Development of Genetic Tools for Managing Populations of the Southern White Rhinoceros (*Ceratotherium simum simum*)

Tarid Purisotayo

Submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy

Institute of Biodiversity Animal Health and Comparative Medicine

College of Medical, Veterinary & Life Sciences University of Glasgow

University of Glasgow

June 2020

Abstract

The southern white rhinoceros (SWR) is one of five extant species of rhinoceros. The species experienced historical bottleneck due to unrestricted hunting and was on the brink of extinction during the end of nineteenth century, with only one population remaining in South Africa. This population was intensively protected and as it subsequently increased, it became the source of SWR for all of Southern Africa. With advances in immobilisation and translocation techniques, a surplus SWRs were relocated to the former range states of the species. Therefore, most or all modern populations of SWRs originated from the single founder population. As in other former range states, Botswana re-established SWR populations, but poaching remained an imminent threat to the national herd and the species was almost wiped out for a second time. In response to this threat, the Department of Wildlife and National Parks of Botswana (DWNP) began to capture free ranging animals and relocate them into enclosed reserves where they could be protected intensively by anti-poaching teams. Subsequently the population size of the national herd has been increasing due to successful breeding together with the continued introduction of SWRs from South Africa. Although this conservation strategy has been successful, it has generated many fragmented populations, which required regular exchanges of animals to prevent inbreeding. However, selection of animals for translocation has been made based on observational data about the relationships among animals and genetic information has rarely been used. The efficient identification of candidates for translocation, requires an accurate and complete pedigree to determine the individuals with high risk of producing inbred progeny. In this thesis, three populations of SWRs in Botswana (Botswana1, Botswana2, and Botswana3) were used as models to develop genetic tools that would facilitate metapopulation management.

The purpose of Chapter 2 was to integrate previously characterised microsatellites (MS) genotypes with an incomplete, field-observed pedigree to make inferences about mean kinship and basic demographic data that could be used to inform translocation programmes for SWR. Level of heterozygosity and genetic diversity of the population were not as low as initially expected based on the severe bottleneck, but the population showed a very low mean number of alleles per locus. Using several different strategies for exclusion of unlikely parents, parental pairs of 29 out of 45 offspring could be assigned confidently. The combined pedigree was constructed from the assignable parent-offspring relationships and subsequently used to estimate kinship coefficients. Based on population mean kinship

(MK), eight bulls with high individual MK could be identified as the best candidates for translocation. The pedigree was further used to estimate population demographic parameters; importantly, the reproductive dominance of the bulls was not as skewed as expected after the original dominant bull was removed, suggesting that closed populations can maintain multiple, simultaneously breeding males.

Because the currently available markers (i.e. microsatellites) did not provide sufficient analytical power to construct a complete pedigree, a sequencing method that would allow marker discovery and genotyping in non-model species was required. A commonly used complexity reduction approach (double digestion restriction-associated DNA sequencing; ddRADseq) for identifying genome-wide single nucleotide polymorphisms (SNPs) was initially attempted. However, signs of DNA degradation were evident for nearly one third of the samples, which made the ddRADseq approach impractical. Thus, in Chapter 3, I tested the efficiency of an approach (RADcapture) that uses hybrid sequence capture to enrich the genome for SNPs identified by ddRADseq conducted on a set of high-quality DNA extractions. A total of 32 samples was chosen based on their molecular weight judged from 1% agarose gel electrophoresis; these were divided into two groups corresponding to their qualities. RADcapture identified 6,481 SNPs and performed equally well in both groups of samples, and there was no relationship between the quality of samples and the performance of the protocol. This suggested that hybrid capture could be useful for resolving SNPs in both high- and low-quality samples.

In Chapter 4, RADcapture was applied to a collection of samples from the three managed populations to assess the utility of applying this approach to population management across metapopulation. Using RADcapture, 302 SNPs could be genotyped consistently across all individuals. For the Botswana1, these markers were used for parentage analysis, for comparison with the combined pedigree in Chapter 2, and for construction a consensus pedigree. Seven offspring for which MS were not effective could be assigned using SNPs, indicating better resolution by SNPs. The consensus pedigree could be constructed and was subsequently used to estimate pedigree-based kinship coefficients that suggested six and eight individuals as the best candidates for translocation and for breeding, respectively. Four of the six candidates for translocation were male; of which one were in agreement with the suggestions made in Chapter 2, the other two were the SNP-assigned fathers that involved four cases that SNPs provided better resolution. This suggested the potential effects of pedigree completeness on the candidates identified. RADcapture data were also used to estimate marker-based kinship coefficients for all three populations. For the

Botswana1 population, marker-based kinships identified different individuals from the candidates suggested by the pedigree-based MK. The disagreements suggested that the latter optimised genetic contributions of animals, whereas the marker-based MK might instead promote the individuals that carried rare alleles. No candidates could be identified for Botswana2 or Botswana3 because there were no individuals with individual MK above and below the thresholds. Based on between-population MKs estimated from RADcapture data of the three populations, exchanges of animals between any pair of them would reduce population MK of the recipients. The principal component analysis revealed no genetic clusters observed across individuals from the three populations. Chapter 4 demonstrated the potential applications of RADcapture for parentage assignment and for identification of the candidates for translocation and breeding; however, completeness of the pedigree and the methods used to estimate kinships could affect the population management regarding the candidates identified.

In summary, the key outcomes of this thesis were 1) resolving the pedigree of a SWR population that has been an important source of animals for the national reintroduction programme in Botswana; 2) development of a sequencing method that allows the retrieval of genetic markers from DNA of various qualities; 3) demonstration of quantifiable methods (i.e. management based on kinship coefficients) that showed the potential to facilitate population management to prevent inbreeding in fragmented populations of SWRs; and 4) initiation of a genomic database obtained from RADcapture (i.e. RADcapture sequences) that could be used as the raw materials for various purposes of future applications (e.g. development of SNP array, wildlife forensics). These tools for genetic-based population management can now be applied to minimise inbreeding which is currently of particular concern for fragmented SWR populations. Most importantly, this thesis demonstrated approaches that are not applicable to only SWR, but can equally be applied in conservation programmes of other endangered species, i.e. sequencing methods for non-model species, methods for identification of candidates for translocation and breeding. The key outcomes present in this thesis should improve efficiency of the conservation of the species as well as other endangered species.

Table of Contents

s

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisors, Professor Nicholas Jonsson and Professor Barbara Mable who guided me through the most challenging time in my life. Thank you for teaching me both academic and writing skills, I could not imagine how difficult it would be to read and give feedbacks to articles of a non-native English speaker. It was such a pleasure and it will always be my honoured working with both of you. I just want to note here that COVID-19 did not affect the productivity of my supervisory team, I am very grateful for your massive efforts to supervise as if it was a normal time.

I also wish to extend my thanks to the Department of Wildlife and National Parks of Botswana for providing an opportunity to work with Botswana SWRs. The study would not be possible to complete without all the Botswana samples collected by Frederick Verreynne who also provided demographic information of the rhino populations. Things would get complicated without the additional samples from the European Endangered Species Programme, National Museums of Scotland, and Science and Advice for Scottish Agriculture (SASA) generously provided by Lucy Webster.

I would also want to thank all the team at the WildGenes Lab, Edinburgh Zoo. In particular, Dr Helen Senn and Dr Alex Ball who helped me designed the laboratory part of the thesis; Jenny Kaden and Jal Ghazali who taught me laboratory skills, without you all the complicated protocol would be impossible to complete; Dr Kara Dick who gave very helpful pieces of advice about bioinformatics.

Thanks must also go to my PhD fund from the Royal Thai Government and Ministry of Science and Technology of Thailand. Huge thanks to my colleagues at the Faculty of Veterinary Sciences at Maha Sarakham University for taking care of my classes during the time that I was in Scotland. Thanks to my friends at the IBAHCM, particularly Kevin Schneider and Arne Jacobs for your bioinformatic advice. Finally, I dedicate this work to my wife for her supports and for putting up with my stress, and most importantly your dishes are the best.

Author’s Declaration

"I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution."

Tarid Purisotayo

Signature:

List of Abbreviations

AfRSG	African Rhino Specialist Group
ACR	Annual calving rate
CI	Calving interval
ddRAD	Double-digest restriction associated DNA
DWNP	Department of Wildlife and National Parks of Botswana
ECC	Ecological Carrying Capacity
F	Inbreeding coefficient
H_o	Observed heterozygosity
H_e	Expected heterozygosity
HG	Herd growth rate
HWE	Hardy-Weinberg Equilibrium
IBD	Identical by descent (IBD).
IUCN	International Union for Conservation of Nature
MS	Microsatellite marker
MK	Mean kinship
N_a	Number of alleles per locus
N_e	Effective population size
NCBI	National Center for Biotechnology Information
NE-PP	Probability of non-exclusion for a candidate parental pair
NGS	Next-generation sequencing
PIC	Polymorphic information content
RADseq	Restriction-site associated DNA sequencing
SNP	Single Nucleotide Polymorphism
SWR	Southern white rhinoceros, *Ceratotherium simum simum*

Chapter 1 Introduction

1.1. Fragmentation of wildlife populations

Habitat loss and fragmentation have been considered important threats to wildlife conservation that limit movements of animals between forest patches (Hughes et al., 1997). Climate change and human-wildlife conflicts are mainly responsible for such events by spoiling and shrinking wild habitats. They consequently limit movements of individuals between previously connected forest patches, which warrants a need of intervention to introduce gene flow between isolated populations to avoid inbreeding and to ensure population viability (Frankham et. al., 2017). Climate change can cause effects on behaviours and distribution of the animals directly due to temperature modulated range shifts in species that specifically inhabit high-latitude or high-altitude habitats (Chen et al., 2011). Thousands of such wildlife species were reported to migrate away from the equator at a rate of 17.6 kilometers/decade and migrate uphill 12.2 meters/decade (Chen et al., 2011). For example, Caucasian snowcock (*Tetraogallus caucasicus*) and Caucasian grouse (*Lyrurus mlokosiewiczi*) were driven to shift their distributions to higher altitudes; given these rates of global warming and migrations, a simulation study suggested that there would be no suitable habitats left for these species by 2070 (Hof and Allen, 2019). Also, climate change can cause catastrophic events such as wildfire, rising of sea levels, flooding, drought, and changing of sea-ice extent, which could contribute to either temporary or permanent habitat loss (Barbraud and Weimerskirch, 2001, Alley and Gartrell, 2019). For instance, the colony and habitat sizes of a population of emperor penguins (*Aptenodytes forsteri*) in the Antarctic region were reduced due to the wider range of temperatures during the winters of 1973 - 1999 compared to the late 1950s (Barbraud and Weimerskirch, 2001). Increase of mortalities in adult penguins was also found to correlate with an increase of sea-surface temperatures and a decrease of sea-ice extent and hatching rates of eggs were reduced in periods with abnormally low temperatures (Barbraud and Weimerskirch, 2001). The Intergovernmental Panel on Climate Change (2014) estimated that an increase of global average temperature as low as ~1.5 °C can increase risk of wildlife extinction by ~20 - 30%. These examples described potential effects of climate change that forces wildlife species to migrate from their current habitats.

Habitat loss and fragmentation can be introduced not only by climate change but also by anthropogenic land uses that possibly cause an even more imminent and major threat to wildlife populations. Wade et al. (2003) used global land cover maps obtained from high-resolution radiometers to identify factors causing forest fragmentation during 1992 and 1993; humans were found to fragment or remove > 50% of temperate broadleaf and mixed forest biomes, and ~25% of tropical rain forests. Anthropogenic land uses can introduce a variety of physical and chemical barriers that prevent movement and interaction between wild animals. For example, construction of wind turbines was found to contribute to collision-induced mortalities and changes in habitat use in populations of birds across Europe and North America (Katzner et. al, 2013); more than 75,000 American dams showed an impact on wild fish and bird species that were reliant on riparian and upstream-flooded habitats, e.g. plain minnows and whooping cranes (Graf, 2006); improvement and construction of road networks in response to expanding human settlements led to an increase in the numbers of Eurasian badgers killed by vehicles in Central Italy (Fabrizio et al., 2019); chemical contamination in rivers of the southwestern region of Spain was reported to generate a chemical barrier, resulting in fragmentation of freshwater shrimp (*Atyaephyra desmarestii*; Araújo et al., 2019). These examples are associated with an increase of human needs (e.g. energy and food needs) and provide insights of how human activities can create a restriction of wildlife migration that would make the issues of climate change more complicated.

The combined effects of climate change and human-wildlife conflicts have posed a challenge in species conservations because the viability of wildlife can be dependent on their ability to migrate to a new favourable habitat when environments change, which can be challenging due to anthropogenic barriers. Hughes et al. (1997) estimated that there were ~1.9 billion isolated eukaryotic populations and~29% of world vertebrates that have been affected by fragmentation. Fragmentation can limit reproductive behaviours due to reduction of movements between forest patches, which consequently lowers population sizes of wild populations and potentially reduce genetic diversity (Fletcher Jr et al., 2018). For example, Schiegg et al. (2006) demonstrated that the inbreeding coefficients and the percentage of inbred individuals were higher in the populations of endangered red-cockaded woodpecker (*Picoides borealis*) that inhabited isolated forest patches compared to other poulations of the same size. A metanalysis study of 83 plant and 52 animal species revealed that number of alleles, expected heterozygosities, and proportions of polymorphic loci negatively responsed to numbers of years and generations of fragmentation

(Schlaepfer et al., 2018).Such high inbreeding coefficients and low genetic diversity were shown to correlate with several traits of inbreeding depression (e.g. juvenile survival rates, body mass, numbers and sizes of egg clutches, rates of hatchling, recapture frequency, seed production) in seven bird, nine mammal, four ectotherm, and 15 plant populations (Crnokrak et al., 1999). Therefore, effects of fragmentation on genetic diversity is an important negative consequence of population fragmentation and may affect the persistence of species.

1.2. Introduction of protected areas in Africa

Africa has been a home of a wide range of animal species, making the continent an important region for conservation of global biodiversity. Wildlife richness is one of the important indicators for monitoring of biodiversity (Pitman et al., 2017) but information about the actual number of African wildlife before and during the arrival of European visitors in the early 18th century (pre-colonisation and colonisation eras) was scarce and rarely recorded in precise detail. Only documents written by groups of early European visitors during the 1830s are available and provide descriptions about the abundance of African wild animals. For example, Harris (1838) described that it was easy to spot large mammalian species such as common warthog (*Phacochoerus africanus*), white rhinoceros (*Rhinoceros sinusus*), waterbuck (*Aigocerus ellipsiprymnus*), pig-faced baboon (*Cynocephalus porcarius*), African elephant (*Elephas africanus*), hippopotamus (*Hippopotamus amphibius*), lion (*Felis leo*) and giraffe (*Camelopardis giraffa*). Note that the scientific names provided for these species were obtained from the original document and might be different from the scientific name used in modern-day terms. Wildlife usage by native Africans at the time was mainly for household consumption as a source of protein, cloth, adornment, and medicines. Therefore, the local communities during the pre-colonial era were not responsible for the significantly reduced wildlife richness (Mkumbukwa, 2008). The arrival of Europeans and the colonialism during the late 19th century led to hunting of wildlife as trophies and sources of luxury products. International trade routes of hunted products had started in the 18th and 19th centuries, when 900,000 kg of ivory were exported yearly, corresponding to ~53,000 elephants being killed annually on average during 1830 and 1930 (Macgregor, 1989).

Although hunting was an important reason behind the decline of African wildlife, there were other factors that contributed to such events. These factors included transmissible

diseases, military conflicts and population fragmentation of wildlife due to an increase of African human populations. For example, an outbreak of Rinderpest disease during the late 1880s and early 1890s killed 90% of Kenya's buffalo populations and further caused an indirect effect on predator populations (Daszak et al., 2000). Intensive military conflicts between 1946 and 2010 were found to correlate with the decline of 36 large mammal species that inhabited 126 protected areas in 19 African countries (Daskin and Pringle, 2018). However, the effects of armed conflicts on wildlife numbers indicated that the conflicts might cause either positive or negative consequences on large mammal species (Gaynor et al., 2016). For example, armed conflicts could make the hunting areas inaccessible to poachers, while displaced villagers might have to hunt for food (Gaynor et al., 2016). Most importantly, rapid increase of the African human population, which is expected to reach 2.5 billion in 2050 (Gerland et al., 2014) has contributed to major threats for terrestrial wildlife species by creating human-influenced habitat losses and population fragmentations. At least 177 terrestrial mammalian species have experienced habitat losses, 56% of which were Africa-based mammals whose distribution ranges declined by more than 80% during the period of 1900 – 2015 (Ceballos et al., 2017). One way to avoid species extinctions driven by these eminent threats is through establishment of secured protected areas for wildlife. The first protected area in Africa was proclaimed in 1895 at the area now known as Hluhluwe-Umfolozi Game Reserve, South Africa (Macgregor, 1989). Since then, wildlife conservation in Africa has been dominated by the establishments of protected areas; as of 2015, the world's protected areas cover 15.4% of the world surface (Barnes et al., 2016). However, to be successful, such areas require intensive management to protect species and reduce risks associated with habitat fragmentation and isolation of populations.

One species that has been particularly impacted by being restricted to small, intensively managed populations due to a variety of threats is the southern white rhinoceros (*Ceratotherium simum simum*; SWR). The species has been long known to be affected by expansion of human settlements in Africa; for example, the SWRs in Kruger National Park had to be removed in the 1960s partly due to the land conflicts with neighbouring human communities (Harthoorn, 1962). Consequently, the ongoing habitat loss has prompted the need to shelter the species in protected areas, which aggravates issues of fragmentation. The species has also suffered from intensive hunting owing to the high value of its body parts, particularly horns (du Toit, 2015). Such effects can be exacerbated by climate change. For example, a reduced birth rate and an increase of the non-poaching related

death rate were reported as the result of an unusual drought that occurred in the rainy season of 2015 – 2016 in Kruger National Park, South Africa (Ferreira et al., 2019). Although, there was no evidence whether the drought was actually caused by climate change, the authors argued that unpredictable rainfall in the park, which has been increasingly frequent (Smalma, 2016), might be the result of climate change (Ferreira et al., 2019). Climate change can also potentially cause an indirect impact on SWR conservation; the incidence of SWR hunting may be increased due to low crop yield and poor production of livestock and therefore people are driven to find alternative sources of incomes (Kideghesho, 2016). It thus makes a good model to investigate how management practices could be changed to mitigate effects of both climate change and habitat fragmentation.

1.3. Conservation history of the southern white rhinoceros

Rhinoceros were one of the many species complexes that were saved by the introduction of protected areas in Africa, just in time when the species were on the brink of extinction. Two subspecies of the African white rhinoceros (WR), the northern (*Ceratotherium simum cottoni*, NWR) and the southern white rhinoceros were separated by ~2,000 kilometers from each other (Rookmaaker and Antoine, 2012); molecular evidence (based on the mitochondrial control region and microsatellites) indicated that the last post-divergence contact (i.e. the latest contact after speciation) between them happened during the last glacial maximum which was ~ 30,000 years ago (Moodley et al., 2018). The NWR historically ranged in Uganda, South Sudan, Chad, Central Africa Republic, and DR Congo (Groves et al., 2017), whereas the SWR inhabited the area between the Orange and Zambezi rivers (Harris, 1838, Pienaar, 1970, Cumming et al., 1990), now designated as Botswana, Namibia, Swaziland, South Africa and Zimbabwe (Rookmaaker and Antoine, 2012). However, the argument regarding whether NWR and SWR should be recognised as two distinct species or two subspecies remains controversial (Harley et al., 2016, Groves et al., 2017, Moodley et al., 2018). The delineation between the two subspecies has become more critical after the death of the world's last NWR male, Sudan, which left only two close-related and post-reproductive females in captivity because the extinction of the NWR would be inevitable if the two subspecies were to be treated as two distinct conservation units. External morphology including dental metrics, cranial growth, and craniometry clearly differentiated the southern from the northern white rhinoceroses (Groves, 1972,

Groves et al., 2010). However, genetic studies initially gave inconsistent results across studies using different techniques for species differentiation. The two species typically have the same number of chromosomes (2n = 82); although, three (out of 9) captive NWRs from San Diego Wild Animal Park and Dvür Králové were found to have a metaphase chromosomal number of 2n = 81 due to size polymorphisms (Houck et al., 1994). Protein polymorphisms of 25 allozyme loci showed very low divergence between seven NWRs originally from Sudan and 23 SWRs from South Africa with Nei's genetic distance (*D*) of only 0.005 compared to *D* = 0.32 between 30 WRs and 9 black rhinoceroses (Merenlender et al., 1989). Neither of these studies could genetically distinguish the two subspecies. These results were in contrast to subspecies status given to NWR and SWR based on mitochondrial DNA (mtDNA) variation at: 10 restriction fragment length recognition sites (George et al., 1993), the 12S ribosomal RNA gene, the D-loop control region, the mitochondrial NADH gene (Groves et al., 2010), and whole mitochondrial genome sequences (Harley et al., 2016). Genetic distinction of the two subspecies appears to be increasingly obvious when a high resolution of techniques/markers used for analyses is employed. This is reinforced by two recent studies based on the mitochondrial control region and 10 microsatellite markers (15 NWRs and 217 SWRs; Moodley et al., 2018) and whole genome resequencing of 25 NWRs and 27 SWRs (Sánchez Barreiro et al., 2020) that clearly identified genetic structure between both subspecies. These studies employed relatively high-resolution genetic markers, and all suggested the separation of the NWR and SWR as two distinct subspecies.

Given the phylogenetic status of the NWR and SWR, a chance to resurrect the northern subspecies is dependent on the two remaining NWR females, Najin (a daughter of Sudan) and Fatu (Najin's daughter). The former is in its post-reproductive age and the latter has very weak hindlimbs; thus, they are not able to breed naturally (Saragusty et al., 2016). There remains a chance to re-establish new populations of NWR by virtue of the advance of assisted reproductive technologies such as sperm, egg, and embryo cryopreservation as well as *in vitro* fertilisation (Roth and Swanson, 2018). Living cells of 13 NWRs have been frozen at the San Diego Zoo and used to induce pluripotent stem cells, nine of which have been generated ready to transform into reproductive cells (Ryder et al., 2020). Establishing of new NWR-SWR hybrid populations is also an option; at least one known hybrid animal was reported (Groves et al., 2010), indicating that there is no issue of mate recognition between the two subspecies. Embryonic stem cells were developed from NWR spermatozoa and 83 SWR oocytes and have been frozen for later transfer to SWR female

surrogates (Hildebrandt et al., 2018). However, the only known female hybrid never calved any offspring; thus, establishing hybrid populations is not necessarily a viable solution. There are also arguments about practicality, economy (i.e. spending massive amount of money to rescue a practically extinct species rather than shifting funds towards conservations of other endangered species), and ethical challenges of bringing back the northern subspecies from laboratories to natural habitats (Ryder et al., 2020). Therefore, in this thesis, the focus is paid to the southern subspecies that has been successfully brought back from the brink of extinction and still requires genetic-based approaches for metapopulation management.

The SWR has a very large geographical distribution; territorial males and females typically occupy 2.5 - 13.9 and 4.7 - 45.23 square kilometres, respectively (White et al., 2007). However, in fenced reserves where SWRs were kept at high density; adult bulls might exclusively occupy territories without overlapping, but in a smaller area than the studies reported in open reserves (Thompson et al., 2016). Conservation of such an umbrella species should also extend protection to other co-inhabitants and help to preserve local biodiversity. The existence of SWRs play an important role in African savanna ecosystems. Grazing megaherbivores such as the SWR can help maintaining short-grass ecosystems, which typically act as barriers preventing the potential spread of wildfires (Waldram et al., 2008). For example, removal of SWRs in Hluhluwe-Umfolozi Park in South Africa contributed to overgrowing of grasses, which subsequently increased wildfire fuel loads and fuel continuity (Waldram et al., 2008). Territorial and resting behaviours of the SWR are also crucial for soil nutrient cycles; an individual typically transports nutrients (via grazing) to its resting areas and middens (via defecations). Such behaviours, together with the works of other herbivores (e.g. buffalo, elephant, Giraffe, impala, and wildebeest) and macrodetritivores (e.g. dung beetle), can introduce spatial nutrient availability to the savanna ecosystems (Veldhuis et al., 2018). Therefore, conservation of the SWR should benefit an ecosystem as a whole.

The SWR experienced a severe population bottleneck during the end of the 1890s due to intensive hunting. Body parts were highly valued for tasty meat, skins suitable for whip crafting, and there is ongoing high demand of horns in Asia. These caused thousands of SWRs being killed annually (du Toit, 2015); as a result, only one population with ~100 individuals remained in the Hluhluwe-Umfolozi Park, Kwazulu-Natal province, South Africa at the end of the 1890s (Emslie and Brooks, 1999). The single remnant population was then provided intensive protection, and consequently the population size started to

increase. During the early 1960s, approximately 650 individuals were estimated to be in the park and the buffer areas between the park and the neighbouring communities (Harthoorn, 1962). Threats from human and domestic livestock invasions, as well as additional threats such as drought and food scarcity, prompted an urgent action of SWR relocations (Harthoorn, 1962). The first translocation of the species was successfully accomplished; four SWRs were relocated over a distance of 560 kilometres to the Kruger National Park in October 1961 (Pienaar, 1970). Following the first translocation, 141 SWRs were additionally translocated to Kruger National Park, only 6 of which died during the transports. The low mortality rate of the translocation was attributable to the advance of capture and immobilisation techniques (Harthoorn, 1962, Pienaar, 1970).

Since then, translocation became the common practice to remove surplus SWRs and to re-establish SWRs to the species' former range states. Essentially, the modern populations of SWRs are the progeny of the single founder population in Hluhluwe-Umfolozi Park. Most of the re-introduced populations have also been protected in secured areas, and the population sizes have rapidly increased. By the end of 2015, the continental number of SWRs reached 20,053 and the conservation of the species exemplified successful wildlife conservation through population management (Emslie et al., 2016; Emslie et al. 2019). However, there has been a recent concern related to ongoing poaching that has caused a decrease in the overall continental number as of 2017 when the population size dropped below 20,000 for the first time since 2010 (Table 1-1). The recent decrease indicated that the rate of death both from natural causes and illegal poaching might already surpass the rate of birth. One study predicted that the species could be extinct in the wild habitats in the next 20 years, given the current birth and death rates; more than 950 SWRs have been illegally poached every year (Di Minin et al., 2015), whereas the average population growth was +7.1% per annum during 1992 - 2010 (Emslie et al., 2019). Captive European populations have also experienced excess death rates, showing a decline 1.19 times faster than the population growth rate between 2001 and 2004 (Reid et al., 2012). Unlike free-ranging SWR populations, a major threat to captive populations has been found to relate to infertility (Hermes et al., 2006, Van der Goot et al., 2015, Tubbs et al., 2017, Roth et al., 2018, Ververs et al., 2018). The recent decline of the overall continental population suggests that despite the species being recognised as an exemplar of conservation success, intensive protection in protected areas is still necessary to conserve the species in the wild. Transborder collaboration is necessary to achieve efficient patrol, animal exchanges,

scientific studies, and an international studbook; the collaboration would result in sustainable population growth and diminish mortality rate.

In Botswana, the species was extinct during the end of the 1890s as in other former range states. However, the successful immobilisations and translocations in South Africa brought about the prospect of bringing SWR back. The Department of Wildlife and National Parks of Botswana (DWNP) implemented a re-introduction scheme by introducing the first collection of four SWRs from South Africa into Chobe Nation Park in July 1967 (vin Richter, 1973). A total number of 60 SWRs were further introduced in northern Botswana between 1970 and 1976: 42 in Chobe Game Reserve, 14 in Moremi Game Reserve, and four in Maun Wildlife Education Park (Gavor, 1988). The additional introduction of 92 SWRs were made between late 1970s and 1980s (Gavor, 1988). At least 150 individuals would have been expected in northern Botswana according to the total number of individuals translocated, but aerial surveys conducted by DWNP in 1987 revealed an approximate number of only 120 individuals (Gavor, 1988).The areas of release were close to several international borders and therefore the introduced animals were targeted by trans-border poaching. The protection of re-introduced populations at the release areas was ineffective and poaching almost wiped out the species for the second time. More recent and intensive aerial surveys in 1992 revealed only 7 individuals in Chobe Game Reserve, which led to a maximum estimate of 17 and 10 individuals in Chobe and Moremi Game Reserves, respectively (Hitchins, 1992). After the aerial surveys in 1992, at least three individuals were known to be killed, and this warranted a need for urgent action. A new conservation strategy that consists of three phases of actions was implemented by DWNP: 1) protect SWRs in secure areas (e.g. fenced reserves subject to close anti-poaching protection); 2) manage and monitor populations to achieve at least 5% annual growth rate (the number was later amended to 6%) across all of the protected areas (i.e. manage the country-wide reserves as a metapopulation); and 3) re-introduce SWRs from the managed populations back into large unfenced national parks in the Okavango delta (Verreynne, 2012).

The capture operations started in February 1993. The first four SWRs were captured and transferred over 700 kilometres to a reserve developed in southern Botswana. Due to ongoing illegal poaching that could threaten the species if the location is publicised, hereafter this reserve is referred to as Botswana1 throughout the thesis. The fence was still in the process of being built at the time that the first group of animals arrived, but completion of construction of the fence in 1995 provided a closed area to protect the

captured SWRs. The combination of capture operation, additional introductions of SWRs from South Africa, and the successful breeding of animals kept in several secured reserves have contributed to a constant increase of SWRs in Botswana (Table 1-1) and resulted in the total national population size increasing to 452 SWRs in 2017 (Emslie et al., 2019).

Re-introduction of SWRs back to unfenced areas started in 2001 on the Chief Island of Moremi Game Reserve, in northern Botswana. The released population has established well and has continuously increased (Verreynne, 2012). The populations of SWRs in Botswana have been divided into two regions: northern populations that are managed by the government and southern populations that are managed by private and community trusts (Verreynne, 2012). The southern populations have been intensively protected and monitored; the Botswana1 population has been growing, with surplus and breeding SWRs being relocated to other reserves to establish new populations. Among the other private management populations, Botswana2 and Botswana3 included in this thesis have been well managed and achieved increasing population sizes as for Botswana1. The successful management to increase population sizes of these three privately managed populations of SWRs indicates the importance of them as genetic sources for the national re-introduction programme (Verreynne, 2012). Therefore, the management of these populations to achieve the targeted population growth as well as to maintain genetic diversity of the current gene pool could be crucial to the conservation of the entire national population. However, genetic variation has not so far been considered to inform management decisions of the metapopulation in Botswana.

Table 1-1 Census population sizes of SWRs in Botswana, South Africa, and the overall continental population. Percentages in the brackets indicate the proportions to the total continental. Note that after the overall number reached 21,316 in 2012, the population started to decrease ~15% over the course of five years as of 2017 numbers (modified from Emslie and Brooks, 1999, Coutts, 2009, Emslie et al., 2016; Emslie et al. 2019).

	1895	1948	1968	1984	1992	2005	2007	2010	2012	2015	2017
Botswana	0	0	0	190 (5%)	27 (0.5%)	99 (0.7%)	106 (0.6%)	135 (0.7%)	185 (0.9%)	239 (1.17%)	452 (2.50%)
South Africa	<100 (100%)	550 (100%)	1,800 (100%)	3,234 (85.2%)	5,297 (91.5%)	13,521 (93.0%)	16,273 (93.1%)	18,796 (93.2%)	18,933 (92.7%)	18,413 (90.37%)	15,625 (86.50%)
Total African SWRs	<100	550	1800	3,798	5,789	14,543	17,470	20,160	21,316	20,053	18,064

1.4. Genetic bottleneck and inbreeding consequences

1.4.1. Genetic studies in SWR

Although conservation of SWRs in closed and intensively protected areas has been a successful strategy, it has created issues related to small and fragmented populations. Reduced genetic diversity and introduction of inbreeding depression are typically the issues of concerns in small and isolated populations (Frankham et al., 2017). Given the historical population bottleneck and the current management practice of the species, low genetic diversity and a high level of inbreeding could be expected in modern SWR populations. Various types of molecular markers have been used in studies that have aimed to determine the level of genetic diversity and its consequences in populations of SWR; these include mitochondrial DNA (mtDNA; Brown and Houlden, 1999, Coutts, 2009, Moodley et al., 2018), microsatellites (MS; Florescu et al., 2003, Moodley et al., 2018), single nucleotide polymorphisms (SNP) (Labuschagne et al., 2013, Labuschagne et al., 2015, Labuschagne et al., 2017), and adaptive molecular markers (Coutts, 2009). However, such data have not yet been applied to inform conservation management of Botswana populations.

According to the database of the National Center for Biotechnology Information (NCBI), only four haplotypes of the mitochondrial control region have been reported in SWR (GenBank accession numbers AF187836 - AF187839). Coutts (2009) reported four haplotypes as present in the NCBI database in 144 SWRs from five southern African populations, but only one haplotype shared among three individuals from the Umfolozi Game Reserve and London Zoological Gardens reported in Brown and Houlden (1999). Moodley et al. (2018) aimed to address evolutionary history and anthropogenic decline of SWRs from three museum specimens collected in 1869 prior to the historical bottleneck and 214 SWRs from eight modern populations (174 SWRs from six wild populations and 40 SWRs from captive populations), the authors reported only two haplotypes from all museum specimens and the modern SWRs from Umfolozi, Songimvelo, Mthethomusha, Loskop, Ohrigstad, Nkomazi, and European and African zoo; there was no significant difference of the haplotype diversity between both groups of samples (Moodley et al., 2018). However, as only three museum specimens were used, the sample size might not be sufficient to represent the entire historical populations. Diversity of the mitochondrial

control region of the SWR was much lower than that found in black rhinoceros (BR; 20 haplotypes), a species that also experienced a historical population bottleneck and subjected to similar conservation strategies, i.e. restriction to heavily managed protected areas (Moodley et al., 2017). Despite the substantial haplotype diversity found in modern BR, comparison between museum and modern DNA samples revealed 69% haplotype loss during the process of population bottlenecks (Moodley et al., 2017). Similarly, lower haplotype diversity (five haplotypes) was also found in modern populations of eight Sumatran rhinoceroses (*Dicerorhinus unicornis*) from Cincinnati Zoo, San Diego Zoo Institute for Conservation Research, and from Malaysia compared to 25 museum specimens (eight haplotype) (Brandt, 2016). These examples in other species might indicate that similar loss of mtDNA haplotypes might have occurred during the population bottleneck of SWR.

Microsatellites have been widely used for various purposes in genetic studies because they can provide highly variable multi-locus markers which are suitable for assessment of individual and population-level variation (Houlden et al., 1996, Stevanović et al, 2009). However, genetic diversity of SWR has been characterised as low based on microsatellites. Florescu et al. (2003) optimised 10 microsatellites in 30 SWRs from the original remnant population from Umfolozi Game Reserve, only five of which were polymorphic with observed (H_o) and expected heterozygosity (H_e) ranging from 0.499 - 0.662 and 0.462 - 0.739, respectively. Nielsen et al. (2008) identified additional 12 polymorphic microsatellites (a total of 17 loci including the loci identified in Florescu et al., 2003) in 22 SWRs from Umfolozi Game Reserve, the markers showed mean $H_o = 0.420$ and $H_e = 0.436$. Coutts (2009) screened a total set of 34 microsatellite loci in 144 SWRs, seven of the markers were developed for SWR (Florescu et al., 2003, Nielsen et al., 2008), 16 were initially screened in black rhinoceros (Brown and Houlden 1999, Cunningham et al., 1999), and 11 were identified for Indian rhinoceroses (Zschokke et al., 2003); only 13 loci of these markers were amplified and polymorphic with mean $H_o = 0.440$ and $H_e = 0.450$. At the time of writing, the largest panel of microsatellites available for SWR consisted of 23 microsatellites (Harper et al., 2103) which have been used for wildlife forensics to facilitate prosecutions, with matching probability of at most 1×10^{-8} between seized samples (Harper et al., 2018); these loci showed mean $H_o = 0.0.363$ and $H_e = 0.393$ across 367 SWRs from South African populations. Low genetic diversity reported across populations of SWRs in South Africa and Namibia was even more obvious in terms of the number of alleles per locus (N_a), which ranged from only 2.6 to 2.8 (Table 1-A1 in

appendix 1) (Florescu et al., 2003, Coutts, 2009, Guerier et al., 2012, Harper et al., 2013). Genetic diversity and numbers of alleles per locus reported in SWRs were relatively low compared to other threatened species; for examples, 31 individuals of endangered Pustertaler Sprinzen cattle breed in Italy and Germany showed H_o = 0.63 and H_e = 0.69, and N_a = 5.3 based on 20 microsatellites (Edwards et al., 2008); 20 polar bears from M'Clintock Channel, Canada were genotyped for 19 microsatellites and showed H_o = 0.63 and H_e = 0.69, and N_a = 5.3 (Brandt et al., 2014). These two parameters typically represent the statistical power of marker sets and low values may hinder studies that want to distinguish particular individuals such as for parentage identifications (Pemberton, 2008) and wildlife forensics (Ogden and Linacre, 2015). Thus, the microsatellites that have been developed for SWR might not provide enough power to inform decisions about which individuals to translocate to maintain or increase levels of genetic variation in managed populations.

Recent advances in sequencing technologies and computational capacities has provided the breakthrough to detect variant sites at the single nucleotide level across entire genomes, which could provide more power for individual-based identification in SWR. The advent of high-throughput sequencing technologies allows scientists to genotype thousands of SNP loci in hundreds of individuals simultaneously, potentially across whole genomes (Ekblom and Galindo, 2010). However, developing SNP loci for SWR remains a challenge and is not straightforward because there have been no high-quality reference genomes or polymorphism databases for the species; only a draft assembly without annotations, published by the Broad Institute in 2012, is available on NCBI databases (cerSim1; GenBank accession GCA_000283155.1). Thus, only a few studies have exclusively developed SNPs for SWR genetic studies. Labuschagne et al. (2013, 2015, 2017) developed a total set of 33 SNPs using Endonuclease V enzyme and Comparative Anchor Tagged Sequence (CATS) primers developed from conserved homologous sequences of other mammalian species. The obtained SNPs in these studies were combined with MS to improve the accuracy of parentage assignment in a population of SWRs in South Africa (Labuschagne et al., 2017); the maternity of all offspring could be assigned but paternity assignments were successful for only six out of 11 offspring. An attempt to exploit prior genomic knowledge of domestic animals was made by using the EquineSNP50 BeadChip targeting 54,000 SNPs developed from the genome of domestic horses (*Equus caballus*) to genotype two NWRs and two SWRs; however, only ~10% of the targeted loci were genotyped consistently (>90%) across individuals (McCue et al., 2012). A study designed

to assess genetic variation among cryopreserved cell lines from wild-born NWR maintained for genetic rescue of this extinct-in-the-wild species included multiplexed shotgun sequencing of four wild-born SWRs (Tunstall et al., 2018); based on alignment to the CerSim1 reference genome, a total of 4,235,589 of polymorphic SNPs could be identified among four SWR individuals. However, the source of these samples was not described in the paper because the main focus was on NWR. With only 33 SNPs currently applied to assess variation in extant populations (Labuschagne et al., 2017) and thousands of loci waiting to be utilised (Tunstall et al., 2018), the species has been far behind from what the sequencing technology can offer in terms of SNP-based genotyping of individuals for conservation management.

1.4.2. Consequences of inbreeding in SWR

An inbred individual is an outcome of mating between closely related animals, the chance of which is commonly increased in small and isolated populations due to the lack of breeding choices. Negative consequences of inbreeding occur when homozygosity of harmful recessive alleles is present at loci responsible for adaptive potential and fitness traits (Lacy, 1992, Li et al., 2014, Xue et al., 2015). The existence of inbreeding within a population could be monitored via observation of the frequencies of observed homozygosity relative to expectations based on allele frequencies within populations (i.e. expected heterozygosity) (Charlesworth et al., 2009). However, detection of inbreeding and its consequences can be challenging for wild populations owing to the difficulties of distinguishing genetic and non-genetic effects and obtaining sufficient numbers of samples to provide reliable population genetic inferences, particularly for populations with low levels of genetic diversity (O'Grady et al., 2006). The introduction of high-throughput sequencing technologies has provided techniques to allow increased precision for prediction of inbreeding at a genome-wide scale; for instance, detection of runs of homozygosity (ROH; Broman and Weber, 1999), defined as the length of consecutive SNPs that are homozygous, can be used to estimate whole-genome inbreeding levels based on continuous sequences of homozygous genotypes in an individual. An uninterrupted series of homozygous SNPs is unlikely to occur by chance; instead, there is a high possibility that both DNA strands of an individual inherited from each of its parents are identical by descent from a common ancestor (Howrigan et al., 2011). Based on the detection of ROH, Kardos et al. (2018) reported the complete or near-complete homozygosity along entire chromosomes for grey wolves (*Canis lupus*) born in an isolated

population that showed strong correlation with inbreeding coefficients estimated from the pedigree (R^2 = 0.86), suggesting the potential of ROH for determine inbreeding in wild populations where no pedigrees are available. However, this approach relies heavily on whole genome sequences of individuals which are cost-prohibitive and may not be practical to use as a tool to estimate variation in inbreeding among individuals from wild populations. Genome-wide averages also might not be informative about how much adaptively important variation has been preserved in endangered populations (reviewed in Mable 2019).

For SWR, there has been very little information about the relationship between low genetic diversity and fitness traits that could be used to assess how much adaptively important variation has been retained. Coutts (2009) aimed to genotype Major Histocompatibility Complex (MHC) genes that encode proteins presented at the surface of immune cells, which play a crucial role in immune responses against pathogens. Variation at the MHC genes is assumed to reflect the adaptive potential to counteract diseases caused by a variety of pathogens (Sommer, 2005). Coutts (2009) reported that two MHC loci (DQA and DQB) were monomorphic in all studied populations, including the original population in Hluhluwe Umfolozi, three seeded populations in the southern part of South Africa, and five samples of NWRs from Garamba National Park in the Democratic Republic of the Congo. The author concluded that neither subspecies of WR may be able to survive if they were to be challenged with an emerging infectious disease. Both DQA and DQB loci were also monomorphic or to contain only few alleles in several populations of bottlenecked species such as Swedish moose (*Alces alces*; Ellegren et al., 1996), northern elephant seal (*Mirounga angustirostris*; Weber et al., 2004); and Pere David's deer (*Elaphurus davidianus*; Wan et al., 2011). However, there has been no evidence suggesting correlation between the low diversity at the DQA and DQB loci and pathogen resistance in SWR and these examples. Another study that aimed to determine the inbreeding consequences of 1,494 captive SWRs present in the international studbook for African white rhinoceros revealed a slight, but not significant, increase in mortality rate of inbred SWRs (Krummenacher and Zschokke, 2007). The authors suggested that this was possibly because of low statistical power of a limited number of samples since it would need more than 640 inbred offspring (only 16 were identified with inbreeding coefficients (F) ranging from 0.125 to 0.25 in the actual dataset) born in the international zoos to detect a statistical difference given the reported mortality rate. The authors also addressed a challenge to define an individual as inbred ($F \geq 0.125$; its parents were closely related with a

relationship at least half-siblings, grandparent and grandchild, aunt/uncle and niece/nephew) or outbred ($F < 0.125$) due to the assumption of unrelated founders that could confound the results because all founders were likely to originate from the same population a couple of generations ago. Based on present knowledge, it may be concluded that the technologies and resources applied to date (i.e. molecular markers, studbook records, number of SWRs included) have not allowed researchers to detect statistically significant consequences of inbreeding in SWR.

Unlike SWR, consequences of inbreeding depression have been widely addressed in black rhinoceros and other mammal species. Le Roex et al. (2018) investigated the differences in genetic and demographic parameters of a small fenced population of BR in South Africa. The comparison was conducted for two consecutive 10-year periods; negative consequences were detected in the latter period, including reduced population growth rate, lower male survival rate, lower genetic diversity and higher relatedness. For other wild and captive mammals, several reproductive and fitness traits have been found to correlate with inbreeding estimated from microsatellite markers; for instance, reduced survival rate and resistance to diseases of inbred offspring in several mammalian species (Ralls et al., 1988, Keller and Waller, 2002). However, an argument that the correlations between marker-estimated inbreeding parameters and these fitness traits might be the result of publication bias in favour of significant correlations was proposed (Hansson and Westerberg, 2002). A metanalysis study comparing published and unpublished data indicated only weak correlations between multilocus heterozygosities of microsatellite loci and negative effects on fitness traits (Coltman and Slate, 2003). This metanalysis suggests that correlations between reduced performance of fitness traits and heterozygosities estimated from a small number of MS loci may be insufficient, and a larger number of markers or more variable markers are required to provide precise and accurate inferences.

