction. Cooperation among these captive facilities is essential to facilitate this process and improve management. This study provided us with a useful set of SNP and MS markers for parentage and relatedness testing among these captive populations. Further assessment of the utility of these markers over multiple (> three) generations and the incorporation of a larger variety of relationships among individuals (e.g. half-siblings or cousins) is strongly suggested.

Introduction

The growing role of captive institutions in the conservation of threatened species requires that they maintain sustainable and genetically diverse *ex-situ* populations that can meaningfully contribute to *in-situ* conservation (Lacy *et al.,* 2013). Molecular tools have the potential to complement and validate traditional studbook-based genetic management of captive populations, with the goal of reducing the negative effects of inbreeding and loss of genetic diversity (Putnam & Ivy, 2013). Complete pedigrees are required to effectively manage the genetic status of captive populations (Ivy & Lacy, 2010), but these are not always available, as the parentage of offspring is often uncertain (Putnam & Ivy, 2013).

The endangered African Penguin (*Spheniscus demersus*) is endemic to southern Africa, with 25 breeding colonies distributed along the coastline between central Namibia and St. Croix Island (Algoa Bay, South Africa). The population is declining despite multiple conservation interventions (IUCN Red List, BirdLife International, 2013) with an estimated 26,000 breeding pairs left (Crawford *et al.,* 2011). Declines have been attributed to excessive egg and guano harvesting (Shelton *et al.,* 1984), competition for food with seals (Crawford *et al.,* 1992) and commercial fisheries (Frost *et al.,* 1976), oil

spills (Morant *et al.,* 1981; Adams, 1994; Underhill *et al.,* 1999), loss of habitat and climate change affecting prey distribution (Boersma, 2008; Crawford *et al.,* 2011).

African Penguins breed well in captivity and are currently held in 11 zoos and aquariums across South Africa. *Ex-situ* populations serve a number of different roles in conservation efforts including public education, resources for scientific discovery and sources for supplementation or restoration of *in-situ* populations (Lacy, 2009). The latter has recently been identified as a potentially valuable conservation action, and looks likely to be implemented in the near future, necessitating a sound understanding of the genetic status of the captive populations. The African Association of Zoos and Aquaria (PAAZAB) established a regional studbook as part of their African Preservation Programme as part of the *ex-situ* management of this species. Similar to other studbooks, it uses the Single Population Analysis and Record Keeping System (SPARKS) developed by the International Species Information System (ISIS) and the PM2000 database programme (Pollack *et al.,* 2002). Studbook-based analyses indicated that 70.9% of the full pedigree information is known, and that the population mean kinship is 0.02 (African Penguin Regional Studbook, 2012). The use of molecular methods to confirm parentage and analyse relatedness among *ex-situ* individuals will complement studbook-based genetic management of the African Penguin captive population

Genealogical relationships among individuals in a population represent a simple concept in biology, but can be powerful when applied to answer evolutionary and ecological questions (Hauser *et al.,* 2011). Pedigree information plays a central role in the study of diverse ecological and evolutionary topics, such as sexual selection, patterns of dispersal and recruitment, quantitative genetic variation, mating systems and managing the conservation of populations of endangered species (Wang & Santure, 2009; Jones *et al.,* 2010). Molecular markers provide new possibilities in establishing genealogical relationships among individuals in populations where such information is difficult to collect from field observations (Pemberton, 2008).

Microsatellites (MS) have been the marker of choice for parental assignment and reconstruction, owing to their high polymorphic information content (PIC) and wide availability (Glowatski-Mullins *et al.,* 1995; Hauser *et al.,* 2011). However, these markers have several disadvantages including homoplasy, complex mutational patterns and data analysis may be affected by genotyping errors (Angers *et al.,* 2000; Hoffman *et al.,* 2005). Despite being bi-allelic, resulting in lower resolving power, single nucleotide polymorphisms (SNPs) are becoming increasingly popular (Baruch and Weller, 2008; Hauser *et al.,* 2011) due to their low genotyping error rate (< 0.1%), high through-put screening applications and the fact that SNPs are easier and cheaper to standardize between laboratories compared to microsatellites (Anderson and Garza, 2006).

In parallel to the advances in genetic markers, many statistical methods have been proposed to analyse marker data for pedigree information (Jones & Ardren, 2003). Jones *et al.,* (2010) categorised parentage analysis techniques into six categories namely Exclusion, Categorical Allocation, Fractional

Allocation, Full Probability parentage analysis, Parental Reconstruction and Sibship Reconstruction. Exclusion based methods compare the compatibility of offspring and parental genotypes with Mendelian inheritance, so that a putative parent is rejected as a true parent if both alleles at one locus mismatch with that of an offspring (Jones *et al.,* 2010). Exclusion methods are appealing since they are simple in concept and implementation, quick in computation and do not require allele frequency information (Wang, 2012). However, exclusion methods suffer from several weaknesses including false exclusion due to genotyping errors, valuable marker information is not fully utilized and exclusion rules are necessary, but insufficient for relationship inference (Jones *et al.,* 2010; Wang, 2012). A range of likelihood methods have been developed that seek to overcome these problems by determining probabilities of parentage assignment from simulations, Monte Carlo permutations or Bayesian approaches (Jones *et al.,* 2010). Likelihood-based methods employ Mendel's laws quantitatively to calculate the likelihoods of different candidate relationships among a set of individuals and choose the relationship that has the highest likelihood as the best inference (Wang, 2012).

In this study, we compare the power of parentage assignment of 31 SNPs and 12 MS markers in isolation and in combination in captive populations of African Penguins. Development of a marker set that accurately determines parentage will provide information on the relationships and relatedness among individuals (e.g. extra-pair mating), contribute to the management of captive African Penguins worldwide and additionally provide insight into mating systems in wild populations.

Materials and Methods:

Pedigrees and Sampling

Blood samples were collected from 33 African Penguins, which are housed in three captive facilities in South Africa: the Two Oceans Aquarium (Cape Town), the National Zoological Gardens of South Africa (Pretoria) and uShaka Sea World (Durban). All penguins are part of the permanent breeding population. Ten family-group pedigrees were constructed based on the regional studbook data (SPARKS) as shown in Figures 12 and 13 (A to J).

Molecular gender verification

For each individual, 30 µL of blood was collected on FTA paper. DNA was extracted using the Qiagen DNeasy® Blood and Tissue Kit. The extraction protocol as outlined in the manufacturer's protocol was followed. Chromo Helicase DNA-binding gene (CHD) based molecular sexing was conducted using the 2550F/2718R (Fridolfsson & Ellegren 1999) primer set. Promega GoTaq® Flexi DNA polymerase (Promega Corporation) was used for amplification in 25 µl reactions. The final reaction conditions were as follows: 1 X PCR buffer, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 5 pmol of each of the forward and reverse primer, 0.25 U Taq DNA polymerase and 10 - 20 ng genomic DNA template. A no template control as well as positive controls for a male and female bird of known sex was included.

The conditions for PCR amplification were as follows; initial denaturation for 2 minutes (min) at 95℃, 30 cycles for 30 seconds (sec) at 95℃, 30 sec at 50℃ and 2 min at 72℃, followed by final extension at 72℃ for 10 min. The PCR reaction was carried out in the BOECO TC-PRO Thermal Cycler. Amplicons were separated by electrophoresis in a 2% agarose gel for 45 min at 100 V in 1× Tris-borate- EDTA buffer. A single band pattern was considered male (CHD-Z) while the two band pattern was considered female (CHD-W/CHD-Z).

Microsatellite genotyping

A total of 12 microsatellite markers were typed as described in Schlosser *et al.,* (2003) and Labuschagne *et al.,* (2013). Promega GoTaq® Flexi DNA polymerase (Promega Corporation) was used for amplification in 12.5 µl reactions. The final reaction conditions were as follows: 1 X PCR buffer, 1 mM $MgCl_2$, 200 µM of each dNTP, 10 pmol of each of the forward and reverse primer, 1 U *Taq* DNA polymerase and 50 ng genomic DNA template. The PCR reaction was carried out in the BOECO TC-PRO Thermal Cycler. The conditions for PCR amplification were as follows: 5 min at 95℃ denaturation, 30 cycles for 30 sec at 95℃, 30 sec at 50-60℃ and 30 sec at 72℃, followed by extension at 72℃ for 40 min. PCR products were pooled and run against an Genescan™ 500 LIZ™ internal size standard on an ABI 3130 Genetic Analyzer (Applied Biosystems, Inc.). Samples were genotyped using GeneMapper v.4.0 (Applied Biosystems, Inc.).

SNP genotyping

A total of 31 SNP markers, distributed across 11 loci were typed as described in Labuschagne *et al.,* (2013). Amplification was achieved using Dream Taq™ Green PCR Master Mix (2×) supplied by Thermo Scientific, Lithuania The PCR mix for each locus contained 12.5 µl of 2× Dream TaqTM PCR master mix (10× Dream Taq™ buffer, dATP, dCTP, dGTP and dTTP, 0.4 mM each, 4 mM $MgCl_2$ and 1.25 U Dream Taq™ polymerase), 1 µl [10 µM] of each primer, 50 ng of template DNA and nuclease free water to reach a final volume of 25 µl. Sequencing of resulting amplicons was conducted by Inqaba Biotechnical Industries (Pty) Ltd using the ABI Big Dye V3.1 kit and the ABI 3500XL genetic analyser. Sequence data was screened and aligned using the Main workbench from CLC Bio (Denmark).

Parentage analysis

Parentage assignment was evaluated with likelihood and exclusion based approaches, using the MS and SNP data sets individually and combined. To assign parentage using a likelihood approach we used the software program CERVUS v3.03 (Kalinowski *et al.,* 2007). The program uses multilocus parental exclusion probabilities (Selvin, 1980) and pair-wise likelihood to assign parent pairs to offspring. CERVUS calculates the log-likelihood of each candidate parent being the true parent relative to an arbitrary individual and then calculates the difference between the two most likely parents (Delta, Δ). Critical values of Δ are determined by computer simulation. Using the real data for

allele frequencies, simulation parameters were set at 10 000 offspring, with 100% of candidate parents sampled and a total proportion of loci typed over all individuals of 0.99, mistyping error rates = 0.01 and likelihood calculation error rates = 0.01, permitting 2 unscored loci. Strict confidence was set to 95% while the relaxed confidence level was 80%. CERVUS was also used to calculate the summary statistics including allele number at each locus (k), Observed heterozygosity (H_{obs}), Expected heterozygosity (H_{exp}), Polymorphic information content (PIC), Average non-exclusion probability for one candidate parent (NE-1P), average non-exclusion probability for one candidate parent given the genotype of a known parent of the opposite sex (NE-2P) and Significance of deviation from Hardy-Weinberg equilibrium (HW). Parentage assignment using exclusion was performed in PARFEX v1.0 (Sekino & Kakehi, 2012). The exclusion method examines incompatibilities between putative parents and offspring genotypes based on Mendelian principles. Parentage assignments were made for zero, one and two mismatches. PARFEX was further used to calculate a minimum marker set required for optimal parentage using the given data set through the PFX_Mchoice macro. The known parental genotypes are used to simulate offspring genotypes, which are then subjected to exclusion based parentage testing with successive one-by-one addition of higher-ranked markers from which the cumulative success rate of parentage allocation is obtained (Sekino & Kakehi, 2012). Markers are ranked through one of three statistics (proportion of unique alleles, polymorphic information content (PIC) and exclusion probability) and the success rate of parentage allocation defined as the number of simulated offspring whose true parental pair is unambiguously identified divided by the total number of offspring (Sekino & Kakehi, 2012).

Results

The 33 individuals used in this study represented 17 males and 16 females according to the studbook data. Molecular sexing using the CHD gene verified the gender of all individuals. All samples were successfully genotyped, with the exception of one MS marker for one sample while the SNP dataset had 5 SNPs missing, affecting three samples. In total 62 alleles were found over all 12 MS loci, with a mean PIC of 0.54 (Table 14). Thirty-one SNPs were identified across 11 loci with a mean PIC of 0.23 (Table 15). For the 33 samples collected, 25 parent/offspring relationships can be made from the studbook data (Figures 12 and 13). Among these relationships, nine are sire/dam/offspring trios (Figure 12-C/F/G/H/I/J), seven single parent/offspring pairs (Figure 12-A/B/C/D/E), four sets have full siblings (Figure 12-A/B/H/J) and two family groups include previous generations (Figure 12-C/H). Using the MS data set in PARFEX (Table 17), only 11 out of the 25 relationships could be correctly assigned using the exclusion method (Figure 14). The SNP data set performed better with 14 out of the 25 relationships being assigned. When combining both data sets 20 of the relationships could be assigned using the exclusion method (Figures 12 and 14). Applying the MS data in PARFEX correct parents were mostly excluded due to a high number of mismatches, while in the SNP data set there were often not enough differences to discern false parents from true parents (Figure 13; Tables 16 and 14). Using the MS data set in CERVUS (Figure12; Table 16), 21 of the relationships could be correctly assigned when using a likelihood method. The SNP data set assigned 22 correct relationships with the same methodology (Figure 12). When combining both data sets in CERVUS all

25 relationships were correctly assigned (Figure 12). Incorrect assignments with the MS data were limited to three family groups (Figure 12: B, D and E), all single parent-offspring groups. All four assignments had low LOD scores (Figure 12). Incorrect assignments with the SNP data were limited to two family groups (Figure 12: I and J). The incorrect assignment in group I was made with 95% confidence, while both assignments in group J had 80% confidence. In contrast with the CERVUS MS data, the correct parent was assigned to PNN156 in group B. Dam PNN149 was the closest match although it contained two mismatches (Table 17). The remaining incorrect CERVUS assignments were also incorrect in PARFEX. A similar disparity was noted in the SNP data set where both parents are correctly assigned in group J for offspring PNN96 using PARFEX (Figure 13). The incorrect assignments for group I and J in CERVUS were non-excluded in PARFEX. Several parents could be assigned without mismatches (Table 17). PFX_Mchoice only reached 99% accumulative success rate when ranking markers through exclusion probability or proportion of unique alleles. Using exclusion probability, 99% accumulative success rate was reached with 15 markers (10 MS and 5 SNPs). Using only these 15 markers 22 out of the 25 relationships could be assigned correctly. Ranking markers through the proportion of unique alleles, 99% accumulative success was achieved with 22 markers (11 MS and 11 SNPs). Using the 22 marker subset, 23 of the 25 relationships could be assigned accurately. Ranking markers using PIC resulted in a 100% accumulative success rate with 34 markers (10 MS and 14 SNPs) (Figure 15). All 25 relationships were assigned correctly when using these markers.

Table 14: Parameters of genetic information content of 12 microsatellite loci estimated from *ex-situ* population of African Penguin. K = number of alleles; N = number of samples; HObs = Observed heterozygosity; HExp = Expected heterozygosity; PIC = polymorphic information content; Ne-1P = Average non-exclusion probability for one candidate parent; Ne-2P = average non-exclusion probability for one candidate parent given the genotype of a known parent of the opposite sex and HW = Hardy-Weinberg equilibrium.

Locus	k	N	HObs	HExp	PIC	Ne-1P	Ne-2P	HW
G2-2	5	33	0.697	0.695	0.627	0.740	0.577	NS
SH1CA9	10	33	0.788	0.779	0.746	0.593	0.409	NS
SH2CA21	7	33	0.667	0.740	0.688	0.672	0.495	NS
B3-2	3	33	0.152	0.172	0.161	0.986	0.915	ND
G3-6	7	33	0.636	0.730	0.669	0.697	0.526	ND
PNN01	4	33	0.727	0.675	0.595	0.773	0.621	NS
PNN03	5	33	0.394	0.424	0.383	0.909	0.773	ND
PNN06	4	33	0.636	0.656	0.578	0.786	0.634	NS
PNN08	4	33	0.697	0.656	0.584	0.781	0.624	NS
PNN09	6	33	0.758	0.769	0.717	0.645	0.468	ND
PNN12	5	32	0.875	0.730	0.671	0.695	0.523	ND
PNN05	2	33	0.121	0.116	0.107	0.994	0.946	ND
Mean	**5.17**			**0.5952**	**0.5439**			

Table 15: Parameters of genetic information content of 31 Single Nucleotide Polymorphisms estimated from *ex-situ* population of African Penguin. K = number of alleles; N = number of samples; HObs = Observed heterozygosity; HExp = Expected heterozygosity; PIC = polymorphic information content; Ne-1P = Average non-exclusion probability for one candidate parent; Ne-2P = average non-exclusion probability for one candidate parent given the genotype of a known parent of the opposite sex and HW = Hardy-Weinberg equilibrium.

Locus	SNP	k	N	H_{Obs}	H_{Exp}	PIC	Ne-1P	Ne-2P	HW
PG Ne 15	P110 NE-15-1	2	33	0.061	0.060	0.057	0.998	0.971	ND
	P110 NE-15-2	2	33	0.303	0.339	0.278	0.944	0.861	ND
PG Ne 12	P110 NE-12-1	2	31	0.032	0.032	0.031	0.999	0.984	ND
	P110 NE-12-2	2	31	0.129	0.228	0.200	0.975	0.900	ND
PG Ne 11	P110 NE-11-1	2	33	0.333	0.416	0.326	0.916	0.837	ND
	P110 NE-11-2	2	33	0.394	0.357	0.290	0.938	0.855	ND
	P110 NE-11-3	2	33	0.273	0.239	0.208	0.972	0.896	ND
PG NE 1	P110 NE 1	2	33	0.485	0.451	0.346	0.901	0.827	ND
PG EVE 5	P110 EVE 5-1	2	33	0.333	0.416	0.326	0.916	0.837	ND
	P110 EVE 5-2	2	33	0.061	0.060	0.057	0.998	0.971	ND
	P110 EVE 5-3	2	33	0.485	0.429	0.333	0.911	0.833	ND
	P110 EVE 5-4	2	33	0.515	0.441	0.340	0.906	0.830	ND
C6 306	P110 C6-306-1	2	33	0.030	0.030	0.029	1.000	0.985	ND
	P110 C6-306-2	2	33	0.273	0.282	0.239	0.961	0.880	ND
B1 534	P110 B1-534-1	2	33	0.424	0.403	0.318	0.921	0.841	ND
	P110 B1-534-2	2	33	0.303	0.261	0.224	0.967	0.888	ND
PG L	P110-L-1	2	33	0.273	0.239	0.208	0.972	0.896	ND
	P110-L-2	2	33	0.242	0.373	0.300	0.933	0.850	ND
	P110-L-3	2	33	0.515	0.478	0.360	0.889	0.820	ND
	P110-L-4	2	33	0.576	0.506	0.374	0.876	0.813	NS
	P110-L-5	2	33	0.152	0.142	0.130	0.990	0.935	ND
	P110-L-6	2	33	0.091	0.088	0.083	0.996	0.958	ND
	P110-L-7	2	33	0.121	0.168	0.152	0.986	0.924	ND
	P110-L-8	2	33	0.242	0.216	0.190	0.977	0.905	ND
PG I	P110 I-1	2	33	0.424	0.373	0.300	0.933	0.850	ND
	P110 I-2	2	33	0.364	0.302	0.253	0.956	0.873	ND
PG A	P110-A1	2	32	0.094	0.091	0.085	0.996	0.957	ND
PG EVE 10	P110 EVE10-1	2	33	0.455	0.416	0.326	0.916	0.837	ND
	P110 EVE10-2	2	33	0.394	0.388	0.309	0.927	0.845	ND
	P110 EVE10-3	2	33	0.152	0.142	0.130	0.990	0.935	ND
	P110 EVE10-4	2	33	0.333	0.321	0.266	0.950	0.867	ND
Mean					**0.280**	**0.228**			

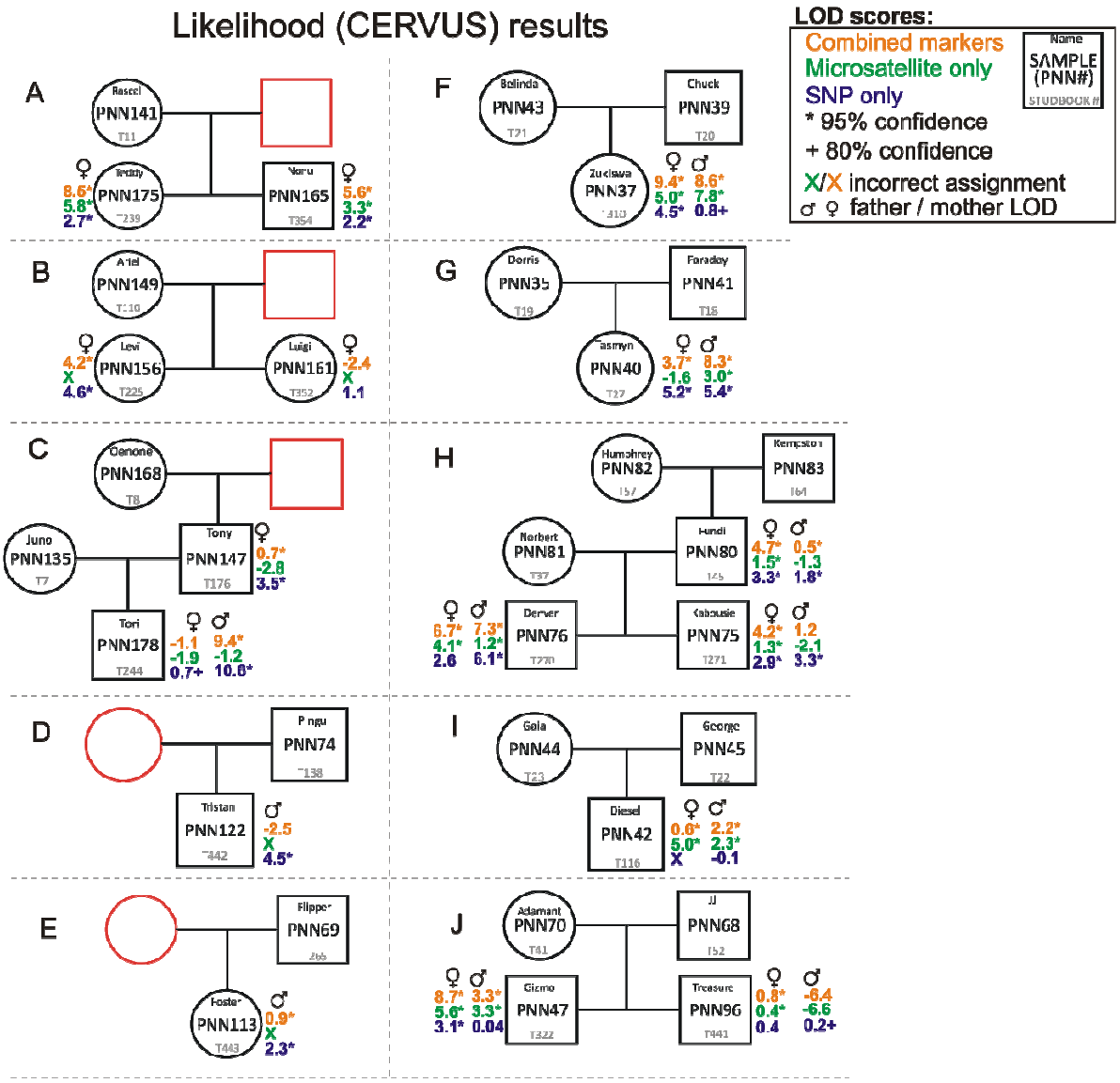


Figure 12: Studbook-based pedigrees of ten families of African Penguins (*Spheniscus demersus*) based on data from Single Population Analysis and Record Keeping System (SPARKS) superimposed with parentage assignment data from CERVUS (likelihood). Squares indicate males, circles indicate females, red shapes indicate un-sampled individuals.

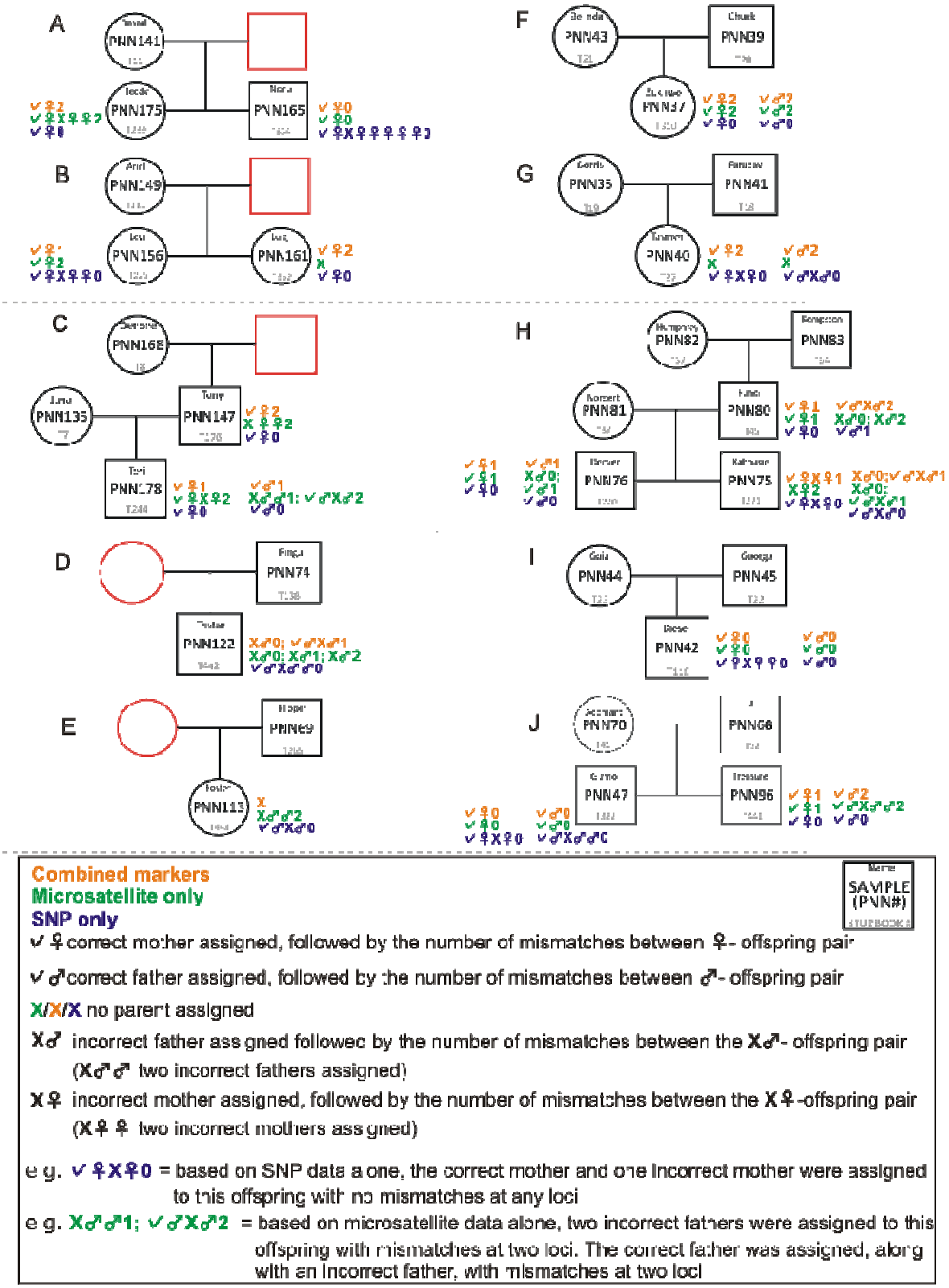


Figure 13: Studbook-based pedigrees of ten families of African Penguins (*Spheniscus demersus*) based on data from SPARKS superimposed with parentage assignment data from PARFEX (exclusion). Squares indicate males, circles indicate females, red shapes represent un-sampled individuals

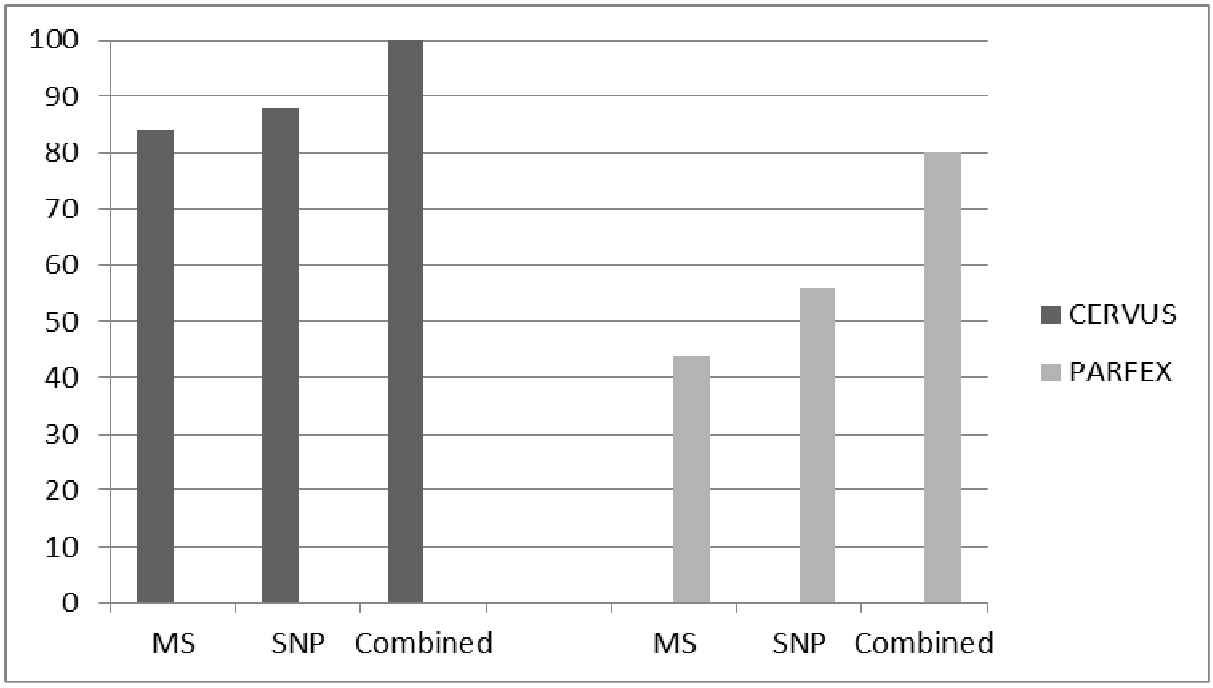


Figure 14: Percentage correct parent-offspring assignments for all datasets using CERVUS and PARFEX