1.4.3. Social behaviours of SWR exacerbating low genetic diversity

In addition to the management strategies that conserve SWR in protected areas, social structure and behaviours of the species may also exacerbate the currently low level of genetic diversity. In a SWR population, one or a small number of bulls are expected to successfully mate with multiple oestrous cows (Owen-Smith, 1975, Rachlow et al., 1998, Kingdon and Hoffmann, 2013,); hence, the effective population size is expected to be

much lower than the census size. Generally, a non-captive SWR population consists of one dominant bull, multiple sub-ordinate bulls, multiple cows with or without calves, and juveniles (Owen-Smith, 1975). A dominant bull solitarily occupies a territory that may be shared with other subordinate bulls; the dominant bull normally does not challenge the subordinates as long as no offending gestures are posed. Each adult bull, either dominant or subordinate, may exclusively occupy a territory without overlapping; however, in a small reserve, overlapping between bull territories has been reported (Thompson et al., 2016). A cow normally holds a larger territory which normally overlaps with other SWRs (Rachlow et al., 1998). When a dominant bull encounters a cow in estrous, the cow is confined within the territory until successfully mated (Kingdon and Hoffmann, 2013). If the conception succeeds, the gestation period of the species takes approximately 16 – 17 months (Rachlow et al., 1998). A juvenile remains with its mother until it is driven away at 2-3 years of age (Owen-Smith, 1975). Male and female juveniles are considered sexual adults at six and five years old, respectively, but most males do not successfully reproduce until several years after (Rachlow et al., 1998), possibly due to the reproductive behaviours of a dominant bull. According to social structure and behaviours of the species, a behaviorally dominant bull is therefore expected to spend more time with breeding females and to be more reproductively successful than the others (Owen-Smith, 1977). A parentage study based on 11 microsatellites supported that only a small number of bulls contributed to the majority of 23 offspring born in a population in northern Namibia (Guerier et al., 2012); the population that initially held two founder bulls revealed that only one bull fathered 10 of 13 offspring during a nine-year period (1993 - 2001) (Guerier et al., 2012). A combination between reproductively dominant behaviours and restriction of gene flow due to habitat loss and the conservation management strategies potentially exacerbates the currently low level of genetic diversity of the species.

1.4.4. Using genetic tools to identify individuals for translocation

Introduction of gene flow between populations is advised to maintain or maximise levels of genetic diversity of fragmented populations. In the context of SWR, corridors linking habitat patches and other means of allowing natural dispersal are not the most pragmatic resolutions because the distance between populations can be hundreds of kilometres apart. For such distances, protection of animals travelling in the corridors from poachers is a challenging task and would demand massive resources. The IUCN suggests that

translocation of SWRs is a more pragmatic resolution to maintain levels of genetic diversity and it has become a routine management practice to artificially introduce gene flow between SWR populations (Emslie et al., 2009). Translocations also offer a means to remove surplus bulls from a growing population. For example, the prevalence of fighting and possibly killing was found to increase when a high number of adult bulls were kept within a small reserve (Du Toit, 2006); consequently, the population growth rate may decrease (Emslie et al., 2009). The Guidelines for the *in-situ* Re-introduction and Translocation of African and Asian Rhinoceros (Emslie et al., 2009) provide instructions and considerations needed to implement prior to translocation. These considerations are divided into three steps of actions: 1) pre-translocation; 2) translocation; and 3) post-release steps. During the first step, the guidelines state that introduced animals should originate from a reserve where the population size is approaching or has exceeded an estimated Ecological Carrying Capacity (ECC), which is defined as the maximum number of the animals that can be sustainably maintained by the resources available in a reserve (Du Toit, 2006). A candidate for translocation should be unrelated to individuals in a recipient population, it should also be an individual that has a high probability to reproduce an inbred offspring in the current population. Therefore, identification of pairwise relationships among individuals (i.e. construction of a pedigree) within a population is necessary to efficiently identify candidates for translocation. Currently, the candidates are identified based on observational records and processes of identification for SWR rarely take genetic information into considerations (Emslie et al., 2009). Observational approaches to identify relationships among individuals are prone to errors, which are particularly common in wild populations (Pemberton, 2008). Even in closed and managed populations, observational parentage assignments can be difficult to obtain due to many factors; in the case of SWR, for example, calves may separate from their mothers before they can be individually marked for later identification and can lead to incorrect assignments. To be most effective, the identification of candidates for translocation requires a more reliable and quantifiable method to find the most appropriate individuals. Constructions of pedigrees based on molecular markers has the potential to improve the accuracy of parentage assignments to facilitate the identification of candidates for translocation and to predict the inbreeding coefficient of an offspring of all possible mating pairs. However, issues of low genetic diversity and the low number of available molecular markers so far have hindered marker-based parentage assignment in SWR.

The ultimate goal of exchanges of individuals between isolated populations is to minimise inbreeding and prevent its consequences but how this is assessed can alter management decisions. Typically, levels of inbreeding are measured using the inbreeding coefficient (F), defined as the probability that two alleles at a locus of a diploid individual are identical by descent (Wright, 1922). An alternative measure is the kinship coefficient, also known as the coefficient of coancestry (Wang, 2011), described as the probability that alleles at a locus randomly selected from "a pair of individuals" are identical by descent (IBD) which is different from inbreeding coefficient that describes the probability of IBD of two alleles at a locus of "an individual" (Rousset, 2002). Alternatively, it is the expected inbreeding coefficient of an offspring reproduced by a given pair of individuals (Ballou and Lacy, 1995). Which measures of relationships among individuals and population differentiation are applied could affect management decisions. The *F* derivatives, including F_{ST}, F_{IT}, and F_{IS} were introduced to provide descriptive approaches to summarise population structure (Wright, 1950). The parameters F_{IS} and F_{IT} offer means to measure the deviation of heterozygosity from Hardy-Weinberg Equilibrium (HWE) of subpopulations and total inbreeding of the entire base population, respectively. F_{ST} describes genetic differentiation among subpopulations relative to the total base population: $F_{ST} = 1 - (H_S/H_T)$, where H_T is the total expected heterozygosity of the entire population and H_S is that expected within-subpopulations. It has been used to identify genetically distinct populations that can be used as sources of introduced animals, in order to maximise genetic diversity within populations (Frankham et al., 2017) but there has been considerable discussion about its reliability (Jost, 2008, Whitlock, 2011) because it is sensitive to levels of genetic diversity present in subpopulations (Jost, 2008).

Firstly, genetic differences between subpopulations may be too subtle to be detected by F_{ST}, which is likely to be the issue in SWR since most of the animals in modern populations originated from the single remnant population in South Africa. Secondly, F_{ST} can be influenced by mutation which can be relatively common in hypervariable markers such as microsatellites (Eckert and Hile, 2009, Fischer et al., 2017). Mutation may either converge two isolated subpopulations and make them more genetically similar (decrease F_{ST}) or diverge them and generate two sister populations (Whitlock, 2011). Either case, using F_{ST} for identification of candidates for translocation can be problematic because in such situations F_{ST} does not reliably represent relationship between two subpopulations. While alternative measures to F_{ST} have been developed (reviewed by Jost et al., 2018), all suffer from the same limitation for population with low levels of variation. Importantly,

management of fragmented subpopulations using F_{ST} is likely to preserve the gene pool of the most genetically different population rather than preserve the gene pool of the entire founder population (Eding and Meuwissen, 2001) which may not be practical for conservation of fragmented species.

An alternative to using F statistics to assess population differentiation is to use mean kinship (MK), estimated at both individual and population levels. An individual MK (MK_{ID}) is calculated by averaging all pairwise kinship coefficients between that individual and other individuals within the population including itself. An individual with low MK_{ID} is considered a genetically important individual and is unlikely to share IBD alleles with the others (Ballou and Lacy, 1995). Population MK ($MK_{population}$) is simply a population mean of MK_{ID} averaging across all individuals within the population (Sekino et al., 2004, Wang, 2011, Frankham et al., 2017); it allows prediction of the expected inbreeding coefficient of an offspring born from parents randomly chosen from the same population. Thus, if an individual with relatively high MK_{ID} is removed, $MK_{population}$ as well as the inbreeding coefficient of the next generations should essentially be reduced. The concept of $MK_{population}$ can also be applied at the inter-population level (hereafter referred to as between-population MK, $MK_{A\text{-}B}$) to predict the expected inbreeding coefficient of an offspring born from parents randomly chosen from populations A and B. Therefore, it can be used to identify sources of introduced animals that would minimise inbreeding coefficients of the subsequent generations of the recipient population (Frankham et al., 2017). Two populations with low $MK_{A\text{-}B}$ are considered genetically distant to each other and translocation of individuals between these populations would theoretically reproduce a progeny with low inbreeding coefficient (Finger et al., 2011, Mickelberg, 2011, Garbe et al., 2016).

Unlike F_{ST}, MK is less sensitive to levels of genetic diversity and provides a means to preserve the gene pool of a founder population (Frankham et al., 2017). Frankham et al. (2017) used molecular and demographic data provided in Culver et al. (2000) to demonstrate the advantages of using MK over F_{ST} in regard to identifying donor populations of puma (*Puma concolor*). The recipient population has been isolated in Florida and showed a very low number of alleles per locus (N_a = 1.2) and low genetic diversity (H_e = 0.041) for 10 microsatellites. Using MK and F_{ST} approaches suggested two different donor populations. Simulations showed that introduction of animals based on MK resulted in higher H_e in the recipient populations compared to the introduction of animals based on F_{ST}. The difference of H_e estimated from different methods indicated that

introduction of the most genetically distinct individuals (high F_{ST}) was not necessarily the most effective approach to improve genetic diversity in the recipient population. However, one possible complication of using MK for population management is that once translocations (or any kind of population changes) are made MK_{ID} and $MK_{population}$ are also changed, which would require regular updates of kinship coefficients (Lacy, 1995).

While F_{ST} is estimated from heterozygosity and allele frequencies of subpopulations, kinship coefficients can be calculated from either population pedigrees or molecular markers. Multilocus genotypes have been used to estimate F_{ST} and MK in wild populations when pedigrees are not available (Rodriguez-Rodriguez et al., 2015, Garbe et al., 2016). However, only heterozygosities calculated from a large set of loci (~200 loci) were reported to show a strong correlation with marker-estimated F (Balloux et al., 2004). This finding raises a question about how many loci are required to reliably estimate F_{ST} and MK. Due to this limitation of the marker-estimated parameters, pedigree-based approaches have been considered the gold standard for estimation of kinship coefficients; however, construction of an accurate wild pedigree can be challenging due to many reasons (Pemberton, 2008). For example, extra pair paternity was reported in a variety of wild species: mammals (Cohas and Allainé, 2009), birds (Grinkov et al., 2018), and fishes (Bose et al., 2018). In a behaviourally monogamous fish species (*Variabilichromis moorii*) for which both parents show bi-parental care behaviour, pair-bonded males sired only 63% of the brood based on 14 MS loci, and the remaining fry were sired by multiple unpaired males (Bose et al., 2018). Mean percentages of extra pair paternity were reported in 22 monogamous mammal and the means of percent extra-pair parentage in these species ranged from 8% to 92% (Cohas and Allainé, 2009). These proportions of extra pair paternity are substantial and can confound estimates of inbreeding and kinship coefficients due to mis-identified parents observed merely from observational pedigrees. Differences in social structures and behaviours such as egg dumping, adoption, extra pair copulation, and male-male competition were also responsible for the variation in extra pair paternity reported in these studies. In the SWR populations included in this thesis, most observational records offer only lists of putative parents, which often involves multiple fathers and mothers. With this limitation, the application of pedigree-based MK for identification of candidates for translocation is challenging for SWR.

The advance of next-generation sequencing offers an opportunity to genotype thousands of molecular markers simultaneously in hundreds of individuals. Such a high-resolution marker panel should improve the accuracy of pairwise kinship estimates to predict the

inbreeding coefficient of an offspring reproduced by two individuals in consideration (Goudet et al., 2018). For example, two full-sib individuals have a kinship coefficient of 0.25; at a given locus, the actual offspring may or may not receive the same allele by descent, thus the actual kinship at a locus is either zero or one. However, as the number of tested loci is increased, the average kinship coefficient over all loci will approach 0.25 (Weir and Goudet, 2017). However, the currently available resources for SWR (i.e. number of markers available) does not allow a reliable estimate of MK directly from molecular markers. Given the limitations posed by both pedigree-based and marker-based MKs in SWR, one approach that provides a resolution is to develop a new set of markers. A new and high-resolution marker panel should improve analytical power of parentage analyses and improve the accuracy of estimations designed to predict which individuals should be targeted for translocation.

1.5. Thesis overview and aims

The overall research aims of this thesis were initially developed to improve the population of management of SWR in situations where complete pedigrees could not be obtained by field observations, such as in privately managed populations in Botswana. Pedigrees are crucial to genetic-based identification of candidates for translocation and breeding that would maximise genetic diversity and preserve the gene pool currently present in the populations. The overall aims of the thesis were to: 1) test the usefulness of existing and newly developed genetic markers as tools to facilitate genetic-based population management and 2) develop an analytical framework for the identification of candidates for translocation and breeding that would maximise genetic diversity of SWR populations.

The first aim (Chapter 2) was to assess whether existing genetic markers (a panel of 23 microsatellite loci as described in Harper et al. 2013) could improve resolution of parentage assignment based on the incomplete observational pedigree available for one of the larger populations (referred to here as Botswana1). Specifically, the proportion of successful assignments made from different approaches to estimating were compared and the combined pedigree used to make predictions about which individuals would make the best candidates for translocation. Only 29/45 parental pairs could be identified which warrant the need to develop a new panel of markers to improve analytical power of parentage analysis. This chapter was published in *Conservation Genetics* in 2019 (Purisotayo et al. 2019).

Given the low genetic diversity estimated from microsatellite markers, in Chapter 3, a sequencing approach that would enable population-wide sequencing for thousands of molecular markers (i.e. SNPs) across the genome was developed. Since the initial restriction associated DNA sequencing (RADseq) approach attempted was limited by degradation of DNA in a large number of the samples obtained, an alternative method that would allow genotyping of SNPs from DNA samples with various degrees of degradation was optimized and tested. The RADcapture approach uses hybrid sequence capture to target SNPs identified based on RADseq. A subset of samples from the total collection were chosen to represent low- and high-quality samples for testing the effects of degradation on RADcapture performance. The sequencing method performed equally well in both groups of samples, suggesting the potential of RADcapture to retrieve genetic information of DNA of different qualities. This chapter has been prepared for publication to demonstrate this sequencing method in non-model species.

In Chapter 4, the newly developed sequencing approach (RADcapture) was applied to SWRs from three managed populations in Botswana to: 1) test whether the new marker set could improve resolution of the pedigree for Botswana1 (Chapter 2), which was the population with the largest population size among the three populations; and 2) evaluate the use of this marker set for a genetic-based approach for identification of candidates for translocation and breeding. A consensus pedigree with improved completeness could be constructed based on the pedigree present in Chapter 2 and the SNP-assigned parents; the consensus pedigree should offer insight about relationships among individuals in the important SWR population of Botswana which should facilitate genetic-based metapopulation management of the national herd. Both consensus pedigree and RADcapture data allowed demonstration of the methods for identification of candidates for translocation and breeding that would prevent inbreeding in fragmented populations of endangered species. This chapter is planned for publication to demonstrate the genetic-based methods for management of isolated populations when pedigrees is either available or unavailable.

In Chapter 5, I discussed the limitations occurring during my study including the issues of sample quality and the lack of reference genome or polymorphisms database of the species that complicated the development of sequencing and genotyping methods; per-sample cost of the RADcapture that was relatively high and might not be practical for inferences of population parameters, the drawbacks of probabilistic methods that might confound parentage assignment when the putative parents were not completely sampled, and the

closely related origins of the three populations included in the thesis that made the populations not an ideal to demonstrate the genetic-based meta-population management. Finally, I further discussed about the topic related to the future applications of this thesis.

Chapter 2 Combining molecular and incomplete observational data to inform management of southern white rhinoceros (*Ceratotherium simum simum*)

*This chapter was published in Conservation Genetics and was co-written with Professor Nicholas N. Jonsson, Professor Barbara K Mable, and Frederick J. Verreynne. I completed all the experimental design and analyses, and wrote most of the manuscript under the supervision of all co-authors. Please note that the population used in this chapter was the Botswana1.

"PURISOTAYO, T., JONSSON, N. N., MABLE, B. K. & VERREYNNE, F. J. 2019. Combining molecular and incomplete observational data to inform management of southern white rhinoceros (*Ceratotherium simum simum*). *Conservation Genetics*, 20, 639-652."

2.1. Abstract

Conservation efforts have preserved the southern white rhinoceros (SWR) in protected areas and have resulted in substantial overall growth in population size, but in small, fragmented populations in which inbreeding is an important risk. However, field observation of breeding often lacks sufficient accuracy to inform translocation strategies that are intended to increase genetic variation. The purpose of this study was to integrate microsatellite genotypes with an incomplete, field-observed pedigree to make inferences about mean kinship and basic demographic data that could be used to inform translocation programmes for SWR in a confined population in Botswana. Using this approach, we identified parents for 29 out of 45 offspring born in the reserve between 1993 and 2013 and detected eight non-breeding bulls with high mean kinship as candidates for translocation. The method also allowed inferences about demographic parameters that could influence the effectiveness of intervention strategies, such as age and timing of reproduction, and natal sex ratios. Importantly, the reproductive dominance of the bulls was not as skewed as expected after the original dominant bull was removed from the population, suggesting that closed populations can maintain multiple, simultaneously breeding males. The genetic data also confirmed that the accuracy of field-based parentage assignment was increased after implementation of an ear-notching programme. This study demonstrates the value of combining genetic information with ongoing surveillance to inform management of threatened populations, and of using mean kinship to inform metapopulation management by identifying candidates for translocation.

Keywords: Kinship coefficient • Microsatellite • Parentage assignment • Pedigree • Translocation • White rhinoceros

2.2. Introduction

The southern white rhinoceros (*Ceratotherium simum simum*; SWR) was historically distributed over the land now designated as Botswana, Namibia, Swaziland, South Africa and Zimbabwe. The continental population was on the edge of extinction in the 1890s, when approximately 50 to 100 individuals were all that remained in a single population at Hluhluwe-Umfolozi Park, Kwazulu-Natal, South Africa. Subsequent intensive protection efforts saw a rapid increase in the size of the population, which reached 1,800 individuals by 1968. With the application of newly developed protocols for translocation in the early 1960s (Player, 1967), the population in Kwazulu-Natal became the founder of all African populations (Emslie and Brooks, 1999). By 2015, the number of SWRs in Africa had reached 20,375 (Emslie et al., 2016).

In Botswana, the population of SWRs became extinct in the late nineteenth century as in all other countries within the species' former range. In 1967, four SWRs were relocated from Natal Park in South Africa into the Moremi Game Reserve in Okavango delta, and between 1974 and 1980, 91 more were relocated to Moremi and Chobe National Park (Tjibae, 2001). However, the release areas, which are near several international frontiers and therefore subject to cross border poaching, were effectively unprotected and poaching almost wiped out the population. An intensive aerial survey of northern Botswana in September 1992 revealed only seven SWRs (leading to an estimate of a maximum of 10 animals in Chobe and 17 in Moremi); and between August and November of the same year, poachers were known to have taken another six SWRs (Department of Wildlife and National Parks Botswana, 2002). Subsequently, a new conservation strategy consisting of three phases of action was employed by the Department of Wildlife and National Parks of Botswana (DWNP). The main pillars of the new strategy were to: a) protect SWRs in secure areas (confined reserves subject to close anti-poaching protection); b) manage and monitor populations to achieve 5% annual growth rates; and c) re-introduce SWRs into large, unfenced national parks in the Okavango delta (Verreynne, 2012). Initially, between 1994 and 1996, seven SWRs were captured at Chobe National Park and Moremi Game Reserve and were translocated into fenced sanctuaries elsewhere in Botswana (Tjibae, 2001) .Combined with further introductions from South Africa, this contributed to an increase in the number of SWRs in Botswana to 239 animals in 2015 (Emslie et al., 2016).

However, the population bottleneck of just over a century ago resulted in inbreeding and loss of genetic diversity, with consequent low number of alleles per locus of microsatellite markers (Florescu et al., 2003, Coutts, 2009, Guerier et al., 2012). The mating system of the species, in which one dominant bull is expected to sire the majority of offspring in a population, results in low effective population size, exacerbating the problem of genetic drift, and generating many surplus bulls (Owen-Smith, 1975). The International Union for Conservation of Nature (IUCN) suggested that translocation of SWRs is crucial to maintain levels of genetic diversity and it has become a routine management practice (Emslie et al., 2009). Translocation has been considered to be especially useful for the species because other means of enabling gene flow such as building habitat corridors are often not feasible, due to risk of poaching in unprotected areas. The population in this study is also located hundreds of kilometers away from the closest neighbouring reserves. Building corridors of such a distance as well as protecting SWRs in migration would be difficult and would demand a massive amount of resources. To be most effective, translocation requires a method for identifying the most appropriate individuals to be relocated, which is dependent on establishing a pedigree of relationships. However, an accurate field-observed pedigree can be difficult to obtain, particularly for wild animal species (Pemberton, 2008). In the case of SWR, for example, calves may separate from their dams before they can be individually marked for later identification. The construction of pedigrees based on molecular markers has the potential to increase accuracy of parentage assignments to aid in the identification of candidates for translocations and to predict the inbreeding coefficient for all possible mating pairs. The kinship coefficient, also known as coancestry, of a given pair of individuals is the average probability that alleles at a locus randomly selected from those individuals are identical by descent (Ballou and Lacy, 1995, Wang, 2011). Alternatively, the kinship coefficient between a pair of individuals can be described as the expected inbreeding coefficient of their progeny. Individuals with high numbers of relatives generally show high mean kinship values (Mickelberg, 2011); thus, young bulls with high mean kinship values would be expected to have a high risk of inbreeding in a population if they were retained and were to become dominant. Therefore, translocation of young bull with high mean kinship combined with regular removal of dominant bulls would help to prevent inbreeding within the population.

The purpose of this study was to integrate genotyping based on microsatellite markers with field observations to build a pedigree to allow inferences about mean kinship and basic demographic data of the population that could be used to inform translocation programmes

for SWRs in a confined population in Botswana. The efficiency of permanent individual marking through ear-notching was also evaluated by determining the difference between observational assignment rates before and after the introduction of the approach.

2.3. Material and Methods

2.3.1. Samples and population genetic parameters

Due to security sensitivities and the risk of poaching, we are unable to explicitly provide the name and location of the studied reserve, which initially comprised a fenced area of approximately 4,000 hectares and has been expanded to 8,600 hectares since 2012. It is constantly patrolled by guards and protects a critically important population of about 55 SWRs. Between 1993 and 2013, 14 animals were introduced to the reserve and were considered to be founders, from 10 of which DNA samples were available from either blood samples or tissue samples retained after ear-notching for individual marking. The other four were either relocated or died before commencement of sampling; thus, no samples were available. Samples were collected during the routine procedures of individual marking or health checks from an additional 45/48 animals born in the park between 1993 and 2013. Three animals had died before they could be sampled. DNA was extracted using a phenol-chloroform protocol (Sambrook and Russell, 2006) and genotyping was carried out in the Onderstepoort Veterinary Genetics Laboratory in South Africa, using 23 microsatellite loci, as previously reported (Harper et al., 2013).

To describe the genetic status of the population, we estimated population genetic parameters for 55 animals (10 founders and 45 offspring) in the population, including number of alleles per locus (N_a), observed (H_o) and expected (H_e) heterozygosities, polymorphic information content (*PIC*), frequency of null alleles (*F-null*) and the probability of non-exclusion for a candidate parental pair (*NE-PP*), using CERVUS 3.0.7 (Kalinowski et al., 2007). The *NE-PP* estimates the probability that a locus cannot exclude a randomly chosen parental pair within a population. Loci that deviated significantly from Hardy-Weinberg Equilibrium (HWE) as tested using GENEPOP 4.2 (Raymond and Rousset, 1995) were excluded. Inbreeding coefficient (F_{IS}) and global deficit and excess of heterozygotes across loci were estimated to determine global deviation from HWE using Fisher's exact tests, as implemented in GENEPOP 4.2.

2.3.2. Construction of population pedigree

The workflow used in the construction of the pedigree for the study population is illustrated in Figure 2-1. Briefly, the field observational records were used to construct a precursor pedigree (Pedigree A). The microsatellite genotypes were then incorporated to test whether the genetically possible parents from among the field-observed parents qualified (Pedigree B). In cases where the field-observed records and genotypes were unable to identify the parents for an offspring, maximum likelihood-based assignments were employed (Pedigree C). Bayesian-based parentage assignment was also implemented to confirm the results of Pedigree C. The final pedigree (Pedigree D) was subsequently used to make inferences about mean kinship and to estimate demographic parameters.

Using field observation to construct Pedigree A

Between 1993 and 2013, the maximum population size in the reserve was 60 but the actual size at any given period was dynamic, being dependent on ongoing translocations. Field observations of mating, births and associations of females with calves were recorded by rangers and the veterinarian responsible for the reserve. The database of field observations included birth date and location of birth, introduction and relocation dates, and suggested a number of potential parents for each offspring. Parentage assignments for the offspring born in the reserve were constructed based on observed dam-offspring relationships. Dam was assigned to an offspring with high confidence when they were present together at the time of capture for individual marking. The level of confidence was reduced when an offspring was marked after it had separated from the dam. In such a case, the offspring was assigned to a set of possible dams. During the early period (1993 - 1995) following the establishment of the reserve, only one adult male and three adult females occupied the site. Although samples were not available for microsatellite genotyping, field-observed relationships between offspring and any of these founders were considered to have a high degree of confidence. To construct Pedigree A, unambiguous assignments were made if only one parent or one parental pair could be assigned to a given offspring. The efficiency of field-observed assignment was quantified as the proportion of unambiguous assignments divided by the total number of offspring. Offspring that were assigned to multiple sires or dams were left unassigned in this initial phase.

Using observational and genotypic data to construct Pedigree B (Exclusion)

To narrow down possible parents for offspring who had been assigned multiple parents based on observations, genetically unlikely parents were excluded from the list of suggested parents of each offspring using *R*-package SOLOMON (Christie, 2013). Parents were excluded (considered to be impossible) whenever there was a mismatch between potential parents and offspring at one or more loci. A new, combined pedigree was then constructed (Pedigree B). The proportion of unambiguous assignments of Pedigree A and Pedigree B was determined and compared. Any change in the proportion assigned would suggest the extent to which the microsatellite markers improved the observational pedigree.

Using maximum likelihood and Bayesian approaches to construct Pedigree C

Prior to conducting parentage assignments using maximum likelihood and Bayes' theorem, information from field records was used to determine the theoretical reproductive window for each individual. The reproductive window defined the period during which an individual was at post-pubertal age and was present at the site. This restricted the number of possible parents for both parentage assignment approaches to only those that were logistically possible. Six and five years of age were considered to be the ages of puberty for males and females, respectively (Rachlow et al., 1998). Sixteen months was used as the average gestation period (Rachlow and Berger, 1998); a potential sire was excluded from the list of candidate parents for an offspring when it had been introduced to the reserve less than sixteen months prior to the birth of the offspring, and an individual was excluded from the list if it had been removed from the reserve more than sixteen months before.

We used CERVUS 3.0.7 (Kalinowski et al., 2007) to identify parents of 33 offspring whose possible parents did not include unsampled animals, as a high confidence level of assignment is achieved when all possible parents are sampled (Marshall et al., 1998). Likelihood-based assignment was conducted using the 18 polymorphic microsatellite loci that were in HWE. The LOD score between a parental pair and an offspring was interpreted as the natural logarithm of the likelihood ratio between the first and second hypothesis. The first hypothesis was that a tested trio comprised true parents and offspring, while the second hypothesis was that the trio was unrelated (Marshall et al., 1998). CERVUS then determined two types of delta scores: 1) trio delta, which was the difference between the LOD score of the most likely trio and the second most likely trio; and 2)

critical delta obtained from computer simulation of parentage inference to identify the proper delta for the population in study. The simulation was conducted to obtain critical delta using the following parameters: 10,000 offspring with 90% of individuals in the population sampled; 98% of loci typed; allele frequencies of the population; and the confidence levels were set at 95% for strict critical delta. A trio was assigned unambiguously when the delta of the trio was greater than strict critical deltas that gave 95% confidence level.

To confirm the results obtained from likelihood-based parentage assignment and to allow estimation of parentage for duos, we used the Bayesian *R*-package SOLOMON (Christie et al., 2013) to assign parents for 76 parent-offspring pairs for which no unsampled founder was in the observational lists of possible parents (36 sire-offspring and 40 dam-offspring pairs). This package can incorporate prior probabilities based on genotype information alone. In this case, the prior probability was defined as the probability that at least one allele at each locus is shared between a randomly selected parent-offspring pair by chance. The prior probability was equal to the expected number of false parent-offspring pairs divided by the total number of possible parent-offspring pairs. A simulation was conducted to estimate the expected number of false pairs; here we set the number of simulations to 1,000 and 50,000,000 for simulated data sets and genotypes, respectively (Christie et al., 2013). Parent-offspring pairs that had a prior probability equal to one were not further used to estimate the posterior probability. The posterior probability can be described as the probability that a possible parent-offspring pair is false given the allele frequencies of shared alleles. In this study, parent-offspring pairs that showed posterior probabilities of at most 0.05 were considered genuine relationships. If an offspring had multiple genuine relationships, the parent with the lowest posterior probability was considered a genuine parent. Pedigree C was then constructed using the results from likelihood-based assignments unless the results were contradicted by the Bayesian-based assignments, in which case the assignment was excluded.

Using combination of pedigrees to construct the final Pedigree D

The final pedigree was constructed using the combination of results obtained from Pedigree B and C to maximise the number of unambiguous assignments. Only in the situation that the Pedigree B failed to unambiguously assign a parent or a parental pair to an offspring, the result obtained from Pedigree C was implemented. In a case when more

than one offspring born in the same year were assigned to the same mother, they all were considered to be unassignable.

2.3.3. Efficiency of ear-notching to aid parentage assignment

Ear-notching for individual identification and collection of samples for genotyping was introduced in 2006 and continuously conducted as newly born and introduced animals were later marked. The change in efficiency of observational-based assignment after the introduction of ear-notching was tested. Twenty offspring observationally assigned to a mother whose genotype was available were used to compare the difference between assignment rates of: 1) offspring born before ear notching was implemented (1993 - 2008; $n = 7$); and 2) offspring born after this change in management practice (2009 - 2013; $n = 13$), using Fisher's exact tests (Agresti, 1992). The implementation of ear-notching had started in 2006, but 2008 was used as a cutoff to ensure that a substantial proportion of animals were marked, and the benefit of the approach could be detected.

2.3.4. Demographic parameters

Pedigree D was subsequently used to estimate population demographic parameters, including: 1) annual calving rate (*ACR*), defined as the annual percentage of dams that gave birth of the total number of reproductive age dams; 2) percentage of herd growth (*HG* – see equation 2-1, below), described as net increase in size of the herd as a result of newborn calves divided by the size of the herd at the beginning of the year (Ververs et al., 2017); 3) mean age of first calving of dams born in the reserve; 4) mean total number of offspring produced per sire and dam over the period of observation; 5) natal sex ratio; 6) the effective population (N_e); and 7) calving interval (*CI*) for multiparous females. The estimation of percentage of *HG* shown in equation 1 accounted for only the effect of newborn SWRs to the annual population growth, with six and five years old considered adult ages for males and females, respectively (Rachlow et al., 1998).

$$\%HG = \frac{no.of\ calves\ born - no.of\ deaths}{no.of\ resident\ adults + no.of\ adults\ arriving\ that\ year} \times 100 \qquad (2\text{-}1)$$

The age of first calving for each of the dams born in the reserve was determined by estimating the interval between its birthdate and the date of its first calving. The duration

between two consecutive calvings was used to determine *CI* values. Bulls were considered to be dominant when they sired more than 50% of all the offspring born in a given year. Natal sex ratio was estimated based on an expectation of 50:50 males to females, and skewness was evaluated using *G-test* statistics (Kretzschmar, 2001). The effective population size (N_e) is here described as the number of breeding individuals (Falconer, 1960) and was estimated by taking the sum of the number of individuals that contributed to offspring born in the reserve. For parameters that required the birthdate for estimation we used the first of January or the first of a given month in cases where the field observations for an individual recorded only a year or month of birth, respectively.

2.3.5. Pedigree plot, kinship coefficients, mean kinship

We used the *R*-package PEDANTICS (Morrissey and Wilson, 2010) to draw Pedigree D. The *R*-package Kinship2 (Sinnwell et al., 2014) was then used to estimate all pairwise kinships based on the assumption that all founder animals were unrelated. The elements in kinship matrices showed pairwise kinships between individuals that were computed by identifying the probabilities that alleles randomly drawn from a pair of individuals are identical by descent. Mean kinship of an individual was estimated by averaging of all pairwise kinship between the individual and other individuals within the population, including itself (Ballou and Lacy, 1995). Bulls with no evidence of contributing paternity, whose individual mean kinship was higher than the population mean kinship, were considered the best candidates for translocation.

2.4. Results

2.4.1. Population summary statistics

Three out of 23 loci were found to be monomorphic (DB23, IR22, and SR74). For the 20 polymorphic microsatellite loci genotyped, three deviated from HWE - two of them highly significantly (DB66, IR12; *p-value* < 0.01) - and so were excluded from the estimation of means of population genetic parameters. Note that locus IR22 was found to be monomorphic in this population but was reported to be polymorphic elsewhere (Scott, 2008). Based on the final set of 18 loci, means of population genetic parameters did not indicate high levels of inbreeding: H_o and H_e were 0.426 and 0.409, respectively; *PIC* = 0.340; and F_{is} = -0.0406 (Table 2-1).

2.4.2. Efficiency of parentage assignments using the combination of approaches and the usefulness of ear-notching for parentage assignments

The proportions of unambious assignments for distinct parentage assignment approaches are provided in Table 2-2 and the assignments made for all offspring are provided in Table 2-A1 in Appendix 2 (please refer to electronic version at https://doi.org/10.1007/s10592-019-01166-4). Using field observation alone: five offspring could be assigned to a parental pair; one and 20 offspring were assigned to a sire or dam, respectively. The application of the exclusion approach (Pedigree B) increased the number of assignable offspring: 12 offspring were assigned to a parental pair; 10 and nine offspring were assigned to a sire and dam, respectively. Using the likelihood-based parentage assignment approach, we could identify a parental pair for 23 out of 33 offspring. Only five out of 76 possible parent-offspring pairs were considered genuine using Bayesian-based assignment, three of which were consistent with the likelihood-based assignment. One was inconsistent (ID 172) but the parent with the second lowest posterior probability (not statistically significant) was the same mother that was assigned using the likelihood-based assignment. The other inconsistent assignment was made for an offspring that was not examined using CERVUS but the assigned parent was identical to the parent suggested by Pedigree B (ID 121). The final pedigree of 45 offspring revealed 29 assignable trios, six sire-offspring duos, and four dam-offspring duos. After the introduction of ear-notching, the assignment rate of observational dam-offspring relationships was significantly improved (*p-value* = 0.02), with only 1/7 compatible relationships prior to 2008 and 10/13 after 2008.

2.4.3. Using mean kinship to identify individuals for translocation

Pedigree D, drawn using PEDANTICS, is shown in Figure 2-2. A population mean kinship of 0.0483 was estimated and all pairwise kinship coefficients are shown in Table 2-A2 in Appendix 2 (please refer to electronic version at https://doi.org/10.1007/s10592-019-01166-4). Unassignable offspring were given no contribution to the estimation of population mean kinship. Twenty parents contributed to the sampled offspring, five and two of which were dams and sires, respectively, that were born in the reserve. No mating between parent-offspring, or any of the aunt-uncle-nephew-niece pairs or first cousins was suggested by the assignments. However, we found that one offspring with ID 146 was produced by a half-sib parental pair (IDs 124 and 131). The individuals considered to be

candidates for translocation were the bulls with IDs 176, 156, 320, 167, 144, 111, 145 and 271.

2.4.4. Population demographic parameters

Reproductive and demographic parameters of the population are shown in Table 2-3. The mean age at first calving among dams born in the reserve was 6.8 years, while two sires that were born in the reserve sired their first offspring at the age of 10 and 14 years. Multiparous dams had an average *CI* of 3.7 years. The mean total number of offspring produced between 1993 and 2013 was 5.8 per sire and 2.4 per dam. The total of 45 sampled offspring out of 48 that were born in the reserve during the period of this study contributed to 15.9% and 18.6% of mean *HG* and *ACR*, respectively. There was no deviation from a 50:50 natal sex ratio, with 22 males and 23 females.

Four bulls (ID 999, 130, 124, and 120) sired more than 50% of offspring for at least one year; however, only two bulls (999 and 130) showed the expected pattern of reproductive dominance since they successively showed exclusive paternity for consecutive years between 1996 and 2005. Following the translocations of 130 in 2005, four competing bulls (ID 120, 124, 133, and 170) sired offspring born between 2006 and 2012 (Figure 2-3).

2.5. Discussion

Our study demonstrates the value of combining genetic information with field observations to construct pedigrees to estimate relatedness and infer population demographic parameters, even when markers are not variable enough to produce distinct multilocus genotypes for every individual. Moreover, we found that management practices that include ear notching for individual identification significantly improved the field-observed assignments, particularly when combined with exclusion of incompatible molecular marker combinations. We could not find previous studies that quantified the effect of ear-notching and close observation on maternity assignment, but this is an encouraging finding. Importantly, we also found that multiple subordinate bulls were able to reproduce simultaneously, when formerly dominant bulls were removed. This has important implications for management practices, since it has been assumed that a single behaviourally dominant bull contributes to offspring born in a population of SWRs (Owen-Smith, 1977, Rachlow et al., 1998).

2.5.1. The efficiency of parentage assignment based on combining approaches

Pedigrees obtained from observational data of wild populations are commonly compromised by inaccuracy and incompleteness of the observations (Bérénos et al., 2014). Similarly, pedigrees of wild animals inferred from molecular data can suffer from low statistical power of the molecular markers (Wang, 2007) and missing data due to incomplete sample collection (Pemberton, 2008). In this study, we combined incomplete observational and molecular data to maximise the rate of successful assignments. Using observational records alone could unambiguously assign parental pairs to only 11% of the total offspring; while the incorporation of genotypic exclusion and the combination of different parentage assignment approaches increased the assignment rate to 27% and 64%, respectively. This is despite the presence of only 2-4 alleles per microsatellite loci used. These results suggest that, even for populations with low genetic diversity, the combination of observational records and molecular markers could significantly improve the population pedigree regarding the proportion of unambiguous parental assignments. The rates of successful parentage assignment obtained in this study were relatively lower than previous SWR studies. Guerier et al. (2012) constructed a complete pedigree of a managed SWR population of 31 individuals by achieving 100% assignment rate of both parents for all 23 offspring using a combination of 11 microsatellite loci sampled from all individuals (H_e = 0.450, H_o = 0.450, N_a = 2.8) and well-maintained historical records. Labuschagne et al. (2017) employed nine microsatellite loci (H_e = 0.508, H_o = 0.478, N_a = 2.8) and 33 SNPs (H_e = 0.350, H_o = 0.357) to confirm maternity of all 11 dam-offspring pairs known from historical records in a managed SWR population; however, paternity assignment could be obtained with confidence for only 6 offspring. The results of these studies suggested that smaller number of candidate parents and the capacity to sample all individuals in the populations were the important factors to achieve a high assignment rate. Although the number of markers used in parentage assignment has also been recognised as another succeeding factor for parentage assignment (Pemberton, 2008); the study of Guerier et al. (2012) demonstrated that, even with low N_a, the complete assignment could be obtained by incorporating a well-maintained observation record. However, given the level of genetic diversity of the species, the number of currently available markers, and the capacity to maintain observational records with high accuracy; either observational-based or molecular-based assignment alone is insufficient to obtain a complete pedigree. One possible means of obtaining a complete pedigree would be to include more genetic markers

in the parentage assignments, which recently has become more feasible because the introduction of next-generation sequencing allows the discovery of thousands of markers for non-model species. However, even with a larger set of markers, DNA samples from all animals in a population are still necessary to minimise erroneous assignment even though incomplete sampling can be taken into account in many parentage assignment software (Kalinowski et al., 2007, Walling et al., 2010). Here we further demonstrated the improvement of field-observed parentage assignments following the introduction of ear-notching; however, the overlap of three years between the starting of ear-notching (2006) and the cutoff year (2008) might lead to an upward bias in the assignment rate of the former period, as some animals were already marked. Regarding the construction of Pedigree C, the Bayesian approach identified many fewer parentage assignments than CERVUS (only 5/76 pairs considered, compared to 23/33 trios considered, respectively). When we deliberately increased the critical value of the posterior probability in Bayesian-based assignments from 0.05 to 0.2, this resulted in an increase of assignable parent-offspring pairs to 13/76. Ten of these pairs were still consistent with the assignments made by likelihood-based assignments. So, even with a more relaxed threshold for SOLOMON, CERVUS was more informative for this dataset. CERVUS takes the genotypes of the second most likely animal into consideration whereas SOLOMON determines confidence levels based on the expected probability of false parent-offspring pairs simulated from genotypes of the population (Walling et al., 2010). The outperformance of CERVUS may suggest that the likelihood-based approach is less sensitive to the low genetic variation in our data.

2.5.2. Population genetic parameters.

Based on the genetic parameters determined in this study, we found that the level of genetic diversity was not as low as might be expected from the historical bottleneck. Means of parameters that described population heterozygosity such as H_o, H_e, and F_{is} in this study were similar to those found in other SWR genetic studies (Coutts, 2009, Guerier et al., 2012, Harper et al., 2013, Labuschagne et al., 2017). The mean H_o estimated from microsatellite markers in other SWR studies ranged from 0.440 to 0.478; while average H_o found in this study was 0.426. Lower heterozygosities have been reported in other mammal species that also experienced historical bottlenecks (Corti et al., 2011, Fitak, 2014, Pertoldi et al., 2010). We did not find evidence of inbreeding but instead found a significant global excess of heterozygotes across loci. The negative mean value of F_{is} across loci also

indicated that individuals in the population were less related than we would expect, although this was not significantly different from zero. The effects of metapopulation management by mixing previously separated populations could be responsible for the low inbreeding level found in this study. For example, observed heterozygosity was shown to be higher for F0 and F1 animals than for F2 animals in a managed SWR population in Namibia for which F0 animals were translocated from different locations (Guerier et al., 2012). Despite the fact that no evidence for inbreeding was found in this study, gradual loss of heterozygosity is likely to occur in managed populations of SWRs that implement similar conservation strategies (Guerier et al., 2012). With a limited number of individuals in the population, the inbreeding coefficient would inevitably increase; thus, regular introductions of new animals and monitoring of genetic diversity in subsequent generations are still necessary to prevent inbreeding. In addition to heterozygosity, the number of alleles per locus is not only an indicator for measuring genetic variation but it is also an important factor to achieve high parentage assignment rate (Bernatchez and Duchesne, 2000). The N_a of 2.5 in this study was severely low and could have hindered successful parentage assignment. This warrants the effort of developing a larger set of markers to compensate for the low N_a, polymorphisms of markers, and confidence level of assignment in SWR populations.

2.5.3. Using mean kinship to identify individuals for translocation

Translocation of individuals among populations has been demonstrated to reduce inbreeding coefficients and increase genetic variation in a range of endangered populations (Bouzat et al., 2009, Mickelberg, 2011, Moraes et al., 2017). The family of F-statistics have been used to monitor inbreeding levels and to consider sources of introduced animals; however, they are likely to preserve the gene pool of the most genetically distinct population (Eding and Meuwissen, 2001). In contrast, using mean kinship at the individual level tends to preserve the gene pool of a founder population and is less sensitive to levels of genetic diversity (Jost, 2008); hence, monitoring of mean kinship would provide the means to maintain a current level of genetic diversity (Willoughby et al., 2017) and prevent negative effects of fitness traits (Lacy et al., 2018). Further, mean kinship can also be estimated at a population level from molecular markers (Wang, 2011) and has been used to identify donor populations that contain valuable genetic resources in a variety of fragmented species (Finger et al., 2011, Frankham et al., 2017, Garbe et al., 2016, Mickelberg, 2011). Outbreeding depression could be expected after translocation made

between subpopulations with no recent geneflow (Frankham et al., 2017). To prevent the harmful effects of outbreeding, genetic status of populations of interest should be carefully studied and taken into consideration in any genetic rescue programme. However, we argue that outbreeding depression is unlikely the case for SWR as most modern populations originated from the same founder population just a couple of generations ago. Here we identified eight non-breeding bulls as candidates for translocation based on their high mean kinship. These bulls would be expected to provide a high risk of inbreeding in the population if they were retained and were to become reproductively active. Five out of the eight candidates already have been relocated to other reserves, two (ID 176 and 156) have been killed by other bulls, and only one candidate (ID 271) currently remains at the reserve. Of the six reproductive bulls (Figure 2-3), all four of the dominant bulls had already been relocated (one in 1999, one in 2005 and two in 2012), leaving two bulls that have left offspring on the reserve. The four relocated bulls were removed from the population when they were 15 to 18 years of age; i.e. IDs 999 (unknown age at relocation), 130 (18 years old), 120 (15 years old), and 124, (16 years old). Given a post-reproductive age of 35 years (Reid et al., 2012), they would probably continue to be dominant and breed for many years, which might exacerbate the issue of inbreeding and prevent other bulls from genetic contributions. The proposal for male rather than female translocations (or other means of removal, including culling) is based on evidence that the prevalence of fighting and possibly killing other animals is increased when a higher number of adult males are kept within a particular area (Du Toit, 2006). Moreover, translocation of males is a more cost-efficient strategy for introducing new diversity into a population than moving females, because males are likely to have a more substantial genetic contribution, as demonstrated by the average number of offspring per bull and dam observed in this study. We suggest that translocation of young bulls with high kinship together with regular relocating of dominant bulls could retard the rate of inbreeding in SWR populations. The determination of a justifiable upper limit to population mean kinship that should be maintained in wild populations is difficult. Frankham et al. (2017) suggested using the value of 0.1 as a practical guideline. Although the population mean kinship of 0.0483 reported in this study was substantially lower than the suggested level, the value was likely to be underestimated, based on the assumption that founder animals were unrelated.

2.5.4. **Male dominance**

Our study also demonstrated the value of pedigrees for assessing reproductive success rather than relying on observed behaviour to infer the social structure of animals with suspected dominance hierarchies. A dominant bull solitarily occupies a territory that may be shared with other subordinate bulls that are normally constrained within a single territory (Owen-Smith, 1975, Thompson et al., 2016). In a large reserve such as Kruger National Park (19,485 square kilometres), the territory size of a bull can be as large as 14 square kilometres (Owen-Smith, 1975, Pienaar et al., 1993); however, in a high-density population, bulls may occupy a small territory without overlapping. For example, all 147 territorial bulls in Welgevonden Game Reserve (~360 square kilometres) exclusively occupied a territory with an overall mean of only 3.46 (1.14 - 5.17) square kilometres. When a dominant bull encounters a cow in oestrous, the cow is confined within the territory until successfully mated (Kingdon and Hoffmann, 2013). Therefore, a dominant bull is expected to spend more time with reproductive females (Rachlow et al., 1998) and to be more reproductively successful than sub-ordinate bulls (Owen-Smith, 1977). However, in this study, we found that the reproductive dominance of bulls was not as complete as expected following the removal of two dominant founder bulls. Sub-territories within the reserve might have developed as a result of the population growing and following the construction of a new waterhole in the reserve that would allow multiple non-overlapping territories of the bulls. There are few genetic studies that have successfully obtained paternity assignment in SWR populations that would allow researchers to determine the dominant reproductive behaviour of the species because most studies have been unable to assign the paternity of offspring with statistical confidence (Coutts, 2009, Labuschagne et al., 2017). However, one study conducted in a limited free-ranging population that held two founder bulls reported that the bull that was believed to be subordinate had actually sired 10 of 13 offspring during a nine-year period. Additionally, two newly introduced bulls had succeeded in breeding before they established their territories (Guerier et al., 2012). Our finding and that of Guerier et al. (2012) contradict the hypothesis that only a single behaviourally dominant bull contributes to the offspring born in a population. With more contributing bulls presented in our study, the results provided clearer insight into the breeding pattern in limited free-ranging populations. We and Guerier et al. (2012) did not find any evidence of female choice biases. This was reinforced by a recent study conducted in a large population of SWRs (104 known parent-offspring relationships) that revealed no skewed mating success across

individuals (Kretzschmar et al., 2019). However, Rachlow et al. (1998) reported that territorial bulls spent more time with females that were not pregnant or with calves older that 10 months of age. In regard to multiple contributing bulls, this also provided evidence to support the hypothesis that regular translocations of dominant bulls could encourage sub-ordinate bulls to breed; hence, slowing rates of inbreeding.

2.5.5. Population demographic parameters

The final pedigree was useful for estimating other population demographic parameters that could influence the effectiveness of intervention strategies. No evidence of natal sex-ratio skewness was found, which supports previous studies. Natal sex-ratio skewness toward males has been reported in captive SWRs (Kretzschmar, 2001, Linklater, 2007, Zschokke et al., 1998), whereas non-captive populations kept at lower stocking densities have been reported to be in the expected 50:50 sex ratio (Ververs et al., 2017). The average *ACR* and *CI* observed in this study were 18.6% and 3.7 years, respectively; which could be interpreted as a very poor fecundity population based on the guidelines provided by the Southern African Development Community Rhino Management Group (Du Toit, 2006). The reproductive performance of the population in this study was less than that reported in other non-captive populations, for which *ACR* and *CI* ranged from 20% to 50% and 2.4 to 3.3 years, respectively (Kretzschmar, 2001, Rachlow and Berger, 1998, Ververs et al., 2017). The *ACR* target is a value greater than 33%, which corresponds to a *CI* of three years (Du Toit, 2006). The high variability of observed *ACR* in this study, which ranged from 0% to 50%, resulted from a small number of breeding dams and was the main reason for the low apparent fecundity of the population in this study. Aberrant ovarian cycles and pathological lesions of female reproductive tracts have been reported in SWRs kept in both captivity and non-captivity, and these defects are believed to cause low reproductive performance (Hermes et al., 2006, Roth et al., 2018, Ververs et al., 2018). However, there was no evidence that these defects could be linked to potential inbreeding. Studies conducted in captive populations led to the hypothesis that prolonged estrogenic exposure from phytoestrogens found in plants that the cows fed on may be responsible for the aberrant ovarian function (Tubbs et al., 2017). Although the accessibility of non-captive SWRs to phytoestrogen-rich plants is limited, it may be worth investigating whether such plant species exist in the habitats of populations with poor fecundity. Another possible explanation for low fecundity is the effect of inbreeding as a consequence of a historical population bottleneck; however, it may be difficult or even impossible to confirm the

effects of inbreeding with sufficient statistical confidence in real SWR populations. Krummenacher and Zschokke (2007) found a slight, but not significant, increase in mortality rate of inbred SWRs using data from international studbook records. However, a power analysis indicated that they would have needed more than 640 inbred offspring born in the international zoos to detect a statistical difference given the reported mortality rate. For the estimation of *CI*, we additionally estimated the adjusted *CI,* which accounted for the unassignable offspring and three offspring that died before being sampled. However, the adjusted *CI* of 2.0 was not consistent with the *ACR* that was estimated from all 45 offspring and indicated low fecundity of the population. The unequal contribution of females could be responsible for this inconsistency; therefore, using *CI* estimated only from the 33 assignable dam-offspring pairs was more reasonable to represent the reproductive success of the population. To estimate mean kinship of a population with incomplete pedigree, either using of average mean kinship or assuming no contribution for a missing relationship has been used (Cassell et al., 2003). Here, we assumed no input from the missing data. This approach would limit the extent of overestimated mean kinship, while the extent of underestimated mean kinship could be substantial given the historical genetic bottleneck of the species.

2.6. Conclusions

Conservation strategies that aim to protect SWRs within secured areas has contributed to an increase in the total number of SWR, but have generated small and fragmented populations, raising concerns about the genetic viability of the species. Translocation has been implemented as a tool to increase gene flow among populations to maintain the current level of genetic diversity of the species. Here we have demonstrated that records of field-observations with a high degree of ambiguity could be improved by the incorporation of genetic markers, even for populations with low levels of diversity. The pedigree thus allowed us to use the kinship coefficient to quantitatively identify the best candidates for translocations to maintain the current genetic diversity of the population. Indications from the pedigree that multiple bulls contributed to a group of offspring born in the same year is valuable information and can be used in genetic management of SWR populations. Although we have demonstrated that using mean kinship to monitor level of inbreeding provided a tool to incorporate genotypes and observations records for metapopulation management, a larger set of markers is necessary to maximise the assignment rate.

Table 2-1 Population summary statistics of 55 animals in the reserve: N_a= number of alleles at a locus; *N* = number of typed individuals at the locus; H_o = observed heterozygosity; H_e = expected heterozygosity; *PIC* = polymorphic information content; *NE-PP* = non-exclusion probability of a parental pair at the locus; *HWE* = Significance of deviation from HWE, *NS* = not significant, *ND* = not done, *p* = statistic *p-value*; F_{is} = inbreeding coefficient; *F-null* = frequency of null alleles.

Locus	N_a	*N*	H_o	H_e	*PIC*	*NE-PP*	*HWE*	F_{is}	*F(Null)*
32A	3	55	0.564	0.558	0.455	0.614	NS	-0.0109	-0.0075
DB44	3	55	0.309	0.316	0.290	0.724	NS	0.0229	0.0471
7B	3	55	0.545	0.492	0.433	0.605	$p = 0.04$	-0.1088	-0.0328
7C	3	54	0.704	0.578	0.509	0.540	NS	-0.2202	-0.0966
BlRh1B	2	55	0.418	0.481	0.363	0.726	NS	0.1309	0.0649
DB66	4	55	0.182	0.428	0.375	0.659	$p < 0.01$	0.5776	0.3908
DB52	3	55	0.636	0.615	0.531	0.531	NS	-0.0350	-0.0320
BR6	2	55	0.400	0.400	0.318	0.753	NS	0.0008	-0.0041
DB1	2	55	0.273	0.238	0.208	0.824	NS	-0.1489	-0.0692
BlRh1C	2	55	0.400	0.416	0.327	0.748	NS	0.0396	0.0155
12F	2	48	0.521	0.495	0.370	0.722	NS	-0.0538	-0.0311
BlRh37D	2	55	0.018	0.018	0.018	0.982	ND	0.0000	-0.0008
32F	3	55	0.327	0.282	0.249	0.780	NS	-0.1641	-0.0847
SR63	2	55	0.509	0.476	0.361	0.728	NS	-0.0693	-0.0377
IR10	2	55	0.291	0.323	0.269	0.783	NS	0.1000	0.0476
IR12	2	55	0.218	0.364	0.296	0.767	$p < 0.01$	0.4033	0.2466
SR262	2	55	0.382	0.409	0.323	0.750	NS	0.0659	0.0292
SR268	3	54	0.259	0.251	0.234	0.773	NS	-0.0327	-0.0078
SR281	3	55	0.655	0.638	0.561	0.503	NS	-0.0256	-0.0133
RH12	2	52	0.462	0.379	0.305	0.761	NS	-0.2216	-0.1034
Mean	2.5	54.4	0.426[a]	0.409[a]	0.340[a]	0.0018[a,b]	$p < 0.01$ [a,c]	-0.0406[a]	0.0160

Table 2-2 The proportion of unambiguous assignments for trios, sire-, and dam-offspring duos for each approach.

Approaches	Proportion of unambiguous assignments		
	Trio	**Sire**	**Dam**
Field-observed	5/45	1/45	20/45
Field-observed plus Exclusion	12/45	10/45	9/45
Likelihood	23[a]/33[b]	-	-
Bayesian	-	4/36[c]	1/40[c]
Combination	29/45	6/45	4/45

[a] the assignments that exceeded the strict critical delta were considered unambiguous

[b] likelihood assignments were conducted for 33 offspring for which genotypes of all possible sires and dams were available

[c] Bayesian assignments were conducted for offspring for which genotypes of all possible sires and dams were available but excluding duos involving unsampled founder

Table 2-3 Demographic information of the population determined using the final pedigree, showing the predicted values and their standard deviations, along with the number of observations used in estimation of the parameters (N).

Parameter	Value	S.D.	N
Mean dam age at first calving	*6.8*	*1.0*	5 dams born in reserve
Mean total no. of calves per sire	*5.8*	*2.1*	6 reproductive bulls
Mean total no. of calves per dam	*2.4*	*1.2*	14 reproductive dams
Mean calving interval (years) (CI)	*3.7*	*1.7*	10 multiparous dams, 19 intervals
Natal sex ratio (male:female)	*22:23*	-	45 offspring
Mean percentage of herd growth (HG)	*15.9*	*13.7*	21 years
Mean annual calving rate (ACR)	*18.6*	*15.1*	21 years
Effective population size (N_e)[b]	*20*	-	-

[a] Deviation from an expected 50:50 sex ratio

[b] Estimated using the number of breeding individual

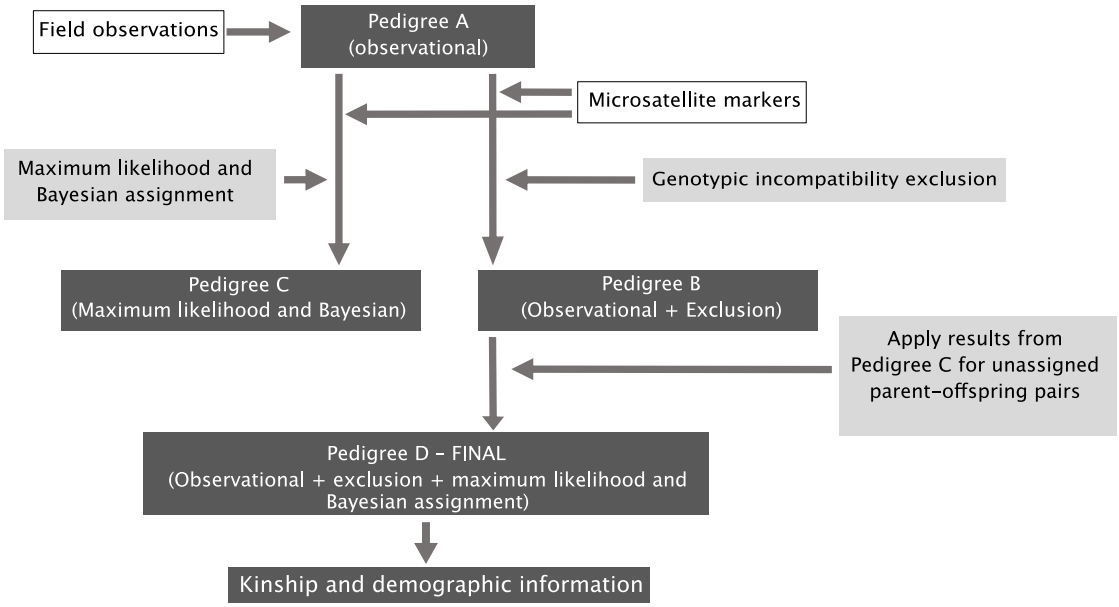


Figure 2-1 Diagram of the workflow used in this study. Two types of inputs, field observations and microsatellite markers, are shown in white boxes; analytical processes and their outputs are presented in the light and dark grey boxes, respectively. Field observations were used to construct Pedigree A and genotypes were subsequently used to qualify the genetically possible parents (Pedigree B). Maximum likelihood-based and Bayesian-based assignments were incorporated (Pedigree C) in cases for which Pedigree B failed to unambiguously identify a parental pair for a particular offspring. The final pedigree based on the combination of all three approaches (Pedigree D) was used to make inferences about kinship and demographic information.

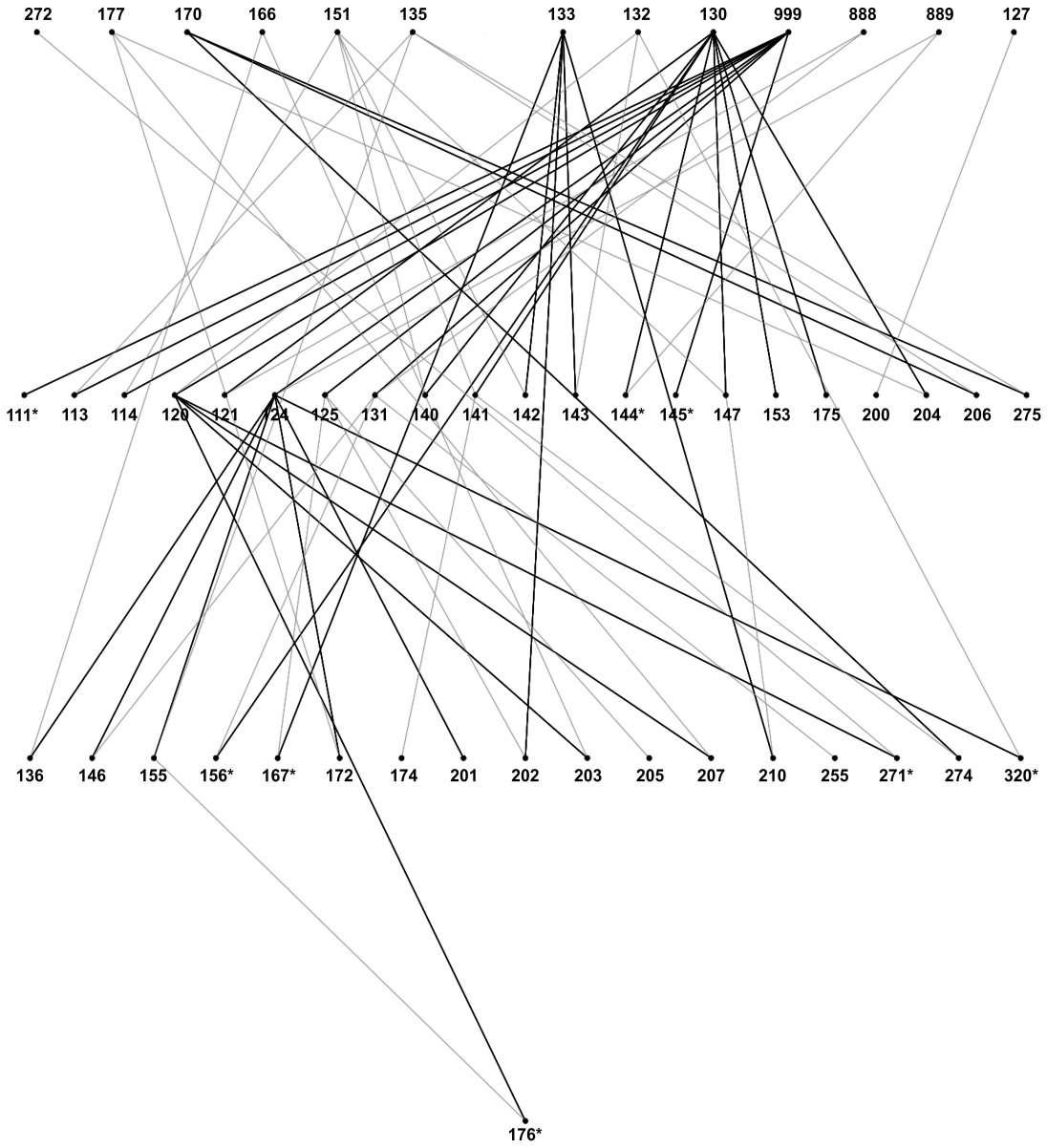


Figure 2-2 Pedigree D illustrating all assignable parent-offspring pairs. Black and grey lines indicate paternity and maternity relationships, respectively. Asterisks indicate candidates suggested for translocation based on mean kinship. Note that the unsampled founders are individuals 999, 888, 889, and 127.

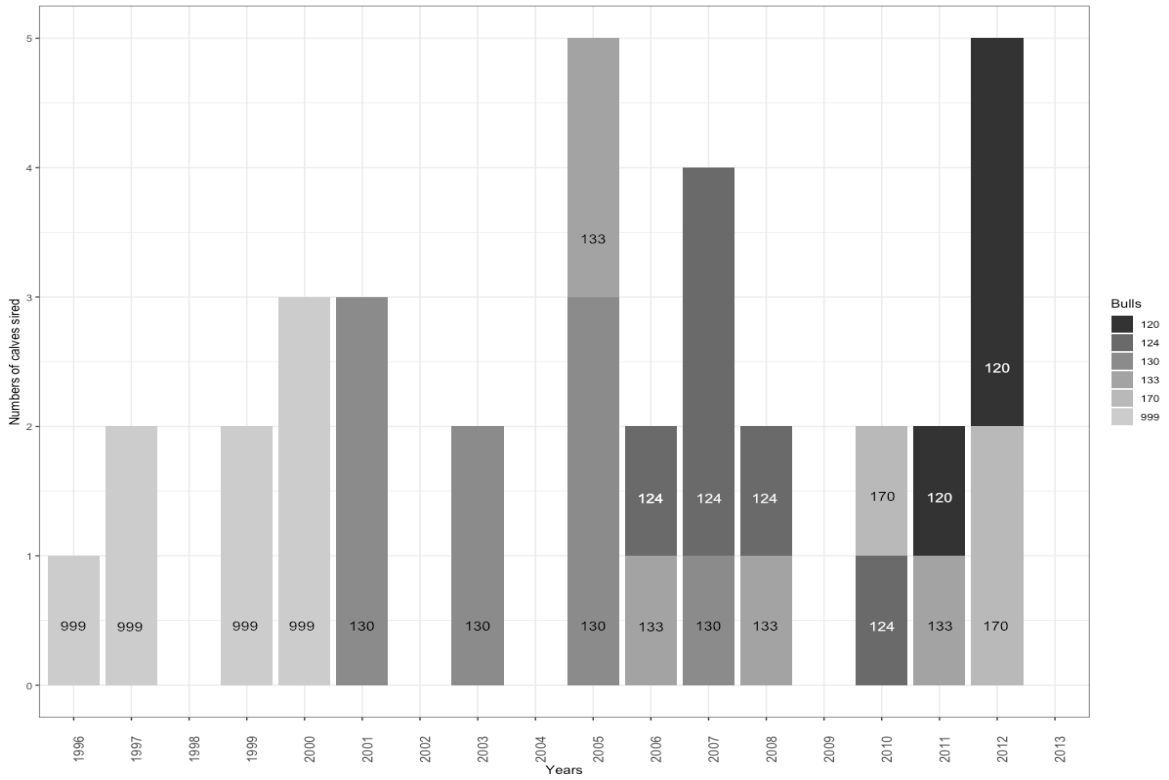


Figure 2-3 The number of calves sired by six breeding bulls (grey scale and ID numbers represent the contribution of different bulls) reflecting the pattern of reproductive dominance in the population. The original dominant bulls have been translocated to other reserves: individual 999 in 1999; 130 in 2005 and 120 and 124 in 2012.

Chapter 3 Methods to discover single nucleotide polymorphisms from samples of high- and low-quality DNA

3.1. Abstract

Many of the remaining southern white rhinoceros (SWR) are managed intensively in protected areas, resulting in small and fragmented populations characterised by low genetic diversity. Genetic markers are required to facilitate population management, to minimise inbreeding within populations and to assist forensic investigation in support of criminal prosecutions. However, the application of molecular markers for population genetics analyses in SWR so far has been limited to only a small set of low-diversity microsatellite markers which are insufficient to completely resolve population management problems such as parentage and population identification. The purpose of this study was to combine double-digest Restriction-Associated DNA sequencing (ddRAD) with hybrid sequence capture, with the ultimate goal of developing a panel of single nucleotide polymorphisms (SNPs) that could be screened effectively for pedigree analyses and other applications that require a reproducible set of high-throughput markers, even for low quality samples. High-molecular weight DNA samples from 20 SWRs were initially screened for variation using ddRAD and used to design probes (baits) for sequence capture, using two baits per ddRAD SNP. After bioinformatic filtering for specificity, baits were tested for sensitivity by hybrid capture and Illumina sequencing of an additional 32 SWRs that were chosen to represent different DNA qualities (low quality = 16, moderate and high quality = 16 samples). There was no relationship between the DNA quality of samples and the performance of the protocol, suggesting that the combination of ddRAD and hybrid capture could be useful for resolving SNPs in both high- and low-quality samples. Given that non-invasive sampling of wildlife often produces samples of varying quality, the approach should enhance applications of genetic analyses in many research questions such as population genetics and forensic science.

3.2. Introduction

Population fragmentation has been an important concern in wildlife conservation; approximately 29% of all vertebrates have been affected by habitat loss and human-wildlife conflicts (Hughes et al., 1997, Frankham et al., 2017). High-resolution satellite photography taken in 2014 revealed that > 70% of global forests were within 1-kilometre distance to their boundaries adjoining to agricultural or other human-modified lands (Haddad et al., 2015). Avoiding human modifications such as road networks can result in population isolation and subsequently reduce genetic diversity, as has been found for wolverine (*Gulo gulo*) populations in the Rocky mountain (Sawaya et al., 2019). Fragmentation can lower genetic diversity (Proctor et al., 2005, Frankham et al., 2014) because it restricts gene flow between populations and results in artificial ceilings on effective population sizes (Pavlova et al., 2017, Baas et al., 2018, Parra et al., 2018). Such loss of genetic variation within isolated populations can cause fixation towards homozygosity of an allele at a locus, which can reduce fitness of individuals in isolated populations (Frankham, 2012). The sizes of small and isolated populations are often lower than the threshold required for adaptive potential; for example, fragmented populations of endangered Australian fish *Macaquaria australasica* had an effective population size (N_e) < 100 and simulations suggested that these populations thus might encounter inbreeding depression in the next few generations (Pavlova et al., 2017). Quantifying the extent of loss of genetic variation due to habitat fragmentation is thus an important goal of conservation measures designed to mitigate the effects of human activities.

In an African wildlife conservation context, human-induced habitat loss has posed a substantial threat and the extent of fragmentation has been accelerated by the rapid increase of human population, which is expected to reach 2.5 billion in 2050 (Gerland et al., 2014). Together with intensive hunting during the late nineteenth century, which warranted the need to conserve African wildlife in protected areas, these factors have raised concerns about the genetic viability of many threatened species (Miller et al., 2019). This is particularly true for highly persecuted animals such as rhinoceros, in which poaching for desirable products exacerbates the problems of small, isolated populations. To reduce threats due to poaching, more than one quarter of the approximately 20,000 southern white rhinoceros (SWR; *Ceratotherium simum simum*) in Southern Africa are managed under private ownership, and many more are managed in heavily protected, small and fragmented populations (IUCN, 2012).

The management of small, isolated populations requires complete and accurate pedigrees to identify the most useful candidates for translocations, thereby reducing the risk of inbreeding in the source population and increasing the genetic diversity of the destination population. Field observations and microsatellite genotypes have been used to construct pedigrees in SWR populations (Coutts, 2009, Guerier et al., 2012). However, in a recent study on a fenced SWR population in Botswana, neither observational data nor molecular genotypes based on microsatellite markers were sufficient to construct a completely resolved pedigree, even when they were combined, (Chapter 2 of this thesis; published as Purisotayo et al., 2019). The observational pedigree contained some inaccuracy due to ambiguities in parentage assignment and were incomplete, while the molecular-based pedigree was compromised by low marker diversity and lack of access to samples from founders. Although 23 microsatellites have been developed specifically for SWR, several of them are monomorphic in entire populations and other loci show low mean numbers of alleles per locus, ranging only from 2.5 to 2.8 (Scott, 2008, Coutts, 2009, Harper et al., 2013, Purisotayo et al., 2019). So, even though most other published parentage assignment studies in SWR have also combined field-observations and molecular data to construct pedigrees; only one study conducted in a managed population of 23 SWRs in Namibia (Guerier et al., 2012) that employed robust observational records (all offspring-dam pairs were known) could obtain complete pedigrees. Lack of polymorphism in the microsatellite markers also resulted in failure to fully resolve pedigrees for which observational data was incomplete for the other studies that have been complete to date, Welgevonden Game Reserve in South Africa and Matobo National Park in Zimbabwe (Coutts, 2009), Songimvelo Nature Reserve (Labuschagne et al., 2017) and a private game park in South Africa (Kretzschmar et al., 2019), As an alternative to microsatellites, Labuschagne et al. (2013, 2015, 2017) developed a single nucleotide polymorphism (SNP)-based approach for parentage assignments using comparative anchored tagged sequence (CATS) primers developed from 16 mammalian genomes previously described in Aitken et al. (2004). Only 33 polymorphic loci were reported in total (H_o ranging from 0.065 – 0.656), which was insufficient to achieve a complete pedigree even though seven out of 11 mother-offspring relationships were known (Labuschagne et al., 2017). Given the low number of markers used so far and their low diversity, one possible and straightforward solution to achieve higher confidence in parentage assignments for conservation management would be to increase the number of markers that could be included in analyses. CATs, which are typically developed from ultra-conserved DNA regions across orders/families, may lead to

insufficient statistical power of a marker panel, especially when the populations of interest were expected to have low genetic diversity.

In addition to the concerns about inbreeding within small and fragmented populations, poaching for the horns of SWR continues to be a threat to the conservation of the species. A census in 2015 revealed a slight decrease in the continental population size of 0.4% per annum from 20,604 to 20,375 since 2012 (Emslie et al., 2016). Ongoing illegal poaching has contributed to this decline, as ~1,000 African SWRs are killed annually (Emslie et al., 2016). Given the current levels of poaching; combined with low population growth rates, wild populations of SWR are expected to become extinct before 2027 under the current management strategy, which includes a ban on trade in rhino horns (Di Minin et al., 2015). While the conservation of SWR becomes more expensive and the costs of anti-poaching potentially outweigh the benefits (Biggs et al., 2013, Di Minin et al., 2015), legalisation of the trade in rhino horn could produce up to $717,000,000 per year, which could be used to cover costs for anti-poaching actions (Di Minin et al., 2015). Whether this is an ethically appropriate solution remains controversial (Biggs et al., 2013). Nevertheless, a reliable means of determining the origins of horns on the market would be required to distinguish "approved" from black market trade.

While forensics for wildlife crime so far has been based largely on microsatellites or other fragment analyses for DNA fingerprinting (Harper et al., 2013, Harper et al., 2018), recent advances in sequencing technology mean that alternative and more sensitive molecular markers could provide more rigorous assignment probabilities. An advantage of sequence-based databases, particularly for standardisation of forensic methods, is that they are interchangeable and reproducible across laboratories. In contrast, allele sizes of microsatellites are called relative to particular size-standards, which can differ among laboratories, and reproducibility is challenging due to amplification biases that can result in null alleles, stutter bands due to errors during replication, and allele shifting (reviewed in Mable, 2019). Lack of reproducibility across laboratories of microsatellite markers would limit the application of the markers for forensic work. Although very cheap to screen once developed, the investment required to identify microsatellites is also nontrivial; for example, they must be developed, optimized and tested individually for each target species (Zane et al., 2002). The development of a panel of SNP markers would thus have definite benefits for multiple conservation initiatives, not only for metapopulation management.

Before widespread adoption of next generation sequencing (NGS) for population genetics, several sequencing approaches allowed utilisation of genomic knowledge of related species to discover SNPs in non-model species. Aitken et al. (2004) developed a targeted gene approach involving comparative anchor tagged sequence (CATS) primers designed from gene regions conserved between primates and rodents to amplify variation present in introns. Alternatively, random sequencing of genomic regions allowed *de novo* SNP discovery of the species of interest; for example, Bensch et al. (2002) screened for SNPs from Amplified Fragment Length Polymorphism (AFLP) fragments and re-sequenced the identified SNPs in willow warblers (*Phylloscopus trochilis*). However, the discovery of SNPs using traditional sequencing (i.e. Sanger sequencing) technology has been limited to a relatively small number of SNPs (<100 loci; Seeb et al., 2011). The advent of NGS technologies enables the discovery of a large number of SNPs in non-model species. Although SNP identification could be most thorough when based on whole genome resequencing, reduced-representation sequencing techniques have been widely used as a cost-effective alternative for population genetic studies in species for which the genomes have not been resolved (Miller et al., 2007, Baird et al., 2008, Ekblom and Galindo, 2010, Davey and Blaxter, 2011). These sequencing techniques improve the coverage of sequenced genomic regions while reducing parts of the genome to be sequenced, which improves confidence in marker discovery (Miller et al., 2007, Baird et al., 2008). Such techniques are particularly useful for studies focusing on non-model species requiring large numbers of molecular markers scored at a population-wide scale.

Although a number of "genotype by sequencing" approaches have been developed, restriction-associated DNA sequencing (RADseq) has dominated this field of research (Davey et al., 2011). In the original RADseq protocols (Miller et al., 2007, Baird et al., 2008), fragments of genomic DNA are randomly sampled using a restriction enzyme, which cuts DNA at a specific recognition site and generates thousands of DNA fragments of diverse lengths, from across the whole genome. The fragments are then ligated to barcoding indices and adaptors to allow sample multiplexing, so that all DNA sequences from an individual organism contain a unique barcode and can be traced back during the downstream analysis. The fragments are then pooled, randomly sheared, selected according to size, purified, amplified and sequenced, normally using short-read deep sequencing technologies (e.g. Illumina). The sequenced reads can then either be aligned to a reference genome or assembled *de novo* to build a set of polymorphic loci for downstream analyses.

However, the original RADseq approach contains some important drawbacks. Above all, by using a single restriction enzyme, approximately 30% - 50% of the sequenced reads may be discarded because they are not of the desired length (Emerson et al., 2010, Hohenlohe et al., 2011). Several modifications have been devised, including double digest RADseq (ddRAD), which employs two restriction enzymes (low- and high-frequency cutters) followed by gel-based size selection to eliminate the random shearing step (Peterson et al., 2012). The original ddRAD method therefore requires high-quality DNA to provide appropriate fragment lengths that would allow two restriction enzymes to cut evenly across the genome. This requirement limits the application of ddRAD in studies of wild animal species from which high-quality DNA can be difficult to obtain because ideal conditions for DNA preservation may not be met in the field (Camacho-Sanchez et al., 2013). Additionally, forensic studies often have to deal with decayed specimens, from which it can be challenging to retrieve high-quality genomic data (Graham et al., 2015).

An alternative method of SNP discovery that would potentially increase sensitivity for low quantity and quality DNA is targeted sequence capture. This sequencing approach is based on hybridisation of oligonucleotide probes, which is increasing in popularity as a means of specifically enriching genomic libraries for particular regions of interest, to develop SNPs for a wide range of applications; e.g. environmental DNA (eDNA) sequencing (Bohmann et al., 2014) and genome-wide exome capture (Cosart et al., 2011). Particularly when combined with deep sequencing approaches, hybrid capture requires relatively low quality and quantity of DNA and thus has been applied to even highly degraded samples, such as from museum specimens, noninvasively collected faeces or oral swabs, or eDNA (Kollias et al., 2015, Bose et al., 2018, McGuire et al., 2018). However, such sequencing techniques require prior genomic knowledge to enable the design of probes complementary to the targeted sequences (Jones and Good, 2016). This can be particularly challenging for non-model organisms for which good reference genomes and data about within- and between-population polymorphisms are lacking. For SWR, although there is a draft genome sequence available from individuals from the Center for Reproduction of Endangered Species, San Diego Zoo, the genome is fragmented into 57,824 contigs, with incomplete annotation (CerSimSim1.0, GenBank accession GCA_000283155.1). Developing a sequence capture panel based on the draft genome without a polymorphism database is thus not yet feasible for the species.

One approach to identify variable SNPs for population genetic studies in non-model species or those with poor quality DNA has been to combine RADseq (or its derivatives)

with sequence capture (Ali et al., 2016, Hoffberg et al., 2016, Delgado et al., 2019, Dorant et al., 2019, Komoroske et al., 2019). The primary aim of the present study was to test this approach for use on samples of SWR that were not of sufficient quality for standard ddRAD, in order to develop a novel set of SNP markers that could be used for pedigree analysis, population genetics analyses or forensic identification of poached samples. After probe development and optimisation, sequencing and mapping results obtained from samples of different quality were compared to determine the thresholds of sensitivity and specificity of the approach. The potential and reproducibility of ddRAD for conservation genetic studies were also addressed using genetic admixture analysis to test repeatability of duplicated samples.

3.3. Materials and Methods

3.3.1. Samples and DNA quality designation

The overall workflow of this study is illustrated in Figure 3-1. For initial screening of variation across ddRAD markers, 16 DNA samples from four zoological parks (Zoo 1 – 4 in Table 3-1) that are members of the European Endangered Species Programme (EEP) and two samples from the National Museums of Scotland were obtained. The samples had been used previously to develop a panel of microsatellite markers for forensic species identification by the Science and Advice for Scottish Agriculture (SASA) wildlife crime laboratory (Ewart et al., 2018). For wild SWR, DNA was extracted using DNeasy Blood and Tissue kits according to the manufacturer's protocol (Qiagen, Manchester, UK) from tissue or blood samples obtained from 110 individuals after ear-notching for individual marking or health checks, respectively. These individuals were residents of three semi-captive populations in the Republic of Botswana (Table 3-A1 in Appendix 3): 1) Botswana1 = 53; 2) Botswana2 = 46; 3) Botswana3= 11). The names have been anonymised to avoid publishing locations of these sensitive populations.

The quality and concentration of DNA of all zoo, museum and wild samples were evaluated using 1% agarose gel electrophoresis and Nanodrop spectrophotometry (Thermofisher Scientific, Cambridge, UK), respectively. Samples were defined into categories based on DNA quality: 1) a tight band of high molecular weight DNA (> 1000 bp position) = high DNA quality; 2) a smeared band but relatively denser at > 1,000 bp position = moderate DNA quality; and 3) a completely smeared band (no noticeable band)

= low DNA quality. Samples classified as high quality were used for marker discovery using ddRAD sequencing, whereas a subset from different quality categories were used for targeted hybrid capture. Samples in category 3 that showed 260/280 nm absorbance ratio ≤1.7 (Thermo Scientific, 2010) and DNA quantity ≤ 10 ng/µl were not considered for testing.

3.3.2. ddRAD library preparation

DNA samples that met the gel-based quality thresholds described above were quantified prior to ddRAD library preparation using Qubit Broad Range dsDNA Assays (Thermofisher Scientific, Cambridge, UK). The final concentration of individual genomic DNA was normalised to 7 ng/µl. Fragments of genomic DNA were obtained using the *Sbf*I (restriction site - 5' CCTGCA|GG 3') and *Sph*I (restriction site - 5' GCATG|C 3') enzymes. To test repeatability of ddRAD, two samples from Botswana and two museum samples were each divided into two (IDs 318A-318B, 328A-328B, IDs M01-M02, and M03-M04) and the libraries of these duplicates were prepared independently. Illumina sequencing P1 and P2 adaptors were ligated to the ends of the DNA fragments (Table 3-1). Samples were pooled and size-selected using gel electrophoresis to obtain ~320 – 590 bp fragments, which were amplified using Illumina P1- and P2-adaptor specific primers (Brown et al., 2016) for 15 cycles of PCR. The amplicons (ddRAD library) were further purified using the MinElute PCR purification kit (Qiagen, Manchester, UK) and quantified using Qubit Broad Range dsDNA Assays (Thermofisher Scientific, Cambridge, UK). Samples were sequenced on an Illumina Miseq (v2 chemistry, paired-end run 2 × 160 nucleotides). Because we obtained and processed the samples from Botswana and museum samples before the EEP samples, the former were sequenced on one lane and the latter were later sequenced on another lane on a different run.

3.3.3. Bioinformatics and Analyses for ddRAD data

Reads were demultiplexed and adaptor sequences were removed for individual samples via the *process_radtags* module of the STACKS bioinformatics pipeline (Catchen et al., 2013). Forward and reverse reads were merged for each individual, and quality of the reads was assessed using FastQC (Andrews, 2010). Reads were trimmed before proceeding to further analyses to obtain >20 per-base quality scores. The retained reads were subsequently trimmed to a standard length of 135 because STACKS required identical

length of all reads (Catchen et al., 2013). Next, reads were assembled *de novo* using STACKS' *denovo_map.pl* module (*-m 4, -M 2, -n 2*). To choose the assembly parameters for *denovo_map.pl*, the parameter *m*, controls the minimum number of identical reads of a sample required to form a putative allele was trialled from one to six. The lowest value of *m* that provided ≥ 20× mean coverage across all stacks was chosen to avoid considering genotyping errors as putative alleles (Paris et al., 2017, Rochette and Catchen, 2017). Parameters *M* and *n* control the allowed numbers of mismatches between two alleles of a sample and the allowed numbers of mismatches between two alleles across samples, respectively. Increasing *M* and *n* values normally decreases the total number of assembled loci, while increasing the total number of polymorphic loci due to merging of duplicate loci. The values of *M* and *n* were chosen based on trial *denovo_map.pl* runs that varied *M* and *n* (*M* = *n*) from one to nine; the values that corresponded to the points where the number of assembled loci did not decrease further as *M* and *n* increased were chosen.

Following the *de novo* assembly, loci that met the following criteria were retained for bait design and hereafter are referred to as a set of potential SNPs (Rochette and Catchen, 2017, Bourgeois et al., 2018): 1) must contain exactly one SNP to filter out physically linked variants and to obtain conserved regions flanking individual SNPs that would facilitate primer/bait designs; 2) must contain exactly two alleles; 3) heterozygosity must be present in at least one sample but not in all samples, as loci containing more than two alleles and loci fixing or lacking heterozygotes might indicate the presence of repetitive sequences (heterozygotes were called when the depth of the minor allele was greater than 1/10th of that of the major allele; STACKS default); and 4) must be present in at least 70% of samples to minimise missing loci across the samples. Mean nucleotide diversity (π), described as the average number of per-site nucleotide differences between a pair of DNA sequences (Nei and Takahata, 1993), observed (H_o) and expected (H_e) heterozygosity (Catchen et al., 2011) of the retained loci were estimated using the *populations* module of STACKS (Catchen et al., 2017).

To visualise genetic structure among the samples used for ddRAD and to test for repeatability of ddRAD techniques of the duplicated samples, likelihood model-based estimation of ancestry was conducted to estimate population clustering using ADMIXTURE version 1.3.0 (Alexander et al., 2009) based on the retained SNPs. The analyses were iterated with varying numbers of possible genetic clusters (*K* = 1 - 12), with the *K* value that gave the lowest 10-fold cross validation error considered to be the most appropriate clustering (Alexander et al., 2015); results were visualised using R (R Core

Team, 2013). Clustering of the samples run in duplicate was used as a test of reproducibility of the ddRAD technique, given the quality of the tested samples. If the technique provided good reproducibility, duplicated runs of a sample would consistently produce read yields; hence, they must show identical ancestry membership proportions.

3.3.4. Bait design for sequence capture

A set of sequences containing potential SNPs developed from the ddRAD run were sent to Arbor Biosciences (Ann Arbor, USA) for bait design for sequence capture. For this purpose, SNPs that were present in at least 10% of samples were considered, rather than the more stringent threshold of 80% used for the admixture analysis. This was intended to allow the possibility of detection of rare SNPs using sequence capture. Two 80-nucleotide baits were designed complementary to each 135-bp sequence of the RAD locus containing the potential SNP. The team at Arbor Biosciences filtered the initial bait set *in silico* prior to manufacture of probes. Baits were soft-masked using Repeat Masker (Smit et al., 2015) for simple repeats and then BLASTed against a scaffold-level assembly of the northern white rhinoceros (NWR; *Ceratotherium simum cottoni*; GenBank accession number GCA_004027795.1). Baits were screened based on distributions of BLAST hits and corresponding melting temperatures (T_m – described as the energy required to separate hydrogen bonds between nucleotide bases). The bait screening was intended to select only baits that had a minimal number of hits around 65 C°, which would be used as the hybridisation temperature. For each bait sequence, the hit with the highest T_m was first excluded, and only the remaining top 500 hits as judged by bit scores were further considered for T_m distributions. Baits that had at most 10 hits with $Tm = 60 – 65$ C° or had at most two hits with $Tm > 65$ C° were retained (Table 3-2). Only baits that did not match to the mitochondrial genome sequence, and for which <25% of the sequences were masked for repeats, were retained and used to manufacture baits for use in the hybrid capture reactions.

3.3.5. Library preparation for bait capture

Libraries were prepared individually using NEBNext Ultra II FS DNA Library Prep Kits for Illumina according to the manufacturer's protocol (New England BioLabs Inc., Massachusetts, USA). Concentrations of DNA samples were normalised to 240 ng/26 µl, and fragmentations were performed at 37 ℃ for 10 minutes using NEBNext Ultra II FS

Enzyme Mix (New England Biolabs Inc., Massachusetts, USA) followed by incubation at 65 ℃ for 30 minutes. DNA fragments were then ligated to barcoded adaptors for Illumina sequencing (NEBNext Multiplex Oligos for Illumina Dual Index Primers Set 1; New England Biolabs Inc., Massachusetts, USA). Size selection for the adaptor-ligated fragments was performed using AMPure XP magnetic beads (Beckman Coulter Inc., California, USA). Bead:library ratios of 0.30× and 0.15× were used to bind and remove unwanted large and small DNA fragments from the libraries, respectively. To obtain the final library sizes between 320 – 470 bp; first, 30 µl of magnetic beads (0.30×) were added to an individual library, then the library was pelleted on a magnetic plate concentrator, large-unwanted DNA fragments bound to the beads were discarded, and the supernatant was retained. Next, 15 µl beads were added to an individual library (0.15×), then the libraries were pelleted on a magnetic plate concentrator, and the supernatant containing small-unwanted DNA fragments was discarded. Size-selected libraries were amplified for eight cycles using NEBNext Ultra II Q5 Master Mix and NEBNext i5 and i7 primers (New England Biolabs Inc., Massachusetts, USA) under the following PCR conditions: 1) initial denaturation at 98 ℃ for 30 seconds; 2) denaturation at 98 ℃ for 10 seconds; 3) annealing and extension at 65 ℃ for 75 seconds; and 4) final extension at 65 ℃ for 5 minutes.