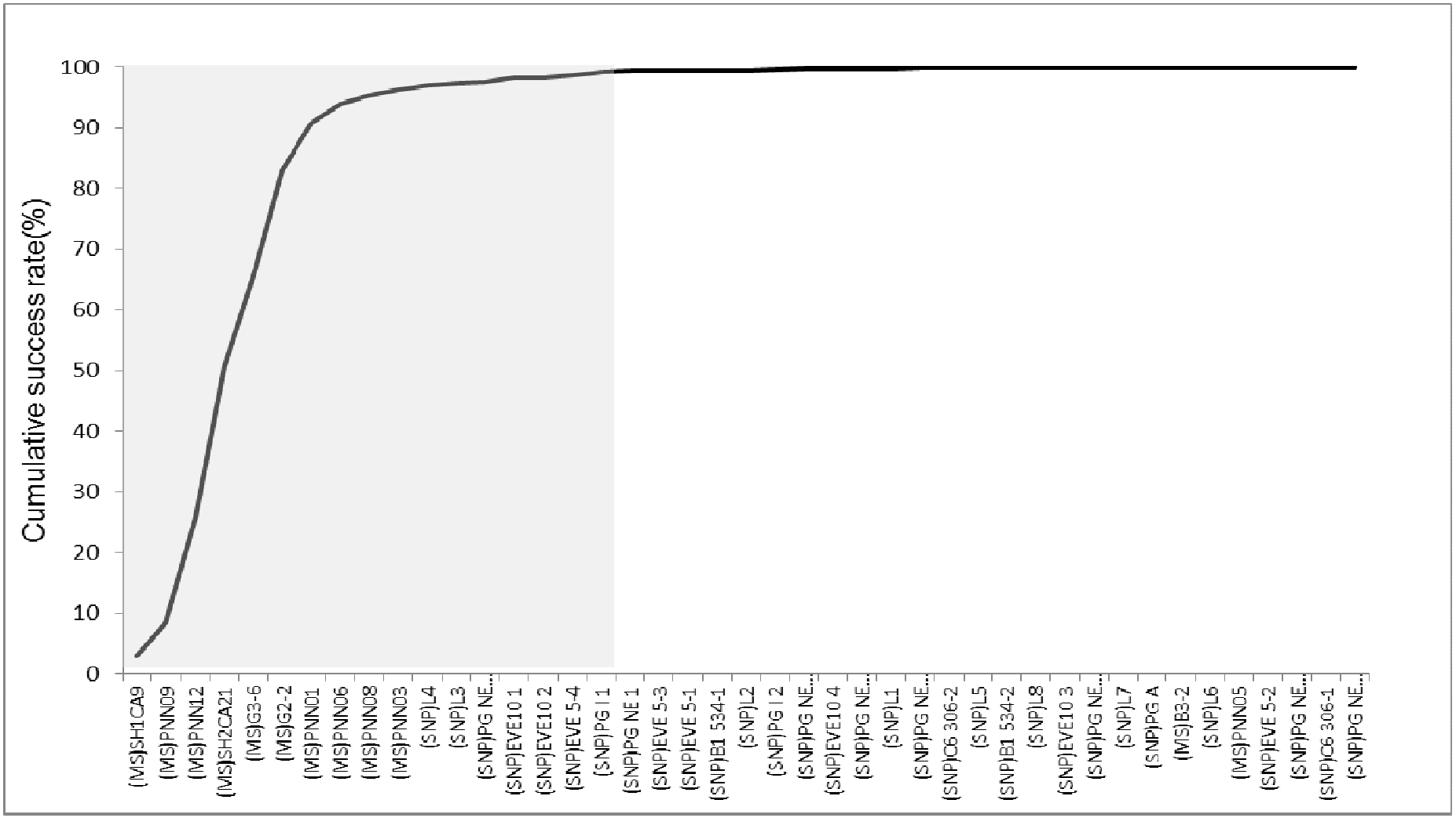


Figure 15: The cumulative success rate of parentage assignment based on exclusion with markers ranked on PIC value. The grey area encompasses all loci required to reach a 100% probability of assigning a correct parent-offspring relationship.

Table 16: CERVUS parentage assignments. Brackets indicate correct assignment; *=95% confidence; +=80% confidence; Incorrect assignments marked in grey.

	Combined Data Sets				Microsatellites				SNPs			
Offspring	**Candidate mother**	**Pair LOD score**	**Candidate father**	**Pair LOD score**	**Candidate mother**	**Pair LOD score**	**Candidate father**	**Pair LOD score**	**Candidate mother**	**Pair LOD score**	**Candidate father**	**Pair LOD score**
PNN147	(PNN168)*	0.65	n/a	n/a	(PNN168)	-2.81	n/a	n/a	(PNN168)*	3.46	n/a	n/a
PNN156	(PNN149)*	4.18	n/a	n/a	PNN135	-3.94	n/a	n/a	(PNN149)*	4.60	n/a	n/a
PNN165	(PNN141)*	5.57	n/a	n/a	(PNN141)*	3.33	n/a	n/a	(PNN141)*	2.24	n/a	n/a
PNN161	(PNN149)	-2.36	n/a	n/a	PNN**168**	-7.12	n/a	n/a	(PNN149)	1.10	n/a	n/a
PNN175	(PNN141)*	8.49	n/a	n/a	(PNN141)*	5.79	n/a	n/a	(PNN141)*	2.70	n/a	n/a
PNN113	n/a	n/a	(PNN69)*	0.85	n/a	n/a	PNN80	-2.08	n/a	n/a	(PNN69)*	2.30
PNN122	n/a	n/a	(PNN74)	-2.48	n/a	n/a	PNN80	-3.84	n/a	n/a	(PNN74)*	4.46
PNN37	(PNN43)*	9.44	(PNN39)*	8.60	(PNN43)*	4.97	(PNN39)*	7.76	(PNN43)*	4.48	(PNN39)+	0.84
PNN40	(PNN35)*	3.68	(PNN41)*	8.34	(PNN35)	-1.55	(PNN41)*	2.97	(PNN35)*	5.24	(PNN41)*	5.37
PNN42	(PNN44)*	0.60	(PNN45)*	2.21	(PNN44)*	5.04	(PNN45)*	2.34	PNN168*	1.73	(PNN45)	-0.13
PNN75	(PNN81)*	4.19	(PNN80)	1.18	(PNN81)*	1.34	(PNN80)	-2.13	(PNN81)*	2.86	(PNN80)*	3.31
PNN80	(PNN82)*	4.70	(PNN83)*	0.48	(PNN82)*	1.45	(PNN83)	-1.32	(PNN82)*	3.25	(PNN83)*	1.80
PNN47	(PNN70)*	8.72	(PNN68)*	3.30	(PNN70)*	5.59	(PNN68)*	3.26	(PNN70)*	3.13	PNN69+	0.94
PNN76	(PNN81)*	6.69	(PNN80)*	7.31	(PNN81)*	4.11	(PNN80)*	1.18	(PNN81)	2.58	(PNN80)*	6.13
PNN96	(PNN70)*	0.78	(PNN68)	-6.39	(PNN70)*	0.39	(PNN68)	-6.60	PNN44+	1.42	(PNN68)+	0.21
PNN178	(PNN135)	-1.14	(PNN147)*	9.42	(PNN135)	-1.85	(PNN147)	-1.15	(PNN135)+	0.72	(PNN147)*	10.57

Table 17: Exclusion-based (PARFEX) parentage assignments. Brackets indicate true parent; incorrect assignments marked in grey.

		Combined Data Sets		Microsatellites		SNPs	
Offspring	**Mismatches**	**Candidate mother**	**Candidate father**	**Candidate mother**	**Candidate father**	**Candidate mother**	**Candidate father**
PNN147	0		n/a		n/a	(PNN168)	n/a
	1		n/a		n/a		n/a
	2	(PNN168)	n/a	PNN35, PNN135	n/a	PNN43	n/a
PNN156	0		n/a		n/a	PNN81, PNN135, (PNN149)	n/a
	1	(PNN149)	n/a		n/a	PNN44, PNN141,	n/a
	2	PNN135	n/a	(PNN149)	n/a		n/a
PNN165	0	(PNN141)	n/a	(PNN141)	n/a	PNN35, PNN44, PNN81, PNN135, (PNN141), PNN168	n/a
	1		n/a		n/a		n/a
	2	PNN135	n/a		n/a		n/a
PNN161	0		n/a		n/a	(PNN149)	n/a
	1		n/a		n/a	PNN35, PNN82	n/a
	2	(PNN149)	n/a		n/a	PNN43, PNN81, PNN135, PNN168	n/a
PNN175	0		n/a		n/a	(PNN141)	n/a
	1		n/a		n/a		n/a
	2	(PNN141)	n/a	PNN43, PNN135, (PNN141)	n/a		n/a

Table 17 Continued: Exclusion-based (PARFEX) parentage assignments. Brackets indicate true parent; incorrect assignments marked in grey.

		Combined Data Sets	Microsatellites	SNPs			Combined Data Sets
Offspring	**Mismatches**	**Candidate mother**	**Candidate father**	**Candidate mother**	**Offspring**	**Mismatches**	**Candidate mother**
PNN113	0	n/a		n/a		n/a	PNN68, (PNN69)
	1	n/a		n/a		n/a	
	2	n/a		n/a	PNN45, PNN80	n/a	PNN39, PNN41, PNN45
PNN122	0	n/a	PNN45	n/a	PNN45	n/a	PNN39, PNN45, (PNN74)
	1	n/a	PNN39, (PNN74)	n/a	PNN39	n/a	
	2	n/a		n/a	PNN80	n/a	
PNN37	0					(PNN43)	(PNN39)
	1					PNN44, PNN168	
	2	(PNN43)	(PNN39)	(PNN43)	(PNN39)	PNN70, PNN81	PNN68, PNN69
PNN40	0					(PNN35), PNN82	(PNN41), PNN45
	1					PNN44, PNN168	PNN68
	2	(PNN35)	(PNN41)				PNN39, PNN74
PNN42	0	(PNN44),	(PNN45)	(PNN44)	(PNN45)	PNN43, (PNN44), PNN168	(PNN45)
	1					PNN81, PNN135	PNN68
	2	PNN81;	PNN41	PNN81, PNN82	PNN68	PNN70, PNN82	PNN69, PNN83, PNN147

Table 17 Continued: Exclusion-based (PARFEX) parentage assignments. Brackets indicate true parent; incorrect assignments marked in grey.

		Combined Data Sets	Microsatellites	SNPs			Combined Data Sets
Offspring	**Mismatches**	**Candidate mother**	**Candidate father**	**Candidate mother**	**Offspring**	**Mismatches**	**Candidate mother**
PNN75	0		PNN83		PNN83	(PNN81), PNN149	(PNN80), PNN83
	1	(PNN81); PNN68	(PNN80)		PNN147, (PNN80)	PNN82, PNN135	PNN45, PNN68
	2			PNN43		PNN44	PNN39, PNN41, PNN45, PNN74
PNN80	0				PNN68	(PNN82)	
	1	(PNN82)		(PNN82)		PNN81, PNN168	(PNN83)
	2		(PNN83), PNN68		PNN74	PNN43, PNN44, PNN135	PNN68
PNN47	0	(PNN70)	(PNN68)	(PNN70)	(PNN68)	(PNN70), PNN168	PNN45, (PNN68), PNN69
	1	PNN81	PNN69			PNN43, PNN44	
	2	PNN43, PNN135, PNN168	PNN45	PNN82, PNN168	PNN39, PNN45, PNN74, PNN83	PNN81, PNN82, PNN135, PNN141	PNN39, PNN41
PNN76	0				PNN83	(PNN81)	(PNN80)
	1	(PNN81)	(PNN80)	(PNN81)	(PNN80)	PNN82, PNN135, PNN149	PNN68
	2	PNN82	PNN68			PNN44	PNN41, PNN45, PNN69, PNN83

Table 17 Continued: Exclusion-based (PARFEX) parentage assignments. Brackets indicate true parent; incorrect assignments marked in grey.

		Combined Data Sets	Microsatellites	SNPs			Combined Data Sets
Offspring	**Mismatches**	**Candidate mother**	**Candidate father**	**Candidate mother**	**Offspring**	**Mismatches**	**Candidate mother**
PNN96	0					(PNN70)	(PNN68)
	1	(PNN70)		(PNN70)		PNN44, PNN81, PNN168	PNN45, PNN74
	2		(PNN68)		(PNN68), PNN69, PNN74	PNN82, PNN135, PNN141, PNN149	PNN69
PNN178	0					(PNN135)	(PNN147)
	1	(PNN135)	(PNN147),		PNN45;PNN68	PNN34, PNN44, PNN81, PNN141, PNN149, PNN168	PNN41, PNN45, PNN68, PNN69
	2		PNN68	(PNN135); PNN141	(PNN147); PNN83		

Discussion

Since inaccuracies in the studbook can have implications on future genetic and demographic analysis and management of the captive population, a suitable validated marker set for genetic parentage verification is an important tool for captive management (Ivy & Lacy, 2010). Such a marker set may not only exclude incorrectly recorded parents, but also help in assigning the correct individuals if sampled. We have described and verified a set of genetic markers for ascertaining parentage and sibling relationships in African Penguins. Few published studies have investigated parentage or paternity in penguins, and to our knowledge none have used SNP markers. Seven MS markers (including one, B3-2, employed in the present study) yielded a general exclusion probability (mother known) of 0.99 for Little Penguins (Billing *et al.,* 2007), and eight MS markers (including one used in the present study - Sh1Ca9) yielded paternity exclusions of 0.94 to 0.99 for captive Adelie Penguins (Sakaoka *et al.,* 2014).

Concerning the discrimination power of both types of markers, MS and SNP, as expected the MS markers with multiple alleles possible at each locus had an overall higher PIC value. Both marker sets had 62 independent alleles. However, with more loci, the optimized SNP marker set performed better than the MS marker set using both the exclusion and likelihood parental assignment methods. This study has indicated that the number of loci and their heterozygosity level, may influence the power of markers for parentage exclusion approaches more than the number of independent alleles (Morin *et al.,* 2004; Hauser *et al.,* 2011). The power of molecular markers is also influenced by genotyping error (Kalinowski *et al.,* 2007). The generally low error rate for SNPs is a definite advantage for parentage over the higher rates reported for MS markers (Walling *et al.,* 2010; Hauser *et al.,* 2011). However, since each locus adds linearly to the multilocus error, but provides diminishing information for parentage, even low error rates may become problematic as the number of loci screened becomes very large (Christie, 2010; Hauser *et al.,* 2011). The optimum number of loci should therefore be determined in preliminary experiments where the number of SNPs required may be less than commonly assumed (Christie, 2010; Hauser *et al.,* 2011). In the current study we used PFX_Mchoice to establish if a smaller subset of markers would achieve the same assignment power over the full combined marker set. A subset of 34 markers consisting of 10 MS markers and 14 SNP markers were identified that could accurately allocate all 25 parent-offspring relationships identified. Such *a priori* knowledge about a minimum set of markers providing a high resolution of parentage assignment helps reduce the experimental cost and labour involved in the subsequent parentage testing.

Since parentage inference is not concerned with inference of evolutionary history, ascertainment bias through discovery in particular populations or genomic regions, does not bias the results of parentage inference (Anderson and Garza, 2006). In effect, such ascertainment typically leads to an overrepresentation of SNPs at intermediate allele frequencies, an advantage in parentage inference (Anderson and Garza, 2006). Those SNP markers with minor allele frequencies of 0.5 provide the most power for parentage inference, although little additional power is gained above frequencies of 0.4 (Anders and Garza, 2006). Choosing SNP markers with allele frequencies above 0.2 can achieve

higher assignment power with fewer loci. Among the current 34 SNP markers only 16 have heterozygosity above 0.3. Replacing the markers falling below these ranges with new marker with higher ranges may greatly improve the number of loci *vs* assignment power ration as well as provide a SNP only marker set that takes full advantage of SNP marker benefits over MS markers. Advantages including low error rates, ease of typing, low-cost high-throughput genotyping and SNP genotypes that are easily standardized across laboratories, all important factors for a multi institutional studbook.

Conclusion

The aim of this study was to generate molecular genetic information to verify/complement studbook-based pedigree data from *ex-situ* populations of African Penguins. In addition, we compared the relative and combined utility of MS and SNP markers for parentage assignment. We found that a combined subset of these two types of markers attained a > 99% correct cumulative parentage assignment probability. Information derived from this "optimal" marker set will be useful for future captive management of African Penguins and for investigating their mating system (e.g. the frequency of extra-pair copulation and fertilization and mate-choice) in wild populations.

References

Adams N.J. 1994. Patterns and impacts of oiling on African Penguins *Spheniscus* demersus: 1981-1991. *Biological Conservation* 68: 35-41.

Anderson E.C. and Garza J.C. 2006. The power of single nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172: 2567-2582.

Baruch E. and Weller J.I. 2008. Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics* 39: 474-479.

Billing T.M., Guay P.J., Peucker A.J., Mulder R.A. 2007. Isolation and characterization of polymorphic microsatellite loci for the study of paternity and population structure in the little penguin *Eudyptula minor. Molecular Ecology Notes*. 7: 425-427.

BirdLife International. 2013. *Spheniscus demersus*. The IUCN Red List of Threatened Species. Version 2014.2. Available from www.iucnredlist.org.

Boersma P.D. 2008. Penguins as marine sentinels. *BioScience* 58: 597-607.

Christie M.R. 2010. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. *Molecular Ecology Resources*. 10: 115-128.

Crawford R.J.M., Underhill L.G., Raubenheimer C.M., Dyer B.M., Martin J. 1992. Top predators in the Benguela ecosystem: implications of their trophic position. *South African Journal of Marine Science*. 12: 675-687.

Crawford R.J.M., Altwegg R., Barham B.J., Barham P.J., Durant J.M., Dyer B.M., Whittington P.A. 2011. Collapse of South Africa's penguins in the early 21st century. *African Journal of Marine Science.* 33: 139-156.

Fridolfsson A.K. and Ellegren H. 1999. A simple and universal method for molecular sexing of non-ratite birds. *Journal of Avian Biology*: 30: 116-121.

Frost P.G.H., Siegfried W.R. Cooper J. 1976. Conservation of the Jackass Penguin (*Spheniscus* demersus (L.)). *Biological Conservation*. 9: 79-99.

Glowatzki-Mullins M.L., Gaillard C., Wigger G., Fries R. 1995. Microsatellite based parentage control in cattle. *Animal Genetics*. 26: 7-12.

Hauser L., Baird M., Hilborn R., Seeb L.W., 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: 150-161.

IUCN. 2011. IUCN Red List of Threatened Species. Version 2011.2. Available from www.iucnredlist.org.

Jones A.G. and Ardren W.R. 2003. Methods of parentage analysis in natural populations. *Mol. Ecol*. 12: 2511-2523.

Jones A.G., Small C.M., Paczolt K.A., Ratterman N.L. 2010. A practical guide to methods of parentage analysis. *Molecular Ecology Resources*. 10: 6-30.

Kalinowski S.T., Taper M.L., Marshall T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol*. 16: 1099-1106.

Lacy R.C. 2009. Stopping evolution: Genetic management of captive populations. Pages 58-81 in: Amato G, DeSalle R, Ryder OA, Rosenbaum HC. Conservation genetics in the age of genomics. Columbia University Press, New York.

Lacy R.C., Traylor-Holzer K., Ballou J.D. 2013. Managing for True Sustainability of Species. *WAZA magazine*. 14: 10-14.

Morant P.D., Cooper J., Randall R.M. 1981. The rehabilitation of oiled Jackass Penguins *Spheniscus* demersus 1970-1980. In: Cooper, J. (Ed.). Proceedings of the Symposium of Birds of the Sea and Shore. Cape Town; *African Seabird Group* : 267-30.

Morin P.A., Luikart G., Wayne R.K. SNP-Workshop_Grp. 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*. 19: 208-216.

Pemberton J.M. 2008. Wild pedigrees: the way forward. *Proc. Biol. Sci*. 275: 613-621.

Ivy J.A. and Lacy R.C. 2010. Using Molecular methods to improve the genetic management of captive breeding programs for threatened species. Chapter 11 in: Molecular Approaches in Natural Resource Conservation and Management. De Woody J.A., Bickham J.W., Michler C.H., Nichols K.M., Rhodes G.E., Woeste K.E. (Eds). Cambridge University Press.

Putnam A.S., Ivy J.A. 2013. Kinship-Based Management Strategies for captive breeding programs when pedigrees are unknown or uncertain. *Journal of Heredity*. 105(3): 303-11.

Sakaoka K., Suzuki I., Kasugai N., Fukumoto Y. 2014. Paternity Testing Using Microsatellite DNA Markers in Captive Adélie Penguins (*Pygoscelis adeliae*). *Zoo Biology*. 1999: 1-8.

Sekino M. and Kakehi S. 2012. PARFEX v1.0: an EXCELTM-based software package for parentage allocation. *Conservation Genetics Resources*, 4: 275-278.

Selvin S. 1980. Probability of nonpaternity determined by multiple allele codominant systems. *Am J Hum Genet.* 32: 276-278.

Shelton P.A., Crawford R.J.M., Cooper J., Brooke R.K. 1984. Distribution, population size and conservation of the Jackass Penguin *Spheniscus demersus*. *S. Afr, J. Mar. Sci*. 2: 217-257.

Underhill L.G., Bartlett P.A., Baumann L., Crawford R.J.M., Dyer B.M., Gildenhuys A., Nel D.C., Oatley T.B., Thornton M., Upfold L., Williams A.J., Whittington P.A., Wolfaardt A.C. 1999. Mortality and survival of African Penguins *Spheniscus demersus* involved in the Apollo Sea oil spill: an evaluation of rehabilitation efforts. *Ibis,* 141: 29-37.

Walling C.A., Pemberton J.M., Hadfield J.D., Kruuk L.E.B. 2010. Comparing parentage inference software: reanalysis of a red deer pedigree. *Molecular Ecology*. 19: 1914-1928.

Wang J. 2012. Computationally efficient sibship and parentage assignment from mutilocus marker data. *Genetics*. 191: 183-194.

Wang J. and Santure A.W. 2009. Parentage and sibship inference from mutilocus genotype data under polygamy. *Genetics*. 181: 1579-1594.

PART III: DEVELOPMENT AND APPLICATION OF CONSERVATION GENETICS TECHNIQUES IN RHINOCEROS

CHAPTER 8

A targeted gene approach to SNP discovery in the White Rhino (*Ceratotherium simum*)

Abstract

We report the characterization of 10 single nucleotide polymorphism (SNP) markers for the White Rhino *(Ceratotherium simum*), based on a targeted gene approach. The polymorphisms of these SNP loci were assessed using a captive population comprising 30 individuals. The minor allele frequency ranged from 0.256 to 0.413 and the observed and expected heterozygosity ranged from 0.05 to 0.37 and from 0.05 to 0.49, respectively. An understanding of genetic population structure is required to effectively formulate strategies for conservation and/or management. These SNP markers could be employed to provide estimates of parameters such as population structure, Relatedness and current and historical gene flow.

The African White Rhino population has suffered a decline over the past 150 years as a result of overhunting, habitat destruction and poaching (Seror *et al.,* 2002; Florescu *et al.,* 2003). Currently the estimated population of white rhinos comprises 20,170 individuals (Emslie, 2011). The trade in rhinoceros horns is a problem in many parts of the world especially in parts of Asia where the rhinoceros horns are used traditionally as material in sculptures or as drug products for medicinal purposes (Hsieh *et al.,* 2003) adding constant pressure on remaining populations. There is thus a real need for markers that can identify the region of origin of rhino products. Genetic diversity and relatedness data for both captive and wild populations also form an important tool in successful reproductive management, population viability assessments and diversity conservation with regards to translocation of animals and establishing breeding programmes (Seror *et al.,* 2002; Harley *et al.,* 2005). The approach described here made use of currently available conserved primer sets designed to amplify from exons across an intervening intron. These CATS primers were designed from other vertebrate genomes to amplify mammalian genes and have been used successfully by Morin *et al.,* (2007) to characterize 18 SNPs for the sperm whale (*Physeter macrocephalus*) and by Li *et al.,* (2009) to describe 51 SNPs in the finless porpoise (*Neophocaena phocaenoides*). The current study is the first to present SNP markers for the white rhino. Fifteen CATS primers previous described by Aitken *et al.,* (2004) were introduced to discover SNPs from five randomly selected samples. The utilisation of CATS primers allows identification of SNPs in genes of known function so that some genomic information is associated with the identified loci even without prior genomic characterization of the target species (Aitken *et. al*., 2004). The PCR reactions were conducted with Thermo Scientifics' DreamTaq™ according to manufacturer's instructions. Out of the fifteen sets, fourteen amplified successfully and these products were subsequently sequenced for each of the thirty selected samples. Sequencing was performed by Inqaba biotec utilising the ABI Big Dye V3.1 kit (Applied Biosystems) and the ABI 3500XL genetic analyser (Applied Biosystems). GENEPOP version

4.0.10 (Raymond and Rousset 1995) was used to calculate observed (Ho) and expected heterozygosity (He) and to test for genotypic linkage disequilibrium (LD) and deviation from Hardy-Weinberg equilibrium (HWE). Sequence data was compared between isolates using CLC Bio Main work bench (Denmark). A total of 7523 base pairs of sequence data were generated across the 14 loci for each of the five isolates. Ten SNPs were identified across five of the loci (MGF, ACTC, BGN, GLUT2 and KIT) relating to a discovery rate of one SNP every 752 bp. SNPs were not identified in the following loci: C5, CHY, COL10A1, COL9A1, FES, GHR, HOXD, LPL and WT1. Previous studies utilising CATS loci describe SNP discovery rates of 1SNP/400bp in chimpanzees (Aitken *et. al*., 2004), 1SNP/540bp in sperm whale (Morin *et al.,* 2007) and 1SNP/551bp in finless porpoise (Li *et al.,* 2009). The current study reported a somewhat lower discovery rate which may be attributed to the reported low genetic variation in white rhino populations (Florescu *et al.,* 2003; Harley *et al.,* 2005). The frequencies of the minor alleles ranged from 0.256-0.413. The observed and expected heterozygosity ranged from 0.05 to 0.37 and from 0.05 to 0.49, respectively. The BGN marker deviated from Hardy-Weinberg equilibrium which may be attributed to small sample size. Linkage disequilibrium was observed with those SNPs identified in locus ACTC as well as GLUT2. This was expected since these SNPs were in close proximity on the same locus. These markers should be further investigated for applications in other species such as the endangered black rhino (*Diceros bicornis*).

Table 18: Characterization of 10 SNPS in White Rhino (*Ceratotherium simum*): F = forward primer; R = reverse primer; bp = base pairs; He = expected heterozygosity and Ho = observed heterozygosity. GenBank accession numbers are NCBI_ss#538305377, 538786572-81.

Locus	SNP Name	Sequencing length (bp)	Sequence (5'-3') (Aitken *et al.,* 2004)	Minor allele frequency	Heterozygosity	
					He	Ho
MGF	MGF-1	820	F-ATCCATTGATGCCTTCAAGG R-CTGTCATTCCTAAGGGAGCTG	41.03	0.4902	0.3668
	MGF-2			2.56	0.0506	0.0487
ACTC	ACTC-1	875	F-GCCCTGGATTTTGAGAATGAGAT R-ACGATCAGCAATACCAGGGTACA	35.90	0.4662	0.3543
	ACTC-2			29.49	0.4212	0.3294
	ACTC-3			38.46	0.4795	0.3613
BGN	BGN	647	F-CTCCAAGAACCACCTGGTG R-TTCAAAGCCACTGTTCTCCAG	33.75	0.4528	0.3472
GLUT 2	GLUT2F-1	301	F-TGGATGAGTTATGTGAGCAT R-GACTTTCCTTTGGTTTCTGG	41.25	0.4908	0.3672
	GLUT2F-2			41.25	0.4908	0.3672
KIT	KIT-1	641	F-CCTGTGAAGTGGATGGCACC R-GCATCCCAGCAAGTCTTCAT	13.75	0.2402	0.2091
	KIT-2			10	0.1823	0.1638

References

Aitken N., Smith S., Schwarz C., Morin P.A. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology,* 13: 1423-12431.

Emslie R. 2011. *Ceratotherium simum*. In: IUCN 2012. IUCN Red List of Threatened Species. Version 2012.1. <www.iucnredlist.org>. Downloaded on 06 September 2012.

Florescu A., Davila J.A., Scott C., Fernando P., Kellners K., Morales J.C., Melnick D., Boag P.T., van Coeverden de Groot P. 2003. Polymorphic microsatellites in white rhinoceros. *Molecular Ecology Notes,* 3: 344-345.

Harley E.H., Baumgarten I., Cunningham J., O'Ryan P. 2005. Genetic variation and population structure in remnant populations of Black Rhinoceros, *Diceros bicornis*, in Africa. *Molecular Ecology,* 14: 2981-90.

Hsieh H., Huang L., Tsai L., Kuo Y., Meng H., Linacre A., Lee J.C. 2003. Special identification of rhinoceros horns using the cytochrome b gene. *Forensic Science International,* 136: 1-11.

Li S., Wan H., Ji H., Zhou K., Yang G. 2009. SNP discovery based on CATS and genotyping in the finless porpoise (*Neophocaena phocaenoides*). *Conservation Genetics,* 6: 2013-2019.

Morin P.A., Aitken N.C., Rubio-Cisneros N., Dizon A.E., Mesnick S. 2007. Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). *Molecular Ecology Notes,* 7: 626-630.

Raymond M. and Rousset F. 1995. Genepop (version 1.2), population genetics software for exact tests and ecumenicism. *J Hered*. 86: 248-249.

Seror B., Zorovsky Y., Terkel A., Katcoff D.J. 2002. Genetic Kinship and Social Structure in a Herd of Square-Lipped Rhinoceroses (*Ceratotherium simum simum*) at the Zoological Center, Tel Aviv/Ramat-Gan, Israel. *Zoo Biology* 21: 561-571.

CHAPTER 9

Endonuclease V digestion for SNP discovery and marker development in South African white rhinoceros (*Ceratotherium simum*)

Abstract

Single nucleotide polymorphism (SNP) markers are a promising new tool that can be used to study evolutionary processes, population genetic parameters, forensic cases and parentage. However, application of SNP marker analysis to wildlife has been limited, due to the lack of available sequence data in non-model organisms. Here, we describe a simple, rapid and cost effective method to isolate candidate SNPs in non-model organisms using the commercially available Endonuclease V enzyme. In a first application of this method, this SNP isolation strategy resulted in the identification of 12 new SNPs for white rhinoceros (*Ceratotherium simum*). This species has low reported genetic variability and has suffered severe bottlenecks over the last 150 years. Developed SNP markers in white rhinoceros could be used to define the genetic mating system of this species, for forensic applications and to determine population structure and variability when other markers prove problematic.