3.3.6. Hybrid capture

Amplified libraries were hybridised to the baits according to the manufacturer's protocol (MyBaits, Arbor bioscience, Ann Arbor, USA). After normalisation to ~71.5/µl (~125 ng/1.75 µl), four libraries were pooled into each capture reaction, giving a total of 500 ng DNA input. Samples of similar DNA quality were pooled to allow for comparison of the hybrid capture performance by DNA quality, and to avoid possible bias in favour of high-quality samples. Hybrid capture reactions comprised two major steps: 1) hybridisation between libraries and baits; and 2) washing away of non-targeted sequences. In the first step, libraries were mixed with adaptor blockers, denatured at 95 ℃ for five minutes, and then incubated at 65 ℃ for five minutes to allow blockers to hybridise with adaptors ligated to DNA fragments. Baits were then introduced into the reactions and left at 65 ℃ for 16 hours to allow hybridisation between the baits and the libraries (Arbor Biosciences, 2018). DNA-bait hybrids were then transferred to a streptavidin-coated magnetic bead solution and incubated at 65 ℃ for five minutes to allow the beads to bind to the hybridised baits, which were then pelleted using a magnetic particle concentrator (ALPAQUA, Beverly, USA). The supernatants were discarded, and the pelleted beads

were washed three times using Wash Buffer X (Arbor Biosciences, Ann Arbor, USA). Bead-bound baits were resuspended in 10 mM Tris-CL, 0.05% TWEEN-20 solution (pH ~ 8.0 – 8.5). Bead-bound libraries were then detached from the beads; the libraries were heated at 95 ℃ for five minutes, followed by immediate pelleting using a magnetic particle concentrator, with only the supernatants being retained. Resulting libraries were PCR amplified for 11 cycles using reamp P5 and P7 primers (Meyer and Kircher, 2010) and Q5 Hot Start High-Fidelity 2X Master Mix, using the off-bead amplification protocol (New England BioLabs Inc., Massachusetts, USA). The conditions of PCR cycles were: 1) initial denaturation at 98 ℃ for 2 minutes; 2) denaturation at 98 ℃ for 20 seconds, annealing at 60 ℃ for 30 seconds, extension at 72 ℃ for 30 seconds; and 3) final extension at 72 ℃ for 5 minutes. Amplified capture libraries were purified using MinElute PCR purification kits (Qiagen, Manchester, UK). Final concentrations were measured using a Qubit dsDNA High Sensitivity Assay Kit (Thermofisher Scientific, Cambridge, UK). In total, eight capture libraries (32 samples) were pooled in equimolar concentrations and sequenced in a single lane of Illumina Hiseq 4000 (Novogene, Beijing, China), using paired end 160 bp sequencing.

3.3.7. SNP calling and performance for DNA of high and low qualities

SNP calling

Demultiplexing and initial filtering of raw reads were performed by Novogene (Beijing, China), in the following order: 1) raw reads were demultiplexed for each individual sample; 2) reads containing adaptor sequences were removed; 3) reads containing N bases (bases that could not be determined) > 10% were removed; and 4) reads for which > 50% of the total bases were of low quality ($Q_{phred} \leq 5$) were removed. FastQC (Andrews, 2010) was used to double-check for over-represented sequences that were present in the reads (i.e. leftover adaptors and Illumina primers), which were subsequently removed using Trimmomatic version 0.38 (Bolger et al., 2014). For each read, bases at the 3-prime end were trimmed when the mean quality score of four-base sliding windows dropped below 20 (Bolger et al., 2014). Forward and reverse reads of each individual were merged together using PEAR - Paired-End Read merger version 0.9.6.1 (Zhang et al., 2013). Cleaned reads were screened for contamination from other possible sources of DNA using FastQ Screen version 0.14.0 (Wingett and Andrews, 2018); mapping reads against a

mammal-size genome could be time-consuming, thus only subsets of ~2,000,000 reads randomly taken from each sample (--*subset 2,000,000*) were mapped against the draft genome of northern white rhinoceros (*Ceratotherium simum cottoni*, cerCot_v1_BIUU, GenBank Accession number GCA_004027795.1), as well as probable sources of exogenous DNA contamination: human (*Homo sapiens*, GRCh38, GenBank accession number GCA_000001405.15), mouse (*Mus musculus*, GRCm38, GenBank accession number GCA_000001635.2), and bacteria (*E. coli*, NCBI Reference Sequence number NC_000913.3).

Because there was no reference-quality genome of the SWR available in public databases at the time this work was conducted, I built a reference assembly *de novo* from the reads of all 32 individuals, using ABySS version 2.0.1 (Simpson et al., 2009). This assembly would be used as a reference for calling SNPs to allow comparison of the effect of DNA quality on the number of called SNPs, and hereafter referred to as the reference assembly. K-mer sizes of 41, 51, and 61 were trialled; 41 was selected based on the assembly's contiguity statistic, i.e. the longest N50 length. If all assembled contigs were sorted from the longest to the shortest contigs, N50 can be described as the minimum contig length required to cover 50% of the assembly, which means that at least half of the nucleotides in the assembly belong to contigs with the N50 length or longer (Gurevich et al., 2013). A minimum overlap of 50 nucleotides between two unitigs (small contigs) was required to form a longer contig (Jackman et al., 2017). A summary of the reference assembly metrics was obtained using Quast version 4.6.3 (Gurevich et al., 2013), including: the total length of the assembly; contig size distributions; N50; and mean %GC content (Table 3-3).

Reads from each individual sample were aligned against the reference assembly using the BWA MEM algorithm of the Burrows-Wheeler Alignment tool (Li and Durbin, 2009). The alignments generated Sequence Alignment Map (SAM) files, which were subsequently compressed to Binary Alignment Map (BAM) format; the BAM files were sorted and indexed using SAMtools (Li et al., 2009). MarkDuplicates (Picard tools, Broad Institute) was used to locate and flag duplicate reads present in the BAM files that started at exactly the same genomic position and had exactly the same insert size. Next, ANGSD (a computer software for ANnalyses of Next-Generation Sequencing Data) was used to call SNPs based on allele frequencies, with the null hypothesis that the frequency of minor alleles was zero (Korneliussen et al., 2014). Base alignment quality (*-baq 1*) scores were calibrated to avoid false SNP calls due to the presence of insertions and deletions around SNPs (Li, 2011). Only reads that mapped uniquely to a region on the reference assembly (-

uniqueOnly 1) with mapping quality and base quality scores ≥ 20 (*-minMapQ 20*, *-minQ 20*) and were not tagged as bad reads (*-remove_bads* 1, duplicate reads or not a primary alignment), were considered for SNP calling. SNPs were filtered according to the following criteria: 1) a variant site must have ≥ 10× coverage in at least 16 individuals (50% of samples), with the global depths across all samples between 160X – 1600X (*-minIndDepth 10, -minInd 16, -setMinDepth 160, -setMaxDepth 1600*) to be considered as a SNP locus; and 2) a SNP locus must have *p-value* ≤ 1×10^{-6} and show minor allele frequency (*MAF*) ≥ 0.05. BCFtools (Li et al., 2009) was used to collect per-sample SNP calling statistics from the obtained VCF files, including SNP depths, number of missing loci, and numbers of homozygous and heterozygous loci.

Effects of DNA quality on performance parameters of hybrid capture

Effects of original DNA quality on the performance of hybrid capture were assessed in relation to cleaned read yields, mean read lengths, mean mapping depths, percentages of duplications, and number of called SNPs. Thirty-two samples were chosen from the 110 wild Botswana SWRs for hybrid bait capture, based on the original DNA quality judged by 1% electrophoresis. The strategy was to equally represent each quality group (high and moderate DNA quality = 16, low DNA quality = 16). Parameter estimates were obtained using SAMtools, BCFtools (Li et al., 2009), and shell commands (Command 3-A3 in Appendix 3); between-group comparisons were made using independent-sample *t-tests* (R Core Team, 2013) and visualised using ggplot2 package for R (Wickham, 2016).

Capture sensitivity and specificity

The variable ddRAD loci used to design the baits were used as a reference (hereafter referred as the bait-reference) for measuring the sensitivity and specificity of the bait set. Cleaned reads were mapped against the bait-reference using BWA MEM (Li and Durbin, 2009), the outputs were then sorted and indexed using SAMtools (Li et al., 2009). The efficiency of the protocol was assessed based on mapping sensitivity and specificity. Sensitivity was defined as the per-sample percentage of the targeted bases (in the bait-reference) that were mapped by at least 1× (López-Domingo et al., 2014). To calculate this value, the total number of bases in the bait-reference was counted using Quast 4.6.3 (Gurevich et al., 2013). A Browser Extensible Data (BED) file containing only zero-coverage regions was then created for each BAM file of an individual sample, and the number of bases present in the zero-coverage regions were counted using shell commands

in BEDtools version 2.29.0 (Quinlan and Hall, 2010). For each sample, total zero-coverage bases were subtracted from the total bases in the bait-reference, the obtained value was then divided by the total bases and multiplied by 100 to acquire %sensitivity (Equation 3-1 and Command 3-A4 in Appendix 3).

$$\%sensitivity = \frac{(\#total\ bases - \#bases\ lacking\ coverage)}{\#total\ bases} \times 100 \quad \text{(Equation 3-1)}$$

Specificity was defined as the percentage of reads that mapped to the bait-reference, the inverse of which indicates resources wasted on non-targeted sequences (López-Domingo et al., 2014). The numbers of unmapped reads and total read numbers were determined using SAMtools (Li et al., 2009) and shell commands (Command 3-A4 in Appendix 3). Specificity was calculated as 100 - % unmapped reads compared to the total number of reads for each sample (Equation 3-2).

$$\%specificity = 100 - (\frac{\#unmapped\ reads}{\#total\ reads} \times 100) \quad \text{(Equation 3-2)}$$

3.4. Results

3.4.1. Sample quality

Both individuals obtained from the National Museum of Scotland and the 16 samples obtained from the EEP programme were classified as high quality based on gel electrophoresis and were included in the pilot ddRAD run. Two of the wild SWR samples from Botswana (Botswana1 = 1, Botswana2 = 1) that were deemed of high quality (Table 3-1) were also included in this initial ddRAD SNP discovery. Both of these samples were run in duplicate. However, sample ID 318A (the sample from Botswana1) showed a mean stack coverage (i.e. mean depth of potential alleles) of 12.50×, which was well below the cut-off value of 20× that was implemented to avoid genotyping errors (Fountain et al., 2016, Paris et al., 2017); therefore, it was removed from the analyses.

Following quality control and merging between forward and reverse reads, read yields for the first ddRAD run (Botswana and museum samples) were between 635,803 - 2,019,313, and the yields of the second ddRAD run (zoo samples) were between 394,950 - 4,208,919. Means of stack depths, i.e. depth of reads supporting each putative allele, ranged from 32.43× to 84.26× and 19.95× to 147.19× reads, respectively (Table 3-1).

The pilot ddRAD run generated 72,642 ddRAD loci across samples, 2,792 of which were retained after the filtering for potential SNPs ($\pi = 0.357$, $H_o = 0.355$, $H_e = 0.348$). Admixture analyses revealed $K = 3$ as the most suitable value with the lowest cross-validation error of 0.63751 (Figure 3-A2 in Appendix 3). This K value indicated that samples could have originated from three possible genetic clusters of ancestries; however, there was no evidence that the identified clusters were related to provenances of the zoo samples due to the lack of data about original sources of the individuals or their parents (in the case of captive-born individuals). Genetic admixture was revealed in most samples that involved three genetic clusters, except two of the zoo samples (IDs Z12 and Z16), two museum samples (IDs M01 - M02, M03 - M04), and the Botswana sample from Botswana2 (IDs 328A-328B) (Figure 3-2). The three pairs of duplicated samples sequenced with sufficient read depth to test revealed good repeatability; all were assigned to single genetic clusters with >0.99 ancestry membership proportion (Figure 3-2).

A total of 10,751 SNPs, with an overall length of 1,481,527 nucleotides (~0.06% of the genome size), were submitted to Arbor Scientific for bait design (Figure 3-1). Based on T_m distributions of the two baits per SNP screened, 13,330 baits were retained for hybrid capture, covering 8,896 of the initial potential SNPs by at least one bait.

3.4.2. SNP calling and performance of the RADcapture protocol for DNA of different qualities

Post-sequencing quality control

Overall, 97% of reads were retained after the initial quality control performed by Novogene (Beijing, China). The results from screening for DNA contamination revealed that, across samples, 55.14 – 71.46% of reads were mapped to the assembly of NWR; < 0.2%, < 0.1%, and ~ 0% were mapped uniquely to the genomes of human, mouse, and *E.coli*, respectively (Table 3-A5 in Appendix 3).

Reference assembly and mapping statistics

A total of 18,824 contigs were obtained from the *de novo* assembly of hybrid capture sequences that contributed to a total length of 30,315,726 nucleotides (~1% of the total length of genome) with average GC content and N50 of 41.66% and 2,508, respectively.

The largest contig contained 45,830 nucleotides; the distribution of contig sizes is provided in Table 3-3.

SNP calling

Post-sequencing quality control suggested minimal contamination from exogenous DNA; thus, the entire reference assembly was used as a reference for SNP calling. Overall 6,481 SNP loci were identified from hybrid capture sequencing for the 32 individual samples from Botswana. Mean locus depth across samples was 15.80×, and ranged from 6.60× to 27.75×; numbers of per-sample missing loci ranged from 60 to 5,125; numbers of homozygous and heterozygous loci ranged from 1,104 to 5,633 and 252 to 930, respectively (Table 3-4).

Effects of DNA quality on performance parameters of hybrid capture

Overall means of performance parameters were (Table 3-5 and Table 3-A6 in Appendix 3): 1) read numbers for Groups A = 27,992,072 and B =29,22,9659; 2) read lengths for Groups A = 141.38 and B = 140.06; 3) depths of coverage for Groups A = 5.97 and B = 6.61; 4) percentages of duplications for Groups A = 68.67 and B = 70.61; and 5) numbers of called SNPs for Groups A = 4,652.94 and B = 4,735.81. There were no significant differences between the two groups for any of the parameters considered (Figure 3-3). However, there was substantial variation among individuals in terms of read depths and number of missing loci. All individuals showed an excess of homozygous loci, with 5× as many homozygous to heterozygous sites, on average (ranging from 4.1× to 7.2×).

Capture sensitivity and specificity

The reference assembly built *de novo* from capture reads of the 32 samples contained 30,315,726 nucleotides (~1% of the entire SWR genome) whereas the originally targeted ddRAD locus bait set was only $1/20^{th}$ of this length (1,481,527 nucleotides). For comparisons with other studies and because of the baits capturing beyond the original targeted genomic regions, efficiency of the the protocol was calculated by comparison with the original bait set. Regarding sensitivity, the percentage of bases in the bait-reference with ≥ 1× coverage ranged from 51% to 65%, with an overall mean of 60%. For specificity, the percentage of reads that were mapped to the bait-reference ranged from 2.66 to 11.98 (Table 3-A6 in Appendix 3).

3.5. Discussion

Despite poor DNA quality in some samples, a total of 6,481 SNP loci were identified using targeted capture sequencing of 32 SWRs from Botswana, based on variants initially identified using ddRAD on a small number of samples from a wider range of geographic sources. Even in the 16 samples assigned as low quality that were included in the bait capture, 1,356 - 6,421 SNPs could be identified. Statistical tests showed no differences between high- and low-quality samples regarding the numbers of cleaned reads, means of read lengths, percentage of duplicates, mean depths of coverage, and numbers of called SNPs. The issue of low-quality DNA was particularly crucial because the chief aims of the thesis were to develop and genotype thousands of genetic markers that can be used in SWR studies from field-collected samples. Loci identified from the pilot ddRAD revealed good reproducibility based on the admixture analysis, indicating that ddRAD could provide cost- and time-efficient approaches to genotype non-model animals at a population-wide scale. However, only 22/110 of the field-collected samples screened here would have met the threshold quality for ddRAD sequencing. Moreover, even though high-quality samples were used exclusively for ddRAD, inconsistent genotyping and read yields was found for one duplicated sample, emphasising the risk of lost information using lower quality samples. Using ddRAD for SNP discovery and genotyping based on sequence capture of these SNPs appears to hold promise for this and other studies where sample preservation could be problematic.

3.5.1. ddRAD data and SNP discovery

Heterogeneity of ddRAD data

The primary aim of the study was development of reproducible SNPs that could be used for genotyping of molecular markers at a population-wide scale in non-model species. Genotyping SNPs using ddRAD provides a means to simultaneously discover and genotype a novel set of SNPs, but the obtained SNPs may be genotyped inconsistently across samples of varying quality due to heterogeneity in read numbers, especially when they are called *de novo* for non-model species (Maroso et al., 2018). In this study, even though only samples that met the threshold for ddRAD quality were included in the pilot study, substantial differences in read numbers between the lowest and the highest yield samples were revealed (ranging between 394,950 and 4,208,919 reads), which in turn led

to substantial difference in read coverages, as has been reported elsewhere (Graham et al., 2015, Maroso et al., 2018, Ewart et al., 2019). Moreover, sample ID 318A was dropped from further analyses due to lack of coverage and low read yield, whereas its duplicate (ID 318B) yielded sufficient read number and coverage to resolve a high assignment probability in the admixture analysis. This discrepancy highlights the potential inconsistency and reproducibility of ddRAD, particularly if lower quality samples are included. The variation in this study was possibly due both to variation in sample quality and the diverse provenance of sample sources, from zoos, museum samples and wild populations. Given that only 20% of the wild samples would have met the thresholds recommended for ddRAD (Table 3-A1 in Appendix 3), there would be a risk of losing genotypic information for substantial numbers of individuals if ddRAD were applied to all samples. Thus, it may not be suitable for studies that require reproducible SNPs that rely on low-quality DNA samples.

Ascertainment bias

Individuals included in the initial ddRAD screening for genetic variation originated from different populations and most individuals were not members of the targeted Botswanan SWR populations. The necessity to include outsiders was driven by the quality issue of the DNA samples but might have biased genetic parameter inferences made in the future for the targeted populations due to ascertainment biases. There are two major types of ascertainment biases: 1) minor allele frequency, MAF bias (Malomane et al., 2018); and 2) subpopulation bias (McTavish and Hillis, 2015). The MAF bias arises because a minimum threshold is set for SNPs to be called as polymorphic loci, which may introduce over-representation and under-representation of high- and low-*MAF* loci, respectively (Nielsen, 2004). Subpopulation bias happens due to only a small group of individuals being used to select/identify variable SNPs, which has been reported to overestimate genetic diversity of the population used in panel development and its closely related populations, particularly when the actual diversity is low (McTavish and Hillis, 2015). Regarding parentage studies, reduced genetic diversity due to ascertainment biases potentially decreases rates of successful assignments (Helyar et al., 2011); this was demonstrated in northwest Atlantic cod (*Gadus morhua*) from 13 locations, where reduced assignment rates were observed in association with increased geographical distance (Bradbury et al., 2011). Given the MAF threshold (≥ 0.05) applied to avoid considering of genotyping errors as minor alleles in this study and the necessity to include SWRs outside the targeted populations, there is a chance that any population genetic parameters estimated based on allele frequencies may be

compromised and possible ascertainment biases should be considered in the interpretation of parameter estimates.

Genetic clusters estimated from ddRAD data

The historical bottleneck during the colonial era of the species suggested that the SWRs present in this study might show no genetic structure. This was supported by a study based on DNA samples of 217 SWRs obtained from eight modern populations in South Africa that revealed only two haplotypes of mitochondrial control region (Moodley et al., 2018). The authors further inferred that the low haplotype diversity might be the result of the historical population bottleneck, which was reinforced by a hierarchical Bayesian model of 10 microsatellites (Moodley et al., 2018). The effects of the historical bottleneck were recently evidenced at the genomic level; based on whole genome re-sequencing of 30 pre- and 22 post-bottleneck NWRs and SWRs, post-bottleneck animals revealed significantly lower heterozygosity and higher inbreeding coefficient than pre-bottleneck animals (Sánchez Barreiro et al., 2020). However, in the present study, three genetic clusters were identified across the samples used for the initial screening for genetic variation using ddRAD. Most samples showed evidence of genetic admixture involving the three identified genetic clusters, except sample IDs 328A-328B, M01-M02, M03-M04, Z12, and Z16. Genetic substructure of the species was also revealed between modern captive and non-captive populations based on 10 microsatellites (Moodley et al., 2018) and whole-genome resequencing (Sánchez Barreiro et al., 2020). However, the samples of modern captive individuals in these studies were obtained from a variety of sources and metadata regarding specific origins of them were not provided. The existence of genetic clusters reported in these studies might be the result of ongoing genetic drift due to population fragmentation, which warrants the urgent need for genetic tools to facilitate metapopulation management to prevent loss of genetic diversity in managed SWR populations.

3.5.2. RADcapture efficiency

The protocol of RADcapture demonstrated in this study could enable screening of SNPs in a large number of samples of varying DNA quality, which is an encouraging result for future studies of endangered species for which it might not always be possible to obtain high-quality samples. Even with the relatively low quality of samples used in this study, the protocol used revealed relatively high per-locus mean SNP depth (ranging from 6.6× to

27.5×; Table 3-4) compared to 7.03× revealed in a study conducted in wild rainbow trout using fresh fin clip samples (Ali et al., 2016) and ≥ 4× reported in a study conducted in a plant genus (*Wisteria*) using frozen leave samples (Hoffberg et al., 2016). The percentage of on-targeted reads (%specificity) of the present study ranged only from 2.66 to 11.98, which was considerably lower than the two previous studies; 52.8% and >80%, respectively in Ali et al. (2016) and Hoffberg et al. (2016), despite lower numbers of per-sample reads (<1,182,936 and <200,000 reads, respectively) than I found (>9,009,770 reads). Low specificity reported in the present study was unsurprising because the hybrid capture not only yielded reads of the original targeted loci, but also reads from off-targeted regions. Screening for exogenous DNA revealed minimal contamination, which indicated that the baits could capture SWR DNA well beyond the original targets. This could be realised from the total size of the reference assembly (~1% of the total SWR genome), which was much larger than the original targets (~0.06% of the genome). Such findings are common in targeted captured sequencing and have been evidenced elsewhere (Portik et al., 2016). For example, a transcriptome capture study for phylogenomics of an African frog family (Hyperoliidae) reported that sequenced reads contained several thousands of flanking bases longer than the targeted transcript sequences (maximum size = 850 nucleotides) (Portik et al., 2016). The concatenated alignment of flanking plus targeted sequences was 631,127 nucleotides, whereas the concatenated alignment of the targets only was only 561,180 bases (Portik et al., 2016). The estimates of sensitivity and specificity in the present study were based on reads mapped to the original bait set, which is conservative but avoids complications with different length regions sequenced in each individual. Although RADcapture could capture well beyond the original targeted regions, it produced reads that were sequenced from relatively short targets without prior knowledge about where exactly the targets were in the SWR genome. With only 0.6% of the genome were targeted, most genomic regions would not be sequenced and might be responsible for the highly fragmented reference assembly.

Despite the promise of combining RADseq with hybrid capture enrichment, there are some limitations to consider. Variation in rates of PCR duplication across samples were found in the present study and were relatively high in some samples (ranging from 45.83 to 91.05%, Table 3-A6 in Appendix 3), which could affect confidence of genotype calls. However, this is comparable to the rates found in the initial RADcapture protocol (>80%), which involved digestion of DNA input prior to hybrid capture (Ali et al., 2016). A protocol to reduce the duplication rate has been developed but involves additional laboratory steps that

separate RAD tag isolation from sequencing library preparation into two distinct steps (both are commonly integrated in the traditional RADseq protocol and the present study) and physically isolate RAD tags from the entire genome prior to hybrid capture (Ali et al., 2016). The modified protocol was able to reduce the duplication rate; it also improved the percentage of on-targeted reads (from 20.5 to 52.8%) and per-locus SNP coverage (from 2.84× to 7.03×), and has been adopted by other studies (Delgado et al., 2019, Dorant et al., 2019, Komoroske et al., 2019). Hoffberg et al. (2016) employed an alternative approach by randomly appending an 8-nucleotide index into RAD fragments of each sample to allow identification and removal of duplicates (6 – 75% duplication rates) during downstream analyses. These studies invested laboratory efforts into an extra sample normalisation step to decrease variance among pooled samples. This suggested that such additional laboratory steps could improve efficiency of the hybrid capture protocol and should be considered for future studies.

3.5.3. SNP calling

Since the additional sequences captured outside of the bait set aligned to the NWR genome, the entire reference assembly built from captured reads was used as a reference for SNP calling to exploit all informative genomic data and to allow comparison of the effects of different DNA qualities on numbers of called SNPs. This approach was different from other RADcapture studies that used the 'radnome' constructed from sequences of RAD loci as a reference for read alignment and SNP calling (Ali et al., 2016, Hoffberg et al., 2016, Delgado et al., 2019, Komoroske et al., 2019), limiting the chance to discover SNPs other than those original targets. The present study did not aim to provide complete genotypes of all studied individuals at this stage, but aimed to screen for SNPs and to examine effects of DNA quality on performance of RADcapture protocol. Based on the criteria used for SNP calling, 6,481 loci could be identified with mean locus depth of 15.80× across samples. However, there was high variance among individuals regarding the per-sample mapping coverage (depth of reads mapped to the entire reference assembly; <15×) which subsequently resulted in high variance of missing SNPs (Table 3-4). Thus, SNP loci were called from a pool of samples rather than an individual sample to accommodate for the issue of low coverage (Korneliussen et al., 2014).

Applications of SNPs for studies on non-model wildlife species can be divided into three main categories (Garvin et al., 2010). The first category concerns the resolution of

population demography and structure, which requires selectively neutral and randomly sampled SNPs; for example, studies about effective population size, gene flow, and population dispersal (Mamoozadeh et al., 2019). The second category aims to classify or delineate individuals into groups, families, or species such as in studies concerning parentage assignment, population assignment, phylogeography, and individual identification (Garvin et al., 2010). This allows more relaxed assumptions of SNP neutrality than for population studies; in many cases, loci affected by local adaptation can provide greater power of distinguishability than selectively neutral SNPs (Bensch et al., 2002, Smith et al., 2005, Anderson et al., 2013, Ogden and Linacre, 2015). The third application for wildlife research is intended to link genotypes with phenotypic traits, which is dependent on prior knowledge about genome annotations. SNPs obtained in this study were intended to be used primarily in studies of the second category, specifically to develop a novel set of SNPs for parentage assignment and forensic work. For future uses of the obtained SNP set in the first category of applications, it may be necessary to remove loci that violate neutrality tests (Mamoozadeh et al., 2019). The set of SNPs present in this study would not be particularly useful for studies in the third category of applications because there is no information about genome annotations available either from this study or in public databases. Further, the baits were not designed to target individual genes or to cover the genome at sufficient depth for genome-wide association studies.

Of the 6,481 identified SNPs, high proportions of homozygous SNPs were found with overall mean ~5× (ranging 4.1× - 7.2×) as many homozygous to heterozygous sites (Table 3-4), which corresponded to ~17% (ranging 13 – 24%) of heterozygotes across all loci. Genome-wide heterozygosity, defined as the proportion of heterozygous sites in the genome of an individual, has been used to estimate genetic variation in various animal species (Dobrynin et al., 2015): 5% in Cheetahs; 21.9% in a Virunga Mountain Gorilla; 37.4% in an inbred Abyssinian cat; 75.9% in an outbred domestic cat. Interestingly, when genome-wide heterozygosity was considered for four captive SWRs, proportion of heterozygous sites was only 9% (Tunstall et al., 2018). The values in the present study might be overestimated due to the initial ddRAD screening that aimed to retain only variant sites for bait design. Additionally, biases of genotype calls towards homozygous genotypes can occur in exceptionally low-coverage sequencing data (Crawford et al., 2016); however, the overall mean of per-sample SNP depth of 15.80× should be sufficient to curb the effect of erroneous homozygous calls. Note that there was a difference between the definitions of ‘the proportion of heterozygous sites’ used in the present study and those

used in Dobrynin et al. (2015) and Tunstall et al. (2018). Herein, a heterozygous site referred to a SNP locus; whereas, it referred to a 50,000-bp sliding window in the others that was used to estimate genome-wide diversity in many mammal species (Dobrynin et al., 2015), and NWRs and SWRs (Tunstall et al., 2018). Despite the different definitions, the results reported in these three studies should provide a brief overview of genome-wide genetic diversity of the species that could lay the groundwork for future NGS-based studies.

3.5.4. Effects of DNA quality on RADcapture

An important contribution of the present study was establishing that the RADcapture approach was not particularly sensitive to DNA quality. The studies of Ali et al. (2016) and Hoffberg et al. (2016) did not explicitly demonstrate the effects of DNA quality on RADcapture performance. However, the latter study demonstrated the effects of low quantity of input DNA, which led to low yields of raw reads, high missing loci, and low mean coverage per million reads per samples. Low-quality DNA samples also have been included in more recent RADcapture studies. In a study that aimed to infer genetic divergence of populations of a fish species (*Galaxias maculatus*) using a modified protocol of RADcapture is described in Ali et al. (2016; i.e. one restriction enzyme without RAD tag isolation), a high percentage of missing loci was found in populations that contained low-quality DNA samples (Delgado et al., 2019). Komoroske et al. (2019) demonstrated the efficiency of a RADcapture approach (with RAD tag isolation exactly as described in Ali et al., (2016)) directed at the assessment of cross-species genetic variation in marine turtles, for both low (<15,000 bp) and high molecular weight DNA samples; although 16% of samples failed quality control and were removed from the study, there were no differences in these failed samples regarding the ages of sample preservation or molecular weight thresholds. The authors argued that RADcapture performed better than traditional RADseq for partially degraded samples (Komoroske et al., 2019). The performance of RADcapture in the present study (Table 3-A6 in Appendix 3) was consistent with Komoroske et al. (2019), supporting the potential and versatility of the RADcapture protocol to recover genomic data from degraded samples.

3.5.5. Conclusions

Southern white rhinoceroses have a very large geographical distribution. Their conservation could extend protection to other co-inhabitants of a vast area of land, and in this sense the SWR can be considered as an umbrella species (Branton and Richardson, 2011). Their horns are worth more than gold or cocaine per unit of weight, so the species has been targeted by criminal syndicates (Biggs et al., 2013). Despite the importance of SWR, very few traditional molecular markers have been available for the species. The 6,481 SNPs reported in this study demonstrated the potential to construct a SNP database to facilitate a wide range of conservation work. Regarding wildlife forensic work, the among-laboratory repeatability of SNPs, together with the advance in NGS technologies should enable database exchanges between laboratories and could speed up criminal prosecutions. For parentage assignment, more molecular markers mean higher statistical power for assigning an offspring to its parents and should improve rates of successful assignments. Southern white rhinoceros has been exemplified for the conservation success demonstrated by the continuous increase of the overall number. There remains the need of genetic tools to keep the success going and the SNP set obtained in this study is expected to contribute to that need.

Although useful for developing a panel of SNPs that could be used for other approaches, the per-sample costs for RADcapture sequencing used in this study can be considered as cost-inefficient from a field-practice point of view and complicated laboratory steps may hinder its popularisation. Even once baits have been designed, there is still a cost for individual library preparation and hybrid capture which is both cost- and time-consuming and may make the technique not suitable for field work. However, the genomic databases developed in this study (i.e. RADcapture sequences, identified SNPs) would provide a laboratory shortcut for developing a highly repeatable panel of markers.

Table 3-1 The final set of samples included in the pilot ddRAD run. Sources of samples indicate the locations from which tissue or blood samples were collected/obtained, Zoo 1 – 4 indicates samples from four different European zoological parks. Provenance provides information about the original origins of the animals (this information was not available for the zoo samples). P1 and P2 barcode sequences were ligated to the ddRAD libraries to allow pooling of samples before sequencing. Mean stack coverage and number of reads were obtained using STACKS *denovo_map.pl (-m 4 -M 2 -n 2)* module.

ID	Source of samples	Provenance	P1 barcodes	P2 barcodes	Mean stack coverage	Numbers of reads
318A	Botswana1	Botswana1	CGATA	CGATC	12.50×	180,779
318B	Botswana1	Botswana1	ATCGA	ATACGGT	32.43×	635,803
328A	Botswana2	South Africa	CTAGGAC	CATCTGT	54.95×	1,234,951
328B	Botswana2	South Africa	GCTAACA	TAGCA	50.25X	1,122,305
M01	Museum	Captive born	GCATT	GCATA	71.24×	1,667,426
M02	Museum	Captive born	ACGTA	GTCAAGT	64.04×	1,479,189
M03	Museum	Captive born	TGCAACA	CTGGT	84.26×	2,019,313
M04	Museum	Captive born	CAGTCAC	GAAGC	83.26×	2,001,104
Z01	Zoo 1	Captive born	TCAGA	GAGATGT	21.79×	419,223
Z02	Zoo 2	Wild born, South Africa	TGCAACA	CGATC	21.98×	436,903

Z03	Zoo 2	Wild born, unknown	GATCG	CATCTGT	22.63×	453,874
Z04	Zoo 3	Captive born	CGTATCA	CTGGT	22.06×	432,788
Z05	Zoo 3	Captive born	CATGA	GTCAAGT	21.96×	434,553
Z06	Zoo 4	Captive born	CACAGAC	GAAGC	23.71×	477,287
Z07	Zoo 1	Wild born, unknown	CGATA	CTGGT	19.95×	394,950
Z08	Zoo 3	Captive born	TCGAG	AGTCA	139.59×	4,156,118
Z09	Zoo 4	Captive born	TCTCTCA	TACGTGT	108.15×	1,479,867
Z10	Zoo 4	Captive born	GTCAC	GCATA	147.19×	4,208,919
Z11	Zoo 4	Captive born	GTACACA	GAGATGT	140.43×	2,089,315
Z12	Zoo 4	Captive born	GCATT	CGATC	129.80×	3,853,821
Z13	Zoo 4	Captive born	CTCTTCA	CATCTGT	143.06×	4,071,816
Z14	Zoo 4	Captive born	GACTA	GAGATGT	143.97×	2,033,262
Z15	Zoo 4	Captive born	ACGTA	ATACGGT	103.28×	1,433,962
Z16	Zoo 4	Captive born	CAGTCAC	TAGC	115.04×	1,558,421

Table 3-2 Filtering criteria applied for classification of bait stringency. Only baits that met one of the stringent or moderate conditions, did not match to the mitochondrial genome sequence, and for which <25% of the bait sequences were masked for simple repeats were included in hybrid capture reactions.

Stringency	Conditions
Stringent	≤10 hits T_m = 62.5 - 65 C°
	≤10 hits T_m = 62.5 - 65 C° and ≤2 hits T_m = 65 – 67.5 C°
	≤2 hits T_m = 62.5 - 65 C° and ≤1 hit T_m = 65 – 67.5 C° and ≤1 hit T_m ≥ 70 C°
	No hits with T_m > 60 C°
Moderate	≤10 hits T_m = 62.5 - 65 C° and ≤2 hits T_m > 65 C°
Relaxed	≤10 hits T_m = 62.5 - 65 C° and ≤4 hits T_m > 65 C°

T_m = melting temperature

Table 3-3 Distributions of contig sizes of the reference assembly reported by Quast 4.6.3 (Gurevich et al., 2013).

length of contigs	Number of contigs
≥ 0 bp	18,824
≥ 1,000 bp	9,443
≥ 5,000 bp	947
≥ 10,000	201
≥ 25,000 bp	6
≥ 50,000 bp	0

Table 3-4 Per-sample SNP statistics. BCFtools (Li et al., 2009) was used to count numbers of homozygous and heterozygous loci, means of per-site SNP depths, and number of missing loci.

Sample IDs	Homozygous	Heterozygous	Mean SNP depths (×)	Missing
131	3,158	705	12.4	2,618
132	4,454	872	17.5	1,155
134	3,704	811	14.4	1,966
135	2,318	556	9.9	3,607
151	1,104	252	6.6	5,125
154	5,343	863	23.3	275
155	3,801	786	13.9	1,894
166	3,415	700	13	2,366
171	4,158	882	15.9	1,441
172	5,603	818	27.1	60
177	4,344	868	17.5	1,269
184	4,056	805	16.5	1,620

185	3,994	844	14.4	1,643
187	5,633	785	27.5	63
188	5,490	867	25	124
201	2,992	622	11.8	2,867
205	3,670	836	13.7	1,975
207	3,865	810	14.2	1,806
238	2,801	637	11	3,043
242	4,656	913	17.3	912
278	3,346	748	12.5	2,387
302	2,873	694	11.2	2,914
304	3,186	730	12.1	2,565
312	5,248	855	21.2	378
315	4,595	890	19.3	996
326	4,183	922	15.5	1,376

328	3,754	807	15.4	1,920
330	4,303	833	17.5	1,345
347	3,186	681	12	2,614
348	4,725	930	18.1	826
352	4,291	889	15.4	1,301
355	3,087	673	12.4	2,721

Table 3-5 Comparison of hybrid capture performance for DNA of different qualities. Group A and B represent low- and moderate- or high-quality samples, respectively. Independent sample t-tests were used to determine significance of differences between the groups for each performance statistics.

Parameters	Mean group A	Mean group B	*p-value*	*S.D.*	*Effect size*
Number of reads	27,992,072	29,229,659	0.76	11,097,340.70	0.11
Read lengths	141.38	140.06	0.27	3.25	-0.41
Depth of coverage	5.97	6.61	0.65	3.78	0.17
% duplications	68.67	70.61	0.65	11.52	0.17
Number of called SNPs	4,652.94	4,735.81	0.84	1,101.93	0.08

S.D. = standard deviation estimated from a pooled value of both groups

Effect size = (mean Group B – mean Group A)/*S.D.*

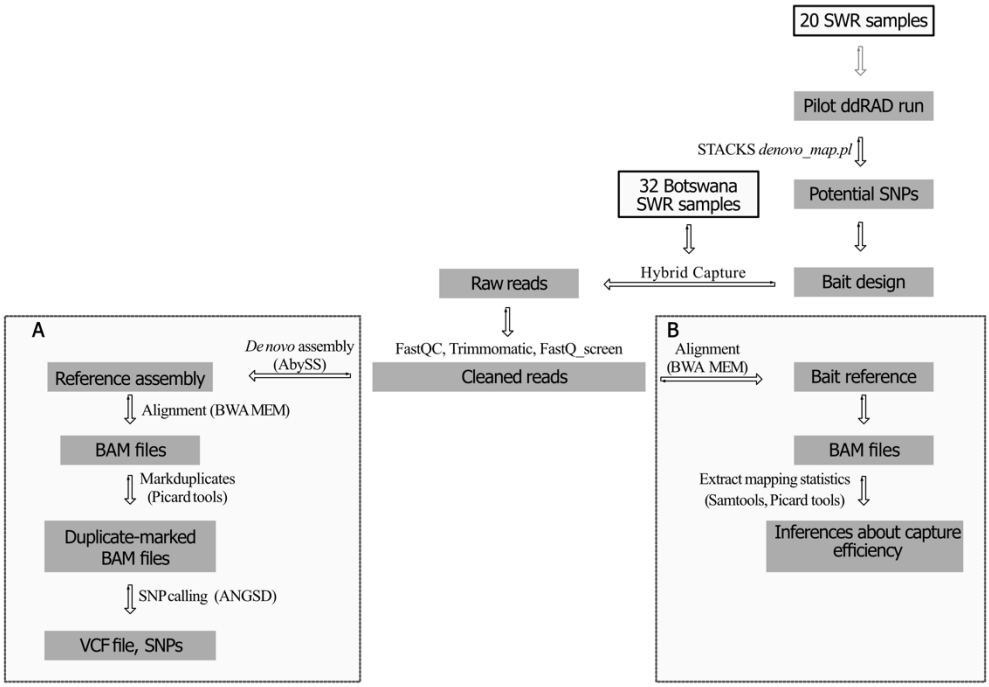


Figure 3-1 Workflow of SNP discovery. Initial screening for potential SNP loci was made for 20 SWR samples that met the high-quality threshold for ddRAD sequencing. The retained SNPs were subsequently used to design baits for hybrid capture for an additional 32 Botswana samples chosen to equally represent low- and high-quality DNA samples. The reference assembly was built *de novo* from sequenced reads of the 32 samples and was used as a reference sequence for SNP calling by mapping reads of each individual sample back to the assembly (Box A). For inferences about capture efficiency (Box B), sequenced reads were mapped against the original ddRAD loci to assessed % sensitivity and specificity.

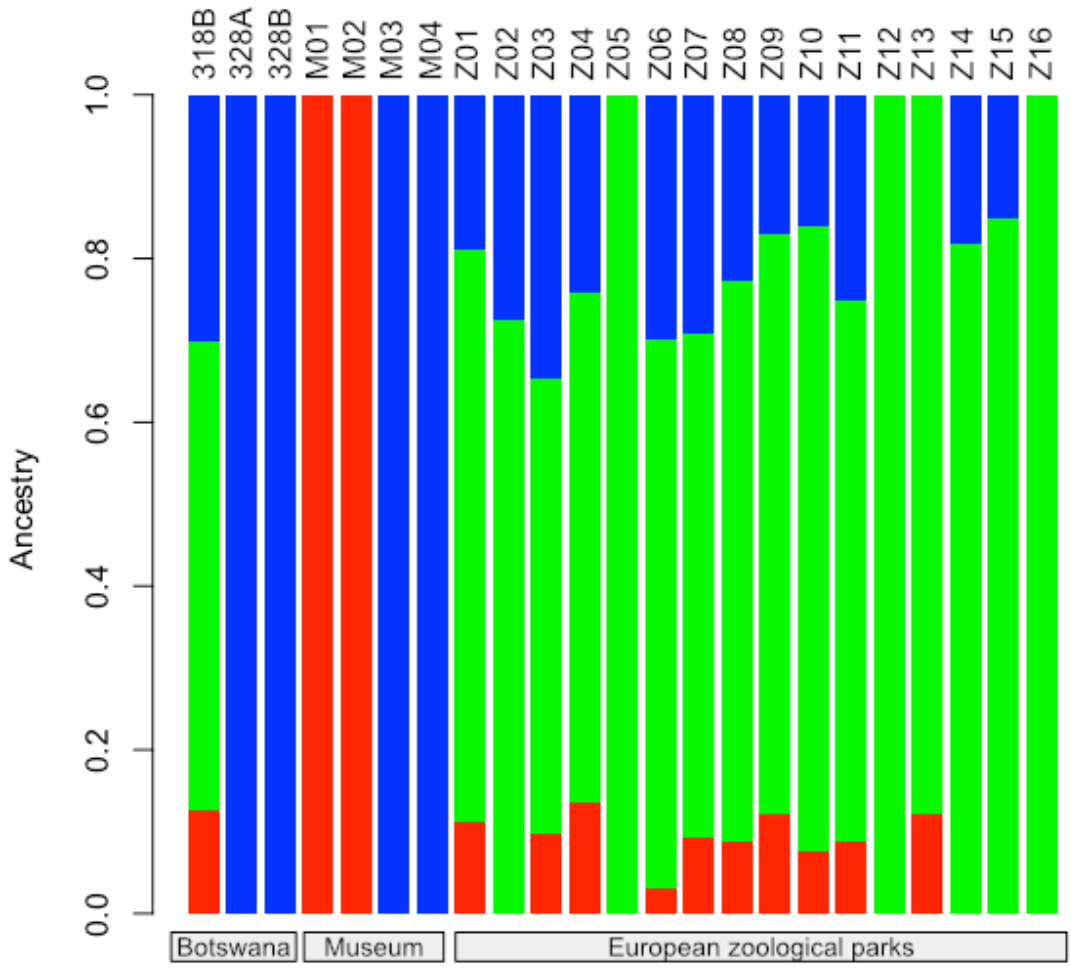


Figure 3-2 Estimates of sample ancestries based on the assumption of three genetic clusters (K = 3), indicated by colours. Each column represents a sample, with the corresponding ID numbers above the bar and the origin of samples indicated on the X axis. Ancestry membership proportions are indicated on the Y axis. Note that individuals 328A & 328B (Botswana), M01 & M02, and M03 & M04 (Museum) are duplicate runs from the same individual; their identical profiles suggest high repeatability of ddRAD genotyping. For the other Botswana sample (318B), only the sequences from one of the duplicates was of sufficient quality to be included.

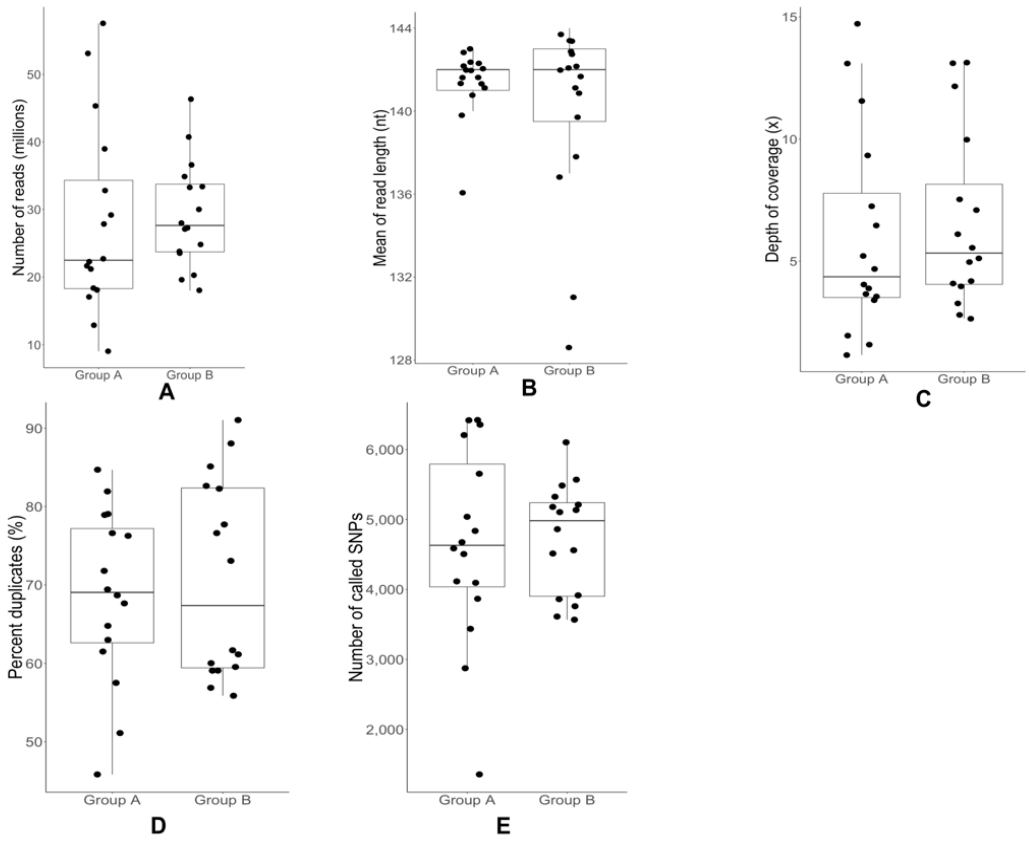


Figure 3-3 Boxplots comparing RADcapture performances between sample groups. (A) Cleaned read number; (B) mean read length; (C) depth of coverage; (D) percent reads flagged as duplicates; and (E) number of called SNPs. Group A (N=16) represents samples with low DNA quality and Group B (N=16) represents samples with high or moderate DNA quality, upper and lower margins of the boxplots indicate upper and lower quartiles of the data, respectively.