Introduction

The white rhinoceros is a species that is affected the most at the hand of illegal trade in rhinoceros horn (Florescu *et al.,* 2003). Highly informative molecular markers are important tools in successful management (Abadia-Cardoso *et al.,* 2011). Single-strand specific nucleases such as CEL I, have been proposed as a simple and rapid method to assay mutations and SNPs (Rungis *et al.,* 2005). The enzyme works with a variety of co-factors to digest heteroduplex DNA immediately 3' of a mismatch site (Oleykowski *et al.,* 1998). CEL I has been widely used in reverse genetics in plants and animals as well as disease diagnostics in human cancers (Colbert *et al.,* 2001; Coghill *et al.,* 2002; Perry *et al.,* 2003; Comai *et al.,* 2004). Xu *et al.,* (2009) applied CEL I in a method to isolate fragments containing SNPs from background DNA in the half-smooth tongue sole (*Cynoglossus semilaevis*). However, CEL I needs to be isolated from celery stems, a time consuming procedure. The aim of the present study was to evaluate the use of commercially available Endonuclease V for the isolation of SNP containing fragments from white rhinoceros DNA.

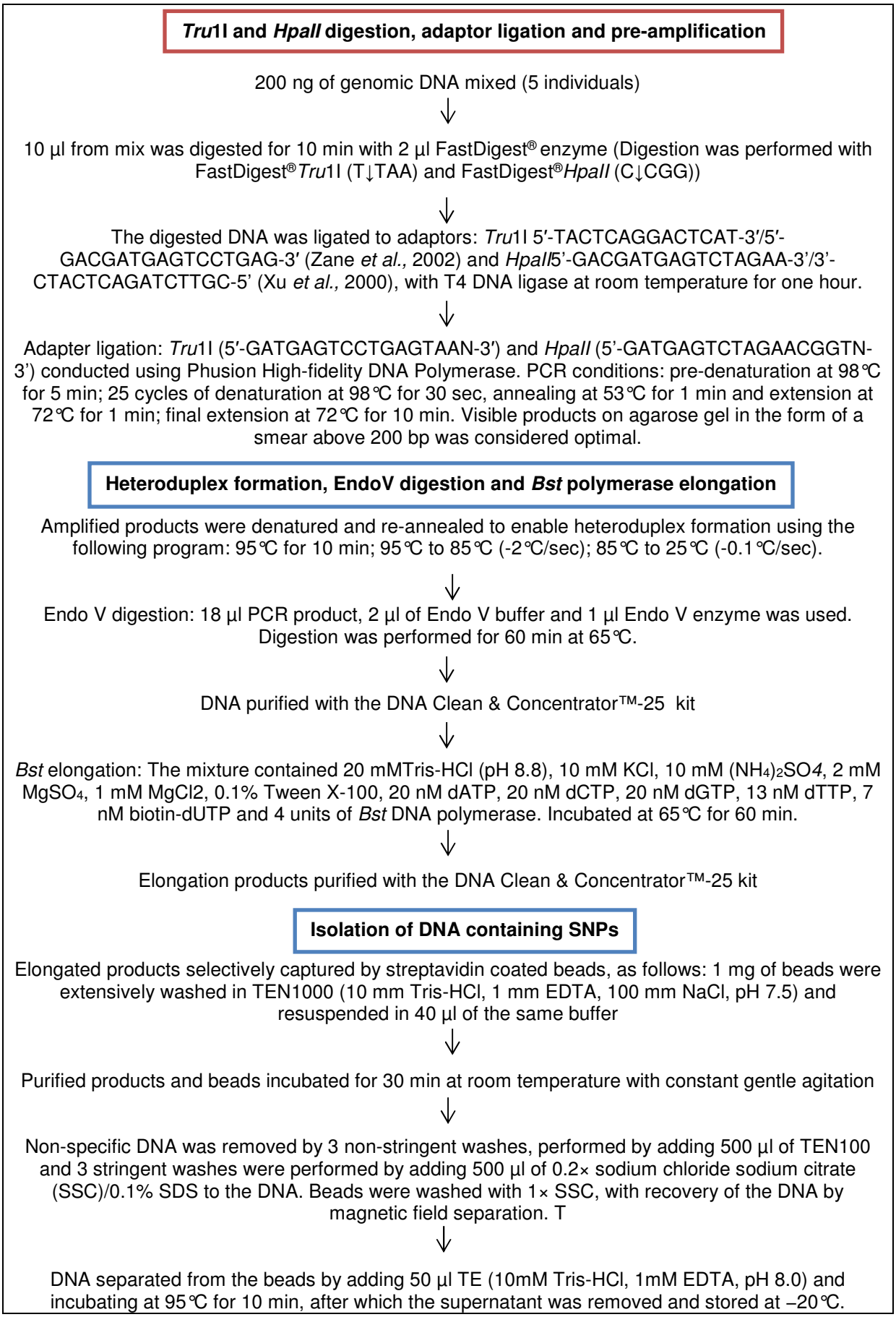
Materials and Methods

Blood samples were collected from 26 rhinoceros. DNA was extracted using the ZR Genomic DNA™-Tissue MiniPrep kit following the manufacturer's instructions. Blood samples were first washed by mixing 100 ul blood with 1000 ul nuclease free water followed by centrifugation at 4000 rpm for 2 minutes to reduce the number of red blood cells and improve DNA yields. The resulting pellet was further extracted using the above mentioned kit. The method used to isolate DNA containing SNPs was performed as shown in Figure 16. Thereafter, Cloning and Sequencing was performed. The

isolated DNA was amplified with either *Tru*1I or *HpaII* primers as described above. Subsequent amplicons were cloned into pJET using the CloneJET PCR Cloning Kit (Thermo Scientific) and Z-Competent™ JM109 *E. Coli* cells (Zymo Research). Ten clones containing fragments ranging from 300bp to 800 bp were selected from the libraries. Cloned fragments were purified utilizing the Zymoclean™ Gel DNA Recovery Kit (Zymo Research) and sequenced utilising a Big Dye V3.1 Terminator Kit and an ABI 3500XL genetic analyser.

SNP validation: Primers were designed according to the obtained sequences. The potential SNP loci were amplified in the 5 isolates used for the initial DNA pool. Amplification reactions were done in a final volume of 25 µl containing 30 ng DNA, 25 pM of each primer and 2X DreamTaq® Green Master Mix using a standard PCR protocol. Resulting amplicons were inspected on 1% agarose gels followed by purification and sequencing as described above. Sequences were inspected and aligned in CLC Bio Genomics work bench 5.0. GENEPOP version 4.0.10 (Raymond and Rousset, 1995) was used to test for deviations from expected Hardy-Weinberg proportions, to evaluate loci for gametic disequilibrium and to determine allelic richness. Differences in mean observed heterozygosity, mean expected heterozygosity and mean number of alleles was determined using GenAlEx6 (Peakall & Smouse 2006).

Figure 16: Method used to isolate DNA containing SNPs



Results

DNA from five rhinoceros was used for SNP discovery. DNA enriched for mismatches through EndoV digestion of *Tru*1I and *HpaII* AFLP libraries were cloned and sequenced. Ten clones were selected for each enzyme. Six out of the 10 *HpaII* clones contained a total of 13 polymorphic sites and were typed in a further 10 isolates. Four of the clones were heterozygous at all polymorphic sites for all individuals typed. These polymorphisms were artefacts from paralogous sequence differences. The remaining two clones both possessed one polymorphic site that presented as both homozygous and heterozygous. These polymorphisms were considered to be true SNPs and were typed through sequencing in the remaining individuals. Nine out of the 10 *Tru*1I clones contained a total of 30 polymorphic sites and were typed in a further 10 isolates. Four of the clones were heterozygous in all and were considered to be artefacts. The remaining 5 clones all showed polymorphic sites that presented as both homozygous and heterozygous alleles in the 10 individuals tested. These were considered to be true SNPs and were typed through sequencing in the remaining individuals. The 12 SNPs, primer sequences and allele frequencies for the 26 individuals are listed in Table 19. SNPs Tru2-1, Tru2-2, Tru2-4 and Tru2-5 are in Linkage disequilibrium (P=0.0000), while Tru2-3 and Tru2-6 are in also linked (P=0.0000). None of the loci deviated significantly from Hardy-Weinberg equilibrium.

Discussion

The preparation of native enzymes from plant material usually involves several steps of purification (Mon *et al.,* 2012). Adopting the SNP enrichment method proposed by Xu *et al.,* (2009), but replacing CEL I with commercially available Endonuclease V, we isolated 12 new SNPs in the white rhinoceros. Endo V is active towards basic sites and urea sites, base pair mismatches, flap and pseudo Y structures, and small insertions/deletions in DNA molecules with the cleavage site generated at the second phosphodiester bond 3' to a lesion (Huang *et al.,* 2001). Endo V recognised the mismatches formed in the heteroduplex library between different isolates creating a nick, which was subsequently recognised by *Bst* DNA polymerase and extended through strand displacement incorporating biotin-dUTP. The DNA strands containing biotin-dUTPs could then be captured with streptavidin-coated beads and separated from background DNA. Xu *et al.,* (2009) reported that 9 out of 10 fragments contained SNPs through CEL I digestion. In the present study only 2 fragments from the *HpaII* library and 5 fragments from the Tru1I library contained true SNPs. Since paralogous regions form mismatches in the heteroduplex in the same way as a polymorphism between isolates would, the *HpaII* library resulted in 6 out of 10 enriched fragments while the Tru1I library resulted in 9 out of 10 enriched fragments, confirming that EndoV can be used to substitute CEL I. The apparent high proportion of paralogs may be a result of the low genetic diversity observed in white rhinoceros.

Table 19: Characterization of 12 SNPs in White Rhinoceros (*Ceratotherium simum*): GenBank accession numbers are 825690401-825690412.

Locus	SNP name	Sequence length (bp)	Primer sequences (5'-3')	Minor allele frequency	Heterozygosity		HWE
					He	Ho	
Hpa-1	Hpa-1-K	605	F- GGGATCATTCATTCATTCAGCTG R- GGAACTCCAGAAGCCACG	0.260	0.385	0.280	0.173 (ns)
Hpa-10	Hpa-2-W	449	F- CCTTGTGTGGATTAAATGAGC R- CTCAGCGGGTGGTTTCTC	0.385	0.473	0.462	0.899 (ns)
Tru-1	Tru-1-K	380	F- GAGAGCTTTCTCTCCTGAT R- GAACTGGAAGTGTGTCAAC	0.058	0.109	0.115	0.755 (ns)
Tru-2	Tru-2-1-S	345	F- CCAGCATGGCTAGCATGC R- CAGCCCTATCCGTGACTTTC	0.423	0.488	0.462	0.781 (ns)
	Tru-2-2-R			0.442	0.493	0.500	0.945 (ns)
	Tru-2-3-Y			0.077	0.142	0.154	0.671 (ns)
	Tru-2-4-Y			0.442	0.493	0.500	0.945 (ns)
	Tru-2-5-R			0.442	0.493	0.500	0.945 (ns)
	Tru-2-6-M			0.077	0.142	0.154	0.671 (ns)
Tru-3	Tru-3-R	335	F- GGCTCTGTTTGCTTGTCTG R- CTTAGTGCTAGATTCTGCATG	0.250	0.375	0.346	0.695 (ns)
Tru-4	Tru-4-K	362	F- GTAGAACCTTCATCTCTGC R- GCAGCTGCATTATATCCAC	0.231	0.355	0.462	0.126 (ns)
Tru-5	Tru-5-W	193	F- CTTGTGCTATTCTTCACTGTC R- CAAGACGTCCACTGCAC	0.280	0.403	0.480	0.341 (ns)

F = forward primer; R = reverse primer; bp = base pairs; He = expected heterozygosity and Ho = observed heterozygosity; HWE = Hardy-Weinburg equilibrium; ns = non-significant.

References

Abadia-Cardoso A., Clemento A.J., Garza J.C. 2011. Discovery and characterization of single-nucleotide polymorphisms in steelhead/rainbow trout, *Oncorhynchus mykiss*. *Mol Ecol Resour*. 11 (supp.1): 31-49.

Butler J.M., Coble M.D., Vollone P.M. 2007. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol*. 3: 200-205.

Coghill E.L., Hugill A., Parkinson N., Davison C., Glenister P., Clements S., Hunter J., Cox R.D., Brown S.D. 2002. A gene-driven approach to the identification of ENU mutants in the mouse. *Nat Genet*. 30: 255-256.

Colbert T., Till B.J., Tompa R., Reynolds S., Steine M.N., Yeung A.T., McCallum C.M., Comai L., Henikoff S. 2001. High-throughput screening for induced point mutations. *Plant Physiol*. 126: 480-484.

Comai L., Young K., Till B.J., Reynolds S.H., Greene E.A., Codomo C.A., Enns L.C., Johnson J.E., Burtner C., Odden A.R., Henikoff S. 2004. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J*. 37: 778-786.

Emslie R. 2012. *Ceratotherium simum*. In: IUCN 2012. IUCN Red List of Threatened Species.Version 2012.2.<www.iucnredlist.org>. Downloaded on 01 May 2013.

Falberg A. 2013. “The Living Are Getting Rarer”: The Causes and Consequences of the International Trade in White Rhinoceros Horns Under the Convention on International Trade in Endangered Species. 2 Penn St. J.L. & Int'l Aff. 182.

Florescu A., Davila J.A., Scott C., Fernando P., Kellners K., Morales J.C., Melnick D., Boag P.T., van Coeverden de Groot P. 2003. Polymorphic microsatellites in white rhinoceros. *Mol Ecol Notes*. **3**: 344-345.

Galeano C.H., Gomez M., Rodriguez L.M., Blair M.W. 2009. CEL I Nuclease digestion for SNP discovery and marker development in common bean (*Phaseolus vulgaris* L.). *Crop Sci*. 49: 381-394.

Glover K.A., Hansen M.M., Lien S., Als T.D., Høyheim B., Øystein S. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genet*. 11: 2.

Huang J., Lu J., Barany F., Cao W. 2001. Multiple cleavage activities of endonuclease V from *Thermotoga maritima*: recognition and strand nicking mechanism. *Biochemistry*. 40: 8738-8748.

Mon H., Lee J., Fukushima M., Nagata Y., Fujii M., Xu J., Nishi O., Iiyama K., Kusakabe T. 2012. Production and characterization of the celery mismatch endonuclease CEL II using baculovirus/silkworm expression system. *Appl Microbiol Biotechnol*. 97 (15): 6813-6822.

Morin P.A., Luikart G., Wayne R.K., SNP workshop group. 2004. SNPs in ecology, evolution and conservation. *Trends Ecol Evol*. 19 (4): 208-216.

Morin P.A., Martien K.K., Taylor B.L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Mol Ecol Resour*. 9: 66-73.

Nielsen L., Meehan-Meola D., Kilbourn A., Alcivar-Warren A. 2008. Characterisation of microsatellite loci in the black rhinoceros (*Diceros bicornis*) and white rhinoceros (*Ceratotherium simum*): their use for cross species amplification and differentiation between the two species. *Conser Genet.* 9: 239-242.