3.6. Data Accessibility

Sequence Read Archive (SRA) accession numbers SRR10902285 - SRR10902372 (Bioproject accession PRJNA601631) for raw hybrid capture reads. The reference assembly used in the alignment of reads to call SNP and VCF file can be found at European Variation Archive (EVA) accession number ERZ1300027 (project accession PRJEB36590).

Chapter 4 Parentage assignment and approaches to identify candidates for translocation and breeding in three privately managed populations of southern white rhinoceros

4.1. Abstract

Most African remaining southern white rhinoceros (SWR) have been kept in protected areas, resulting in small and fragmented populations that require regular exchanges of animals to reduce inbreeding. However, genetic information has been rarely incorporated to identify candidates for translocation and breeding, partly due to low levels of genetic variation caused by extreme historical bottlenecks. Using the RADcapture approach described in chapter 3, the purpose of this study was to: 1) assess whether use of genome-wide single-nucleotide polymorphisms (SNPs) could improve reliability of parentage assignments for pedigree reconstruction compared to microsatellite genotyping (Chapter 2); 2) determine whether molecular marker-based mean kinship estimates could be useful for identifying genetically distinct or inbred individuals, in the absence of a robust pedigree; and 3) compare patterns of genetic diversity, heterozygosity and population differentiation in three privately managed populations in Botswana, in order to assess whether between-population differences could inform translocation decisions. Parentage assignments for the Botswana1 population based on microsatellites (MS; described in Chapter 2) were compared to inferences based on SNP-based probabilities determined using AlphaAssign. A consensus pedigree was then constructed and individual (MK_{ID}) and within-population mean kinships ($MK_{population}$) were estimated from the pedigree (for Botswana1) and from the SNPs only (for Botswana1, Botswana2, Botswana3) to identify potential candidates for translocation and breeding. The MK_{ID} and $MK_{population}$ and between-population MKs estimated from the SNP data were used to identify appropriate sources of introduced animals for each population. Since only four of the Botswana3 samples met the quality threshold, individuals from that population were only considered to assess between-population variation. Principal component analysis (PCA) was used to illustrate genetic structure among the three populations. Using genotypes of 302 SNPs, parentage assignments made for 24 offspring revealed that while the SNP analyses tended

to show the same or better resolution as microsatellites, they did not overcome problems associated with incomplete sampling of putative parents: consensuses between SNPs and MS were revealed for nine offspring; better resolution was shown by SNPs for seven offspring; neither SNPs nor MS could identify parents for one offspring; and SNPs assigned alternative parents when SNP genotypes of the MS-assigned parents were not available for seven offspring. Pedigree-based estimates of mean kinship for the 49 individuals from Botswana 1 showed a lower population mean (pedigree $MK_{Botswana1}$ = 0.051; ±2SD = 0.033 – 0.069) than marker-based estimates (marker $MK_{Botswana1}$ = 0.061; ±2SD = 0.055 – 0.067), but higher variance. However, the two methods for kinship estimation suggested completely different individuals for translocation, based on outliers above a threshold of $MK_{population}$ ± 2SD. Molecular estimates of $MK_{Botswana2}$ was 0.067 (N = 35, ±2SD = 0.059 – 0.075) and no animals showed MK_{ID} beyond the thresholds (±2SD) and so no candidates for translocation or breeding could be identified. No genetic structure across the three populations was observed from the PCA and the between-population MK estimates suggested that exchanges of animals between any pair of them would minimise MKpopulation of the recipient populations. Assignments made to the alternative parents by SNPs when putative parents were not completely sampled indicates the potential limitation of probabilistic methods for parentage assignment in wild populations that complete sampling may be difficult to obtain. The markers developed and methods demonstrated in this study should offer genetic tools for selection of animals for translocation and breeding that should improve the management of isolated populations of SWRs to avoid inbreeding. However, the inability to identify suitable candidates for translocation based on marker-based kinship estimates for the populations without pedigrees available emphasises the benefits of combining observational data with molecular markers for parentage assignments.

Keyword: Botswana • Genetic diversity • Kinship • Parentage • Population management • SNP • Translocation

4.2. Introduction

Ex situ conservation of endangered wildlife populations relies on population genetic parameters that provide insights about population dynamics to allow the most efficient management (How et al., 2009, Henkel et al., 2012). Many population parameters are dependent on accurate pedigrees; missing or false information present in pedigrees may over- or underestimate such parameters of interest (Henkel et al., 2012, Farquharson et al., 2019). The parameters that describe relationships (e.g. coefficients of kinship, relatedness, inbreeding) between animals within a population are particularly important for management of isolated wild populations and *ex situ* insurance populations. Gene flow and mating choices are often limited in such populations, posing a threat of inbreeding depression (decrease of fitness due to mating between close relatives) that may affect genetic viability of the populations (Frankham et al., 2017). For instance, if unequal genetic contributions of adult animals happen unnoticed due to an incomplete or inaccurate pedigree, a substantial number of inbred offspring may be reproduced, resulting in genetic erosion in the subsequent generations (Farquharson et al., 2019). Thus, prevention of mating between closely related animals and selection of genetically valuable animals for breeding (i.e. those who have not contributed many offspring) are crucial to minimise levels of inbreeding in populations with limited gene flow.

Accurate pedigrees can be constructed from observational records about parent-offspring relationships and/or from molecular markers (Blouin, 2003) in order to identify genetically valuable individuals that are predicted to have a low risk of reproducing inbred offspring can be determined. Regular exchanges of animals are often implemented to ensure genetic viability of isolated managed populations, such as: Florida panthers, *Puma concolor coryi* (Benson et al., 2011); golden lion tamarins, *Leontopithecus rosalia* (Kierulff et al., 2012); African Wild dogs, *Lycaon pictus* (Davies-Mostert et al., 2015); Tasmanian devil, *Sarcophilus harrisii* (Hogg et al., 2017); and African lions, *Panthera leo* (Miller et al., 2015, Miller et al., 2019). However, construction of pedigrees merely from observational information is challenging for wild populations due to the difficulty to identify individual animals, let alone identification of their relationships. Initial applications of molecular markers for parentage studies were based on minisatellite markers (DNA fingerprinting) employed to detect extra-pair breeding in behaviourally mono- and polygamous bird species (Westneat and Sherman, 1997, Petrie and Kempenaers, 1998). Applications of microsatellite markers across a wide range of species and the advance of statistical theories

for parentage assignment have popularised the uses of molecular markers for construction of pedigrees for both wild and managed populations (Jones and Ardren, 2003). To date, there have been six major categories of molecular-based approaches for parentage assignment (reviewed in Jones et al., 2010): 1) exclusion; 2) categorical allocation; 3) fractional allocation; 4) full probability parentage analysis; 5) parental reconstruction; and 6) sibship reconstruction. These approaches are by no means mutually exclusive and often parentage studies required combination of approaches to construct pedigrees (Hadfield et al., 2006, Flanagan and Jones, 2019).

The first approach, exclusion, is based on rules of Mendelian inheritance. Given a locus of a diploid animal, a parent must share at least one allele with an offspring; otherwise, the proposed parent can be excluded and considered genetically impossible (Chakraborty et al., 1974). Complete exclusion is obtained when only a single putative parent is retained. The concept of exclusion is straightforward but prone to false conclusions attributable to DNA mutations, which occur at relatively high frequency for microsatellite loci, ~10^{-6} to 10^{-2} substitutions per locus per generation (Eckert and Hile, 2009, Fischer et al., 2017). Null alleles, lack of amplification of particular alleles, can also compromise the exclusion approach. They occur when the polymerase chain reaction (PCR) fails to amplify an allele at a locus and results in incorrect genotype calls (Dakin and Avise, 2004). If this happens for both alleles of a diploid, the locus would be considered missing for the individual; if only one allele fails to amplify, the individual would be considered homozygous regardless of its actual genotype, which could result in false exclusion (Pemberton et al., 1995). Both mutations and null alleles are commonly accommodated by allowing a previously specified number of allele mismatches between the parent and offspring (Jones et al., 2010), which may contribute to incomplete assignment (e.g. if multiple putative parents cannot be excluded), particularly if a panel of markers used has low diversity.

The second approach, categorical allocation, is often applied to a group of retained putative parents if the exclusion fails to obtain a complete assignment. It exploits probabilistic statistical frameworks to determine the most likely parent for an offspring. Maximum likelihood was initially developed to explicitly deal with a situation that there might be many potential parents and social structures were difficult to assess, a situation that is commonly observed in wild populations (Meagher and Thompson, 1986). Bayesian statistics were later introduced to enable inclusion of prior information (e.g. social structure, sexually dominant individuals) into the models (Nielsen et al., 2001, Hadfield et al., 2006). Detailed methods of the categorical approach were described in the materials

and methods in Chapter 2. Multiple software packages have been developed for parentage analysis: 1) likelihood-based approaches, including FAMOZ (Gerber et al., 2003), CERVUS (Kalinowski et al., 2007), COLONY2 (Wang and Santure, 2009), SNPPIT (Anderson Eric, 2012), SEQUOIA package for R (Huisman, 2017), FRANz (Riester et al., 2009), and AlphaAssign (Whalen et al., 2019); and 2) Bayesian approaches, including MasterBayes (Hadfield et al., 2006) and SOLOMON (Christie, 2013). A Major drawback is that most probabilistic methods are based on the assumptions that populations are panmictic (Kalinowski et al., 2007) and individuals in a parental generation are unrelated (Jones and Ardren, 2003) which assumptions are not always met in wild populations (Meagher and Thompson, 1986). Also, estimation of statistical confidence relies on a comparison of the assignment scores of the proposed parents; thus, the probabilistic methods are most efficient when individuals have the same amount of genetic data (i.e. same number of typed loci) and are sensitive to missing genotypes (Dodds et al., 2019, Whalen et al., 2019).

The third approach, fractional allocation, employs a similar strategy as the categorical approach; however, instead of determining a most likely parent, it allocates fractions of paternity and maternity to multiple parents (Devlin et al., 1988). Although it is impossible in reality that an offspring is the progeny of multiple fathers or mothers, it provides a means to estimate genetic parameters at a population level without an absolute pedigree; e.g. to identify variances in reproductive success (Koyano et al., 2013) and patterns of dispersal (Saro et al., 2014). However, limitations of the fractional assignment are that applications of the approach are limited to only studies concerning patterns of population parameters and it does not actually construct pedigrees.

The fourth approach, full probability, allows direct incorporation of explanatory variables other than genotypic data (e.g. status of dominance, geographic locations, phenotypic data) in the construction of pedigrees (Hadfield et al., 2006). For example, probability of parentage might be inversely proportional to spatial distance between a proposed parent and the offspring (Krutovsky et al., 2012). Unlike the former three approaches, full probability employs Bayesian statistics to simultaneously identify parents of individuals and infer population parameters which should introduce less bias in the parameters than other assignment approaches that rely on the assumption about panmictic population structure (reviewed in Pemberton, 2008). The approach was demonstrated to successfully assign parentage by incorporating the information about spatial proximity of Seychelles warblers (*Acrocephalus sechellensis*), a species that had been characterised as having low

genetic diversity and therefore parentage assignment was previously challenging (Hadfield et al., 2006). However, the benefits of the approach come with the cost that the assumptions about the explanatory variables (which may be difficult to determine in wild populations) must be met to obtain accurate pedigrees and unbiased population parameters (Hadfield et al., 2006).

The fifth approach, parental reconstruction, is suitable for a situation in which a cluster of offspring is known to contain only full- and half-siblings (e.g. strings of squid eggs; Emery et al., 2001); thus, the genotypes of offspring can be used to reconstruct the genotypes of their parents without access to parental samples (Emery et al., 2001). If alleles of one parent are known and shared across offspring, genotypes of the unknown parent(s) can be reconstructed and can be used to estimate number of the unknown. This approach has been employed to answer various research questions such as identification of patterns of polyandry in field crickets (*Gryllus bimaculatus*; Bretman and Tregenza, 2005) and Cheetahs (*Acinonyx jubatus*; Gottelli et al., 2007) and patterns of gene flow between demes of amphibians (Jehle et al., 2005). The approach requires highly polymorphic markers and a large cluster of siblings to efficiently construct genotypes of the unknown parents; also, if the unknowns contribute to small number of offspring (e.g. < 8 – 10) in the cluster, the probability to successfully reconstruct genotypes is reduced considerably due to binomial sampling (Jones et al., 2010). For example, given that low-diversity markers are used to identify parentage of a cluster of 100 eggs, if one of the parents contributes to < 8 eggs, there is a reasonable probability that only one allele at a heterozygous locus of the parent will segregate in the cluster.

The sixth approach, sibship reconstruction, aims to cluster a collection of offspring into groups of full- and half-siblings without prior knowledge about whether these offspring originated from the same family (Ashley et al., 2009). Individuals are clustered in a configuration that maximises the probabilities of all sibling families, given that the alleles are following Mendelian segregation patterns and the population is under Hardy-Weinberg equilibrium (Konovalov et al., 2004, Wang, 2004a). The approach is useful for parentage analysis when a large collection of samples can be collected with or without connection to any particular parent; if samples from a pool of parent are available, parentage can be assigned to each group of siblings. More recently, this approach has been combined with the methods for parental reconstruction (i.e. inferring genotypes of parents/grandparents of sibling groups) to connect sibling families to increase the width and depth of pedigrees (Huisman, 2017, Whalen et al., 2019). With sufficient width and depth of pedigrees,

estimation of population parameters should be improved; for example, Balloux et al. (2004) demonstrated in simulated populations that 5-generation pedigrees could capture 90% of variance of inbreeding coefficients estimated from 50-generation pedigrees.

These approaches and their statistical frameworks have been built to serve the applications of microsatellite markers for parentage assignment. Microsatellites contain many beneficial attributes that make them suitable for parentage studies in both domestic and wild species, such as Mendelian inheritance, co-dominance, neutrality, and multiple alleles per locus (Jarne and Lagoda, 1996). Often, they provide sufficient statistical power to achieve a high proportion of successful parental assignments. For example: high parentage exclusion probabilities (> 0.99) were obtained for seven endangered parrot species based on 106 loci with N_a = 6.4 – 8.3, observed heterozygosity (H_o) = 0.65 – 0.80, and expected heterozygosity (H_e) = 0.60 – 0.80 (Jan and Fumagalli, 2016); 12 microsatellites with mean N_a = 9.17, H_o = 0.710, and H_e = 0.722 provided exclusion probabilities > 0.999 in Holstein-Friesian cattle (Ozkan et al., 2009); 25 fathers and 27 mothers of 28 ring-tailed lemurs (*Lemur catta*) could be identified using 10 loci with N_a = 5 – 8, H_o = 0.417 – 0.722, and H_e = 0.438 – 0.795 (Breton et al., 2019). However, in species that experienced a population bottleneck and show low genetic diversity, microsatellite markers can be insufficient to obtain such high successes. For instance, in a modern European bison population (*Bison bonasus*) where >80% of the population were descendants of only two individuals, the application of 17 microsatellite loci (mean N_a = 3.06, H_e = 0.31) could identify fathers for only two out of 92 offspring due to low heterozygosities and low numbers of alleles per locus (Tokarska et al., 2009). Similar results were reported in southern white rhinoceros (SWR) that experienced a historical bottleneck; the species is characterised by low genetic diversity based on the available microsatellites (see Chapter 2), where N_a ranged from 2.5 to 2.8 and H_o and H_e ranged from 0.363 to 0.597 and 0.393 to 0.578, respectively (Florescu et al., 2003, Coutts, 2009, Guerier et al., 2012, Harper et al., 2013, Moodley et al., 2018, Purisotayo et al., 2019). A SWR study that combined robust demographic data (known mother-offspring relationships) and microsatellite genotypes was able to construct complete population pedigrees (Guerier et al., 2012), whereas, studies with incomplete demographic data were unsuccessful to do so (Coutts, 2009, Labuschagne et al., 2017, Kretzschmar et al., 2019, Purisotayo et al., 2019). Since obtaining a complete pedigree for SWR is challenging based on the currently available microsatellites and observational data, identification of candidates for translocation and breeding is particularly challenging.

A simple solution for the lack of statistical power of microsatellites would be increasing the number of markers employed in the analyses. With only a maximum of 23 microsatellites (Harper et al., 2013) and 33 SNP loci (Labuschagne et al., 2017) currently available for SWR and some of them being monomorphic (Purisotayo et al., 2019), there is a need for a novel set of markers to inform conservation management. Single nucleotide polymorphisms are the most abundant molecular markers; in human populations, they are expected to occur once for every 500 – 1000 bp (Wang et al., 1998). Unlike microsatellites, for which genotypes are called relative to size-standards which may be different across laboratories (Moran et al., 2006), each SNP directly represents a variant site in a DNA sequence. Thus, the genetic data are automatically standardised and exchangeable across sequencing platforms and laboratories (Glover et al., 2010). However, SNPs are commonly treated as biallelic and theoretically the analytical power of individual markers are relatively low compared to microsatellites (Krawczak, 1999). Regarding parentage analysis in domestic and wildlife populations, SNPs thus have been reported to require more markers than microsatellites to achieve similar statistical power for resolving parentage in domestic and wildlife populations (Tokarska et al., 2009, Hauser et al., 2011, Buchanan et al., 2016, Kaiser et al., 2017). For SWR, 33 currently available SNPs developed exclusively for the species (Labuschagne et al., 2013, Labuschagne et al., 2015, Labuschagne et al., 2017) were not sufficient to resolve a complete pedigree for a population of 32 SWRs in South Africa (Labuschagne et al., 2017). However, SNP-based approaches can sometimes improve estimates even based on relatively few markers. For example, in European domestic pigs, SNPs from at least 30 polymorphic loci (H_e = 0.150 – 0.950) yielded better exclusion probabilities than 12 microsatellites (N_a = 4 – 9, H_e = 0.711 – 0.889) (Yu et al., 2015). Kaiser et al. (2017) identified paternity of 238 black-throated blue warblers using both microsatellites and SNPs, with 92% consistency (219 juveniles) between the assignments made from six microsatellites (mean N_a = 18.92, H_e = 0.87, and H_o =0.86) and 97 SNPs (mean N_a = 2.0, H_e = 0.22, and H_o =0.19). However, the advance of next-generation sequencing (NGS) has provided the means to generate high-throughput sequencing data, which can facilitate SNP discovery and genotyping across a wider range of loci, to enable increased power for parentage assignment and genome-wide perspectives on variation (reviewed in Garvin et al., 2010).

Discovery of novel SNPs is expected to be most thorough and efficient in model species for which high-quality reference genomes are available. Screening for SNPs in such species is often done *in silico* by selection of putative variant sites from public databases,

then validating the chosen SNPs in a particular group of individuals (Hawken et al., 2004). However, for non-model species, the initial screening for variation has to be done *in vitro* by sequencing DNA fragments of multiple individuals to allow identification of variant sites. Several sequencing approaches have been implemented to ensure that the same genomic regions across individuals are sampled and sequenced. For example: 1) amplicon sequencing employs nucleotide-tagged PCR primers to construct homologous amplicons from multiple individuals and subsequently sequencing the amplicons using NGS sequencing platforms (Binladen et al., 2007); 2) Restriction Associated DNA sequencing (RADseq) selects DNA segments by cutting the genome at specific recognition sites of restriction enzymes, typically followed by size-selection and sequencing of the retained fragments (Baird et al., 2008, Davey and Blaxter, 2011); 3) targeted capture sequencing uses probes designed from genomic sequences of closely related species (or other forms of prior genomic knowledge) to enrich the genome for specific sequences prior to sequencing (Gnirke et al., 2009, McCue et al., 2012). These sequencing approaches combined with deep sequencing using NGS platforms (Van Tassell et al., 2008) can efficiently develop a reproducible set of SNPs without a reference genome.

There are two major challenges for developing a new set of informative SNPs for SWR: 1) the species lacks a high-quality reference genome or polymorphisms databases; and 2) biological samples obtained under field conditions may contain some extent of DNA degradation. Regarding the former challenge, the scaffold assembly available in the National Center for Biotechnology Information (NCBI, CerSimSim1.0, GenBank accession GCA_000283155.1) is fragmented into 57,824 contigs without complete annotations. High-quality genome of domestic horses was employed to target 54,000 SNPs for parsimony analysis of the members of Perissodactyla (odd-toed and members) family, including 2 SWR and 2 NWR individuals (McCue et al., 2012). Nevertheless, only 10% of the targeted loci were consistently genotyped across four white rhinoceroses. Further, genetic parameters estimated from markers that are developed from other populations or from another species may be compromised due to ascertainment biases (McTavish and Hillis, 2015, Malomane et al., 2018). For the second challenge, biological samples of SWR are often obtained from fieldwork where ideal conditions for sample preservation may not be met (Camacho-Sanchez et al., 2013). The condition of samples may be severely degraded in a situation that has to deal with decayed carcasses, such as a forensic case (Ludes et al., 1993). The combination of both challenges limits the applications of targeted capture techniques that require some form of prior genomic knowledge and of RADseq

(and derivatives) that require high molecular weight DNA to provide sufficient fragment length for restriction enzymes to cut at recognition sites evenly across genomes (see Chapter 3; Baird et al., 2008, Peterson et al., 2012). Using degraded DNA samples as RADseq inputs could reduce per-individual read yields, percentage of identical sequenced regions, and the number of variant sites identified (Graham et al., 2015). Given these limitations, either target capture sequencing or RADseq alone is not suitable for SNP discovery and genotyping in SWR.

RADcapture (described in Chapter 3), which combines RADseq and targeted capture sequencing, to integrate different advantages from the two sequencing strategies provides one possible solution (Ali et al., 2016, Hoffberg et al., 2016). Specifically, RADcapture relies on the advantage that RADseq does not require prior genomic knowledge and the benefit that targeted capture sequencing requires relatively short DNA fragments to sample homologous sites across individuals (Carpenter et al., 2013). The ability to recover SNPs from degraded samples without a reference genome can benefit experimental designs to answer a wide range of research questions, such as studies that are dependent on museum or fossilised specimens (Blaimer et al., 2016, Delsuc et al., 2019). In Chapter 3 (3.3.6 Hybrid capture), RADcapture was applied to 32 SWR samples and identified 6,481 SNPs. Despite reports of a high proportion of missing information (i.e. genotypes) in populations with degraded DNA samples has been reported in other species elsewhere (Delgado et al., 2019), the preliminary analyses conducted on SWR performed equally well in both high- and low-quality samples regarding read yields, read lengths, depths of coverage, percentages of duplication, and numbers of identified SNPs. The advantages of RADcapture also extends to parentage assignment studies that preferably involve samples from all members of a population because the approach allows samples to be pooled in a single reaction; thus, pedigree constructions can be done with affordable costs. In Chapter 3, only a small number of individuals were tested to specifically test implications of resolving markers based on samples of varying quality but the sampling was no conducive to assessing impacts on the reliability of population-level inferences from these data.

Reduced representation methods for SNP discovery (genotype by sequencing, GBS, such as RADseq and RADcapture) differ in the particular strategies used to sample genomic regions of interest (Elshire et al., 2011) but they share challenges with genotype calling from the reads sequenced. Data obtained from these methods are simply numbers of read counts supporting reference or alternative sequences, which is different from genotypes obtained from SNP arrays containing embedded oligo probes that target every allele of the

chosen SNPs (De Rycke and Staessen, 2017). Ambiguous genotypes are thus commonly found in low-coverage GBS data (Dodds et al., 2019, Whalen et al., 2019). For example, at a locus of an individual that yields only two reads (2× coverage) of the reference allele; in this case, the actual genotype of the individual can be either homozygous or heterozygous, but by chance only one allele might be sequenced. The issue is particularly crucial for the homozygote of an alternative allele that has low frequency and simply has a higher chance of not being sequenced. Population genetic estimates can be made from low-coverage data based on genotype likelihoods without individual genotype calls, theoretically as low as 1.5×-2× (Nielsen et al., 2012, Han et al., 2013). However, individual genotype calling is still necessary for parentage assignment because many approaches rely on finding opposing homozygous loci between individuals; thus, the issue of genotype uncertainty poses a paramount challenge for parentage studies (Huisman, 2017).

Filtering out low-coverage loci can be applied to retain only the variant sites called using GBS methods that show some threshold of coverage (Andrews et al., 2016), but this strategy for removing uncertain loci would come with the cost of losing information, potentially substantial in low-coverage data. The proportion of missing genotypes is crucial for likelihood-based parentage assignment approaches that rely on comparisons of likelihood scores between putative parents to determine confidence in the assignments (Kalinowski et al., 2007). The likelihood ratio of a proposed trio (i.e. offspring, proposed maternal and paternal parents) is determined by dividing the likelihood of the first hypothesis that a proposed trio contains an offspring and its true parents given their genotypic data by the likelihood of the alternative hypothesis that the trio contains unrelated individuals given the population allele frequencies. For each offspring, the assignment is made for the trio with the highest likelihood ratio, and the confidence in assignment is assessed by determining the difference between the ratios of the most likely and the second most likely trios (Kalinowski et al., 2007). Thus, parentage identifications based on likelihood approaches (e.g. Kalinowski et al., 2007, Anderson Eric, 2012, Huisman, 2017) work best when individuals carry the same amount of genetic data (i.e. same number of typed loci) and may be compromised when different individuals contain missing genotypes at different markers (Marshall et al., 1998).

In addition to missing genotypes, the existence of siblings in a population is another challenge for parentage assignment; for example, full-siblings of the true parent may share one allele across loci with an offspring and possibly have a comparable pairwise likelihood as the actual parent (Marshall et al., 1998, Vandeputte and Haffray, 2014). Missing data

and the presence of siblings provide less of a challenge for pedigree-based assignments since pedigree reconstruction shifts the goal of parentage assignments to classification of relationships among individuals, rather than identification of pairwise relationships (Huisman, 2017). The algorithm behind the approach allocates a proposed parent into one of seven possible relationships with an offspring: parent-offspring; full-siblings; half-siblings; grandparent-grand-offspring; full aunt/uncle-niece-nephew; half aunt/uncle-niece/nephew; and unrelated. The proposed parent is considered the true parent if the likelihood of parent-offspring relationship is greater than the second most likely alternative relationship. If there remain multiple possible parents, likelihoods are estimated for all possible opposite-sex pairs among the remaining parents, the pair with the highest likelihood is assigned (Huisman, 2017). Once parents are assigned to all offspring, groups of half- and full-siblings are clustered. If needed, grandparents can be assigned to the clusters using a method similar to the initial parentage assignment to allow connections between the clusters, i.e. bottom-up pedigree reconstruction. With this approach, the necessity to compare likelihoods between putative parents is no longer required, which makes the approach suitable for data sets with some extent of missing information.

Advances in analytical approaches have the accompanied rapid developments in sequencing technology. For example, the programme AlphaAssign (Whalen et al., 2019) extended the algorithm for parentage assignment described in Huisman (2017) to explicitly deal with low-coverage GBS data. AlphaAssign considers only four classes of relationships including the true parent, full- and half-sibling of the true parent, and unrelated; each potential parent is classified into one of these classes. The potential parent is considered the true parent if it is classified as the parent, has an assignment score above a specified threshold, and has the highest score in the case of multiple potential parents classified as the parent. To deal with potentially noisy genotypes due to low-coverage sequences, all possible genotypes are marginalised based on the observed genetic data; i.e. the number of reads supporting different alleles. Genotype peeling (Elston and Stewart, 1971) is implemented in AlphaAssign to allow genotype imputations of unsampled parents or grandparents to allow connection between cohorts. The genotype imputations can improve efficiency of the assignment because incomplete sampling may be common in wild populations. Ros-Freixedes et al. (2019) demonstrated high accuracy of genotype imputations using AlphaAssign (mean per-individual correlation between the imputed and actual genotypes = 0.97) in four populations of 18,349 – 107,815 pigs for which only ~2% of each population were genotyped for 15,000 to 75,000 genome-wide SNPs. Whalen et al.

(2019) simulated a 5-generation pedigree of initially 1,000 individuals and tested the accuracy of AlphaAssign for parentage assignment of the individuals in the 5th generation; the results revealed that GBS data with ≥ 10× performed equally well as a high-coverage SNP array data. Data with coverage as low as 0.1× could still be used for parentage assignment if a sufficient number of markers (e.g. 5,000 loci) is employed.

For the SWR, translocations are recommended by the International Union for Conservation of Nature's (IUCN) African Rhino Specialist Group (AfRSG) to: 1) re-stock SWRs in suitable habitats; 2) reduce density to avoid population sizes exceeding estimated ecological carrying capacity (ECC); and 3) establish new populations (Emslie et al., 2009). Although translocation has been widely adopted as a routine practice for metapopulation management (Emslie et al., 2009), incorporation of genetic tools into the identification of candidates for translocation and breeding has recently been proposed and remains in its infancy (Purisotayo et al., 2019). Fortunately, most managed SWR populations have been intensively protected and animals are subjected to regular health checkups and individual identification using blood samples and ear notching, respectively (Verreynne, 2012, Ferreira et al., 2017), which offer an opportunity to obtain DNA sequences from most SWRs.

Relocating of SWRs is a challenging task and involves a large amount of costs, staff, and thorough preparation; thus, it is important to choose which genetic parameters to be monitored to allow most efficient translocations. Although pedigree-based approaches have advantages for selecting genetically distinct or inbred individuals, calculations can be made based only on genetic markers, where a pedigree is lacking. The kinship coefficient is defined as the probability that both alleles randomly drawn from two individuals are identical by descent (IBD) (Lacy, 1995). Individual mean kinship (MK) is defined as the average of all pairwise coefficients between that individual and all members of the same population including itself (Ballou and Lacy, 1995); it theoretically predicts the inbreeding coefficient (F) of an offspring born from the animal, given that the population is randomly mating. Therefore, translocation and breeding strategies that aim to reduce population MK should essentially reduce F of the subsequent generations. Such strategies have been demonstrated to be efficient methods to slow down the rate of genetic erosion in small and isolated populations (Willoughby et al., 2015, Willoughby et al., 2017), and should offer tools for genetic-based population management of SWRs.

The overall aim of this chapter was to assess whether the SNP panel developed in Chapter 3 would be useful for making population management decisions for captive SWR populations. Specifically, the purpose was to: 1) assess whether use of genome-wide single-nucleotide polymorphisms (SNPs) could improve reliability of parentage assignments for pedigree reconstruction compared to microsatellite genotyping (Chapter 2); 2) determine whether molecular marker-based mean kinship estimates could be useful for identifying genetically distinct or inbred individuals, in the absence of a robust pedigree; and 3) compare patterns of genetic diversity, heterozygosity and population differentiation in three privately managed populations in Botswana, in order to assess whether between-population differences could inform translocation decisions.

4.3. Materials and Methods

4.3.1. Samples

Samples used in this study were the same group of 110 SWRs described in Chapter 3 (Table 3-A1 in Appendix 3). They were collected during individual identification as a part of "The Central Database and Studbook for all Southern White Rhinoceros (*Ceratotherium simum simum*) Under Private Management in the Republic of Botswana" (Verreynne, 2012). Thus, at the time of sampling, these samples included all individuals (Botswana1 = 53, Botswana2 = 46, Botswana3 = 11) that were already ear-notched for individual marking. The three populations have sheltered ~25% of the total national population (452 SWRs) (Emslie et al., 2019), and have been important source populations for the nation's reintroduction programme (Verreynne, 2012). Due to the continuous increase of illegal poaching, the exact names and locations of the three populations cannot be disclosed in this study. The populations are fragmented hundreds of kilometres away from each other without any means of habitat connection. Six founders of the Botswana2 ~~were~~ originated from Botswana1 (of which two were the breeding bulls), and the rest were obtained from three populations in South Africa. Botswana3 received one breeding bull and one female directly from Botswana1, and the other two founding females were taken from the state-managed populations in Botswana.

In chapter 3, DNA samples of 32 SWRs were chosen to evaluate the efficiency of RADcapture to retrieve genetic information from DNA of different qualities. In the present chapter, as many additional samples as possible were included for RADcapture sequencing

to test the appropriateness of GBS-based parentage assignments and approaches for identification of genetically valuable individuals as candidates for translocation and breeding. DNA extraction and quantification were performed as detailed in Chapter 3 (3.3.1 Samples and DNA quality designation). As in Chapter 3, classification of quality was performed to allow pooling of samples of similar quality into capture reactions to avoid hybridising biases toward high-quality samples. Samples judged as low quality and that yielded ≤10 ng/µl were considered as severely degraded and were not included in the present study (Table 3-A1 in Appendix 3). This resulted in 64 additional samples that were of sufficient quality to proceed with library preparation; combined with the samples from chapter 3, this meant that 96/110 samples (51 from Botswana1; 41 from Botswana2; 4 from Botswana3) could be included. The overall workflow for this study is shown in Figure 4-1.

4.3.2. Library preparation and hybrid capture

The sequencing library for each sample was prepared as detailed in Chapter 3 (3.3.5 Library preparation for bait capture). Briefly, samples were fragmented and ligated to a specific combination (Table 4-A1 in Appendix 4) of i5 and i7 adaptors (NEBNext Multiplex Oligos for Illumina Dual Index Primers Set 1). Libraries were amplified for eight PCR cycles and subsequently size-selected to obtain the final library size of 320 – 470 bp. Concentration of the final libraries was measured using Qubit dsDNA Broad Range Assay Kit (Thermofisher Scientific, Cambridge, UK). Libraries with a concentration ≤ 20 ng/µl were removed from the study because their concentrations would be below the minimum requirement (i.e. ~125 ng/1.75 µl) of the bait capture protocol (MyBaits, Arbor bioscience, Ann Arbor, USA). This resulted in a total of 88 purified libraries that could be included for RADcapture (49 from Botswana1; 35 from Botswana2; 4 from Botswana3). Since so many of the Botswana3 samples were below the quality threshold, individual form that population were only considered to assess between population variation.

The libraries were pooled into 4-sample-pool reactions; each sample contributed ~125 ng/1.75 µl, constituting a total of 500 ng of the input libraries per capture reaction. Hybrid bait capture was performed for each reaction according to the manufacturer's protocol (MyBaits, Arbor bioscience, Ann Arbor, USA) as detailed in Chapter 3 (3.3.6 Hybrid capture). The captured libraries were pooled in equimolar concentrations and sequenced in

a single lane of Illumina Hiseq 4000 (Novogene, Beijing, China), using paired-end 160 bp sequencing.

4.3.3. Bioinformatics for SNPs and genotype calls

Demultiplexing and initial quality control of sequencing data were done by Novogene (Beijing, China). Further quality control and removal of over-represented sequences were performed using FastQC (Andrews, 2010) and Trimmomatic version 0.38 (Bolger et al., 2014), respectively. Please refer to the detailed criteria for quality control in Chapter 3 (3.3.7 SNP calling). For the cleaned reads, forward and reverse reads of each individual were merged using Geneious Prime 2020.0.4 (Biomatters Inc., Auckland, New Zealand).

From this step onwards, cleaned and merged reads from Chapter 3 and the present chapter were combined for subsequent analyses. Contigs of the reference assembly in Chapter 3 were concatenated to form a single dummy chromosome with 350 Ns added between the contigs to reduce computational time; hereafter referred to as the concatenated reference. Reads of each individual were mapped against the concatenated reference using the BWA MEM algorithm of the Burrows-Wheeler Alignment (BWA) tool version 0.7.17 (Li and Durbin, 2009); the resulting Binary Alignment Map (BAM) file was sorted and indexed using SAMtools version 1.8 (Li et al., 2009).

Putative SNPs and their genotypes were called according to the Genome Analysis Tool Kit (GATK) Best Practices Workflows (Broad Institute), except for the Base Quality Score Recalibration (BQSR) because there has been no known variant database of the species. Reads present in the BAM files that had the exact insert size and started at the same genomic position were flagged as duplications using *MarkDuplicates* (Picard tools, Broad Institute). The *HaplotypeCaller* function implemented in GATK version 4.0.8.1 (Poplin et al., 2018) was used to identify SNPs for each individual; this outputted a genomic variant call format (GVCF) file containing genotype likelihoods of all variant sites regardless of whether the they were called as putative SNPs or not. The GVCF files of all individuals were subsequently combined using *CombineGVCFs* (GATK, Broad Institute), generating a multi-sample GVCF file. Joint genotype calling was performed using *GenotypeGVCFs* (GATK, Board Institute) for the pool of 88 samples, with all variant sites analysed simultaneously across samples; this generated a single VCF file containing information about genotypes and positions of SNPs.

Hard filtering (GATK Team, 2020) was applied to the called variant set to minimise the chance of considering a false SNP as a true variant. The *SelectVariants* function implemented in GATK was used to create a new VCF file containing only SNPs because indels might be the result of assembling errors of the draft reference (Franzén et al., 2009). The following criteria for hard filtering were applied (Chandran, 2016, Shultz, 2018): 1) *ReadPosRankSumTest,* which is equivalent to a z-approximation of the Rank Sum Test for testing the hypothesis that a variant allele was not randomly found at the beginning or the end of a sequenced read; 2) *QualByDepth* (*QD*) was obtained by dividing a variant confidence score (QUAL field in the VCF file) by the sequencing depth of the non-reference homozygous allele - application of this parameter aimed to avoid inflation of variant confidence occurring in high-coverage sites; and 3) *RMSMappingQuality* (*MQ*) provided the root mean square of mapping quality estimated over all individuals at a site. SNPs that showed one of the following conditions were removed: *ReadPosRankSumTest* < -8.0, *QD < 2.0*, and *MQ* < 40.0 (Chandran, 2016, Shultz, 2018). Command lines used for the hard filtering and generating a new VCF file that contained only the retained SNPs are provided in Command 4-A3 in Appendix 4. To obtain the final genotypes for parentage assignment, VCFtools (Danecek et al., 2011) was further used to retain only SNPs that were present in all 88 individuals, had minor allele counts > 3 reads, and minor allele frequency (MAF) ≥ 0.05 (Command 4-A4 in Appendix 4). Summary statistics of the retained SNPs including overall and per-population nucleotide diversity (π), observed (H_o) and expected heterozygosity (H_e), and the inbreeding coefficient (F_{is}) were estimated using VCFtools (Danecek et al., 2011). Tests for deviation from Hardy-Weinberg equilibrium, excess and deficit of heterozygosity were performed for each SNP using VCFtools (Danecek et al., 2011). Global tests were done using GENEPOP 4.2 (Raymond and Rousset, 1995).

4.3.4. Parentage assignment and pedigree construction

Individuals from Botswana1 whose SNP genotypes were available from RADcapture sequencing and were included as offspring in the microsatellite-based parentage analysis in Chapter 2 (hereafter referred to as MS-based assignments) were considered for parentage assignment in this chapter (SNP-based assignment). These individuals included 24 offspring and their candidate parents (Table 4-1). As described in Chapter 2, demographic records of the population were used to determine potential parents of an offspring: candidate fathers and mothers had to be at least six and five years of age (Rachlow et al.,

1998), respectively, and present in the Botswana1 population during the expected time of conception, which was ~16 months before the birthdate of an offspring (Figure 4-2). (Rachlow and Berger, 1998).

For the SNP data, AlphaAssign (Whalen et al., 2019) was used to determine parents of the 24 offspring; the input file consisting of SNP genotypes of all animals under consideration was prepared in plink binary format file using PLINK 2.0 (Purcell et al., 2007). Proposed distributions (Elston and Stewart, 1971) were constructed for each putative parent, described as the probabilities of the expected genotypes of the following relatives of a given offspring: 1) a parent; 2) a full-sibling of the true parent; 3) a half-sibling of the true parent; and 4) an unrelated individual. Each putative parent was classified into one of these possible relationships, conditioned on the genotypes of the proposed parent and the offspring. For each offspring, the proposed parent was considered the true parent if it was exclusively classified as parent. In cases where multiple putative parents were classified as a parent, the proposed parent with the highest assignment score was classed as the parent (Whalen et al., 2019). If multiple offspring born in the same year were assigned to the same mother, all of them were considered unassigned due to the 16-month gestation period. Rates of successful assignments were determined as a percentage of assigned parents to the total number of parents (N = 48) of the 24 offspring. To construct the consensus pedigree from the two assignment approaches, where there were discrepancies between the MS-based (Pedigree D in Chapter 2) and SNP-based assignments, the former was considered the correct assignments because they were made based on a combination of multiple parentage methods with high confidence. The consensus pedigree was drawn using Pedantic package for R (Morrissey and Wilson, 2010)

4.3.5. Principal component analysis (PCA)

Principal component analysis (PCA), as implemented in the Ade4 package for R (Dray and Dufour, 2007) was employed to assess genetic structure among 88 SWRs from the three populations. The genotypes of the same set of SNPs used in mean kinship analysis were summarised into principal components (PCs) that explained overall genetic variance among individuals and populations. Please refer to the filtering criteria used to obtain the SNP panel in the section 4.3.6. The PCA analysis was visualised to assess between-population genetic structure using Ade4 package for R (Dray and Dufour, 2007).

4.3.6. Estimation of kinship coefficients, identification of candidates for translocation and breeding, and between-population structure

Population MK ($MK_{population}$) was estimated from the consensus pedigree of the Botswana1 using the Kinship2 package for R (Sinnwell et al., 2014). Individual MK (MK_{ID}) was estimated for each animal by averaging all pairwise kinships between the animal and the others within the population, including itself. Inbred animals were defined as the SWRs that showed self-kinship coefficients > 0.5. The pedigree-based $MK_{Botswana1}$ was estimated as the overall mean of all MK_{ID} within the population, unassignable offspring were given no contribution to the estimation of population mean.

For molecular-based MK, NGSrelate (Korneliussen and Moltke, 2015), a software package for estimation of relatedness and other relationship statistics from low-coverage NGS data implemented in ANGSD (Korneliussen et al., 2014), was used to identify pairwise kinship coefficients of all pairs of animals from the three populations. Two input files containing population allele frequencies (.maf) and genotype likelihoods (.glf) were generated from BAM files using ANGSD. Only the reads that were primary alignment (*-remove_bads 1*, removed reads that were aligned to multiple regions of the concatenated reference) and had minimum base quality and mapping quality scores ≥ 20 (*-minQ 20*, *-minMapQ 20*) were considered. Variant sites that passed the following criteria were included in the estimation of kinship coefficients: 1) *p-value* of sites being variable < 1×10^{-6}; 2) *MAF* ≥ 0.05; 3) per-site depth ≥ 15× in at least 44 (50%) individuals; 4) global per-site depth ≥ 660× but ≤ 6,600× (Command 4-A5 in Appendix 4).

NGSrelate employed Jacquard's coefficients (*J1* - *J9*), which describe nine possible patterns of IBD at a locus between two individuals, together with their corresponding genotype probabilities (Jacquard, 1972). These probabilities are used to estimate pairwise kinship coefficients between two individuals (Korneliussen and Moltke, 2015). The software defines the pairwise kinship coefficient between diploid individuals *X* and *Y* (K_{XY}) as:

$$K_{XY} = J1 + (0.5 \times (J3 + J5 + J7)) + (0.25 \times J8)$$

where *J1, J3, J5, J7, J8* represent joint genotype probabilities for each possible arrangement (each diagram in Table 4-2) of the four alleles at a bi-allelic locus carried by *X* and *Y* (column 'allelic states' in Table 4-2). For example, given that individuals *X* and *Y* carry X_{AiAi}, Y_{AiAj} (with frequency of allele *i* and *j*= 0.75 and 0.25, respectively), the probability that two alleles randomly drawn from them are IBD (kinship coefficient) is 0 + (0.5 × (0.188 + 0 + 0) + (0.25 × 0.141) = ~0.129 which is close to the pairwise kinship coefficient of a pair of half siblings (Weir et al., 2006), please refer to **Table 4-2** for joint genotype probability of each *J*. NGSrelate estimates pairwise kinship coefficients from genotype likelihoods instead of genotypes of individuals (Korneliussen and Moltke, 2015). Thus, for each pairwise kinship, each *J* was estimated as a maximum likelihood conditioned on the observed NGS data and allele frequencies of two individuals, multiplied across all observed loci. Individual and population MKs were estimated using the similar approaches as the pedigree-based kinships described earlier.

A management scheme that encourages mating of individuals with relatively low MK_{ID} and regularly removes individuals with relatively high MK_{ID} should essentially minimise inbreeding levels in the following generations. In this study, animals that showed individual MK_{ID} greater than two standard deviations ($MK_{ID} > MK_{population} + 2SD$) of the population mean for their population were identified as candidates for translocation, because they showed evidence of inbreeding (i.e. high MK). On the contrary, animals that showed individual MK_{ID} less than two standard deviations ($MK_{ID} < MK_{population} - 2SD$) of the population MK for their population were considered as appropriate candidates for breeding because they were less related on average than mean MK and so the most genetically distinct.