Oleykowski C.A., Bronson Mullins C.R., Godwin A.K., Yeung A.T. 1998. Mutation detection using a novel plant endonuclease. *Nucleic Acids Res*. 26: 4597-4602.

Peakall R. and Smouse P.E. 2006. GENALEX 6: geneticanalysis in Excel. Population genetic software for teaching and research*. Mol Ecol Notes*. 6: 288-295.

Perry J.A., Wang T.L., Welham T.J., Gardner S., Pike J.M., Yoshida S., Parniske M. 2003. A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol*. 131: 866-871.

Pertoldi C., Bijlsma R., Loeschcke V. 2007. Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodivers Conserv*. 16: 4147-4163.

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

Rungis D., Hamberger B., Berube Y., Wilkin J., Bohlmann J., Ritland K. 2005. Efficient genetic mapping if single nucleotide polymorphisms based upon DNA mismatch digestion. *Mol Breed*. 16: 261-270.

Ryynänen H.J. and Primmer C.R. 2006. Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primers exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo sala*r) and other salmonid fishes. *BMC Genomics*. 7: 192.

Seror B., Zorovsky Y., Terkel A., Katcoff D.J. 2002. Genetic Kinship and Social Structure in a Herd of Square-Lipped Rhinoceros (*Ceratotherium simum simum*) at the Zoological Center, Tel Aviv/Ramat-Gan, Israel. *Zoo Biol*. 21: 561-571.

Smith C.T., Baker J., Park L., Seeb L.W., Elfstrom C., Abe S., Seeb J.E. 2005. Characterisation of 13 single nucleotide polymorphism markers for chum salmon. *Mol Ecol Notes.* 5**:** 259-262.

Xu M.L., Li X.Q., Korban S.S. 2000. AFLP-based detection of DNA methylation. *Plant Mol Biol Reptr.* 18: 361-368.

Xu J.Y., Xu G.B., Chen S.L. 2009. A new method for SNP discovery. *Biotechniques.* 46: 201-208.

Zane L., Baegelloni L., Patarnello T. 2002. Strategies for microsatellite isolation: a review. *Mol Ecol*. 11**:** 1-16.

CHAPTER 10

SNP discovery and characterisation in White Rhino (*Ceratotherium simum*) with application to parentage assignment

Abstract

The white rhino is one of the great success stories of modern wildlife conservation, growing from as few as 50-100 animals in the 1880s, to approximately 20,000 white rhinoceros remaining today. However, illegal trade in conservational rhinoceros horns is adding constant pressure on remaining populations. Captive management of *ex-situ* populations of endangered species using molecular methods can contribute to improving the management of the species. Here we compare for the first time the utility of 33 Single Nucleotide Polymorphisms (SNPs) and nine microsatellites (MS) in isolation and in combination for assigning parentage in captive White Rhinoceros. We found that a combined dataset of SNPs and microsatellites was most informative with the highest confidence level. This study thus provided us with a useful set of SNP and MS markers for parentage and relatedness testing. Further assessment of the utility of these markers over multiple (> three) generations and the incorporation of a larger variety of relationships among individuals (e.g. half-siblings or cousins) is strongly suggested.

Introduction

Due to intensive protection and conservation efforts, the Southern white rhinoceros (*Ceratotherium simum simum*) have increased from a population of less than 100 at the end of the 19th century, to an estimated population of over 20,000 (Emslie 2012). However, the illegal trade in rhinoceros horn in many parts of the world especially in Asia where the rhinoceros horns are used traditionally as material in sculptures or as drug products for medicinal purposes (Hsieh *et al*., 2003) is adding constant pressure on remaining populations. Currently, the remaining white rhino populations are being managed as small isolated groups thus monitoring and maintaining genetic diversity is a key concern for long term survival of this species (Emslie and Brooks, 1999). Potential consequences of a reduction in genetic variability include (1) the inability of the species to adapt to changes in their environment and (2) inbreeding, whereby the expression of rare deleterious alleles may contribute to developmental, reproductive and immunological impairments (Pertoldi *et al*., 2007; Väli *et al*., 2008). In order to maintain genetic diversity as well as reduce the effects of inbreeding, a rescue-strategy can be employed whereby gene flow amongst populations is encouraged (Pertoldi *et al*., 2007). However, an analysis of genetic structure is required in order to ensure that outbreeding depression due to the introduction of mal-adapted genes does not occur (Pertoldi *et al*., 2007; Väli *et al*., 2008).

Single nucleotide polymorphisms (SNPs) represent the most abundant type of DNA variation in the vertebrate genome and are distributed across the entire genome providing broader genome coverage as compared to mitochondrial DNA or microsatellites (MS) (Pertoldi *et al*., 2007; Ryynänen and

Primmer, 2006). In addition, SNPs offer higher recovery of information from degraded DNA samples since the DNA target sequence in SNP-based genotyping is appreciably shorter (50-70 bp) than that in microsatellite-based genotyping (80-300 bp) (Butler *et al*., 2007; Morin *et al*., 2004; Pertoldi *et al*., 2007; Ryynänen and Primmer, 2006). In contrast to microsatellites, SNP genotyping reveals polymorphisms directly on the DNA sequence, and thus data is automatically standardized across chemistries, hardware platforms and laboratories (Glover *et al*., 2010; Smith *et al*., 2005). Furthermore, the development of high through-put genotyping platforms permits simultaneous genotyping of thousands of loci, enabling the identifications of highly diagnostic panels (Glover *et al*., 2010).

In this study, we compare the power of parentage assignment of 33 SNPs and 9 MS markers in isolation and in combination in a captive population of white rhinoceros. Development of a marker set that accurately determine parentage will provide information on the relationships and relatedness among individuals, contribute to the management of captive white rhinoceros worldwide and additionally provide insight into mating systems in wild populations.

Materials and methods

Blood samples were collected from 32 white rhinoceros in South Africa. Blood aliquots were first treated by mixing 100 ul blood with 1000 ul nuclease free water followed by centrifugation at 4000 rpm for 2 minutes, to reduce the number of red blood cells and improve DNA yields. Genomic DNA was extracted from the resulting pellet using the ZR Genomic DNA™-Tissue Mini-Prep kit (Zymo Research) following the manufacturer's instructions. A SNP enriched library was constructed using DNA from 5 individuals and digestion with Endonuclease V as previously described (Labuschagne *et al*., 2015). This protocol was used without any changes. Subsequent SNP enriched amplicons were cloned into pJET using the CloneJET PCR Cloning Kit (Thermo Scientific) and Z-Competent™ JM109 *E. Coli* cells (Zymo Research). Clones containing fragments ranging from 200bp to 700 bp were selected and sequenced utilising a Big Dye V3.1 Terminator Kit and an ABI 3500XL genetic analyser. The potential SNP loci were amplified in the 5 isolates used for the initial DNA pool. Amplification reactions were done in a final volume of 25 µl containing 30 ng DNA, 25 pM of each primer and 2X DreamTaq® Green Master Mix (Thermo Scientific). Thermal cycling consisted of initial denaturation at 95°C for 5 minutes, 45 cycles of denaturation at 95°C for 30 seconds, annealing at 55-59°C for 30 sec, extension at 72°C for 1.5 min, followed by final extension at 72°C for 10 min. Resulting amplicons were inspected on 1% agarose gels followed by purification and sequencing as described above. Sequences were inspected and aligned in CLC Bio Genomics work bench 8.0.1 (CLC bio, Denmark). Twelve resulting SNP markers were further typed in the remaining 27 isolates. GENEPOP version 4.0.10 (Raymond and Rousset, 1995) was used to test for deviations from expected Hardy-Weinberg (HW) proportions, to evaluate loci for gametic disequilibrium and to determine allelic richness. Differences in mean observed heterozygosity (Ho), mean expected heterozygosity (He) and mean number of alleles was determined using CERVUS v3.03 (Kalinowski *et al*., 2007). All 32 samples

were further typed for 21 previously described SNP markers through Sanger sequencing (Labuschagne *et al*., 2013, Labuschagne *et al*., 2015).

Nine microsatellite loci: BR6 (Cunningham *et al*., 1999), DB44, DB66, DB49, DB1 (Brown and Houlden, 1999), RHI7C, RHI32A, RHI7B (Florescu *et al*., 2003), SW35 (Rohrer *et al*., 1994) were also used. Markers were selected based on previously reported polymorphism in white rhinoceros. The PCR optimization for each locus was as follows: 2 ng of template DNA, 1.5 – 2.5 mM $MgCl_2$, 2 mM dNTP's, 1 µM forward and 1 µM reverse primer, 0.10 U *Taq* DNA polymerase and ddH_2O to a final volume of 15 µl. PCR cycles were as follows: initial denaturing stage at 94ºC for 3 min, followed by 30 cycles of 94ºC for 30 sec, annealing at 50-55ºC for 30 sec, extension at 72 ºC for 30 sec and a final step of 72 ºC for 20 min. Products were electrophoresed on an ABI Prism 3130 DNA sequencer (Applied Biosystems). Allele sizes were estimated by comparison with a Genescan™ 500 LIZ™ internal size standard (ABI, Foster City, CA) using the ABI programs GENESCAN (version 1.2.2.1) and GENOTYPER (version 1.1). Sanger sequencing was performed in both directions and SNP calls were only made on bases with quality scores higher than Q>20. The SNPs all fall within the central region of the fragments where sequencing quality is the highest. In order to ensure accurate genotyping, the samples were repeated if they were homozygous, aberrant stutter patterns or spurious peaks were observed or if the profiles were below the quality score. Differences in mean observed heterozygosity (Ho), mean expected heterozygosity (He) and mean number of alleles was determined as mentioned above.

Parentage assignment was evaluated using the MS and SNP data sets individually and as a combined dataset. The software program CERVUS v3.03 (Kalinowski *et al*., 2007) was implemented for parentage assignment using likelihood. The program uses multilocus parental exclusion probabilities (Selvin, 1980) and pair-wise likelihood to assign parent pairs to offspring. CERVUS calculates the log-likelihood of each candidate parent being the true parent relative to an arbitrary individual and then calculates the difference between the two most likely parents (Delta, Δ). Critical values of Δ are determined by computer simulation. Using the real data for allele frequencies, simulation parameters were set at 10 000 offspring, with 100% of candidate parents sampled and a total proportion of loci typed over all individuals of 0.99, mistyping error rates = 0.01 and likelihood calculation error rates = 0.01, permitting 2 unscored loci. Strict confidence was set to 95% while the relaxed confidence level was 80%.

Results

Twelve SNPs (GenBank accession numbers 1416044499-1416044509) were identified in this study across 11 loci (WR1-WR11). The primer sequences and allele frequencies of the 12 SNPs developed here together with a further 21 SNPs for the 32 individuals are listed in Table 20. The PIC ranged from 0.060 to 0.396 with a mean of 0.2742. The observed and expected heterozygosity ranged from 0.065 to 0.656 and from 0.063 to 0.520, respectively. Marked BGN deviated from Hardy-Weinberg equilibrium. Large differences between the observed and expected heterozygosity was also observed

for four markers namely; WR1, WR8-Y, WR11 and Tru-3. The observed deviations may be attributed to small sample size. Linkage disequilibrium was observed between markers ACTC-2/ACTC-3, GLUT2F-1/GLUT2F-2 and Tru2-1/Tru2-2/Tru2-4/Tru2-5. Such linkage is not unexpected since these SNPs are in close proximity on the same locus.

The nine MS markers, primer sequences and allele frequencies for the 32 individuals are listed in Table 21. The PIC ranged from 0.259 to 0.578 with a mean of 0.4282, while observed and expected heterozygosity ranged from 0.273 to 0.654 and from 0.298 to 0.655, respectively. None of the MS loci deviated significantly from Hardy-Weinberg equilibrium and no linkage disequilibrium was observed. Only two alleles were observed in four of the markers, while four markers exhibited three alleles and one marker, five alleles, resulting in a mean allele number (Na) of 2.7.

Table 20: Summary statistics for 33 SNPs in White Rhino (*Ceratotherium simum*). PIC Mean polymorphic information content, F forward primer, R reverse primer, bp base pairs, Ho observed heterozygosity, He expected heterozygosity, *F(Null)* the *F* score for the null hypothesis that the locus is in Hardy-Weinberg Equilibrium.

Locus	SNP name	Sequence length (bp)	Primer sequences (5'-3')	PIC	Heterozygosity		*F(Null)*
					Ho	He	
WR1[a]	WR1-Y	136	F-GCAACTGAGGAGCAATCA R-AGAAGCAAACTCATAAGATA	0.354	0.656	0.468	-0.175
WR2[a]	WR2-S	173	F-GTATTATGCTGAGTGATACAG R-CAGGTGTAGATGCTGGA	0.110	0.125	0.119	-0.023
WR3[a]	WR3-W	562	F-CACTCACTCACCTGAGGCAC R-CTGTGGAGTATATAGTCCTAGC	0.314	0.406	0.396	-0.020
WR4[a]	WR4-M	358	F- CCTGAGTAATATGACAGCAGTCC R-GTAAGGCCTGCTGCTCTTAG	0.330	0.531	0.424	-0.119
WR5[a]	WR5-K	349	F- CTTCTCCTGTTACTGCATGGTCAC R-GTCAGTGGTGCCAATATGCAAG	0.176	0.219	0.198	-0.052
WR6[a]	WR6-Y	586	F-GACTCGCCCTTTGTGAAAGTG R-CTGCATTGTTGCCTGGTTC	0.134	0.156	0.146	-0.032
WR7[a]	WR7-R	406	F-GAGCTGCTGCTCAGCAGAG R-GTACCTCTGAGAAGCCACTAG	0.314	0.469	0.396	-0.091
WR8[a]	WR8-Y	485	F-GTGCTTCTTCACAGCTGTAG R-GATACGTGTGTTTGGAGTGG	0.244	0.344	0.289	-0.091
	WR8-R			0.134	0.156	0.146	-0.032
WR9[a]	WR9-K	197	F-GACTTCCAAATGTAAGAAGGTG R- CAAGTTTCTTTGCTGAATGTTTGC	0.314	0.344	0.396	0.063

Table 20: Summary statistics for 33 SNPs in White Rhino (Continued)

WR10[a]	WR10-M	333	F-CACTGTATACCAAACAAAATGG R-CTCACAATTCTGCAATCTGG	0.349	0.500	0.458	-0.051
WR11[a]	WR11-W	296	F-GGGTCACCTTAGGTAGG R-GAGGAATAACACAAGTAACAACG	0.359	0.250	0.476	0.304
MGF[b]	MGF-1	820	F-ATCCATTGATGCCTTCAAGG R-CTGTCATTCCTAAGGGAGCTG	0.362	0.516	0.482	-0.042
	MGF-2			0.060	0.065	0.063	-0.007
ACTC[b]	ACTC-1	875	F-GCCCTGGATTTTGAGAATGAGAT R-ACGATCAGCAATACCAGGGTACA	0.353	0.516	0.465	-0.059
	ACTC-2			0.310	0.452	0.389	-0.082
	ACTC-3			0.358	0.484	0.474	-0.018
BGN[b]	BGN	647	F-CTCCAAGAACCACCTGGTG R-TTCAAAGCCACTGTTCTCCAG	0.363	0.156	0.484	0.505
GLUT2[b]	GLUT2F-1	301	F-TGGATGAGTTATGTGAGCAT R-GACTTTCCTTTGGTTTCTGG	0.369	0.594	0.496	-0.098
	GLUT2F-2			0.369	0.594	0.496	-0.098
KIT[b]	KIT-1	641	F-CCTGTGAAGTGGATGGCACC R-GCATCCCAGCAAGTCTTCAT	0.176	0.156	0.198	0.109
	KIT-2			0.155	0.188	0.173	-0.042
Hpa-1[c]	Hpa-1-K	605	F-GGGATCATTCATTCATTCAGCTG R- GGAACTCCAGAAGCCACG	0.310	0.258	0.389	0.194
Tru-1[c]	Tru-1-K	380	F- GAGAGCTTTCTCTCCTGAT R- GAACTGGAAGTGTGTCAAC	0.085	0.094	0.091	-0.014

Table 20: Summary statistics for 33 SNPs in White Rhino (Continued)

Tru-2[c]	Tru-2-1-S	345	F- CCAGCATGGCTAGCATGC R- CAGCCCTATCCGTGACTTTC	0.396	0.531	0.520	-0.017
	Tru-2-2-R			0.375	0.531	0.507	-0.030
	Tru-2-3-Y			0.134	0.156	0.146	-0.032
	Tru-2-4-Y			0.375	0.531	0.507	-0.030
	Tru-2-5-R			0.375	0.531	0.507	-0.030
	Tru-2-6-M			0.134	0.156	0.146	-0.032
Tru-3[c]	Tru-3-M	335	F- GGCTCTGTTTGCTTGTCTG R- CTTAGTGCTAGATTCTGCATG	0.294	0.281	0.365	0.121
Tru-4[c]	Tru-4-K	362	F- GTAGAACCTTCATCTCTGC R- GCAGCTGCATTATATCCAC	0.258	0.375	0.310	-0.101
Tru-5[c]	Tru-5-W	193	F- CTTGTGCTATTCTTCACTGTC R- CAAGACGTCCACTGCAC	0.327	0.452	0.419	-0.045

a)This study; b) Labuschagne *et al*., 2013; c) Labuschagne *et al*., 2015.