Next, between-population kinship was estimated for each pair of populations, i.e. Botswana1 and 2 ($MK_{Botswana1\text{-}2}$), Botswana1 and 3 ($MK_{Botswana1\text{-}3}$), Botswana2 and 3 ($MK_{Botswana2\text{-}3}$). Between-population kinship of populations A and B, denoted as $MK_{A\text{-}B}$, is estimated as the mean of pairwise kinships between all individuals in A and all individuals in B (Frankham et al., 2017):

$$MK_{A\text{-}B} = \frac{\left(\sum_{i=1}^{N_A}\sum_{j=1}^{N_B} K_{XY}\right)}{(N_A \times N_B)}$$

where the K_{XY} is the pairwise kinship of individual *X* from population A and individual *Y* from population B; N_A and N_B represent population sizes of A and B, respectively.

Between-population MK was used to identify the appropriate sources of introduced animals because it would theoretically predict the inbreeding coefficient of an offspring born from two individuals that were randomly selected from each population. Thus, for each recipient population, proposed donor populations (the other two populations) that showed lower between-population MK than 2SD of the recipient's MK ($MK_{A\text{-}B} < MK_{recipient}$ - 2SD) were considered the appropriate sources of introduced animals.

4.4. Results

4.4.1. Samples, library preparation, and hybrid bait capture

DNA extractions obtained from 11 individuals were severely degraded and removed from the study prior to library preparation including: IDs 130 and 146 from Botswana1; 128 and 182 from Botswana2; and 251, 286, 300, 308, 327, 331, 353 from Botswana3. Additionally, three low-quality samples from the Botswana2 population (IDs 263, 268, and 333) were excluded to allow sequencing of all Botswana1 individuals that exceeded the quality threshold to fit the 96-sample library preparation kits. After library preparations, the libraries of eight SWRs with IDs 125, 137, 140, 142, 180, 360, 361, and 362 yielded final library concentrations ≤ 20 ng/µl; thus, they were not included for bait capture (Table 4-A1 in Appendix 4). A total of 88 libraries were retained, giving a total of 22 × 4-sample-pool reactions (Table 4-A1 in Appendix 4). The first eight pools were previously sequenced in Chapter 3, and 14 more were sequenced for the present chapter.

After hard filtering, a total of 302 SNPs was retained (Table 4-A2 in Appendix 4). Overall mean per-site π was 0.216: Botswana1 π = 0.214, Botswana2 π = 0.222, and Botswana3 π = 0.197. Overall means of H_o and H_e across three population were 0.254 and 0.215, respectively: Botswana1 H_o = 0.249, H_e = 0.211; Botswana2 H_o = 0.265, H_e = 0.219; Botswana3 H_o = 0.216, H_e = 0.173. Based on this SNP panel, these populations showed excess heterozygosity (*p-value* = 0.01) with overall mean F_{is} of -0.1766; details for each locus are shown in Table 4-A2.

4.4.2. Parentage assignment and construction of the consensus pedigree

Of the 24 offspring with SNP data resolved, 43 out of 48 parents (two parents for each offspring) could be identified using the 302 SNPs, which corresponded to a successful assignment rate of 89.6% (Table 4-1). Consensus assignments between the SNP-based and MS-based analyses were obtained in nine offspring (IDs 155, 172, 176, 202, 203, 206, 207, 271, 320), but SNP-based assignments showed better resolution than microsatellites in seven cases (IDs 154, 171, 200, 201, 242, 255, 316), which were not able to be assigned using MS (i.e. identification of one or both parents when this remained ambiguous based on MS). There was one case for which neither SNPs nor MS could resolve the assignment (ID 238), which might indicate that the parents were not sampled. In seven cases, SNP data for the MS-assigned parents (IDs 130, 125, 170) were not available and SNP-based assignments suggested the alternative parents, so the reliability of the results for these offspring (IDs 141, 156, 204, 205, 210, 274, 275) could not be interpreted (Table 4-1) which were not included in the consensus pedigree (Figure 4-3). Overall, 10 parent-offspring relationships of seven offspring that SNP-based assignments showed better resolution were added to the consensus pedigree (Figure 4-3).

4.4.3. PCA analysis

In the PCA analysis, the first two PCs explained ~73% of the total genetic variance among the three populations (Figure 4-5). However, most individuals were clustered into a single group without clear differentiation observed among the pre-defined clusters (populations).

4.4.4. Estimation of kinship coefficients, identification of candidates for translocation and breeding, and between-population structure

Using pedigree-based and marker-based MKs to identify candidates for translocation and breeding

The pedigree-based MK of the Botswana1 population (N = 55) estimated from the consensus pedigree was 0.051 (±2SD = 0.033 – 0.069). Two inbred offspring with IDs 146 and 255 were born from a half-sibling parental pair (IDs 124 and 131) and showed self-

kinship coefficients > 0.50; so were classified as inbred individuals. The marker-based kinship estimate for this population was slightly higher but showed less variance than for the pedigree-based analysis (marker-based $MK_{Botswana1}$ = 0.061; ±2SD = 0.055 – 0.067).

Using ±2SD as a cut-off value identified different individuals as outliers in the two sets of analyses. Pedigree-based $MK_{Botswana1}$ identified six individuals (IDs 120, 124, 146, 176, 255, and 999) with MK_{ID} above the threshold as candidates for translocation based on above average relatedness (i.e. inbreeding) and identified eight individuals (IDs 127, 170, 177, 206, 272, 275, 316, and 888) with MK_{ID} below the threshold as candidates for breeding based on their relative genetic distinctiveness (Table 4-A6 in Appendix 4). In contrast, marker-based $MK_{Botswana1}$ suggested four individuals (IDs 154, 172, 187, 188) as candidates for translocation, and only one individual (ID 151) was considered the most suitable candidate for breeding (Figure 4-4). The candidates for translocation suggested in this chapter based on the pedigree-based $MK_{Botswana1}$ were partly different from the candidate bulls suggested in Chapter 2; four of the six candidates in this chapter were male (IDs 120, 124, 176, 999); of which one (ID 176) were in agreement with the suggestions made in Chapter 2, the other two (IDs 120, 124) were the SNP-assigned fathers that involved four cases that SNPs provided better resolution. This suggested the potential effects of pedigree completeness on the candidates identified

The marker-based population MK estimated for Botswana2 was similar ($MK_{Botswana2}$ = 0.067, N = 35) to that for Botswana1, but showed a higher variance (±2SD = 0.059 – 0.075). Although only four individuals could be genotyped for Botswana3, the individuals showed substantially higher MK than in the other populations ($MK_{Botswana3}$ = 0.161, ±2SD = 0.153 – 0.169), respectively, with two individuals (IDs126, 208) showing high MK (> 0.163) and two (IDs 128, 150) showing lower MK (< 0.159). However, they all showed more evidence of inbreeding than in the other two populations. No individuals in Botswana2 and Botswana3 could be identified as candidates for translocation and breeding, based on a threshold of 2SD (Figure 4-4).

Using between-population MKs to identify sources of introduced animals

Between-population kinships were estimated as 0.052, 0.050, and 0.052 for Botswana1 and 2, Botswana1 and 3, and Botswana2 and 3, respectively (Table 4-3). Using $MK_{recipient}$ - 2SD as a threshold, Botswana 2 ($MK_{Botswana1\text{-}2}$ = 0.052) and 3 ($MK_{Botswana1\text{-}3}$ = 0.050) were considered suitable donor populations for Botswana1 ($MK_{Botswana1}$ – 2SD = 0.055).

Likewise, Botswana1 ($MK_{Botswana2\text{-}1}$ = 0.052) and Botswana3 ($MK_{Botswana2\text{-}3}$ = 0.052) were deemed the appropriate sources for Botswana2 ($MK_{Botswana2}$ – 2SD = 0.059), and Botswana1 ($MK_{Botswana3\text{-}1}$ = 0.050) and Botswana2 ($MK_{Botswana3\text{-}2}$ = 0.052) were considered the sources for Botswana3 ($MK_{Botswana3}$ – 2SD = 0.153).

4.5. Discussion

This study demonstrated the applications of RADcapture sequencing data for population management in three privately managed populations of SWR in Botswana. The RADcapture sequencing technique retrieved genetic information from DNA of different qualities and consistently genotyped 302 SNP loci across 88 SWRs, despite degradation of about one third of the samples. Since these samples would have been below the threshold for reliable ddRAD analyses, combining sequence capture to target the RAD-based SNPs enabled a much more complete picture of variation within these managed populations than would have been possible based on ddRAD alone (Graham et al., 2015). For the pedigree analysis, analyses based on the panel of SNPs (using AlphaAssign) identified by RADcapture allowed more resolution of parental genotypes than microsatellites (based on a combination of methods described in Chapter 2) in seven out of 24 cases and yielded almost 90% successful assignments. Since the panel could resolve parentage of offspring that were previously unassignable using microsatellites, it essentially deepened and widened the pedigree which should improve the inference of population parameters (Pemberton, 2008). However, the importance of sampling was emphasised by identification of alternative parents using the SNP markers when those identified by MS were not included. Comparing pedigree and marker-based mean kinship estimates revealed no overlap between individuals identified as candidates for translocation based on the outlier analyses using the two types of markers but suggested more individuals when pedigree information was considered. Given that the marker-based approach did not identify any candidates for translocation or breeding in Botswana2, indicating that the individuals might be so genetically similar. This suggests that establishment of robust pedigree is critical for genetic-based population management of SWR, as has been suggested in the management of other species, e.g. North American bison (*Bison bison*; Giglio et al., 2018) and endangered black stilt (*Himantopus novaezelandiae*; Galla et al., 2020). This emphasised the need of pedigrees that would allow individual-wise management to equalise genetic contributions of all individuals. The between-population MK estimates and the PCA suggested little genetic differentiation between the three

populations sampled, as might be expected given the historical bottleneck (Pang-Ching et al., 2018), but it does suggest that candidates could be moved between any of the populations. Overall, applications of the genetic tools demonstrated in this chapter should offer a mean to inform management of the isolated SWR populations as well as other fragmented species to increase genetic diversity, which is a crucial component for species to adapt to their ever-changing environments.

4.5.1. Parentage assignment

In this study, based on a set of 302 SNPs called from RADcapture sequencing data, the percentage of successful assignments for the captive population with the largest number of individuals available (Botswana1) was higher than that made based on microsatellites for the same families. This suggested that 302 biallelic SNPs with mean per-site $\pi = 0.216$ and global $H_e = 0.215$ offered better resolution than 18 microsatellite loci with $N_a = 2.5$ and $H_e = 0.409$. Note that the global excess of heterozygosity of the marker panel used in this chapter was similar to the results reported in Chapter 2, possibly due to the management strategy that mixed individuals from multiple provenances. Labuschagne et al. (2017) compared the power of 33 SNPs ($H_e = 0.063 – 0.520$) and nine microsatellites ($H_e = 0.298 – 0.655$, $N_a = 2.7$) for parentage assignment in a population of SWRs consisting of 11 known mother-offspring pairs and four pubertal males. Using the SNPs, the authors could correctly identify mothers of all offspring and identify fathers of six with high confidence, whereas the microsatellites incorrectly assigned the mother of one offspring and could identify fathers for only two offspring. However, in my study, several critical putative parents that had previously been identified by microsatellite analyses were not available for testing using RADcapture, which made it difficult to interpret whether the latter was more accurate or just more likely to identify a parent. AlphaAssign identified an alternative parent in each of these cases. Given the high degree of agreement in other cases where both candidate parents were available for MS and SNP genotyping (9 consensus, 1 disagreement), it seems likely that in these seven cases AlphaAssign had assigned the next most likely parent rather than a more correct parent than the MS-assigned parent. The concern about false assignment due to incomplete sampling has arisen because in such cases the SNP-based assignments tended to assign the alternative individuals to the offspring. Harrison et al. (2013) demonstrated the increased frequencies of false assignments when the true parents were not sampled using likelihood and Bayesian frameworks (i.e. FAMOZ, Gerber et al., 2003; COLONY2, Wang and Sancture, 2009;

SOLOMON, Christie, 2013). This issue emphasises a potential limitation of probabilistic methods for parentage assignment in a study system for which complete sampling may be difficult to obtain.

Apart from the completeness of sampling, selection of a marker panel for parentage analyses that varies in type, numbers, and degrees of polymorphisms is key to successful parentage assignments (Flanagan et al., 2019). In this chapter SNP-based assignments provided better resolution than microsatellites for seven offspring. The results were in line with many parentage studies in the NGS era in that SNPs provided at least equal (Buchanan et al., 2014, Weinman et al., 2016, Kaiser et al., 2017) or better analytical power (Hauser et al., 2011, Nguyen et al., 2014, Zhao et al., 2018) compared to microsatellites. These were possible because SNPs could be sequenced in a high throughput scale using NGS technologies. However, the better resolution of SNPs over microsatellites in this thesis did not directly reflect the statistical power of the two marker systems. Instead, it was a comparison between different parentage approaches; i.e. the combined approach vs a single probabilistic approach. Also, given that I could not obtain sequenced data for all putative parents (Table 4-1), I argue that it was impossible to compare analytical power of the two marker systems.

4.5.2. Estimation of kinship coefficients and identification of candidates for translocation and breeding

Estimation of kinship coefficients

Pairwise kinship between a pair of individuals can be estimated either from pedigrees or molecular markers. In my study, different values were obtained for population MKs estimated from the consensus pedigree and SNPs for the Botswana1 population. The consensus pedigree-based MK of 0.051 was slightly increased from 0.048 estimated using microsatellites in Chapter 2 but was still lower than the marker-based kinship of 0.061 estimated in this chapter. The low MK previously reported in Chapter 2 could possibly be explained by the relatively less complete pedigree; for example, two inbred individuals (IDs 146 and 255) could be identified from the new consensus pedigree, whereas only one individual (146) was identified previously. Molecular-based kinships reported for the three populations were 0.061, 0.067, and 0.161, which were approximately equivalent to the individuals (on average) being related between the level of first cousins (0.0625) and half-

siblings (0.125) (Weir et al., 2006). In regard to the discrepancy between population MKs estimated from pedigree and molecular markers, biases of mean kinships can occur in both approaches. The mean kinship determined from the consensus pedigree could be underestimated because it relies on the assumption that the founder individuals were unrelated which is possibly untrue in the species that experienced such a severed historical bottleneck as SWR. On the other hand, both over- and under-estimation of marker-based kinships can occur due to biases during the selection of the SNP panel; for example, the geographically restricted SNP panel developed from 10 eastern fence lizards (*Sceloporus undulatus*) was reported to introduce downward biases of allele frequencies and genetic diversity when applied to animals from other populations (Rosenblum and Novembre, 2007). Similar biases would also apply to the estimation of pairwise kinship coefficients because it normally involves sample-level allele frequencies (Thornton et al., 2012), which possibly be the case in this study because the individuals included in the ascertainment panel (i.e. samples included in the initial ddRAD, Table 3-1) were mostly from the captive populations, with varying histories since they were introduced from wild populations. Both approaches for estimation of kinship coefficients have their own pros and cons. Pedigree-based kinship indicates the probability of identity by descent of alleles that are actually shared from a common ancestor known from pedigrees, but it is time- and budget-consuming to construct an accurate and complete pedigree via observations. Conversely, marker-based kinship does not require demanding observations, but it relies on a statistical framework to interpret observed identity by state of alleles to make an inference about identity by descent. Please refer to the section "5.3.2 Management to minimise marker-based and pedigree-based kinships can lead to different outcomes" for further discussion.

Differences between pedigree-based and marker-based kinships also have been reported elsewhere (e.g. Bömcke and Gengler, 2009, Tienderen et al., 2013). For example, in three *ex situ* populations of golden-headed lion tamarins (*Leontopithecus chrysomelas*); marker-based MKs estimated from 21 microsatellite loci were greater than the pedigree-based MKs in all three populations considered (0.0628 vs 0.0398, 0.0843 vs 0.0515, 0.1435 vs 0.0398) (Tienderen et al., 2013). However, positive correlations between marker-based and pedigree-based kinships were evidenced (R^2 = 0.596) in these populations (Tienderen et al., 2013). Similarly, in an isolated population of robins (*Petroica australis rakiura*), correlation between kinships estimated from the pedigree and the markers was observed; additionally, traits affected by inbreeding depression (including hatching, fledging, and juvenile survival rates) were found to negatively correlate with the kinships estimated from

both approaches (Townsend and Jamieson, 2013). The extent of correlations was stronger when a high-density SNP chip (35,519 loci) was applied (R^2 = 0.76) in a herd of Iberian pigs (*Sus scrofa meridionalis*) compared to 56 microsatellites (R^2 = 0.47) (Saura et al., 2013). These examples suggest that population management to reduce population MKs should provide similar outcomes (e.g. maximise genetic diversity and qualitative traits), regardless of methods used to estimate the parameter.

Types of markers are crucial to obtain precise estimates of kinship coefficients because upward biases may occur due to alleles that are identical by state between individuals (Eding and Meuwissen, 2001). These biases are particularly pronounced between distantly related individuals/populations that share the same but non-IBD alleles (Bernardo, 1993), unless a large panel of markers are used for the estimations; e.g. 200 microsatellite loci (Eding and Meuwissen, 2001). This recommendation is in agreement with other studies that suggest using a large marker panel to minimise biases occurring in kinship estimation when pedigrees are not available (Bernardo, 1993, Li et al., 1993, Lynch and Ritland, 1999). Theoretically marker-based kinships can be estimated from any type of markers such as restriction fragment length polymorphisms (RFLPs; Bernardo, 1993), microsatellites (Eding and Meuwissen, 2001, Lenstra, 2006), and SNPs (Goudet et al., 2018). The suggestion to use a large marker panel indicates that markers suitable for kinship estimation should allow high throughput sequencing and/or genotyping. Variations of minisatellites and RFLPs have been shown to provide an insufficient resolution to differentiate relationships between individuals more distant than first-order relationships (e.g. parent-offspring and full-siblings) (Gilbert et al., 1991; Bernardo, 1993). For example, in a study that used minisatellites for parentage and kinship analyses of 12 prides of lions in the Serengeti, based on percentages of band sharing, kinship coefficients could differentiate only the first-order from unrelated relationships (Gilbert et al., 1991). For microsatellites, although they provide relatively high analytical power, laboratory processes for isolating new microsatellites can be challenging and require a substantial time period (Selkoe and Toonen, 2006), which may limit their use for kinship estimation. With the advance of molecular technologies, I argue that soon (if not already) SNPs will be the marker of choice for kinship estimation.

Identification of candidates for translocation and breeding: a justifiable level of within-population MK

Kinship coefficients offers a means to predict the inbreeding level of the subsequent generations; however, methods to determine a maximum acceptable value of kinship to be maintained in a population have not been developed. To restore genetic diversity of isolated populations, Frankham et al. (2017) suggested keeping the population mean kinship below 0.1 by translocation of animals between a pair of populations that show between-population MK below the threshold. Expected kinship coefficients of different classes of relatives could also be used as a cut-off value; for example, a goal to allow mating between two individuals with pairwise kinship < 0.125, which is equal to the pairwise kinship between half-siblings (Weir et al., 2006) would limit the chances of breeding between closely related animals (i.e. parent-offspring, full-siblings, and half-siblings). This strategy was adopted in the metapopulation management of Tasmanian devils (*Sarcophilius harissii*) (CBSG, 2008). The programme successfully preserved 98.68% of the gene pool of the founders over the course of 14 years (Hogg et al., 2013). Alternatively, Ralls and Ballou (2004) adopted a similar strategy as used in the present study by giving high breeding priority to Californian condors that showed kinships below the population mean. The strategy was able to preserve more than 90% of the founding gene pool in three captive populations. Population kinships of the three SWR populations in the present study were 0.050 – 0.052, well below the suggested value of 0.1 (Frankham et al., 2017) and 0.125 (Hogg et al., 2013). Nevertheless, I argue that translocation and breeding of SWRs based on kinship data remain necessary to avoid potential inbreeding depression and to preserve the gene pool currently present in the three populations.

In the Botswana1 population, where the comparison between the population MKs estimated from pedigree (Table 4-A6 in Appendix 4) and SNPs could be made, two methods suggested different numbers and different individuals for translocation (IDs 120, 124, 146, 176, 255, 999 suggested by the pedigree versus IDs 154, 172, 187, 188 suggested by molecular markers) and breeding (IDs 127, 170, 177, 206, 272, 275, 316, 888 suggested by the pedigree vs ID 151 suggested by molecular markers). This was possibly because a difference in sample sizes used to determine kinships (i.e. N = 55 for pedigree-based $MK_{Botswana1}$ and N = 49 for marker-based $MK_{Botswana1}$) that affected the estimations of MK_{ID} as well as population means. Also, population management based on kinship coefficients that were estimated from different methods was reported to introduce different managing decisions (de Cara et al., 2013). Pedigree-based kinship estimations tend to equalise

genetic contributions of all individuals and are intended to promote breeding of the individuals that have relatively low numbers of relatives within the population (Meuwissen, 1997), which shoud fit the original aim of this thesis (i.e equalise genetic contribution of individuals). In contrast, marker-based kinship estimations are intended to equalise allele frequencies; thus, it encourages mating of the animals that carry rare alleles (de Cara et al., 2013). Therefore, both factors (i.e. different sample sizes and potentially different outcomes introduced by different methods of kinship estimation) might be responsible for the different candidates for translocation and breeding suggested by the two methods. For Botswana2 and Botswana3, no SWRs showed MK_{ID} above or below their $MK_{population} \pm 2SD$; so, no individuals were identified as better candidates than others for translocation and breeding based on these criteria. To determine candidates for translocation, reduction of the thresholds can be used as an alternative approach; for example, $MK_{population}$ or $MK_{population} + SD$ can be applied instead of using $MK_{population} + 2SD$ as the cut-off value (Miller, 1995, Ralls and Ballou, 2004). An alternative approach for identification of candidates for breeding in these populations would be using pairwise kinship coefficients to determine a specific breeding pair (Hogg et al., 2013), which should also allow minimising of inbreeding and preservation of the current gene pool of the populations. However, promoting breeding between a specific pair may be obtainable only in small and managed populations. For a relatively large population such as Botswana1, I argue that regular relocation of dominant bulls is still necessary to reduce kinship coefficients and prevent the potential consequences of inbreeding depression.

4.5.3. Between-population MKs and genetic structure

In this study, between-population MKs among the three populations allowed the identification of the sources of introduced animals that would reduce $MK_{population}$ of the recipient populations and should consequently reduce the overall mean of inbreeding coefficients in the next generations. However, the PCA combined with the between-population MKs ($MK_{Botswana1\text{-}2} = 0.052$; $MK_{Botswana1\text{-}3} = 0.050$; $MK_{Botswana2\text{-}3} = 0.052$) suggested little genetic differentiation between the three populations, perhaps because some of the founders of Botswana2 and 3 were originated from the Botswana1. Interestingly, six individuals (IDs 184, 301, 302, 303, 304, 328) that originally came from a particular population in South Africa (data not shown) were somewhat explained by PC1 as they were clustered around +1 of the PC1 axis (Figure 4-5). However, because the provenance of these individuals could not be traced further back (i.e. whether the

individuals were born in the population or translocated from elsewhere), it was not possible to make any inference about whether or not this was an indication of spatially genetic variance. Exchanges of SWRs between isolated but not genetically differentiated populations are simply the attempts to increase effective population sizes (if the introduced individuals reproduce); if no new genetic stocks are regularly imported, genetic diversity and allelic diversity can decrease over time (Sigg, 2006). Since relocating of SWRs typically involves a large amount of costs, staff, and thorough preparation, this may not be the most cost-efficient approach to manage the three populations included in this study. Instead, introduction of animals from other sources is worth considering to cost-efficiently minimise inbreeding and maximise genetic diversity. Additional genetic studies are advised to determine genetic similarity between all populations in consideration before introduction of animals from alternative sources to allow the most efficient management (Frankham et al., 2017). However, there is a value in minimising inbreeding regardless of the genetic similarity of the populations, i.e. the main concern is to keep as low as possible the probability that alleles at any locus are identical by descent from a common ancestor. Negative consequences of outbreeding depression, defined as reduced fitness of due to a separation of an allele combination that favour a particular trait (Frankham et al., 2011), can be occurred as a result of mating between dissimilar animals (e.g. individuals adapted to different environments) which may happen unintentionally after translocation. Banes et al. (2016) addressed a potential outbreeding depression in an orang-utan rehabilitation site in Borneo that two confiscated females of non-native subspecies were introduced which subsequently reproduced 22 hybrid offspring. However, prevalence of outbreeding depression is rare (Ralls et al., 2018) and has been shown to persist for only a few generations (Edmands, 2007); with only little genetic differentiation between the three populations, the chance that outbreeding depression occurred after translocation between the three populations would be negligible.

Regarding genetic structure, due to historical movements of animals between the three populations; my study suggested the opposite to a study conducted in six wild populations in South Africa and two (European and African) captive populations of SWRs based on sequences of mitochondrial control region and 10 microsatellites (Moodley et al., 2018). This study revealed ongoing genetic drift (i.e. reduced genetic diversity in the captive-born SWRs compared of their wild-born founders) within the two captive populations and also revealed indications (i.e. genetic substructure) of genetic differentiation between captive and non-captive animals (Moodley et al., 2018). This suggested that, after a half-century

(from 1960s to 2010s) following the initial translocation of surplus SWRs from the founder population at Hluhluwe-Umfolozi Park, genetic differentiation has already occurred to the extent that is detectable among at least some of the modern populations. This finding (i.e. genetic structure between captive and non-captive populations) may be useful for population assignment because it should improve the probability of assignment, which can facilitate forensic work; for example, identification of origins of seized biological samples.

4.6. Conclusions

In this study, 302 SNPs with moderate diversity provided better resolution for parentage assignment than the currently available 18 microsatellite loci regarding the proportion of successful parentage assignments for a highly managed and closed population of SWR. However, this could be due to the ability to use likelihood-based genotype imputation for SNP-based data, which is not as feasible for multi-allelic microsatellite data, rather than simply having more markers. Comparison of pedigree-based and marker-based methods to infer kinship suggested that the choice of analysis and thresholds for selecting outliers could strongly influence identification of candidates for translocation and breeding to reduce inbreeding and increase genetic variation within captive populations. I thus suggest caution in use of these approaches, particularly when comparing captive populations that have originated from the same original sources, such as described here. In a broad sense, the methods described here to simultaneously discover and genotype SNPs in a non-model species together with the approaches to identify candidates for breeding and translocation demonstrate the insights that can be gained by using genome-wide approaches to assess genetic variation. Further development of marker panels such as this should offer a means to monitor and minimise inbreeding levels in small and fragmented populations, which are a common circumstance for endangered species.

4.7. Data Accessibility

Sequence Read Archive (SRA) accession numbers SRR10902285 - SRR10902372 (Bioproject accession PRJNA601631) for raw hybrid capture reads.

Table 4-1 Parentage assignments made for 24 offspring from Botswana1. The offpring, previously included in the MS-based assignment in Chapter 2 and genotyped for SNPs in the present chapter, were used to compare the efficiency of analyses based on different types of markers. SNP-based assignments were made using AlphaAssign (Whaelen et al. 2019) based on 302 SNPs.

Offspring ID/year of birth	Candidate father IDs	Candidate mother IDs	MS-based father	MS-based mother	SNP-based father	SNP-based mother	Comment[c]
141/2001	124, 130[a], 133	132, 134, 135, 151, 166, 177	130	151	124	151	Missing SNPs for MS-based parent but SNP identified alternative
154/2007	120, 124, 133, 145[a], 170[a]	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 151, 166, 177	NA[b]	NA	133	135	SNP identified parents but unresolved by MS
155/2006	120, 124, 130[a], 133, 145[a]	131, 132, 134, 135, 151, 166, 177	124	135	124	135	Consensus

156/2003	124, 130[a],133	131, 132, 134, 135, 151, 166, 177	130	131	124	131	Missing SNPs for MS-based parent but SNP identified alternative
171/2003	124, 130[a],133	131, 132, 134, 135, 151, 166, 177	NA	NA	124	166	SNP identified parents but unresolved by MS
172/2008	120, 124, 133, 144[a], 145[a], 170[a]	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 151, 166, 177	124	177	124	177	Consensus
176/2012	120, 124, 133, 156, 167[a], 170[a], 171, 204	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 147[a], 151, 155, 166, 177, 272	120	155	120	155	Consensus

200/2011	120, 124, 133, 145[a], 156, 170[a], 171	113[a], 114[a], 125[a], 127, 131, 132, 134, 135, 141, 147[a], 151, 166, 177, 272	NA	127	120	127	SNPs allowed more resolution than MS
201/2007	120, 124, 133, 145[a], 170[a]	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 151, 166, 177	124	NA	124	166	SNPs allowed more resolution than MS
202/2008	120, 124, 133, 144[a], 145[a], 170[a]	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 151, 166, 177	133	125	133	141	Consensus
203/2012	120, 124, 133, 156, 167[a], 170[a], 171, 204	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 147[a], 151, 155, 166, 177, 272	120	166	120	166	Consensus

204/2005	120, 124, 130[a], 133	125[a], 131, 132, 134, 135, 151, 166, 177	130	177	120	177	Missing SNPs for MS-based parent but SNP identified alternative
205/2012	120, 124, 133, 156, 167[a], 170[a], 171, 204	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 147[a], 151, 155, 166, 177, 272	NA	125	NA	NA	Missing SNPs for MS-based parent and alternatives not resolved by SNPs
206/2010	120, 124, 133, 145[a], 156, 170[a], 171	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 147[a], 151, 166, 177, 272	170	135	120	135	Consensus
207/2012	120, 124, 133, 156, 167[a], 170[a], 171, 204	113[a], 114[a], 125[a], 131, 132, 134, 135,	120	177	120	177	Consensus

		141, 147[a], 151, 155, 166, 177, 272					
210/2011	120, 124, 133, 145[a], 156, 170[a], 171	113[a], 114[a], 125[a], 127, 131, 132, 134, 135, 141, 147[a], 151, 166, 177, 272	133	147	133	135	Missing SNPs for MS-based parent but SNP identified alternative
238/2013	120, 124, 133, 156, 167[a], 170[a], 171, 204	113[a], 114[a], 125[a], 131, 132, 134, 135, 136[a], 141, 146[a], 147[a], 151, 154, 155, 166, 175[a], 177, 201, 272	NA	NA	NA	NA	Parents not resolved
242/2008	120, 124, 133, 144[a], 145[a], 170[a]	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 151, 166, 177	NA	NA	124	166	SNP identified parents but unresolved by MS

255/2011	120, 124, 133, 145[a], 156, 170[a], 171	113[a], 114[a], 125[a], 127, 131, 132, 134, 135, 141, 147[a], 151, 166, 177, 272	NA	131	124	131	SNPs allowed more resolution than MS
271/2011	120, 124, 133, 145[a], 156, 170[a], 171	113[a], 114[a], 125[a], 127, 131, 132, 134, 135, 141, 147[a], 151, 166, 177, 272	120	272	120	272	Consensus
274/2012	120, 124, 133, 156, 167[a], 170[a], 171, 204	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 147[a], 151, 155, 166, 177, 272	170	141	120	141	Missing SNPs for MS-based parent but SNP identified alternative

275/2012	120, 124, 133, 156, 167[a], 170[a], 171, 204	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 147[a], 151, 155, 166, 177, 272	170	135	124	**134**	Missing SNPs for MS-based parent but SNP identified alternative; only case where different parent predicted by MS and SNP when both genotyped
316/2010	120, 124, 133, 145[a], 156, 170[a], 171	113[a], 114[a], 125[a], 131, 132, 134, 135, 141, 147[a], 151, 166, 177, 272	NA	NA	NA	166	SNPs allowed more resolution than MS but only one parent resolved
320/2010	120, 124, 133, 145[a], 156, 170[a], 171	113[a], 114[a], 125[a], 131, 132, 134, 135,	124	132	124	132	Consensus

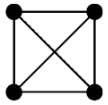
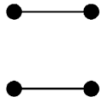
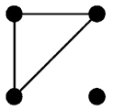
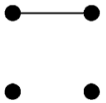
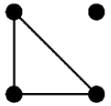
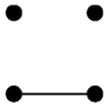
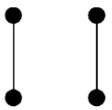
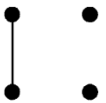
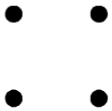
		141, 147[a], 151, 166, 177, 272					

[a] candidate parent lacking a SNP genotype

[b] NA indicates unassigned parents

[c] 9/24 cases showed consensus between predicted parents based on MS and SNP genotypes; 7/24 cases showed more resolution based on SNPs than MS; 1/24 cases showed no parents resolved by either marker type; 7/24 missing SNP data for MS assigned parent so couldn't interpret

Table 4-2 Joint probabilities of different patterns of identity states between a pair of diploid individuals *X* and *Y*. Each group of four dots indicates states (*J1 – J9*) of identity at a biallelic locus *A* that possesses alleles *i* and *j*. Top and bottom row dots represent two alleles of *X* and *Y*, respectively. Alleles that are identical are connected with lines. P_i and P_j represent allele frequencies of allele *i* and *j*, respectively. Each cell in the table represents the joint genotype probability for each possible arrangement (dot diagram in the top panel) of the four alleles at a bi-allelic locus carried by *X* and *Y*. (modified from Weir et al., 2006, Hanghøj et al., 2019).

Allelic states	Genotype probabilities of two individuals that are corresponded to different patterns of identical by descent								
	 J1	 J2	 J3	 J4	 J5	 J6	 J7	 J8	 J9
X_{AiAi}, Y_{AiAi}	Pi	P_i^2	P_i^2	P_i^3	P_i^2	P_i^3	P_i^2	P_i^3	P_i^4

X_{AiAi}, Y_{AjAj}	0	P_iP_j	0	P_iP_j	0	$P_i^2P_j$	0	0	$P_i^2P_j^2$
X_{AiAi}, Y_{AiAj}	0	0	P_iP_j	$2P_i^2P_j$	0	0	0	$P_i^2P_j$	$2P_i^3P_j$
X_{AiAj}, Y_{AiAi}	0	0	0	0	P_iP_j	$2P_i^2P_j$	0	$P_i^2P_j$	$2P_i^3P_j$
X_{AiAj}, Y_{AiAj}	0	0	0	0	0	0	$2P_iP_j$	P_iP_j	$4P_i^2P_j^2$

Table 4-3 Kinship matrix illustrating mean kinships (MK) within and between populations. The 3×3 matrix shows all possible MK between Botswana1 and 2, Botswana1 and 3, Botswana2 and 3. The right of the matrix shows: mean within-population MK ($MK_{population}$), $MK_{population}$ ± 2SD, and the appropriate donor population identified by between-population MKs.

	Botswana1	**Botswana2**	**Botswana3**	$MK_{population}$	$MK_{population} \pm 2SD$	Sources of introduced animals
Botswana1		0.052	0.050	0.061	0.055 – 0.067	Botswana2, Botswana3
Botswana2	0.052		0.052	0.067	0.059 – 0.075	Botswana1, Botswana3
Botswana3	0.050	0.052		0.161	0.153 – 0.169	Botswana1, Botswana2

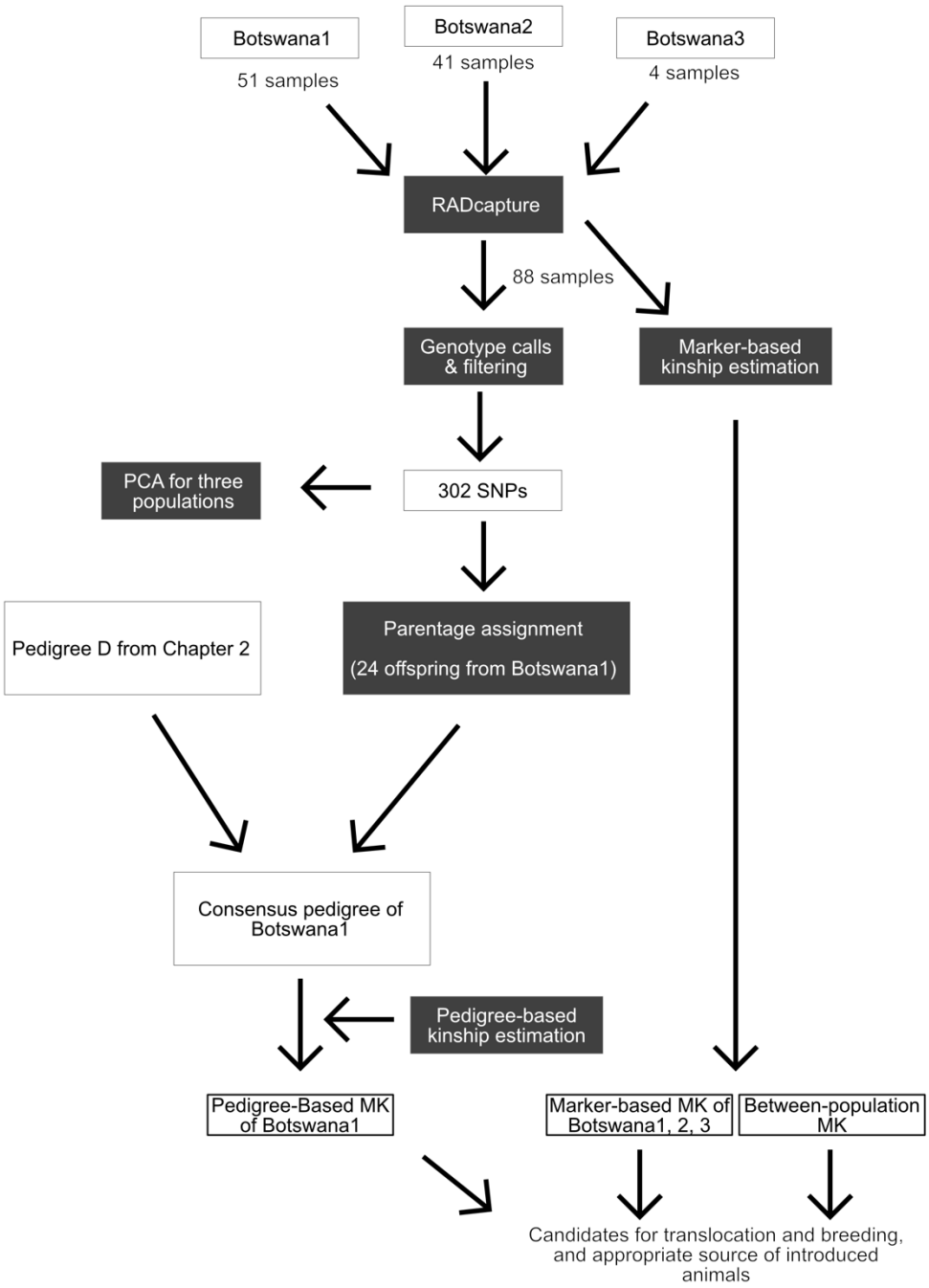
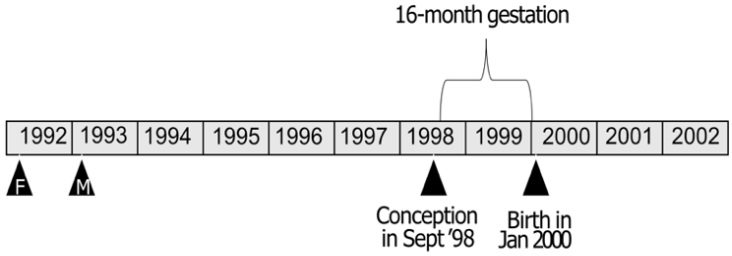


Figure 4-1 Diagram illustrating the workflow of this study. After removal of samples classed as severely degraded, libraries of 96 samples (both Chapter 3 and 4) were prepared but only 88 yielded sufficient concentration to be included for RADcapture sequencing. A final set of 302 SNPs was used for parentage assignment of 24 offspring from Botswana1, for comparison with Pedigree D from Chapter 2 and to identify individuals for translocation based on mean kinships (MKs) estimated based on a consensus pedigree based on both sets of markers. RADcapture data of 88 samples (from three populations, including Botswana1) were used to visualise variation within an between populations based on principal components analyses, and to estimate marker-based individual, within-population and between-population MKs, to allow identification of candidates for translocation and breeding, and appropriate sources of introduced animals.



.

Figure 4-2 Example of methods used to determine candidate parents for an offspring born in January 2000. Black F and M triangles indicate the birth dates of a candidate father and mother, respectively. The father/mother is considered a putative parent if it is at least six/five years old at the expected date of conception (16 months before the offspring's birth date) and is present at the reserve during that time.

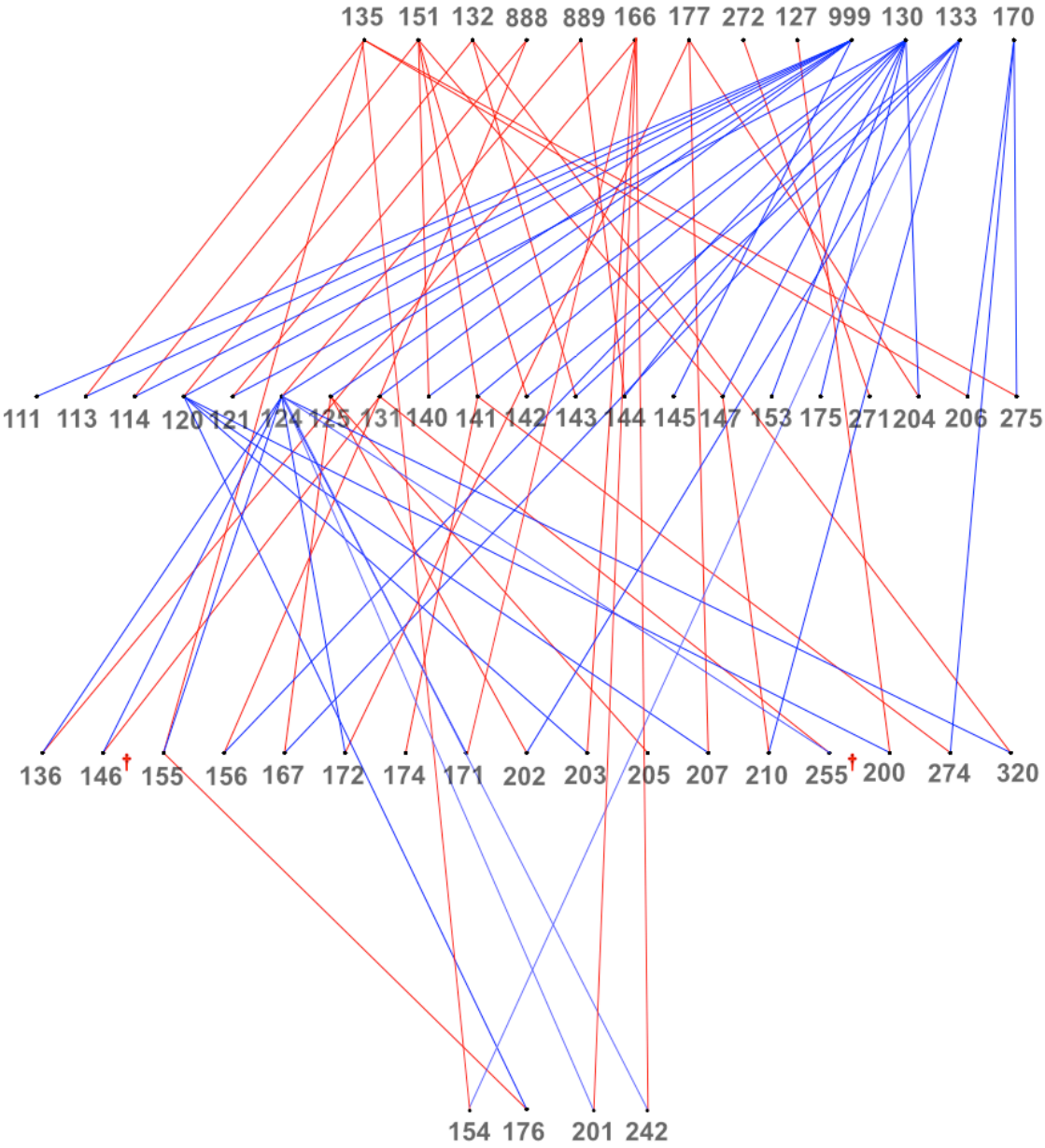


Figure 4-3 Consensus pedigree drawn from MS-based (Pedigree D in Chapter 2) and SNP-based parental assignments. Blue and red lines represent paternal and maternal relationships, respectively. Each number represents individual identification number of the SWRs. Two inbred individuals born from closely related parents are marked with †.