Table 21: List of nine microsatellite loci used for DNA profiling in white rhino.

Locus	Primer sequences (5'-3')	Allele size range	Allele no.	PIC	Heterozygosity		*F(Null)*	Reference
					Ho	He		
SW35	F-TCAAGTTGGAGAGTCTGAGGC R-AAGACTGCCCACCAAATGAG	127-133	2	0.417	0.545	0.535	-0.0310	Rohrer *et al*., 1994
BR6	F-TCATTTCTTTGTTCCCCATAGCAC R-AGCAATATCCACGATATGTGAAGG	133-153	3	0.474	0.424	0.529	+0.0894	Cunningham *et al*. 1999
DB44	F-GGTGGAATGTCAAGTAGCGG R-CTTGTTGCCCCATCCCTG	173-181	2	0.363	0.469	0.441	-0.0525	Brown & Houlden 1999
DB66	F-CCAGGTGAAGGGTCTTATTATTAGC R-GGATTGGCATGGATGTTACC	201-203	3	0.416	0.452	0.531	+0.0595	Brown & Houlden 1999
RHI7C	F-TGAACTCTGATGGAAATGAG R-AAACAGGTCTTGATTAGTGC	247-255	3	0.480	0.500	0.555	+0.0145	Florescu *et al*., 2003
DB49	F-GTCAGGCATTGGCAGGAAG R-CAGGGTAAGTGGGGGTGC	159-163	3	0.578	0.654	0.655	-0.0266	Brown & Houlden 1999
RHI32A	F-CAGTCCTGCTGCATAAATCTC GCAGTACAGCTAGAATCACC	234-248	2	0.406	0.548	0.513	-0.0559	Florescu *et al*., 2003
RHI7B	F-CCTCTGTGATTAAGCAAGGC R-ATGAACAGGAAGGAAGACGC	261-269	5	0.461	0.438	0.519	+0.0799	Florescu *et al*., 2003
DB1	F-AGATAATAATAGGACCCTGCTCCC R-GGAGGTTTATTGTGAATGAGGC	129-131	2	0.259	0.273	0.298	+0.0120	Brown & Houlden 1999

The 32 individuals consisted of 11 known mother/offspring groups with two mothers having two offspring as illustrated in Figure 17. There were a further seven juvenile samples, which did not have known mothers in the data set as well as one adult female without any offspring. The data set included four adult male samples presumed from observational data to be the possible paternal candidates for all 11 juvenile samples with known mothers. Parentage analysis was conducted with all ten adult females as maternal candidates group, all four adult males as paternal candidate group against all 18 juveniles as offspring set. The summary of parentage assignment for maternal candidates is given in Table 22 and paternal candidates in Table 23. The SNP dataset achieved a combined first parent non-exclusion probability of 0.0889, the MS data set 0.2755 and the combined data sets 0.0153. The combined second parent non-exclusion probabilities were 0.0072, 0.0735 and 0.0001 for SNP, MS and combined data sets respectively. Using the SNP data set, all 11 juveniles were correctly assigned to their mothers with no pair loci mismatching noted. The MS data set correctly assigned ten out of the 11 parent offspring pairs. No pair loci mismatching was noted in any MS assignments including the wrong assignment of maternal candidate WR-22 to juvenile WR-110. In order to assess the effect of missing MS loci on the assignment of parentage, analysis on the assignment of mothers to a subset of samples; WR101 and WR44.1 was conducted. In both cases a reduction of MS loci from nine to five maintained correct assignment (positive LOD scores), however an absence of three markers resulted in a drop of the pair confidence from 95% to 80%. All 11 juveniles were correctly assigned using the combined data set with ten assignments having confidence of 95% and one with 80%.

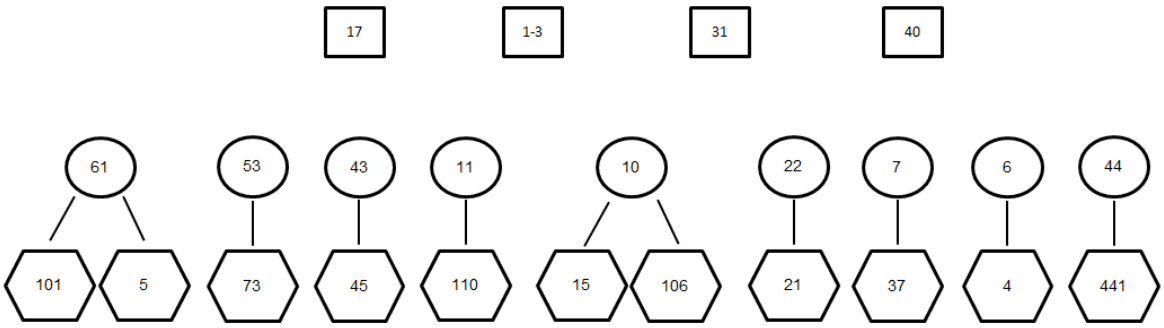


Figure 17: Diagram illustrating the known relationships between 24 rhino samples. Rectangles indicate potential paternal candidates, ovals the maternal samples and hexagons the offspring.

Using the SNP data set, six paternal allocations could be made with 95% confidence. Using the MS data set, two paternal allocations can be made with 95% confidence and six with 80% confidence. Two of the allocations with positive scores correspond between the two data sets. Using the combined data set, five paternal allocations can be made with 95% confidence. Table 24 includes the parentage assignments when siblings WR-101/WR-5 and WR-15/WR-106 is included in the pool of

maternal candidates. Using only SNP data, the correct maternal candidates are assigned to WR-5 and WR-106. WR-5 was wrongly assigned as best maternal candidate for both WR-101 and WR-15. Using only MS data the assignments are correct except for WR-101 which has a higher LOD score than the true mother for WR-106. Using the combined data sets the assignments are all correct with 95% confidence. All other assignments remained as stated in Table 22.

Table 22: CERVUS parentage assignment for maternal candidates showing the two most likely candidates. *=95% confidence; +=80% confidence; incorrect assignments marked in grey.

	SNP				MS				Combined			
Offspring ID	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score
WR-5	WR-61	33	0	6.97 (*)	WR-61	8	0	0.69	WR-61	41	0	7.66 (*)
	WR-61	33	0	6.97 (*)	WR-61	8	0	0.69	WR-61	41	0	7.66 (*)
WR-101	WR-61	33	0	4.29 (*)	WR-61	9	0	5.76 (*)	WR-61	42	0	10.05 (*)
	WR-22	33	1	-2.12	WR-61	9	0	5.76 (*)	WR-22	42	2	-5.21
WR-73	WR-53	33	0	5.08 (*)	WR-53	7	0	0.32	WR-53	40	0	5.43 (*)
	WR-53	33	0	5.08 (*)	WR-43	7	0	1.90	WR-53	40	0	5.43 (*)
WR-45	WR-43	33	0	6.34 (*)	WR-43	8	0	1.25 (+)	WR-43	41	0	7.62 (*)
	WR-53	33	0	3.92	WR-11	8	0	-0.58	WR-53	41	1	0.21
WR-110	WR-11	33	0	1.81 (*)	WR-22	6	0	0.94	WR-11	39	0	0.06 (+)
	WR-53	33	1	-1.80	WR-22	6	0	0.94	WR-53	39	1	-2.82
WR-15	WR-7	33	3	-9.02	WR-10	8	0	0.77	WR-10	41	0	4.73 (*)
	WR-10	33	0	3.97 (*)	WR-10	8	0	0.77	WR-10	41	0	4.73 (*)
WR-106	WR-10	33	0	3.42 (*)	WR-10	8	0	1.98 (+)	WR-10	41	0	5.40 (*)
	WR-10	33	0	3.42 (*)	WR-61	9	1	-1.07	WR-10	41	0	5.40 (*)
WR-21	WR-44	33	0	4.61	WR-22	7	0	1.31	WR-22	40	0	7.16 (*)
	WR-22	33	0	5.84 (*)	WR-22	7	0	1.31	WR-22	40	0	7.16 (*)
WR-37	WR-7	33	0	4.72 (*)	WR-7	7	0	2.06	WR-7	40	0	6.80 (*)
	WR-7	33	0	4.72 (*)	WR-7	7	0	2.06	WR-7	40	0	6.80 (*)
WR-4	WR-6	33	0	4.79 (*)	WR-6	7	0	1.10	WR-6	40	0	4.91(*)
	WR-11	33	1	-1.48	WR-6	7	0	1.10	WR-43	40	3	-10.37
WR-44.1	WR-44	33	0	5.70 (*)	WR-44	9	0	3.94 (*)	WR-44	42	0	9.00 (*)
	WR-53	33	0	3.56	WR-57	9	0	0.57	WR-44	42	0	9.00 (*)

Table 23: CERVUS parentage assignment for paternal candidates showing the two most likely candidates. *=95% confidence; +=80% confidence

	SNP				MS				Combined			
Offspring ID	Candidate father ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Candidate father ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Candidate father ID	Pair loci compared	Pair loci mismatching	Pair LOD score
WR-5	WR-17	33	1	-0.38	WR-17	8	0	0.32	WR-17	41	1	-0.06
	WR-1-3	33	1	-3.93	WR-1-3	8	0	0.95 (+)	WR-1-3	41	1	-2.98
WR-101	WR-31	32	0	4.89 (*)	WR-31	9	0	-0.67	WR-31	41	0	4.22 (*)
	WR-31	32	0	4.89 (*)	WR-40	9	1	-5.88	WR-31	41	0	4.22 (*)
WR-73	WR-17	33	1	-1.88	WR-17	7	0	1.08 (+)	WR-17	40	1	-0.80
	WR-31	32	1	-0.99	WR-31	7	0	0.86	WR-31	39	1	-0.11
WR-45	WR-17	33	0	1.86 (*)	WR-17	8	0	1.70 (*)	WR-17	41	0	3.57 (*)
	WR-17	33	0	1.86 (*)	WR-17	8	0	1.70 (*)	WR-17	41	0	3.57 (*)
WR-110	WR-31	32	0	6.01 (*)	WR-1-3	6	0	-0.02	WR-31	38	0	5.43(*)
	WR-31	32	0	6.01 (*)	WR-17	9	0	-0.64	WR-31	38	0	5.43(*)
WR-15	WR-17	33	3	-9.91	WR-17	9	0	0.07 (+)	WR-17	42	3	-9.84
	WR-31	32	3	-15.33	WR-1-3	9	0	-1.43	WR-1-3	42	2	-9.02
WR-106	WR-17	33	1	0.90 (*)	WR-40	9	0	0.73 (+)	WR-31	41	3	-13.40
	WR-31	32	1	-5.71	WR-40	9	0	0.73 (+)	WR-17	42	4	-10.45
WR-21	WR-17	33	0	3.63 (*)	WR-31	7	0	1.52 (+)	WR-17	40	1	-0.45
	WR-17	33	0	3.63 (*)	WR-40	7	0	0.93	WR-31	39	2	-6.23
WR-37	WR-31	32	0	1.63	WR-31	9	0	0.21 (+)	WR-31	39	0	1.86 (*)
	WR-1-3	33	0	2.63 (*)	WR-40	7	0	-0.37	WR-1-3	40	1	-1.82
WR-4	WR-31	32	1	0.10	WR-31	7	0	1.17	WR-31	39	1	0.34(*)
	WR-31	32	1	0.10	WR-40	7	0	0.48	WR-31	39	1	0.34(*)
WR-44.1	WR-17	33	1	-1.73	WR-1-3	9	0	2.36 (*)	WR-17	42	1	-2.73
	WR-17	33	1	-1.73	WR-1-3	9	0	2.36 (*)	WR-1-3	42	4	-12.38

Table 24: CERVUS parentage assignment for maternal candidates including siblings showing the two most likely candidates. *=95% confidence; +=80% confidence; incorrect assignments marked in grey.

	SNP				**MS**				**Combined**			
Offspring ID	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score	Candidate mother ID	Pair loci compared	Pair loci mismatching	Pair LOD score
WR-5	WR-61	33	0	6.97 (*)	WR-61	8	0	0.69	WR-61	41	0	7.66 (*)
	WR-61	33	0	6.97 (*)	WR-61	8	0	0.69	WR-61	41	0	7.66 (*)
WR-101	WR-5	33	0	4.76	WR-61	9	0	5.76 (*)	WR-61	42	0	10.05 (*)
	WR-61	33	0	4.29	WR-61	9	0	5.76 (*)	WR-5	41	2	-2.90
WR-15	WR-7	33	3	-9.02	WR-10	8	0	0.77	WR-10	41	0	4.73 (*)
	WR-5	33	1	1.18	WR-10	8	0	0.77	WR-10	41	0	4.73 (*)
WR-106	WR-10	33	0	3.42 (*)	WR-101	9	0	2.73 (+)	WR-10	41	0	5.40 (*)
	WR-10	33	0	3.42 (*)	WR-10	8	0	1.98	WR-10	41	0	5.40 (*)

Discussion

Together with the 12 new SNPs identified in this study, 33 SNPs are now available for white rhino (Labuschagne *et al*., 2013, Labuschagne *et al*., 2015). The SNPs, were discovered through random selection and sequencing of cloned fragments from a SNP enriched library. Ascertainment bias is often a concern when using SNPs in population studies, with bias introduced by heterogeneity in the SNP discovery process, varying sample sizes or differences in sample composition leading to underestimation or overestimation of the frequency of SNPs (Nielsen and Signorovitch, 2003; Clark *et al*., 2005). Ascertainment of SNPs through discovery in particular populations or genomic regions does not bias the results of parentage inference in any way since the parentage analysis is not concerned with the inference of evolutionary history (Anderson and Garza, 2005). In effect, SNP ascertainment leads to an advantage in parentage inference, since ascertainment typically leads to an overrepresentation of intermediate allele frequency SNPs, the type of loci that are most powerful for parentage (Anderson and Garza, 2005). The SNP loci presented here contain extra flanking data to allow for Sanger sequencing. Shorter amplicons may be designed in the future to optimise their utility in degraded DNA samples.