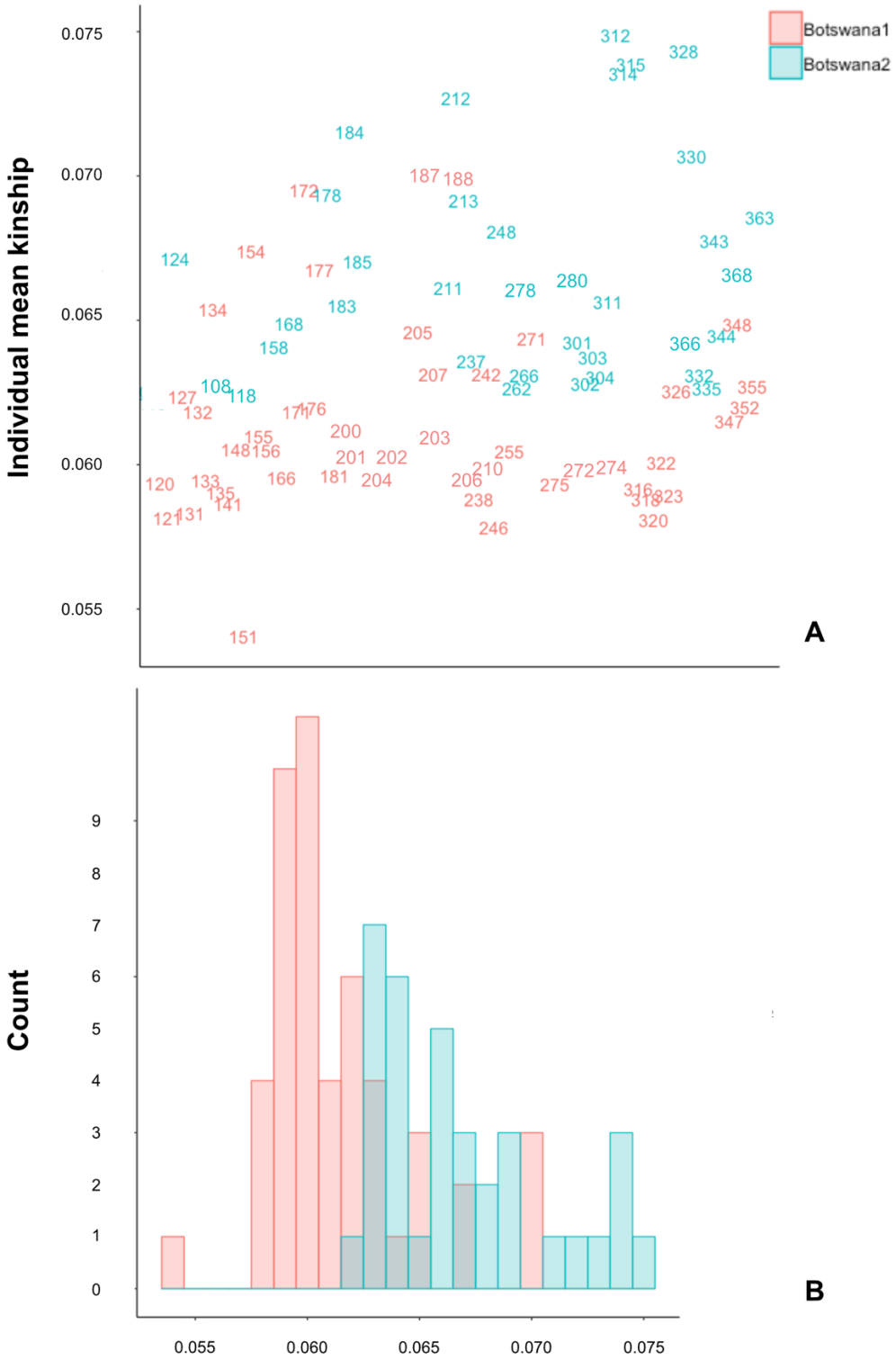


Figure 4-4 Individual mean kinships (MK) of 84 SWRs estimated from molecular markers. Because only four individuals from Botswana3 were sequenced, these figures show only the individuals from the other two populations. Figure 4-4A illustrates an identification number of each SWR that is plotted corresponding to its individual mean kinship. Figure 4-4B shows a histogram representing the distribution of mean kinships (bin width = 0.001).

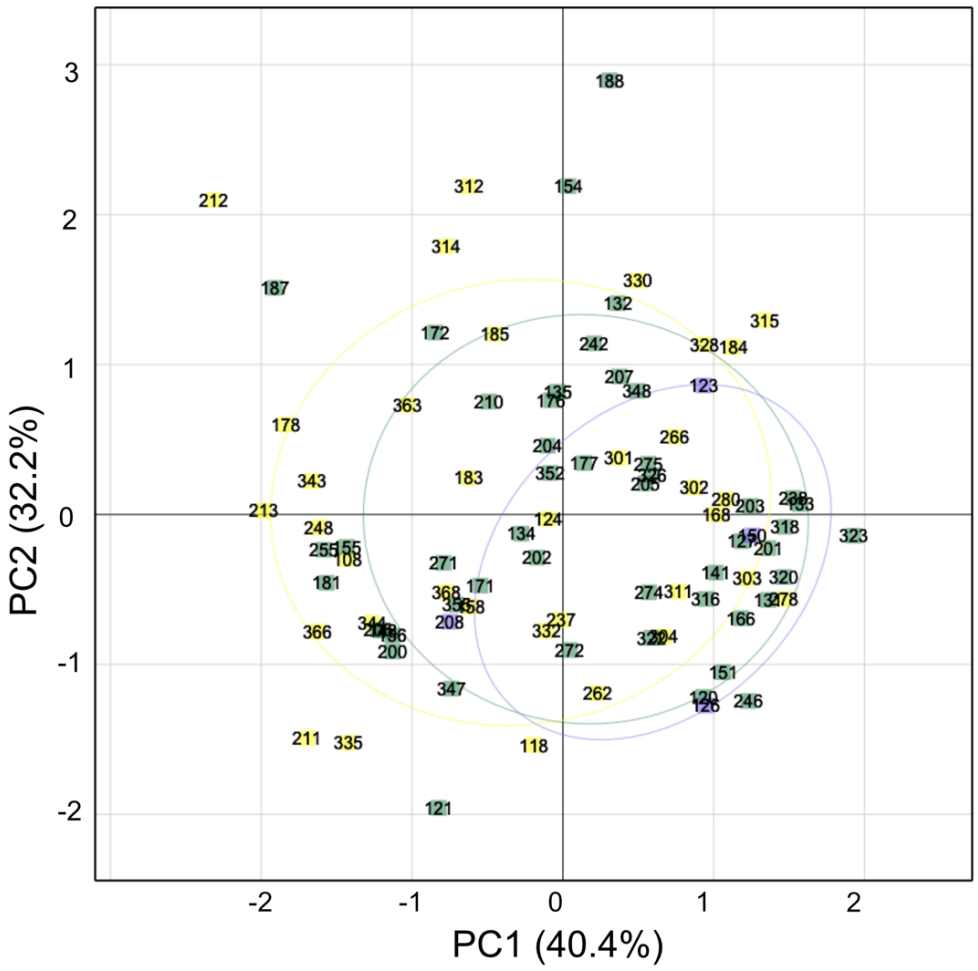


Figure 4-5 PCA analysis. Genetic variance among 88 SWRs from the three rhino populations. Numbers and colours indicate individual identification numbers and populations, respectively: Botswana1 = green, Botswana2 = yellow, Botswana3 = blue. Overall, ~73% of the total genetic variation was explained by the first two PCs (PC1 and PC2). Sizes of the inert ellipses correspond to the total variance of different populations.

Chapter 5 General discussion

5.1. Overview

The research that was conducted throughout this PhD programme was intended to develop genetic tools for the management of SWR populations. Three managed populations in the Republic of Botswana were used as models to develop the tools. In Chapter 2, I demonstrated that existing microsatellites were not sufficiently variable to obtain a complete pedigree, despite combining them with observational records to narrow down the potential parents. Lack of variation in markers and incomplete/inaccurate observational records resulted in inadequate analytical power to obtain a complete pedigree. The parentage analysis using microsatellites (MS) in Chapter 2 showed that several mother-offspring relationships identified by observations (Pedigree A in Chapter 2) were considered to be genetically impossible based on microsatellites. A final, synthetic pedigree (Pedigree D in Chapter 2) based on MS assignments (exclusion and probabilistic), together with field observations, was constructed and then used to estimate kinship coefficients using the Kinship2 package for R (Sinnwell et al., 2014). Candidates for translocation were identified as those SWRs that showed relatively high individual mean kinships (MK) compared to the population mean. Notwithstanding its incompleteness, the pedigree constructed from the combination of microsatellites and observational records revealed unexpected contributions from non-dominant bulls following the removal of the apparently dominant bulls. This suggested the possibility of maintaining or even maximising genetic diversity of the managed SWR populations by regular translocation of appropriate animals among the populations.

Of the total 110 samples, about a third showed signs of DNA degradation, as judged from the molecular weight present in 1% agarose gel electrophoresis. Degradation possibly arose from the conditions used to preserve the samples in the field. This, combined with the lack of high-quality reference genome or polymorphisms database of the species, meant that discovery and genotyping of a new set of molecular markers in SWR was not straightforward (Further discussed in “5.2.1. Quality of samples and lack of reference genome”). A useful method for DNA sequencing and marker genotyping should be effective with DNA of different qualities. I thus tested the robustness of RADcapture sequencing, which integrates the benefits of using double-digest restriction associated DNA (ddRAD), which has no requirement for a reference genome, applied together with

targeted capture sequencing, which requires relatively short DNA fragments and genotypes a reproducible set of markers, to simultaneously discover and genotype single nucleotide polymorphisms (SNPs). I demonstrated that RADcapture could retrieve genetic information from DNA samples that were of high and low quality, although there was wide variation in read yields across samples. This finding enabled the application of RADcapture to a larger group of potentially degraded samples. However, RADcapture involves multiple laboratory steps that subsequently contribute to a relatively high per-sample cost. This limitation may limit the applications of RADcapture in field practice (discussed in "5.2.2 Per-sample costs").

In Chapter 4, RADcapture was applied to a total of 96 samples, only 88 libraries of which yielded a sufficient concentration for hybrid capture. After genotype calls and hard filtering, parentage assignment with 302 SNPs using a method for likelihood-based estimation of probable genotypes showed better resolution than a more traditional method for inferring exclusion probabilities based on 18 MS loci for seven out of 24 offspring but predict the same in nine cases. However, when the MS-assigned parents of some offspring were not sampled for SNP genotyping, SNP-based assignments for these offspring were made to alternative parents. Interpretation of the results of these offspring is challenging. It is likely that AlphaAssign has simply assigned a possible parent in these cases because the MS-based assignments were made from the consensus of multiple parentage methods with high confidence in Chapter 2 (i.e. observational, exclusion, probabilistic approaches) and the sampling of parents was more complete. The assignments made to the alternative parents by the SNP panel when not all putative parents were sampled, raised concerns about false assignments potentially caused by probabilistic parentage assignment methods. This concern is particularly important for studies that rely on genotypes that have been called from low-coverage genotyping-by-sequencing (GBS) data, which are characterised by some degree of genotype uncertainty (discussed in "5.2.3 Exclusion or probabilistic parentage assignment).

By combining the Pedigree D in Chapter 2 with the SNP-based assignments, a consensus pedigree of Botswana1 could be constructed. Kinships estimated from pedigree and molecular markers suggested different individuals as the best candidates for translocation and breeding. Differing recommendations suggested by estimation of MK from pedigree or markers are further discussed in "5.3.2 Management to minimise marker-based and pedigree-based kinships can lead to different outcomes". The markers-based approach also did not identify evidence for genetic differentiation between the populations (based on

between-population MKs and PCA) that could be used to facilitate metapopulation management that would minimise inbreeding and maximise genetic diversity of the fragmented populations. This could be due to the severe historical bottleneck, which means that high relatedness might be expected among all of the managed Botswana populations; while the low genetic diversity limits statistical power for identifying appropriate candidates for translocation, it does mean that there are fewer risks of introducing individuals that are so genetically distinct that they might cause outbreeding depression. Section "5.2.4 Low variance of between population kinships and genetic structure" is devoted to discussing this in detail.

5.2. Limitations of the study

5.2.1. Quality of samples and lack of reference genome

Ear-notch tissue samples used for RADcapture sequencing were obtained in two different batches, with 77 and 85 samples each. Some of the samples were collected from the same animals as the results of sampling overlap, leaving a total of 110 unique SWR samples. Only 72 of the samples classed as high quality; the remaining samples were judged as low (n=22) or moderate (n=16). Such a considerable proportion of low-quality samples warranted genotyping methods that would allow sequencing of the degraded samples. Most genotyping methods for non-model species (e.g. RADseq and its derivatives) require high molecular weight DNA to provide sufficient length of DNA fragments for restriction enzymes to cut consistently across the entire genome (Graham et al., 2015); the potential in terms of inconsistent genotyping was demonstrated in Chapter 3, in that one of two duplicates of one sample from Botswana1 failed the ddRAD quality controls post-sequencing. Also, in non-model species, building a *de novo* reference database to align sequenced reads still requires an adequate length of reads to provide sufficient confidence in likelihood-based genetoypes calls (Paris et al., 2017). In model-species with high quality reference genomes available, sequenced reads containing SNPs with flanking regions as short as 25 nucleotides on either side can be aligned against the reference genomes and allow genotype calls with more confidence (Cornelis et al., 2017). The problem of DNA quality, combined with the lack of a reference genome for SWR, posed two important challenges to this project, with the result that a modified sequencing method was required.

Reduced representation methods such as ddRAD, when applied to the discovery and genotyping of molecular markers in non-model species commonly leverage restriction enzymes to sample a subset of genomic regions, allowing the sequencing of thousands of loci at a population-wide scale. In 2016 my study started with a limited number of available SNP genotyping methods for non-model species, particularly restricted for degraded samples (Graham et al., 2015). Two appealing approaches at the time that were demonstrated to perform well in degraded samples were HyRAD (Suchan et al., 2016) and RADcapture (Ali et al., 2016, Hoffberg et al., 2016). Both allowed retrieval of genomic data from somewhat degraded samples and performed well even in museum specimens. They are derivatives of RADseq but incorporate the benefits of targeted capture sequencing, which is more tolerant of poor-quality DNA. Both HyRAD and RADcapture develop in-solution probes from an initial RADseq run performed on high quality samples to identify putative markers, and these probes are subsequently used in a larger pool of query samples. The main difference lies in that HyRAD involves in-house probe generation via amplification of the biotinylated ddRAD library (Suchan et al., 2016). Probes used in RADcapture are subjected to multiple steps of *in silico* screening prior to hybrid capture, including soft-masking of repetitive sequences, and BLASTing against the draft reference genome (in this case, *Ceratotherium simum cottoni*; GenBank Accession number GCA_004027795.1) to filter out baits with too many hits (Ali et al., 2016). Given the processes of bait design and screening, RADcapture should produce more reproducible baits and consequently it should yield more reproducible markers compared to HyRAD. I therefore adopted RADcapture for this thesis.

Over the course of my study, sequencing technologies have advanced progressively. More sequencing approaches that show the potential to overcome the issue of DNA degradation in non-model species have recently become available. For example: Mobiseq uses highly conserved transposable elements as primers to anchor and sequence their flanking regions (Rey-Iglesia et al., 2019); a modified HyRAD approach has been developed that improves reproducibility of SNPs (Lang et al., 2020); Genotyping-in-Thousands sequencing (GT-seq) sequences multiplexed amplicons that are amplified using primers targeting SNPs identified from RADseq (Schmidt et al., 2020); and low coverage (≤ 10×) whole-genome sequencing (WGS; Bowen et al., 2012). The potential of the last approach is worth special attention as per-sample costs have declined. For example, at the time of writing (May 2020), the commercial charge that was quoted by Novogene (Beijing, China) for 10× WGS of an individual with genome size ~2.5 Gb (such as SWR) was ~£300. As an alternative

approach to the initial screening for SNPs with ddRAD, low-coverage WGS in a small pool of high-quality samples could be performed at a reasonable cost, potentially also increasing SNP discovery. However, it has not yet been tested for parentage assignment, when similar issues with the necessity to consider only loci present in all individuals as for RADcapture could apply, reducing the number of markers

5.2.2. Per-sample costs

RADcapture sequencing described in this thesis involved many laboratory steps and so costs per sample were high (Table 5-1). The sequencing library for each individual sample was prepared by 1) fragmentation and ligation of DNA fragments to the adaptors; 2) size-selection; and 3) amplification. Four libraries were pooled into a single reaction, followed by hybrid capture and sequencing. The per-sample cost for library preparation was ~£24 using the 96-sample NEBNext Ultra II FS DNA Library Preparation Kit for Illumina and NEBNext Multiplex Oligos for Illumina Dual Index Primers Set 1 (New England BioLabs Inc., Massachusetts, USA). The cost for bait construction and hybrid capture reagents was an additional ~£25 per sample based on 4-sample pooling using MyBaits Custom 20K kit (Arbor bioscience, Ann Arbor, USA). Finally, consumables, reagents and sequencing of the final capture libraries was estimated at ~£52 per sample, constituting an overall cost of ~£101 per sample. Such high per-individual price might limit the applications of RADcapture as a tool for population management when genomic information is required from a large number of individuals. Alternative approaches that can be used to genotype molecular markers with relatively cheap per-sample cost should promote further implementation of genetic-based population management. For example, 50,000 SNPs can be genotyped for dairy cows using a SNP-chip at ~£35 (Neogen, Ayr, UK). Additionally, SNP-chips consistently yields better coverage and provides more reliable genotypes than GBS-based methods (Flanagan and Jones, 2019). The technique also allows collaboration between laboratories to build a genetic database of the species because it provides good reproducibility (Hong et al., 2012) and does not rely on laboratory-specific size standard for genotype calls (Moran et al., 2006). High resolution SNP-chips have the potential to employ knowledge of candidate genes from other organisms to investigate adaptive variations (Pardo-Diaz et. al., 2015), which could include markers that could be used to assess adaptive potential of the SWR. With these advantages of SNP-chip and as costs continue to come down, the potential for genotyping all individuals in captive populations

might become feasible, which would be the most reliable way to identify individuals for translocation.

Table 5-1 Break down of per-sample cost in Great British Pounds (£) for RADcapture sequencing.

Laboratory protocols/items	Per-sample price (£)	Products
DNA extraction	~3	DNeasy Blood & Tissue kit (Qiagen, Manchester, UK)
Library preparation	~19	NEBNext Ultra II FS DNA Library Prep Kits for Illumina (New England BioLabs Inc., Massachusetts, USA).
Molecular adaptors	~5	NEBNext Multiplex Oligos for Illumina Dual Index Primers Set (New England BioLabs Inc., Massachusetts, USA).
Magnetic beads	~4	AMPure XP 2 × 5 mL (Beckman Coulter Inc., California, USA).
Baits and hybrid capture (4-sample pool)	~25	MyBaits Custom 20K (Arbor bioscience, Ann Arbor, USA).
Capture library purification	~2	MinElute PCR purification kits (Qiagen, Manchester, UK)
Sequencing of final capture libraries	~43	2 lanes of Illumina Hiseq400 for 88 samples (Novogene, Beijing, China)
Total	~101	

5.2.3. Exclusion or probabilistic parentage assignment

Most of the probabilistic-based methods currently available for parentage assignment are based on the assumption that a population is random mating (reviewed in Flanagan and Jones, 2019) which is not always met in wild populations (Kretzschmar et al., 2019). This is particularly a problem in fragmented populations of a species that shows mating behaviours with hierarchical breeding dominance (such as SWR) for which breeding choices are limited, and animals are more likely to share common ancestors. Probabilistic parentage identification is prone to false assignment (false positives) when a population contains a substantial number of closely related members or genetically similar individuals; in such a population, there is a chance that relatives of the true parent of an offspring are proposed as putative parents (Huisman, 2017). Also, false positives might occur when the pool of putative parents is not complete. Unlike probabilistic methods, exclusion has been considered the gold standard to exclude genetically impossible putative parents (Jones et al., 2010). Regardless of the platform used (MS or SNP), it is necessary to provide for the possibility of genotyping errors when using exclusion methods. A pre-specified number of mismatched loci between parents and offspring is typically allowed to reduce the rate of false exclusions, the number of permitted mismatches being determined by the number of markers in the panel and the probability of a mis-call (Strucken et al., 2016). In the work described in this thesis, a challenge arose because the SNPs called from RADcapture had high heterogeneity of read yield and sequence coverage (Table 3-5 and Figure 3-3). These are common problems that are shared with all GBS-derived data, relative to MS and SNP arrays (Attard et al., 2018). Therefore, exclusion-based parentage assignment using RADcapture data alone would not be a sensible option on its own. In Chapter 4, the SNP-assignment was combined with demographic records to filter ineligible parents on age and location, and this was followed by the probabilistic AlphaAssign method to identify the most likely parents. Exclusion was not applied because of concerns about read heterogeneity and sequencing coverage. Parents of most offspring could be identified and were largely consistent with the MS-assignment in Chapter 2; however, discrepancies occurred in the assignments of some offspring whose MS-assigned parents had no SNP genotypes, raising a concern about false assignment when sampling was incomplete. The seven cases in which AlphaAssign made alternative assignments to the previously assigned parents for which there was no SNP genotype involved the substitution of individuals 120, 124, 134, 135, in place of the missing IDs 130, 147 (daughter of 130), and 170. All the alternatives were either a population founder (134) or

the progeny of the founders (120, 124, 134, 135), which suggested potential relationships between the alternatives and the previously assigned parents. However, based on the consensus pedigree, no relationship between the SNP-assigned alternatives and the MS-assigned parents could be identified to support the statement about false assignment due to the presence of relatives in this study.

In domestic species such as cattle, for which methods have been well defined and standardized at an international level (International Society for Animal Genetics, 2012), exclusion-based assignment remains the gold standard method, currently using SNP in most cases (Fisher et al., 2009, Strucken et al., 2014, Brenig and Schütz, 2016, Strucken et al., 2016, Panetto et al., 2017). Low tolerance for mis-assignment is common for commercial domestic cattle because incorrect calls may cause substantial economic loss. Genotypes of SNPs in cattle are often obtained using SNP-arrays or using PCR primers to amplify genomic regions containing SNPs (SNP-PCR); these approaches produce SNPs with fewer missing data and genotyping errors than GBS methods (Hoffman et al., 2012, Flanagan and Jones, 2019). Often, 1 – 2% of mismatching loci are allowed to prevent false exclusion; Strucken et al. (2016) studied the effects of varying percentages (0-2%) of mismatch loci on the rates of mis-assignments and showed that 1% provided the best compromise between false assignments and false exclusion. To eliminate concerns about false positives from the probabilistic assignment using RADcapture in SWR, I argue that an alternative SNP genotyping method that can consistently obtain high sequencing coverage across alleles and across animals is required to yield reliable and usable genotypes for parentage exclusion. I used a conservative approach that considered only SNP loci found in all individuals but this meant that only 302 markers were available, which reduces statistical power for SNP-based analyses. Without a reference genome, it is also not possible to test whether the markers used are clustered or dispersed throughout the genome.

With the aim of developing a SNP genotyping method for SWR, SNP-PCR and SNP-array are particularly interesting because they should consistently produce reliable genotypes across individuals with relatively low per-sample cost (compared to RADcapture). The SNP-PCR methods use prior knowledge about known SNP databases and their flanking sequences to design PCR primers (Beacham et al., 2017). However, as with any PCR-based approach, it would be expected to suffer from limitations of microsatellites, such as biases in amplification, potentially resulting in null alleles. Often a SNP-PCR panel is chosen from a validated SNP chip to reduce overall costs for sequencing (Fisher et al.,

2009), but could also be identified from the SNPs developed from RADcapture described in this thesis or using other GBS-based methods (e.g. RADseq, Holman et al., 2017; RNAseq, Kaiser et al., 2017). A SNP-PCR panel generates more reliable genotypes but produces a much lower number (< 500) of loci than the GBS methods. Similar to SNP-PCR, development of a SNP array requires well-established genomic knowledge, but ultra-deep sequencing of RADseq and RNAseq have also been used to develop high-density SNP arrays in non-model species (Liu et al., 2011, Houston et al., 2014). Initial costs for array development are high; however, once an array is available, sequencing can be done at a smaller cost than using GBS methods but higher than the SNP-PCR (Flanagan and Jones, 2019). A panel of SNPs can be selected from RADcapture data, making SNP array and SNP-PCR interesting options for application to species like SWR. Evidently, there is no technological barrier to develop a SNP array based on the RADcapture data, except for the initially prohibitive cost. However, design of a SNP panel for a non-model species is technically challenging and involves multiple validations (Liu et al., 2011, Houston et al., 2014, Holman et al., 2017). The SNP-PCR approach is promising and may be useful in applications that require only hundreds of loci such as parentage studies.

5.2.4. Low variance of between-population kinships and genetic structure

Between-population MK should predict the inbreeding coefficient of the progeny of any two randomly chosen animals from two populations under consideration. Selection of individuals to introduce from a potential donor population with low between-population MK should reduce inbreeding in the following generations of the recipient population. In Chapter 4, I demonstrated that translocation of SWRs between any pair among the three populations would reduce the within-population MK of the recipient. However, the assumption was based on a very small difference in between-population kinships, i.e. $MK_{Botswana1\text{-}3} = 0.050$, $MK_{Botswana1\text{-}2} = 0.052$, $MK_{Botswana2\text{-}3} = 0.052$. It should be noted that the small difference in between-population *MK* estimates did not necessarily suggest no genetic differentiation; instead, it indicated that animals from three populations on average were likely related more distantly than first cousins (0.0625). In contrast to the *MK*, *F*-statistics (F_{ST}) describe genetic differentiation between populations, such that a high F_{ST} suggests high differentiation (Wright, 1950). To test whether the small difference in between-population MK estimates was caused by low genetic differentiation and/or the use of insufficient markers, I estimated population-wise F_{ST} using Nei's estimator (Nei, 1973),

implemented in the Hierfstat package for R (Goudet, 2005). The estimates of Botswana1-Botswana2, Botswana1-Botswana3, Botswana2-Botswana3 were as following: 0.0137, 0.0101, 0.0150. This suggested little (if any) genetic differentiation between populations. This was supported by the PCA in Chapter 4. Lack of population structure was not an unexpected discovery because the Botswana1 population included some of the founders used to establish Botswana2 and Botswana3. A minimal difference in between-population MKs suggests that outsider populations may be necessary to maximise genetic diversity of the populations included in this thesis.

5.3. Future applications

The primary aim of this thesis was to develop genetic tools that would be useful for population management. Throughout the study, I have developed genomic data that can be used as fundamental resources; nevertheless, further development of genetic tools to serve the conservation of SWR remains necessary and is in progress.

5.3.1. Maintaining highly accuracy pedigrees

To serve the conservation of species, pedigrees are used to provide insights into heritage and to predict future viability of extant populations. In this thesis, I proposed using kinship coefficient as a key parameter for monitoring and prediction of inbreeding levels. Traditionally, kinships are estimated from pedigrees (Wright, 1922) with the limitation that complete multi-generational pedigrees are rarely achievable, even in intensively managed populations such as SWRs. With the rapid development of genotyping techniques, genome-wide markers are increasingly used to estimate kinship coefficients in non-model species (Städele and Vigilant, 2016, Goudet et al., 2018). Estimation of marker-based kinships is arguably faster and demands less effort for observations and keep the records of all animals in a population; together with the advance in sequencing of non-invasively collected samples, kinship and inbreeding can be monitored without visual observations of the animals (Quinn et al., 2019). However, the quantity and diversity of the markers used to estimate kinship should be high to avoid biases because small numbers of low variant markers may be insufficient to capture genetic variance within and between populations and do not represent the entire genome (Fernández et al., 2005). For example, in a species with a genome size of ~2.5 Gb (about the same size as SWR), marker-based kinships estimated from ≥ 10,000 unlinked SNPs were sufficient to yield the same accuracy as

pedigree-based kinships (Wang, 2016). With NGS genotyping methods available nowadays, obtaining genotypes of thousands of markers in hundreds of individuals is not impossible. Whether a high accuracy pedigree is worth maintaining, the answer to this question would depend on the aims of management. The ultimate goals that I proposed using kinship as a tool for management were to: 1) optimise genetic contributions of parents to the subsequent generations via translocation and promoting mating between particular animals; and 2) identify an appropriate source for introduced animals. The latter can be achieved by sampling only a subset of animals across metapopulations, but the former is impossible without genotypes of all animals or complete pedigrees. I argue that until the development of a high-density SNP array for SWR is completed and complete genotype databases are available, high-quality pedigrees remain necessary for kinship estimations and molecular markers should be used only as a tool to facilitate pedigree constructions.

5.3.2. Management to minimise marker-based and pedigree-based kinships can lead to different outcomes

Estimation of kinship relies on determination of the probability that two alleles randomly drawn from two individuals are identical by descent. Kinships estimated from molecular markers reflect identity of the alleles whether they are shared by common ancestors (identical by descent) or identical by state. Nei (1973) described expected heterozygosity (H_e) as the probability that two alleles at a locus randomly chosen from a panmictic population are different. In this sense, a population mean kinship equals to 1 - H_e, given that a locus under consideration is in Hardy-Weinberg equilibrium (Lacy, 1995). In contrast, pedigree-based kinships reflect only alleles that are shared via common ancestors (Toro et al., 2014); in this sense, an individual *MK* suggests how many relatives the animal has in the population. Hence, it is an indication of inbreeding across the pedigree and breeding programmes that aim to minimise pedigree-based *MK* would equalise genetic contributions across individuals (Caballero and Toro, 2002).

In contrast, breeding to reduce marker-based *MK* attempts to equalise allele frequencies, thereby promoting the mating of animals carrying rare alleles (Frankham et al., 2017). Marker-based approaches should allow preservation of the gene pool present in the populations and maximisation of genetic diversity in the subsequent generations. However, problems may occur if rare alleles are present in an inbred animal or an unrecognised

hybrid (Halbert and Derr, 2006, Feulner et al., 2013). Such a circumstance is likely to occur in a species whose territory is encroached on by domestic animals, with which it might interbreed. For example, marker-based management of North American bison (*Bison bison*) populations was conducted without knowledge of genetic introgression from domestic cattle (*Bos taurus*), alleles of which were present in seven out of 11 bison populations (Halbert and Derr, 2006). As a precaution for the case that rare deleterious alleles might be present unnoticed in a population, removal of loci that have low minor allele frequencies (< 0.1) from a marker panel used to estimate kinships was found to increase fitness of simulated populations because this method removed close-to-fixation markers possibly linked to deleterious mutations (de Cara et al., 2013). Application of marker-based kinship remains useful for population management, although there is a risk of transmitting rare deleterious alleles to the next generations. If introgression of rare alleles is expected, checks for kinship outliers and/or morphological characteristics of hybrids may be necessary.

5.3.3. Number and frequency of translocations

Translocation of the candidates identified in this thesis aimed to minimise kinship coefficients either at an individual, a population, or between-population levels by trying to equalise genetic contributions of all individuals. In practice there will be a challenge to determine a justifiable number of translocations over a certain period of time. If there are too few migrants, the strategy would be inefficient to prevent inbreeding and may fail to maximise genetic diversity. Conversely, if too many, a gene pool of locally adapted populations can be swamped by the newcomers (Whiteley et al., 2015), especially if the population size of the recipient population is relatively small (Pickup et. al., 2013). In the case of the SWR, unless a candidate for introduction comes from a captive breeding programme, the latter issue should not be something of concern given the historical bottleneck and recent genetic studies that revealed no genetic structure among modern non-captive SWR populations (Moodley et al., 2018, Sánchez Barreiro et al., 2020). However, too many unnecessary translocations means a waste of resources; thus, determining a justifiable level of gene flow is strategically and economically important for SWR conservation.

Strategies for determining the proper number of introductions/translocations have been proposed. First, a rule of thumb suggesting an introduction of at least one individual per

generation has been proposed (Mills and Allendorf, 1996). This strategy is based on the estimation of the rate of genetic differentiation between subpopulations which is proportional to a population size: $F_{ST} \approx \frac{1}{4mN+1}$ where m and N are the proportion of migrants per generation and the census population size, respectively (Wright, 1950). Based on the equation, only one migrant per generation should essentially reduce the rate of genetic differentiation between isolated subpopulations and the rate of loss of genetic diversity due to genetic drift. However, it relies on simplifying assumptions that may not hold in real populations: 1) the effective population size (N_e) must equal to N; 2) equal sex ratio and all individuals have the same chance of breeding; and 3) equal chances of surviving and breeding between migrants and residents (Mills and Allendorf, 1996). In regard to the three Botswana populations, violation of all assumptions was very likely as the results show in Chapter 2 (i.e. skewed N_e:N ratio, unequal chance of breeding). For such an issue, using an *Ne:N* ratio may help to determine a justifiable number of introduced animals (Frankham et al., 2017). Frankham (1995) reviewed 192 studies from 102 species and suggested that for a species with highly skewed *Ne:N* ratio, there might be a need of more than one migrant per generation to effectively maximise genetic diversity; however, how many more was not addressed (Frankham, 1995). Based on the *Ne:N* ratio of ~0.36 estimated for Botswana1 reported in Chapter 2 and a generation time of eight years (Hillman-Smith et al., 1986), I would suggest introduction of at least three individuals every 8-year period to ensure at least one effective migrant per generation (Wang, 2004b).

Second, simulation is an alternative mean for determining the number of migrants to inform metapopulation management. Discussion about software packages for this kind of studies is beyond the scope of this thesis, but they are reviewed in Hoban et al. (2012). Simulation overcomes the issue about oversimplification of the one-migrant-per-generation strategy and allows examining the outcomes of multiple possible scenarios. For example: translocations of only a single sex, and combination of translocations with other management strategies (Bruford et al., 2010); effects of releasing locations (Bretagnolle and Inchausti, 2005); varying ages and sex ratios of translocated individuals (Tocher et al., 2006); and timing of translocations (Facka et al., 2016). These studies were done in many species to facilitate population management; however, each species had its own unique circumstances and the results from one species should not be extrapolated for management of the others. For the SWR, although the idea of individual translocations between populations is not novel and has long been practiced (Emslie et al., 2009), incorporation of

such simulations is rarely done to inform management. Further studies using computational simulation should be very welcome and are needed to improve SWR translocation programmes; indeed they are overdue.

5.3.4. Forensic studies

Population fragmentation causes genetic differentiation due to random sampling in populations with a finite size. The rate of genetic change in populations is critically dependent on their effective population size (N_e) and gene flow between subpopulations (Nunney, 1999). Populations with no gene flow at all are likely to suffer from reduced genetic diversity, inbreeding depression, and increased risk of extinction (Rodríguez and Delibes, 2003, Martinez-Cruz et al., 2007). However, if genetic differentiation exists in heavily poached species, it can be exploited to facilitate wildlife crime prosecutions. The questions asked in wildlife forensics often involve the identification of whether a seized piece of evidence is obtained from: 1) legally protected species; 2) captive or wild-caught animals; 3) a specific population or geographic location; or 4) a particular individual in a protected population (Ogden and Linacre, 2015). For SWR, the first question requires relatively less information and the currently available molecular markers should be able to solve the proposed forensic problem. For example, a panel of currently available microsatellite markers was able to distinguish horns of the SWR from three black rhino subspecies (Harper et al., 2018). The panel has also been used to facilitate prosecution of SWR crimes; e.g. matching the seized horns with illegally poached carcasses or blood stains on the carpet of the offender's vehicles (Harper et al., 2018). Moodley et al. (2018) reported indications of genetic drift within two captive populations of SWRs based on lower genetic diversity (10 microsatellites) revealed in the captive-born individuals compared to their wild-born founders. This finding suggested that genetic differentiation might already occurred across isolated SWR populations which warranted the need of genetic tools to facilitate metapopulation management to avoid genetic drift towards rare deleterious alleles. However, on the bright side, genetic differentiation among SWR populations suggests the potential application of genetic tools to help identification the source of the biological evidence, there is a need for further optimisation of the SNP panel and genotyping method.

The third and fourth forensic questions are often more difficult to answer with existing genomic resources of the SWR, but they have become decisive evidence in wildlife crime

prosecutions for other species (Ogden and Linacre, 2015). Origin assignment of biological evidence requires the collection of reference genotypes from different possible origins to test the possibility that an individual originated from one of the proposed populations (Frantz et al., 2006, Ewart et al., 2018). Construction of reference genomic databases would facilitate large-scale genotyping of many animals from many populations but will require collaboration among laboratories. In such a case, one important challenge is the exchangeability of genotype data, for which SNP arrays are highly suited.

Genotyping of SNPs is automatically standardised because they directly represent variants in DNA sequences and are considerably more straightforward than microsatellites for inter-laboratory transfer. In this thesis, no obvious genetic structure among three populations was revealed which posed a challenge for development of a SNP panel that would be able to distinguish between tested populations. However, given the historical movements of animals between the populations included in this study, individuals in these populations would be expected to be related to each other to some extent and so diagnostics for forensics might have to target the whole metapopulation, rather than individual protected areas. Indications of genetic drift reported in Moodley et al. (2018) gave a glimpse of the potential application of molecular markers for population assignment across isolated populations; together with the 'raw materials' developed in this thesis, they bring the population assignment in SWRs one step closer to practice.

5.4. Conclusions

This thesis revealed insights about genetic diversity of three important SWR populations that shelter ~25% of the Botswana's national herd. Despite the low genetic diversity within these populations, parentage assignment could be successfully made for the majority of offspring in the largest population sampled and revealed unexpected contributions of multiple bulls following the removal of the behaviourally dominant bulls. This questions a fundamental assumption about the mating system of these animals, which could substantially alter management practices, since removing the dominant bull could allow higher genetic diversity even with existing variation within populations. Whether the marker-based mean kinship estimations using the new set of molecular markers developed through RADcapture are sensitive enough to identify the most promising individuals to translocate without a reliable pedigree remains unclear, but the method developed did allow inclusion of much lower quality samples than would have been possible for either

microsatellites or ddRAD genotyping. Although the high cost could be prohibitive, it could provide a useful approach for other endangered species, when only poor-quality DNA is available. Future development of a SNP-chip based on these data could provide a more accessible tool for future conservation management, but it would require extensive validation using a wider range of populations.

References

AGRESTI, A. 1992. A survey of exact inference for contingency tables. *Statistical science,* 7**,** 131-153.

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.

ALEXANDER, D. H., NOVEMBRE, J. & LANGE, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res,* 19**,** 1655-64.

ALEXANDER, D. H., SHRINGARPURE, S. S., NOVEMBRE, J. & LANGE, K. 2015. Admixture 1.3 software manual.

ALI, O. A., O'ROURKE, S. M., AMISH, S. J., MEEK, M. H., LUIKART, G., JEFFRES, C. & MILLER, M. R. 2016. RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics,* 202**,** 389-400.

ALLEY, M. R. & GARTRELL, B. D. 2019. Wildlife diseases in New Zealand: recent findings and future challenges AU - Alley, MR. *New Zealand Veterinary Journal,* 67**,** 1-11.

ANDERSON, J. T., LEE, C.-R., RUSHWORTH, C. A., COLAUTTI, R. I. & MITCHELL-OLDS, T. 2013. Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology,* 22**,** 699-708.

ANDERSON ERIC, C. 2012. Large-scale Parentage Inference with SNPs: an Efficient Algorithm for Statistical Confidence of Parent Pair Allocations. *Statistical Applications in Genetics and Molecular Biology.*

ANDREWS, S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

ANDREWS, K. R., GOOD, J. M., MILLER, M. R., LUIKART, G. & HOHENLOHE, P. A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics,* 17**,** 81.

ARAÚJO, C. V. M., GONZÁLEZ-ORTEGÓN, E., PINTADO-HERRERA, M. G., BIEL-MAESO, M., LARA-MARTÍN, P. A., TOVAR-SÁNCHEZ, A. & BLASCO, J. 2019. Disturbance of ecological habitat distribution driven by a chemical barrier of domestic and agricultural discharges: An experimental approach to test habitat fragmentation. *Science of The Total Environment,* 651**,** 2820-2829.

ARBOR BIOSCIENCES. 2018. Hybridization Capture for Targeted NGS. *In:* BIOSCIENCES, A. (ed.) April 2018 ed. https://arborbiosci.com/wp-content/uploads/2019/08/myBaits-Manual-v4.pdf.

ASHLEY, M. V., BERGER-WOLF, T. Y., CABALLERO, I. C., CHAOVALITWONGSE, W., DASGUPTA, B. & SHEIKH, S. I. 2009. Full sibling reconstruction in wild populations from microsatellite genetic markers. *In:* RUSSE, A. S. (ed.) *Computational biology:New research.* Hauppauge, New York: Nova Publisher.

ATTARD, C. R. M., BEHEREGARAY, L. B. & MÖLLER, L. M. 2018. Genotyping-by-sequencing for estimating relatedness in nonmodel organisms: Avoiding the trap of precise bias. *Molecular Ecology Resources,* 18**,** 381-390.

BAAS, P., VAN DER VALK, T., VIGILANT, L., NGOBOBO, U., BINYINYI, E., NISHULI, R., CAILLAUD, D. & GUSCHANSKI, K. 2018. Population-level assessment of genetic diversity and habitat fragmentation in critically endangered Grauer's gorillas. *American Journal of Physical Anthropology,* 165**,** 565-575.

BANES, G. L., GALDIKAS, B. M. F., & VIGILANT, L. (2016). Reintroduction of confiscated and displaced mammals risks outbreeding and introgression in natural populations, as evidenced by orang-utans of divergent subspecies. *Scientific Reports, 6*(1), 22026.

BAIRD, N. A., ETTER, P. D., ATWOOD, T. S., CURREY, M. C., SHIVER, A. L., LEWIS, Z. A., SELKER, E. U., CRESKO, W. A. & JOHNSON, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one,* 3**,** e3376.

BALLOU, J. D. & LACY, R. C. 1995. Identifying genetically important individuals for management of genetic variation in pedigreed populations. *Population management for survival and recovery***,** 76-111.

BALLOUX, F., AMOS, W. & COULSON, T. 2004. Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology,* 13**,** 3021-3031.

BARBRAUD, C. & WEIMERSKIRCH, H. 2001. Emperor penguins and climate change. *Nature,* 411**,** 183.

BARNES, M. D., CRAIGIE, I. D., HARRISON, L. B., GELDMANN, J., COLLEN, B., WHITMEE, S., BALMFORD, A., BURGESS, N. D., BROOKS, T., HOCKINGS, M. & WOODLEY, S. 2016. Wildlife population trends in protected areas predicted by national socio-economic metrics and body size. *Nature Communications,* 7**,** 12747.

BEACHAM, T. D., WALLACE, C., MACCONNACHIE, C., JONSEN, K., MCINTOSH, B., CANDY, J. R., DEVLIN, R. H. & WITHLER, R. E. 2017. Population and individual identification of coho salmon in British Columbia through parentage-based tagging and genetic stock identification: an alternative to coded-wire tags. *Canadian Journal of Fisheries and Aquatic Sciences,* 74**,** 1391-1410.

BENSCH, S., ÅKESSON, S. & IRWIN, D. E. 2002. The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology,* 11**,** 2359-2366.

BENSON, J. F., HOSTETLER, J. A., ONORATO, D. P., JOHNSON, W. E., ROELKE, M. E., O'BRIEN, S. J., JANSEN, D. & OLI, M. K. 2011. Intentional genetic introgression influences survival of adults and subadults in a small, inbred felid population. *Journal of Animal Ecology,* 80**,** 958-967.

BÉRÉNOS, C., ELLIS, P. A., PILKINGTON, J. G. & PEMBERTON, J. M. 2014. Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular ecology,* 23**,** 3434-3451.

BERNARDO, R. (1993). Estimation of coefficient of coancestry using molecular markers in maize. *Theoretical and Applied Genetics, 85*(8), 1055-1062.

BERNATCHEZ, L. & DUCHESNE, P. 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences,* 57**,** 1-12.

BIGGS, D., COURCHAMP, F., MARTIN, R. & POSSINGHAM, H. P. 2013. Legal Trade of Africa's Rhino Horns. *Science,* 339**,** 1038-1039.

BINLADEN, J., GILBERT, M. T. P., BOLLBACK, J. P., PANITZ, F., BENDIXEN, C., NIELSEN, R. & WILLERSLEV, E. 2007. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLOS ONE,* 2**,** e197.

BLAIMER, B. B., LLOYD, M. W., GUILLORY, W. X. & BRADY, S. G. 2016. Sequence Capture and Phylogenetic Utility of Genomic Ultraconserved Elements Obtained from Pinned Insect Specimens. *PLOS ONE,* 11**,** e0161531.

BLOUIN, M. S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution,* 18**,** 503-511.

BOHMANN, K., EVANS, A., GILBERT, M. T. P., CARVALHO, G. R., CREER, S., KNAPP, M., YU, D. W. & DE BRUYN, M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution,* 29**,** 358-367

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-2120

BÖMCKE, E. & GENGLER, N. 2009. Combining microsatellite and pedigree data to estimate relationships among Skyros ponies. *Journal of Applied Genetics,* 50**,** 133-143.

BOSE, A. P. H., ZIMMERMANN, H., HENSHAW, J. M., FRITZSCHE, K. & SEFC, K. M. 2018. Brood-tending males in a biparental fish suffer high paternity losses but rarely cuckold. *Molecular Ecology,* 27**,** 4309-4321.

BOSE, N., CARLBERG, K., SENSABAUGH, G., ERLICH, H. & CALLOWAY, C. 2018. Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples. *Forensic Science International: Genetics,* 34**,** 186-196.

BOURGEOIS, S., SENN, H., KADEN, J., TAGGART, J. B., OGDEN, R., JEFFERY, K. J., BUNNEFELD, N., ABERNETHY, K. & MCEWING, R. 2018. Single-

nucleotide polymorphism discovery and panel characterization in the African forest elephant. *Ecology and Evolution,* 8**,** 2207-2217.

BOUZAT, J. L., JOHNSON, J. A., TOEPFER, J. E., SIMPSON, S. A., ESKER, T. L. & WESTEMEIER, R. L. 2009. Beyond the beneficial effects of translocations as an effective tool for the genetic restoration of isolated populations. *Conservation Genetics,* 10**,** 191-201.