To our knowledge this is the first study to employ SNP and MS markers for parentage analysis in white rhino. The current SNP set out performed the MS set during maternal assignment, where all assignments were correct while the MS data set allocated one maternal sample incorrectly. In general assignments made with the SNP data set had higher confidence than those with the MS set. Confidence levels increased when combining the two data sets. The increased accuracy of the SNP markers in this study over MS markers can be attributed to the greater marker numbers in the SNP data set and low allele numbers of the MS markers. It is apparent that low levels of genetic diversity characterise white rhino populations and the results from our study (Na=2.7; PIC=0.4282) are consistent with other studies making use of MS markers. Harper and colleagues (2013) reported Na=2.722 and PIC=0.329 for 367 rhino samples, while Guerier and colleagues (2012) reported Na=2.72 and PIC=0.357 in a sample set of 31 individuals. Florescu *et al*., (2003) observed higher values, Na=2.8 and PIC=0.4812 in a sample set of 30 individuals, but selected specifically for highly polymorphic loci, which may account for the elevated in their data. The low levels of genetic diversity observed in white rhinos may be attributed to the small (20-40 individuals) founder population and subsequent bottleneck (Walker and Walker, 2012),

Challenges to parentage assessment can arise when family members other than the parents of the offspring are included in the pool of candidate parents (Jones and Ardren, 2003). Inclusion of either half- or full-siblings in the pool of candidate parents may pose the most problematic situations (Jones and Ardren, 2003). In the current study two pairs of siblings were available to evaluate the effect on assignment when included in the parental pool. Inclusion led to some wrong assignments when using the two marker sets separately, but not in the combined data set. It would seem that the combined data set has enough discriminating power for accurate assignment in the current population even when siblings are included in the parental pool, however as relatedness levels increase so should the

number of markers. In extremely inbred populations this may reach prohibitive numbers. Further assessment of the utility of these markers over multiple (> three) generations and the incorporation of a larger variety of relationships among individuals (e.g. half-siblings or cousins) as well as a larger set of samples is strongly suggested.

References

Anderson E.C. and Garza J.C. 2006. The power of single nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172: 2567-2582.

Brown S.M. and Houlden B.A. 1999. Isolation and characterization of microsatellite markers in the black rhinoceros (*Diceros bicornis*). *Molecular Ecology*, 8: 1559-1561.

Butler J.M., Coble M.D., Vollone P.M. 2007. STRs vs. SNPs: thoughts on the future of forensic DNA testing*. Forensic Sci Med Pathol,* 3: 200-205.

Clark A.G., Hubisz M.J., Bustamante C.D., Williamson S.H., Nielsen R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res,* 15: 1496-1502.

Cunningham J., Harley E.H., O'Ryan C. 1999. Isolation and characterization of microsatellite loci in black rhinoceros (*Diceros bicornis*). *Electrophoresis,* 20:1778-1780.

Emslie R.H. and Brooks M. 1999. African rhino: status survey and conservation action plan. Gland and Cambridge, IUCN/ SSC African Rhino Specialist Group, pp. i-x, 1-92.

Emslie R. 2012. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 21 April 2015.

Florescu A., Davila J.A., Scott C., Fernando P., Kellners K., Morales J.C., Melnick D., Boag P.T., van Coeverden de Groot P. 2003. Polymorphic microsatellites in white rhinoceros. *Mol Ecol Notes,* 3: 344-345.

Glover K.A., Hansen M.M., Lien S., Als T.D., Høyheim B., Øystein S. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics,* 11: 2.

Guerier A.S., Bishop J.M., Crawford S.J., Schmidt-Kuentzel A., Stratford K.J. 2012. Parentage analysis in a managed free ranging population of southern white rhinoceros: genetic diversity, pedigrees and management. *Conservation Genetics,* 2012: 1-14.

Harley E.H., Baumgarten I., Cunningham J., O'Ryan P. 2005. Genetic variation and population structure in remnant populations of Black Rhinoceros, *Diceros bicornis*, in Africa. *Molecular Ecology,* 14: 2981-90.

Harper C.K., Vermeulen G.J., Clarke A.B., De Wet J.I. 2013. Extraction of nuclear DNA from rhinoceros horn and characterization of DNA profiling systems for white (*Ceratotherium simum*) and black (*Diceros bicornis*) rhinoceros. *Forensic Science International: Genetics,* 7: 428-433.

Hsieh H., Huang L., Tsai L., Kuo Y., Meng H., Linacre A., Lee J.C. 2003. Special identification of rhinoceros horns using the cytochrome b gene. *Forensic Science International,* 136: 1-11.

Jones A.G. and Ardren W.R. 2003. Methods of parentage analysis in natural populations. *Mol. Ecol*. 12: 2511-2523.

Kalinowski S.T., Taper M.L., Marshall T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol*. 16: 1099-1106.

Labuschagne C., Kotzé A., Grobler J.P., Dalton D.L. 2013. A targeted gene approach to SNP discovery in the white rhino (*Ceratotherium simum*). *Conservations Genetics* Resource, 5 (1): 265-266.

Labuschagne C., Kotzé A., Grobler J.P., Dalton D.L. 2015. Endonuclease V digestion for SNP discovery and marker development in South African white rhinoceros (*Ceratotherium simum*). *Conservation genetics resources*, 7(1): 53-56.

Morin P.A., Luikart G., Wayne R.K. 2004. SNP workshop group. SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution,* 19(4): 208-216.

Nielsen R. and Signorovitch J. 2003. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theor Popul Biol,* 63: 245-255.

Pertoldi C., Bijlsma R., Loeschcke V. 2007. Conservation genetics in a globally changing environment: present problems, paradoxes and future challenges. *Biodivers Conserv,* 16: 4147-4163.

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

Rohrer G.A., Alexander L.J., Keele J.W., Smith T.P., Beattie C.W. 1994. A microsatellite linkage map of the porcine genome. *Genetics,* 136: 231-244.

Ryynänen H.J. and Primmer C.R. 2006. Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primers exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. *BMC Genomics,* 7: 192.

Seror B., Zorovsky Y., Terkel A., Katcoff D.J. 2002. Genetic Kinship and Social Structure in a Herd of Square-Lipped Rhinoceroses (*Ceratotherium simum simum*) at the Zoological Center, Tel Aviv/Ramat-Gan, Israel. *Zoo Biology,* 21: 561-571.

Selvin S. 1980. Probability of nonpaternity determined by multiple allele codominant systems. *Am J Hum Genet,* 32: 276-278.

Smith C.T., Baker J., Park L., Seeb L.W., Elfstrom C., Abe S., Seeb J.E. 2005. Characterisation of 13 single nucleotide polymorphism marker for chum salmon. *Molecular Ecology Notes,* 5: 259-262.

Väli U., Einarsson A., Waits L., Ellegren H. 2008. To what extent do microsatellites markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology,* 17: 3808-3817.

Walker C. and Walker A. 2012. The Rhino Keeprs, Jacana Media (Pty) Ltd., Johannesburg.

CHAPTER 11

Conclusion

A central issue in conservation genetics is the level of genetic variation present, a prerequisite for evolution. The rates of adaptive evolution need to, at least, match the rate of environmental change in order for a population to persist. Evaluation of levels of genetic diversity is therefore particularly important in conservation genetics. African Penguin (*Spheniscus demersus*) populations have decreased substantially in South Africa and Namibia in the 20th century with fewer than 31,000 breeding pairs left and still declining. Population decline are due to a number of factors including competition for food, oil spills, loss of habitat and environmental change. The African White rhinoceros (*Ceratotherium simum*) population has suffered a decline over the past 150 years as a result of overhunting, habitat destruction and poaching with the current estimated population comprising 20,170 individuals. The trade in rhinoceros horns is a problem in many parts of the world especially in parts of Asia where the rhinoceros horns are used traditionally as material in sculptures or as drug products for medicinal purposes adding constant pressure on remaining populations.Comprehensive management plans for any species of conservation concern including the African penguin and white rhinoceros should contain plans for maintaining existing genetic diversity, both to ensure ability to adapt to changing environments and to preserve the possibility of future speciation.

Firstly, in African Penguin this is the first report of the complete nucleotide sequence for the mitochondrial genome. The *S. demersus* mtDNA genome was found to be very similar, both in composition and length to both the *E. chrysocome* and *E. minor* genomes. The gene content of the African Penguin mitochondrial genome is typical of vertebrates and all three penguin species have the standard gene order originally identified in the chicken. The control region for *S. demersus* is located between tRNA-Glu and tRNA-Phe and all three species of penguins contain two sets of similar repeats with varying copy numbers towards the 3' end of the control region, accounting for the size variance. These results can be subsequently used to provide information for penguin phylogenetic studies and insights into the evolution of genomes. Mitochondrial data however, only allows the tracing of maternal lineages.

Eight microsatellite (MS) markers were developed via pyrosequencing of a microsatellite-enriched library for the African Penguin. These microsatellite loci displayed 2 to 6 alleles with expected heterozygosity values ranging between 0.316 and 0.782 and observed heterozygosity between 0.381 and 0.84. These loci may be suitable for assessing patterns of genetic variability in African Penguin. This is the first development of species-specific markers for the African Penguin. In this study, molecular genetic data based on 12 microsatellite markers from 131 first and second generation penguin samples from four facilities was generated in order to determine the level of genetic variation, population structure and differentiation, and effective population size to assist in the development of

an effective captive management plan. Expected heterozygosity ranged from 0.541 to 0.618 and allelic richness from 5.0 to 5.7. However, based on differences between first and second generation captive birds, it was concluded that the *ex-situ* population is at risk of losing genetic variability in the future and therefore, management programmes should include exchange of birds between captive facilities in order to induce gene flow and increase effective population size. Microsatellite loci may yield unreliable estimates of divergence times and gene flow among populations due to an incomplete understanding of the underlying mutation model. These markers are often subject to high mutation rates which may lead to homoplasy.

I report on the identification of 30 new single nucleotide polymorphism markers for the endangered African Penguin, based on screening of a random genomic library. The SNP loci were assessed using a captive population comprising 34 individuals. These SNP markers will provide a necessary addition to the genetic tools employed for understanding population structure and for developing a conservation management strategy for this endangered species.

The baseline assessment of genetic diversity and population structure is an important first step for the establishment of a genetic monitoring program for the African Penguin. As the African Penguin population continues to decline, *ex-situ* breeding programs are increasingly being considered as potentially valuable conservation management tools and these programs are being implemented on a small scale by a number of rehabilitation centres, zoos, aquaria and conservation authorities in South Africa. The *ex-situ* population may be at risk of losing genetic variability over time due to genetic drift, as seen from the loss of allelic diversity and absence of low frequency alleles in the offspring generation. In order to buffer the potential negative effects of genetic drift, management strategies should focus on increasing the number of individuals in the *ex-situ* populations. Natural or artificial gene flow between the captive and natural populations could enhance effective population size and could potentially restore genetic variation within this small and vulnerable population.

The utility of Single Nucleotide Polymorphisms and microsatellites were compared and two analytical methods for assigning parentage in ten families of captive African Penguins held in South African facilities. It was found that SNPs performed better than microsatellites under both analytical frameworks, but a combination of all markers was most informative. Captive or supportive breeding programmes will play an important role in future African Penguin conservation efforts as a source of individuals suitable for reintroduction. Cooperation among these captive facilities is essential to facilitate this process and improve the genetic management of the species. In general seabirds make good model organisms for DNA-based research into evolution and ecology as they can travel great distances, violating assumptions of many population divergence and speciation models, generally exhibit natal philopatry and breed in large colonies from which sufficient samples can be collection for robust genetic analysis. Seabird reliance on the marine environment also makes them good candidate species to study in relation to climate change and anthropogenic effects.

Utilising the white rhinoceros as subject species, a simple, rapid and cost effective method is described to isolate candidate SNPs in non-model organisms using the commercially available Endonuclease V enzyme. Developed SNP markers in white rhinoceros could be used to define the genetic mating system of this species, for forensic applications and to determine population structure and variability when other markers prove problematic. Utilising next generation sequencing, SNPs may be discovered on a much larger scale from enriched reduction libraries compared to the study presented here. Due to the nature of the heteroduplex formation, the Endo V enrichment protocol may also increase the number of false SNPs identified from paralogous regions.

Using a targeted gene approach and Endo V enrichment as discovery approaches, 33 SNPs are now available for white rhino. Typing these SNPs in 32 individuals showed PIC values ranging from 0.060 to 0.396 with a mean of 0.2742. The observed and expected heterozygosity ranged from 0.065 to 0.656 and from 0.063 to 0.520. Typing nine MS markers in the same population showed PIC values ranging from 0.259 to 0.578 with a mean of 0.4282, while observed and expected heterozygosity ranged from 0.273 to 0.654 and from 0.298 to 0.655, respectively. The low levels of genetic diversity observed in white rhinoceros may be attributed to the small (20-40 individuals) founder population and subsequent bottleneck.

In the studied white rhinoceros population the SNP dataset achieved a combined first parent non-exclusion probability of 0.0889, the MS data set 0.2755 and the combined data sets 0.0153. The combined second parent non-exclusion probabilities were 0.0072, 0.0735 and 0.0001 for SNP, MS and combined data sets respectively. Using the SNP data set, all 11 juveniles were correctly assigned to their mothers with no pair loci mismatching noted. Challenges to parentage assessment can arise when family members other than the parents of the offspring are included in the pool of candidate parents. Inclusion of either half- or full-siblings in the pool of candidate parents may pose the most problematic situations. In the current study two pairs of siblings were available to evaluate the effect on assignment when included in the parental pool. Inclusion led to some wrong assignments when using the two marker sets separately, but not in the combined data set. It would seem that the combined data set has enough discriminating power for accurate assignment in the current population even when siblings are included in the parental pool, however as relatedness levels increase so should the number of markers. In extremely inbred populations this may reach prohibitive numbers. These markers should be further investigated for applications in other species such as the endangered black rhinoceros (*Diceros bicornis*).

Single nucleotide polymorphism (SNP) markers are a promising new tool that can be used to study evolutionary processes, population genetic parameters, forensic cases and parentage. These SNP markers could be employed to provide estimates of parameters such as population structure, relatedness and current and historical gene flow. However, application of SNP marker analysis to wildlife has been limited, due to the lack of available sequence data in non-model organisms.

Information derived with the genetic markers described here in African Penguin and white rhinoceros may be useful for estimations of genetic diversity, examining population structure, investigating mating systems (e.g. the frequency of extra-pair copulation and fertilization and mate-choice for penguins) in wild populations as well as developing a conservation management strategy for these endangered species. The utility of these markers should also be investigated for related species. The management of *ex-situ* populations of endangered species is traditionally based on pedigree information derived from studbook data. Adding molecular techniques can contribute to improving the management of the species as these methods provide a powerful set of complementary tools to verify studbook records and to contribute to an improved understanding of the genetic status of captive populations.