BOWEN, M. E., HENKE, K., SIEGFRIED, K. R., WARMAN, M. L., & HARRIS, M. P. (2012). Efficient Mapping and Cloning of Mutations in Zebrafish by Low-Coverage Whole-Genome Sequencing. *Genetics, 190*(3), 1017-1024.

BRADBURY, I. R., HUBERT, S., HIGGINS, B., BOWMAN, S., PATERSON, I. G., SNELGROVE, P. V. R., BENTZEN, P. (2011). Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, Gadus morhua. *Molecular ecology resources, 11*(s1), 218-225.

BRANDT, J. R. 2016. *Conservation and population genetics of African and Asian Rhinoceros.* University of Illinois at Urbana-Champaign.

BRANDT, J. R., VAN COEVERDEN DE GROOT, P. J., ZHAO, K., DYCK, M. G., BOAG, P. T., & ROCA, A. L. (2014). Development of nineteen polymorphic microsatellite loci in the threatened polar bear (Ursus maritimus) using next generation sequencing. *Conservation Genetics Resources, 6*(1), 59-61.

BRANTON, M. & RICHARDSON, J. S. 2011. Assessing the Value of the Umbrella-Species Concept for Conservation Planning with Meta-Analysis. *Conservation Biology,* 25**,** 9-20.

BRENIG, B. & SCHÜTZ, E. 2016. Recent development of allele frequencies and exclusion probabilities of microsatellites used for parentage control in the German Holstein Friesian cattle population. *BMC Genetics,* 17**,** 18.

BRETAGNOLLE, V., & INCHAUSTI, P. (2005). Modelling population reinforcement at a large spatial scale as a conservation strategy for the declining little bustard (Tetrax tetrax) in agricultural habitats. *Animal Conservation, 8*(1), 59-68.

BRETMAN, A. & TREGENZA, T. 2005. Measuring polyandry in wild populations: a case study using promiscuous crickets. *Molecular Ecology,* 14**,** 2169-2179.

BRETON, G., THIBAULT, S., WERTS, M. & BAUDRY, E. 2019. Captive ring-tailed lemur breeding in semi-free ranging conditions and genetic parentage analysis. *Zoo Biology***,** 1-11.

BROMAN, K. W., & WEBER, J. L. (1999). Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain. *The American Journal of Human Genetics, 65*(6), 1493-1500. doi:10.1086/302661

BROWN, S. M. & HOULDEN, B. A. 1999. Isolation and characterization of microsatellite markers in the black rhinoceros (Diceros bicornis). *Mol Ecol,* 8**,** 1559-61.

BROWN, S. M. & HOULDEN, B. A. 2000. Conservation genetics of the black rhinoceros (Diceros bicornis). *Conservation Genetics,* 1**,** 365-370.

BROWN, J. K., TAGGART, J. B., BEKAERT, M., WEHNER, S., PALAIOKOSTAS, C., SETIAWAN, A. N., SYMONDS, J. E. & PENMAN, D. J. 2016. Mapping the sex determination locus in the hāpuku (Polyprion oxygeneios) using ddRAD sequencing. *BMC Genomics,* 17**,** 448.

BRUFORD, M. W., ANCRENAZ, M., CHIKHI, L., LACKMANN-ANCRENAZ, I., ANDAU, M., AMBU, L., & GOOSSENS, B. (2010). Projecting genetic diversity and population viability for the fragmented orang-utan population in the Kinabatangan floodplain, Sabah, Malaysia. *Endangered Species Research, 12*(3), 249-261.

BUCHANAN, J. W., WORONUK, G. N., MARQUESS, F. L., LANG, K., JAMES, S. T., DEOBALD, H., WELLY, B. T. & VAN EENENNAAM, A. L. 2016. Analysis of validated and population-specific single nucleotide polymorphism parentage panels in pedigreed and commercial beef cattle populations. *Canadian Journal of Animal Science,* 97**,** 231-240.

CABALLERO, A. & TORO, M. A. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics,* 3**,** 289-299.

CAMACHO-SANCHEZ, M., BURRACO, P., GOMEZ-MESTRE, I. & LEONARD, J. A. 2013. Preservation of RNA and DNA from mammal samples under field conditions. *Molecular Ecology Resources,* 13**,** 663-673.

CARPENTER, MEREDITH L., BUENROSTRO, JASON D., VALDIOSERA, C., SCHROEDER, H., ALLENTOFT, MORTEN E., SIKORA, M., RASMUSSEN, M., GRAVEL, S., GUILLÉN, S., NEKHRIZOV, G., LESHTAKOV, K., DIMITROVA, D., THEODOSSIEV, N., PETTENER, D., LUISELLI, D., SANDOVAL, K., MORENO-ESTRADA, A., LI, Y., WANG, J., GILBERT, M. THOMAS P., WILLERSLEV, E., GREENLEAF, WILLIAM J. & BUSTAMANTE, CARLOS D. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *The American Journal of Human Genetics,* 93**,** 852-864.

CASSELL, B. G., ADAMEC, V. & PEARSON, R. E. 2003. Effect of Incomplete Pedigrees on Estimates of Inbreeding and Inbreeding Depression for Days to First Service and Summit Milk Yield in Holsteins and Jerseys. *Journal of Dairy Science,* 86**,** 2967-2976.

CATCHEN, J., HOHENLOHE, P. A., BASSHAM, S., AMORES, A. & CRESKO, W. A. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology,* 22**,** 3124-3140.

CATCHEN, J. M., AMORES, A., HOHENLOHE, P., CRESKO, W. & POSTLETHWAIT, J. H. 2011. Stacks: Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3: Genes|Genomes|Genetics,* 1**,** 171-182.

CATCHEN, J. M., HOHENLOHE, P. A., BERNATCHEZ, L., FUNK, W. C., ANDREWS, K. R. & ALLENDORF, F. W. 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources,* 17**,** 362-365.

CEBALLOS, G., EHRLICH, P. R. & DIRZO, R. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences,* 114**,** E6089.

CHAKRABORTY, R., SHAW, M. & SCHULL, W. J. 1974. Exclusion of paternity: the current state of the art. *American Journal of Human Genetics,* 26**,** 477.

CHANDRAN, S. 2016. *Understanding and adapting the generic hard-filtering recommendation* [Online]. https://gatkforums.broadinstitute.org/: Broad Institute. Available: https://gatkforums.broadinstitute.org/gatk/discussion/6925/understanding-and-adapting-the-generic-hard-filtering-recommendations [Accessed 14 April 2020 2020].

CHEN, I.-C., HILL, J. K., OHLEMÜLLER, R., ROY, D. B. & THOMAS, C. D. 2011. Rapid Range Shifts of Species Associated with High Levels of Climate Warming. 333**,** 1024-1026.

CHRISTIE, M. R. 2013. Bioinformatics in press see http://www.science.oregonstate.edu/~christim/SOLOMON.html for a full user guide and FAQs.

CHRISTIE, M. R., TENNESSEN, J. A. & BLOUIN, M. S. 2013. Bayesian parentage analysis with systematic accountability of genotyping error, missing data and false matching. *Bioinformatics,* 29**,** 725-732.

COHAS, A. & ALLAINÉ, D. 2009. Social structure influences extra-pair paternity in socially monogamous mammals. *Biology Letters,* 5**,** 313-316.

COLTMAN, D. W. & SLATE, J. 2003. Microsatellite measures of inbreeding: a meta-analysis. *Evolution,* 57**,** 971-983.

CONSERVATION BREEDING SPECIALIST GROUP (CBSG). 2008. Tasmanian devil PHVA final report.

COORDINATORS, N. R. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic acids research,* 44**,** D7-D19.

CORNELIS, S., GANSEMANS, Y., DELEYE, L., DEFORCE, D., & VAN NIEUWERBURGH, F. (2017). Forensic SNP Genotyping using Nanopore MinION Sequencing. *Scientific Reports, 7*(1), 41759. doi:https://doi.org/10.1038/srep41759

CORTI, P., SHAFER, A. B. A., COLTMAN, D. W. & FESTA-BIANCHET, M. 2011. Past bottlenecks and current population fragmentation of endangered huemul deer (Hippocamelus bisulcus): implications for preservation of genetic diversity. *Conservation Genetics,* 12**,** 119-128.

COSART, T., BEJA-PEREIRA, A., CHEN, S., NG, S. B., SHENDURE, J. & LUIKART, G. 2011. Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics,* 12**,** 347.

COUTTS, N. J. 2009. *Investigating genetic diversity at neutral and adaptive DNA markers in the severly bottlenecked Southern white Rhinoceros (Ceratotherium simum simum).* Doctoral dissertation Dissertation, University of Cape Town.

CRAWFORD, J. E., RIEHLE, M. M., MARKIANOS, K., BISCHOFF, E., GUELBEOGO, W. M., GNEME, A., SAGNON, N. F., VERNICK, K. D., NIELSEN, R. & LAZZARO, B. P. 2016. Evolution of GOUNDRY, a cryptic subgroup of Anopheles gambiae sl, and its impact on susceptibility to Plasmodium infection. *Molecular ecology,* 25**,** 1494-1510.

Crnokrak, P., & Roff, D. A. (1999). Inbreeding depression in the wild. *Heredity, 83*(3), 260-270.

CULVER, M., JOHNSON, W. E., PECON-SLATTERY, J. & O'BRIEN, S. J. 2000. Genomic ancestry of the American puma (*Puma concolor*). *Journal of Heredity,* 91**,** 186-197.

CUMMING, D. H. M., DU TOIT, R. F. & STUART, S. N. 1990. African Elephants and Rhinos: Status Survey and Conservation Action Plan: Action Plan for the Conservation of Biological Diversity 10. IUCN/SSC-AESG, Gland, Switzerland.

CUNNINGHAM, J., HARLEY, E. H., & O'RYAN, C. (1999). Isolation and characterization of microsatellite loci in black rhinoceros (Diceros bicornis). *Electrophoresis, 20*(8), 1778-1780.

DAKIN, E. E. & AVISE, J. C. 2004. Microsatellite null alleles in parentage analysis. *Heredity,* 93**,** 504-509.

DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C. A., BANKS, E., DEPRISTO, M. A., HANDSAKER, R. E., LUNTER, G., MARTH, G. T., SHERRY, S. T., MCVEAN, G., DURBIN, R. & GENOMES PROJECT ANALYSIS, G. 2011. The variant call format and VCFtools. *Bioinformatics (Oxford, England),* 27**,** 2156-2158.

DASKIN, J. H. & PRINGLE, R. M. 2018. Warfare and wildlife declines in Africa's protected areas. *Nature,* 553**,** 328.

DASZAK, P., CUNNINGHAM, A. A. & HYATT, A. D. 2000. Emerging Infectious Diseases of Wildlife-- Threats to Biodiversity and Human Health. *Science,* 287**,** 443.

DAVEY, J. W. & BLAXTER, M. L. 2011. RADSeq: next-generation population genetics. *Briefings in Functional Genomics,* 9**,** 416-423.

DAVEY, J. W., HOHENLOHE, P. A., ETTER, P. D., BOONE, J. Q., CATCHEN, J. M. & BLAXTER, M. L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics,* 12**,** 499.

DAVIES-MOSTERT, H. T., MILLS, M. G. L. & MACDONALD, D. W. 2015. The Demography and Dynamics of an Expanding, Managed African Wild Dog Metapopulation. *African Journal of Wildlife Research,* 45**,** 258-273, 16.

DE CARA, M. Á. R., VILLANUEVA, B., TORO, M. Á. & FERNÁNDEZ, J. 2013. Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology,* 22**,** 6091-6099.

DE RYCKE, M. & STAESSEN, C. 2017. Chapter 23 - Preimplantation Genetic Diagnosis. *In:* PATRINOS, G. P. (ed.) *Molecular Diagnostics (Third Edition).* Academic Press.

DELGADO, M. L., GÓRSKI, K., HABIT, E. & RUZZANTE, D. E. 2019. The effects of diadromy and its loss on genomic divergence: The case of amphidromous Galaxias maculatus populations. *Molecular Ecology,* 28**,** 5217-5231.

DELSUC, F., KUCH, M., GIBB, G. C., KARPINSKI, E., HACKENBERGER, D., SZPAK, P., MARTÍNEZ, J. G., MEAD, J. I., MCDONALD, H. G., MACPHEE,

R. D. E., BILLET, G., HAUTIER, L. & POINAR, H. N. 2019. Ancient Mitogenomes Reveal the Evolutionary History and Biogeography of Sloths. *Current Biology,* 29**,** 2031-2042.e6.

DEVLIN, B., ROEDER, K. & ELLSTRAND, N. C. 1988. Fractional paternity assignment: theoretical development and comparison to other methods. *Theoretical and Applied Genetics,* 76**,** 369-380.

DEPARTMENT OF WILDLIFE AND NATIONAL PARKS OF BOTSWANA. 2002. Conservation and Management Strategy for the White Rhinoceros *Ceratotherium simum* and the Black Rhinoceros *Diceros bicornis* in Botswana. *In:* Anonymus (ed.). Gaborone: Department of Wildlife and National Parks Botswana.

DI MININ, E., LAITILA, J., MONTESINO-POUZOLS, F., LEADER-WILLIAMS, N., SLOTOW, R., GOODMAN, P. S., CONWAY, A. J. & MOILANEN, A. 2015. Identification of policies for a sustainable legal trade in rhinoceros horn based on population projection and socioeconomic models. 29**,** 545-555.

DOBRYNIN, P., LIU, S., TAMAZIAN, G., XIONG, Z., YURCHENKO, A. A., KRASHENINNIKOVA, K., KLIVER, S., SCHMIDT-KÜNTZEL, A., KOEPFLI, K.-P., JOHNSON, W., KUDERNA, L. F. K., GARCÍA-PÉREZ, R., MANUEL, M. D., GODINEZ, R., KOMISSAROV, A., MAKUNIN, A., BRUKHIN, V., QIU, W., ZHOU, L., LI, F., YI, J., DRISCOLL, C., ANTUNES, A., OLEKSYK, T. K., EIZIRIK, E., PERELMAN, P., ROELKE, M., WILDT, D., DIEKHANS, M., MARQUES-BONET, T., MARKER, L., BHAK, J., WANG, J., ZHANG, G. & O'BRIEN, S. J. 2015. Genomic legacy of the African cheetah, Acinonyx jubatus. *Genome Biology,* 16**,** 277.

DODDS, K. G., McEWAN, J. C., BRAUNING, R., VAN STIJN, T. C., ROWE, S. J., McEWAN, K. M., & CLARKE, S. M. (2019). Exclusion and Genomic Relatedness Methods for Assignment of Parentage Using Genotyping-by-Sequencing Data. *G3: Genes|Genomes|Genetics, 9*(10), 3239-3247.

DORANT, Y., BENESTAN, L., ROUGEMONT, Q., NORMANDEAU, E., BOYLE, B., ROCHETTE, R. & BERNATCHEZ, L. 2019. Comparing Pool-seq, Rapture, and GBS genotyping for inferring weak population structure: The American lobster (Homarus americanus) as a case study. *Ecology and Evolution,* 9**,** 6606-6623.

DRAY, S. & DUFOUR, A.-B. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software,* 22**,** 1-20.

DU TOIT, J. G. 2015. History of Rhinoceros Conservation in South Africa. *In:* DU TOIT, J. G. (ed.) *The Story of the White Rhinoceros.* South Africa: Kejafa Knowledge Works.

DU TOIT, R. F. 2006. *Guidelines for implementing SADC rhino conservation strategies,* Harare (Zimbabwe), WWF-SARPO.

ECKERT, K. A. & HILE, S. E. 2009. Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Molecular Carcinogenesis,* 48**,** 379-388.

EDING, H. & MEUWISSEN, T. H. E. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics,* 118**,** 141-159.

EDMANDS, S. (2007). Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology, 16*(3), 463-475.

EDWARDS, C. J., LOFTUS, R. T., BRADLEY, D. G., DOLF, G., & LOOFT, C. (2000). Relationships between the endangered Pustertaler–Sprinzen and three related European cattle breeds as analysed with 20 microsatellite loci. *Animal genetics, 31*(5), 329-332.

EKBLOM, R. & GALINDO, J. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity,* 107**,** 1-15.

ELLEGREN, H., MIKK, S., WALLIN, K. & ANDERSSON, L. 1996. Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose A. alces. 5**,** 3-9.

ELSHIRE, R. J., GLAUBITZ, J. C., SUN, Q., POLAND, J. A., KAWAMOTO, K., BUCKLER, E. S. & MITCHELL, S. E. 2011. A robust, simple genotyping-by-

sequencing (GBS) approach for high diversity species. *PloS one,* 6**,** e19379-e19379.

ELSTON, R. C. & STEWART, J. 1971. A General Model for the Genetic Analysis of Pedigree Data. *Human Heredity,* 21**,** 523-542.

EMERSON, K. J., MERZ, C. R., CATCHEN, J. M., HOHENLOHE, P. A., CRESKO, W. A., BRADSHAW, W. E. & HOLZAPFEL, C. M. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences,* 107**,** 16196-16200.

EMERY, A. M., WILSON, I. J., CRAIG, S., BOYLE, P. R. & NOBLE, L. R. 2001. Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. *Molecular Ecology,* 10**,** 1265-1278.

EMSLIE, R., AMIN, R. & KOCK, R. 2009. *Guidelines for the in situ re-introduction and translocation of African and Asian rhinoceros*, International Union for Conservation of Nature.

EMSLIE, R. & BROOKS, M. 1999. *African rhino: status survey and conservation action plan*, IUCN.

EMSLIE, R. H., MILLIKEN, T., TALUKDAR, B., ELLIS, S., ADCOCK, K. & KNIGHT, M. H. 2016. African and Asian rhinoceroses—status, conservation and trade: A report from the IUCN Species Survival Commission (IUCN/SSC) African and Asian Rhino Specialist Groups and TRAFFIC to the CITES Secretariat pursuant to Resolution Conf. 9.14 (Rev. CoP15).

EMSLIE, R. H., MILLIKEN, T., TALUKDAR, B., BURGESS, G., ADCOCK, K., BALFOUR, D. & KNIGHT, M. H. 2019. African and Asian Rhinoceroses - Status, Conservation and Trade A Report from the IUCN Species Survival COmmission (IUCN SSC) African and Asian Rhino Specialist Groups and TRAFFIC to the CITES Secretariat Pursuant to Resoultion Conf. 9.14 (Rev CoP17). Colombo.

EWART, K. M., FRANKHAM, G. J., MCEWING, R., WEBSTER, L. M. I., CIAVAGLIA, S. A., LINACRE, A. M. T., THE, D. T., OVOUTHAN, K. & JOHNSON, R. N. 2018. An internationally standardized species identification test

for use on suspected seized rhinoceros horn in the illegal wildlife trade. *Forensic Science International: Genetics,* 32**,** 33-39.

EWART, K. M., JOHNSON, R. N., OGDEN, R., JOSEPH, L., FRANKHAM, G. J. & LO, N. 2019. Museum specimens provide reliable SNP data for population genomic analysis of a widely distributed but threatened cockatoo species. *Molecular Ecology Resources,* 19**,** 1578-1592.

FABRIZIO, M., DI FEBBRARO, M., D'AMICO, M., FRATE, L., ROSCIONI, F. & LOY, A. 2019. Habitat suitability vs landscape connectivity determining roadkill risk at a regional scale: a case study on European badger (Meles meles). *European Journal of Wildlife Research,* 65**,** 7.

FACKA, A. N., LEWIS, J. C., HAPPE, P., JENKINS, K., CALLAS, R., & POWELL, R. A. (2016). Timing of translocation influences birth rate and population dynamics in a forest carnivore. *Ecosphere, 7*(1), e01223.

FALCONER, D. S. 1960. *Introduction to quantitative genetics*, Oliver And Boyd; Edinburgh; London.

FARQUHARSON, K. A., HOGG, C. J. & GRUEBER, C. E. 2019. A case for genetic parentage assignment in captive group housing. *Conservation Genetics,* 20**,** 1187-1193.

FERNÁNDEZ, J., VILLANUEVA, B., PONG-WONG, R. & TORO, M. Á. 2005. Efficiency of the Use of Pedigree and Molecular Marker Information in Conservation Programs. *Genetics,* 170**,** 1313-1321.

FERREIRA, S. M., LE ROEX, N. & GREAVER, C. 2019. Species-specific drought impacts on black and white rhinoceroses. *PLOS ONE,* 14**,** e0209678.

FERREIRA, S. M., BISSETT, C., COWELL, C. R., GAYLARD, A., GREAVER, C., HAYES, J., HOFMEYR, M., MOOLMAN-VAN DER VYVER, L. & ZIMMERMANN, D. 2017. The status of rhinoceroses in South African National Parks. *2017,* 59.

FEULNER, P. G. D., GRATTEN, J., KIJAS, J. W., VISSCHER, P. M., PEMBERTON, J. M. & SLATE, J. 2013. Introgression and the fate of domesticated genes in a wild mammal population. *Molecular Ecology,* 22**,** 4210-4221.

FINGER, A., KETTLE, C. J., KAISER-BUNBURY, C. N., VALENTIN, T., DOUDEE, D., MATATIKEN, D. & GHAZOUL, J. 2011. Back from the brink: potential for genetic rescue in a critically endangered tree. *Mol Ecol,* 20**,** 3773-84.

FISCHER, M. C., RELLSTAB, C., LEUZINGER, M., ROUMET, M., GUGERLI, F., SHIMIZU, K. K., HOLDEREGGER, R. & WIDMER, A. 2017. Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in Arabidopsis halleri. *BMC Genomics,* 18**,** 69.

FISHER, P. J., MALTHUS, B., WALKER, M. C., CORBETT, G. & SPELMAN, R. J. 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. *Journal of Dairy Science,* 92**,** 369-374.

FITAK, R. R. 2014. *Conservation genomics of the endangered Mexican wolf and de novo SNP marker development in pumas using next-generation sequencing.* The University of Arizona.

FLANAGAN, S. P. & JONES, A. G. 2019. The future of parentage analysis: From microsatellites to SNPs and beyond. 28**,** 544-567.

FLETCHER JR, R. J., REICHERT, B. E. & HOLMES, K. 2018. The negative effects of habitat fragmentation operate at the scale of dispersal. *Ecology,* 99**,** 2176-2186.

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. *Molecular Ecology Resources,* 3**,** 344-345.

FOUNTAIN, E. D., PAULI, J. N., REID, B. N., PALSBØLL, P. J., & PEERY, M. Z. (2016). Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Mol Ecol Resour, 16*(4), 966-978.

FRANKHAM, R., BALLOU, J. D., RALLS, K., ELDRIDGE, M., DUDASH, M. R., FENSTER, C. B., LACY, R. C. & SUNNUCKS, P. 2017. *Genetic management of fragmented animal and plant populations*, Oxford University Press.

FRANKHAM, R. (2012). How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. *Heredity, 108*(3), 167-178. doi:https://doi.org/10.1038/hdy.2011.66

FRANKHAM, R., BALLOU, J. D., ELDRIDGE, M. D. B., LACY, R. C., RALLS, K., DUDASH, M. R., & FENSTER, C. B. (2011). Predicting the Probability of Outbreeding Depression. *Conservation Biology, 25*(3), 465-475.

FRANKHAM, R., BRADSHAW, C. J. A. & BROOK, B. W. 2014. Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation,* 170**,** 56-63.

FRANKHAM, R. (1995). Effective population size/adult population size ratios in wildlife: a review. *Genetical Research, 66*(2), 95-107.

FRANZÉN, O., JERLSTRÖM-HULTQVIST, J., CASTRO, E., SHERWOOD, E., ANKARKLEV, J., REINER, D. S., PALM, D., ANDERSSON, J. O., ANDERSSON, B. & SVÄRD, S. G. 2009. Draft genome sequencing of giardia intestinalis assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS pathogens,* 5**,** e1000560-e1000560.

FRANTZ, A. C., POURTOIS, J. T., HEUERTZ, M., SCHLEY, L., FLAMAND, M. C., KRIER, A., BERTOUILLE, S., CHAUMONT, F. & BURKE, T. 2006. Genetic structure and assignment tests demonstrate illegal translocation of red deer (Cervus elaphus) into a continuous population. *Molecular Ecology,* 15**,** 3191-3203.

GALLA, S. J., MORAGA, R., BROWN, L., CLELAND, S., HOEPPNER, M. P., MALONEY, R. F., STEEVES, T. E. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. *Evolutionary Applications, 13*(5), 991-1008.

GARBE, J. R., PRAKAPENKA, D., TAN, C. & DA, Y. 2016. Genomic Inbreeding and relatedness in wild panda populations. *PLoS One,* 11**,** e0160496.

GARVIN, M. R., SAITOH, K. & GHARRET, A. J. 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources,* 10**,** 915-934.

GATK TEAM. 2020. *(How to) Filter varinats either with VQSR or by hard-filtering* [Online]. https://gatk.broadinstitute.org: GATK Team. [Accessed 14 April 2020 2020].

GAVOR, I. K. N. 1988. Status Report: Rhinos in Botswana. Chobe, Botswana.

GAYNOR, K. M., FIORELLA, K. J., GREGORY, G. H., KURZ, D. J., SETO, K. L., WITHEY, L. S. & BRASHARES, J. S. 2016. War and wildlife: linking armed conflict to conservation. 14**,** 533-542.

GEORGE, M., CHEMNICK, L., CISOVA, D., GABRISOVA, E., STRAIL, A., & RYDER, O. (1993). Genetic differentiation of white rhinoceros subspecies: diagnostic differences in mitochondrial DNA and serum proteins. *Rhinoceros Biology and Conservation, OA Ryder, ed.(San Diego: Zoological Society of San Diego)*, 105-113.

GERBER, S., CHABRIER, P. & KREMER, A. 2003. FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes,* 3**,** 479-481.

GERLAND, P., RAFTERY, A. E., ŠEVČÍKOVÁ, H., LI, N., GU, D., SPOORENBERG, T., ALKEMA, L., FOSDICK, B. K., CHUNN, J., LALIC, N., BAY, G., BUETTNER, T., HEILIG, G. K. & WILMOTH, J. 2014. World population stabilization unlikely this century. *Science,* 346**,** 234-237.

GIGLIO, R. M., IVY, J. A., JONES, L. C., & LATCH, E. K. (2018). Pedigree-based genetic management improves bison conservation. *The Journal of Wildlife Management, 82*(4), 766-774.

GILBERT, D. A., PACKER, C., PUSEY, A. E., STEPHENS, J. C., & O'BRIEN, S. J. (1991). Analytical DNA Fingerprinting in Lions: Parentage, Genetic Diversity, and Kinship. *Journal of Heredity, 82*(5), 378-386.

GLOVER, K. A., HANSEN, M. M., LIEN, S., ALS, T. D., HØYHEIM, B. & SKAALA, Ø. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics,* 11**,** 2.

GNIRKE, A., MELNIKOV, A., MAGUIRE, J., ROGOV, P., LEPROUST, E. M., BROCKMAN, W., FENNELL, T., GIANNOUKOS, G., FISHER, S., RUSS, C., GABRIEL, S., JAFFE, D. B., LANDER, E. S. & NUSBAUM, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology,* 27**,** 182-189.

GOTTELLI, D., WANG, J., BASHIR, S. & DURANT, S. M. 2007. Genetic analysis reveals promiscuity among female cheetahs. *Proceedings of the Royal Society B: Biological Sciences,* 274**,** 1993-2001.

GOUDET, J., KAY, T. & WEIR, B. S. 2018. How to estimate kinship. *Molecular Ecology,* 27**,** 4121-4135.

GOUDET, J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes,* 5**,** 184-186.

GRAF, W. L. 2006. Downstream hydrologic and geomorphic effects of large dams on American rivers. *Geomorphology,* 79**,** 336-360.

GRAHAM, C. F., GLENN, T. C., MCARTHUR, A. G., BOREHAM, D. R., KIERAN, T., LANCE, S., MANZON, R. G., MARTINO, J. A., PIERSON, T., ROGERS, S. M., WILSON, J. Y. & SOMERS, C. M. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources,* 15**,** 1304-1315.

GRINKOV, V. G., BAUER, A., GASHKOV, S. I., STERNBERG, H. & WINK, M. 2018. Diversity of social-genetic relationships in the socially monogamous pied flycatcher (Ficedula hypoleuca) breeding in Western Siberia. *PeerJ,* 6**,** e6059.

GROVES, C. P., COTTERILL, F. P. D., GIPPOLITI, S., ROBOVSKY, J., ROOS, C., TAYLOR, P. J. & ZINNER, D. 2017. Species definitions and conservation: a review and case studies from African mammals. *Conservation Genetics,* 18**,** 1247-1256.

GROVES, C. P. (1972). Ceratotherium simum. *Mammalian Species*(8), 1-6.

GROVES, C. P., FERNANDO, P., & ROBOVSKY, J. (2010). The sixth rhino: a taxonomic re-assessment of the critically endangered northern white rhinoceros. *PLoS One, 5*(4), e9703.

GUERIER, A. S., BISHOP, J. M., CRAWFORD, S. J., SCHMIDT-KÜNTZEL, A. & STRATFORD, K. J. 2012. Parentage analysis in a managed free ranging population of southern white rhinoceros: genetic diversity, pedigrees and management. *Conservation genetics,* 13**,** 811-822.

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-1075.

HADDAD, N. M., BRUDVIG, L. A., CLOBERT, J., DAVIES, K. F., GONZALEZ, A., HOLT, R. D., LOVEJOY, T. E., SEXTON, J. O., AUSTIN, M. P., COLLINS, C. D., COOK, W. M., DAMSCHEN, E. I., EWERS, R. M., FOSTER, B. L., JENKINS, C. N., KING, A. J., LAURANCE, W. F., LEVEY, D. J., MARGULES, C. R., MELBOURNE, B. A., NICHOLLS, A. O., ORROCK, J. L., SONG, D.-X. & TOWNSHEND, J. R. 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances,* 1**,** e1500052.

HADFIELD, J. D., RICHARDSON, D. S. & BURKE, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology,* 15**,** 3715-3730.

HALBERT, N. D. & DERR, J. N. 2006. A Comprehensive Evaluation of Cattle Introgression into US Federal Bison Herds. *Journal of Heredity,* 98**,** 1-12.

HAN, E., SINSHEIMER, J. S. & NOVEMBRE, J. 2013. Characterizing Bias in Population Genetic Inferences from Low-Coverage Sequencing Data. *Molecular Biology and Evolution,* 31**,** 723-735.

HANGHØJ, K., MOLTKE, I., ANDERSEN, P. A., MANICA, A. & KORNELIUSSEN, T. S. 2019. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience,* 8.

HANSSON, B. & WESTERBERG, L. 2002. On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology,* 11**,** 2467-2474.

HARLEY, E. H., DE WAAL, M., MURRAY, S. & O'RYAN, C. 2016. Comparison of whole mitochondrial genome sequences of northern and southern white rhinoceroses (*Ceratotherium simum*): the conservation consequences of species definitions. *Conservation Genetics***,** 1285-1291.

HARPER, C., LUDWIG, A., CLARKE, A., MAKGOPELA, K., YURCHENKO, A., GUTHRIE, A., DOBRYNIN, P., TAMAZIAN, G., EMSLIE, R., VAN HEERDEN, M., HOFMEYR, M., POTTER, R., ROETS, J., BEYTELL, P., OTIENDE, M., KARIUKI, L., DU TOIT, R., ANDERSON, N., OKORI, J., ANTONIK, A., KOEPFLI, K.-P., THOMPSON, P. & O'BRIEN, S. J. 2018. Robust forensic matching of confiscated horns to individual poached African rhinoceros. *Current Biology,* 28**,** R13-R14.

HARPER, C. K., VERMEULEN, G. J., CLARKE, A. B., JACOBUS, I. & GUTHRIE, A. J. 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

HARRIS, W. C. 1838. *Narrative of an expedition into Southern Africa, during the years 1836, and 1837, from the Cape of Good Hope, through the territories of the Chief Moselekatse, to the Tropic of Capricorn, with a sketch of the recent emigration of the border colonists, and a zoological appendix*, American Mission Press.

HARRISON, H. B., SAENZ-AGUDELO, P., PLANES, S., JONES, G. P. & BERUMEN, M. L. 2013. Relative accuracy of three common methods of parentage analysis in natural populations. *Molecular Ecology,* 22**,** 1158-1170.

HARTHOORN, A. M. 1962. Capture of the white (square-lipped) rhinoceros *Ceratotherium Simum Simum (Burchell)* with the use of drug immobilization technique. *Canadian Journal of Comparative Medicine and Veterinary Science,* 26**,** 203-208.

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.

HAWKEN, R. J., BARRIS, W. C., MCWILLIAM, S. M. & DALRYMPLE, B. P. 2004. An interactive bovine in silico SNP database (IBISS). *Mammalian Genome,* 15**,** 819-827.

HEDRICK, P. W. (2009). Conservation Genetics and North American Bison (Bison bison). *Journal of Heredity, 100*(4), 411-420.

HELYAR, S. J., HEMMER-HANSEN, J., BEKKEVOLD, D., TAYLOR, M. I., OGDEN, R., LIMBORG, M. T., NIELSEN, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular ecology resources, 11*(s1), 123-136.

HENKEL, J. R., JONES, K. L., HEREFORD, S. G., SAVOIE, M. L., LEIBO, S. P. & HOWARD, J. J. 2012. Integrating microsatellite and pedigree analyses to facilitate the captive management of the endangered Mississippi sandhill crane (Grus canadensis pulla). *Zoo biology,* 31**,** 322-335.

HERMES, R., HILDEBRANDT, T. B., WALZER, C., GÖRITZ, F., PATTON, M. L., SILINSKI, S., ANDERSON, M. J., REID, C. E., WIBBELT, G. & TOMASOVA, K. 2006. The effect of long non-reproductive periods on the genital health in captive female white rhinoceroses (*Ceratotherium simum simum, Cs cottoni*). *Theriogenology,* 65**,** 1492-1515.

HILDEBRANDT, T. B., HERMES, R., COLLEONI, S., DIECKE, S., HOLTZE, S., RENFREE, M. B., GALLI, C. (2018). Embryos and embryonic stem cells from the white rhinoceros. *Nature Communications, 9*(1), 2589.

HILLMAN-SMITH, A. K. K., OWEN-SMITH, N., ANDERSON, J. L., HALL-MARTIN, A. J., & SELALADI, J. P. (1986). Age estimation of the White rhinoceros (Ceratotherium simum). *Journal of Zoology, 210*(3), 355-377.

HITCHINS, P. M. 1992. An aerial census of black and square-lipped rhinoceros in Northern Botswana. Botswana: Department of Wildlife and National Parks of Botswana.

HOBAN, S., BERTORELLE, G., & GAGGIOTTI, O. E. (2012). Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics, 13*(2), 110-122.

HOF, A. R. & ALLEN, A. M. 2019. An uncertain future for the endemic Galliformes of the Caucasus. *Science of The Total Environment,* 651**,** 725-735.

HOFFBERG, S. L., KIERAN, T. J., CATCHEN, J. M., DEVAULT, A., FAIRCLOTH, B. C., MAURICIO, R. & GLENN, T. C. 2016. RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular ecology resources,* 16**,** 1264-1278.

HOFFMAN, J. I., TUCKER, R., BRIDGETT, S. J., CLARK, M. S., FORCADA, J. & SLATE, J. 2012. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the Antarctic fur seal. *Molecular Ecology Resources,* 12**,** 861-872.

HOGG, C., SRB, C. & HOCKLEY, J. 2013. Annual Report for the DPIPWE-ZAA Tasmanian devil insurance population. *Zoo and Aquarium Association***,** 42.

HOGG, C. J., LEE, A. V., SRB, C. & HIBBARD, C. 2017. Metapopulation management of an Endangered species with limited genetic diversity in the presence of disease: the Tasmanian devil Sarcophilus harrisii. *International Zoo Yearbook,* 51**,** 137-153.

HOHENLOHE, P. A., AMISH, S. J., CATCHEN, J. M., ALLENDORF, F. W. & LUIKART, G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources,* 11**,** 117-122.

HOLMAN, L. E., GARCIA DE LA SERRANA, D., ONOUFRIOU, A., HILLESTAD, B. & JOHNSTON, I. A. 2017. A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture,* 476**,** 59-64.

HOUCK, M. L., RYDER, O. A., VAHALA, J., KOCK, R. A., & OOSTERHUIS, J. E. (1994). Diploid Chromosome-Number and Chromosomal Variation in the White Rhinoceros (Ceratotherium-Simum). *Journal of Heredity, 85*(1), 30-34.

HOULDEN, B. A., ENGLAND, P., & SHERWIN, W. B. (1996). Paternity Exclusion in Koalas Using Hypervariable Microsatellites. *Journal of Heredity, 87*(2), 149-152.

HOUSTON, R. D., TAGGART, J. B., CÉZARD, T., BEKAERT, M., LOWE, N. R., DOWNING, A., TALBOT, R., BISHOP, S. C., ARCHIBALD, A. L., BRON, J. E., PENMAN, D. J., DAVASSI, A., BREW, F., TINCH, A. E., GHARBI, K. & HAMILTON, A. 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). *BMC Genomics,* 15**,** 90.

HOW, R. A., SPENCER, P. B. S. & SCHMITT, L. H. 2009. Island populations have high conservation value for northern Australia's top marsupial predator ahead of a threatening process. *Journal of Zoology,* 278**,** 206-217.

HOWRIGAN, D. P., SIMONSON, M. A. & KELLER, M. C. 2011. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics,* 12**,** 460.

HUGHES, J. B., DAILY, G. C. & EHRLICH, P. R. 1997. Population diversity: its extent and extinction. 278**,** 689-692.

HUISMAN, J. 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular ecology resources,* 17**,** 1009-1024.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. 2014. Climate Change 2014 Synthesis Report. Retrieved from https://www.ipcc.ch/site/assets/uploads/2018/02/SYR_AR5_FINAL_full.pdf

INTERNATIONAL SOCIETY FOR ANIMAL GENETICS. 2012. ISAG guidelines for cattle parentage verification based on SNP markers. *ISAG conference 2012.* Retrieved from https://www.isag.us/Docs/Guideline-for-cattle-SNP-use-for-parentage-2012.pdf

INTERNATIONAL UNION FOR CONSEVATION OF NATURE 2012. The IUCN Red List of Threatened Species: Ceratotherium simum. URL: http://dx.doi.org/10.2305/IUCN.UK.2012.RLTS.T4185A16980466.en

JACKMAN, S. D., VANDERVALK, B. P., MOHAMADI, H., CHU, J., YEO, S., HAMMOND, S. A., JAHESH, G., KHAN, H., COOMBE, L., WARREN, R. L. & BIROL, I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome research,* 27**,** 768-777.

JACQUARD, A. 1972. Genetic Information Given by a Relative. *Biometrics,* 28**,** 1101-1114.

JAN, C. & FUMAGALLI, L. 2016. Polymorphic DNA microsatellite markers for forensic individual identification and parentage analyses of seven threatened species of parrots (family Psittacidae). *PeerJ,* 4**,** e2416.

JARNE, P. & LAGODA, P. J. L. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution,* 11**,** 424-429.

JEHLE, R., WILSON, G. A., ARNTZEN, J. W. & BURKE, T. 2005. Contemporary gene flow and the spatio-temporal genetic structure of subdivided newt populations (Triturus cristatus, T. marmoratus). *Journal of Evolutionary Biology,* 18**,** 619-628.

JONES, M. R. & GOOD, J. M. 2016. Targeted capture in evolutionary and ecological genomics. *Molecular Ecology,* 25**,** 185-202.

JONES, A. G. & ARDREN, W. R. 2003. Methods of parentage analysis in natural populations. *Molecular Ecology,* 12**,** 2511-2523.

JONES, A. G., SMALL, C. M., PACZOLT, K. A. & RATTERMAN, N. L. 2010. A practical guide to methods of parentage analysis. *Mol Ecol Resour,* 10**,** 6-30.

JOST, L. O. U. 2008. GST and its relatives do not measure differentiation. *Molecular ecology,* 17**,** 4015-4026.

JOST, L., ARCHER, F., FLANAGAN, S., GAGGIOTTI, O., HOBAN, S., & LATCH, E. (2018). Differentiation measures for conservation genetics. *Evolutionary Applications, 11*(7), 1139-1148.

KAISER, S. A., TAYLOR, S. A., CHEN, N., SILLETT, T. S., BONDRA, E. R. & WEBSTER, M. S. 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular ecology resources,* 17**,** 183-193.

KALINOWSKI, S. T., TAPER, M. L. & MARSHALL, T. C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology,* 16**,** 1099-1106.

KARDOS, M., ÅKESSON, M., FOUNTAIN, T., FLAGSTAD, Ø., LIBERG, O., OLASON, P., SAND, H., WABAKKEN, P., WIKENROS, C. & ELLEGREN, H. 2018. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution,* 2**,** 124-131.

Katzner, T., Johnson, J. A., Evans, D. M., Garner, T. W. J., Gompper, M. E., Altwegg, R., Pettorelli, N. (2013). Challenges and opportunities for animal conservation from renewable energy development. *Animal Conservation, 16*(4), 367-369.

KELLER, L. F. & WALLER, D. M. 2002. Inbreeding effects in wild populations. *Trends in Ecology & Evolution,* 17**,** 230-241.

KIDEGHESHO, J. R. 2016. Reversing the trend of wildlife crime in Tanzania: challenges and opportunities. *Biodiversity and Conservation,* 25**,** 427-449.

KIERULFF, M. C. M., RUIZ-MIRANDA, C. R., DE OLIVEIRA, P. P., BECK, B. B., MARTINS, A., DIETZ, J. M., RAMBALDI, D. M. & BAKER, A. J. 2012. The Golden lion tamarin Leontopithecus rosalia: a conservation success story. *International Zoo Yearbook,* 46**,** 36-45.

KINGDON, J. & HOFFMANN, M. 2013. Mammals of Africa. Volume V: carnivores, pangolins, equids and rhinoceroses. Bloomsbury Publishing, London.

KOLLIAS, S., POORTVLIET, M., SMOLINA, I. & HOARAU, G. 2015. Low cost sequencing of mitogenomes from museum samples using baits capture and Ion Torrent. *Conservation Genetics Resources,* 7**,** 345-348.

KOMOROSKE, L. M., MILLER, M. R., O'ROURKE, S. M., STEWART, K. R., JENSEN, M. P. & DUTTON, P. H. 2019. A versatile Rapture (RAD-Capture) platform for genotyping marine turtles. *Molecular Ecology Resources,* 19**,** 497-511.

KONOVALOV, D. A., MANNING, C. & HENSHAW, M. T. 2004. kingroup: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Notes,* 4**,** 779-782.

KORNELIUSSEN, T. S., ALBRECHTSEN, A. & NIELSEN, R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics,* 15**,** 356.

KORNELIUSSEN, T. S. & MOLTKE, I. 2015. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics,* 31**,** 4009-4011.

KOYANO, H., SERBEZOV, D., KISHINO, H. & SCHWEDER, T. 2013. Fractional parentage analysis and a scale-free reproductive network of brown trout. *Journal of Theoretical Biology,* 336**,** 18-35.

KRAWCZAK, M. 1999. Informativity assessment for biallelic single nucleotide polymorphisms. *ELECTROPHORESIS,* 20**,** 1676-1681.

KRETZSCHMAR, P. Population growth, sex ratio and reproduction of a natural living population of white rhinoceros. 2001. 196-201.

KRETZSCHMAR, P., AULD, H., BOAG, P., GANSLOSSE, U., SCOTT, C., VAN COEVERDEN DE GROOT, P. J. & COURTIOL, A. 2019. Mate choice, reproductive success and inbreeding in white rhinoceros: New insights for conservation management. *Evolutionary Applications,* n00**,** 1-16.

KRUMMENACHER, T. S. & ZSCHOKKE, S. 2007. Inbreeding and outbreeding in African rhinoceros species. *Pachyderm,* 42**,** 108-115.

KRUTOVSKY, K. V., BURCZYK, J., CHYBICKI, I., FINKELDEY, R., PYHÄJÄRVI, T. & ROBLEDO-ARNUNCIO, J. J. 2012. Gene Flow, Spatial Structure, Local Adaptation, and Assisted Migration in Trees. *In:* SCHNELL, R. J. & PRIYADARSHAN, P. M. (eds.) *Genomics of Tree Crops.* New York, NY: Springer New York.

LABUSCHAGNE, C., DALTON, D. L., GROBLER, J. P. & KOTZÉ, A. 2017. SNP discovery and characterisation in white rhino (*Ceratotherium simum*) with application to parentage assignment. *Genetics and molecular biology,* 40**,** 84-92.

LABUSCHAGNE, C., KOTZE, A., GROBLER, J. P. & DALTON, D. L. 2013. A targeted gene approach to SNP discovery in the white rhino (*Ceratotherium simum*). *Conservation Genetics Resources,* 5**,** 265-266.

LABUSCHAGNE, C., KOTZE, 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**,** 53-56.

LACY, R. C. 1992. The effects of inbreeding on isolated populations: are minimum viable population sizes predictable? *Conservation biology.* Springer.

LACY, R. C. 1995. Clarification of genetic terms and their use in the management of captive populations. *Zoo Biology,* 14**,** 565-577.

LACY, R. C., MALO, A. F. & ALAKS, G. 2018. Maintenance of genetic variation in quantitative traits of a woodland rodent during generations of captive breeding. *Conservation Genetics***,** 1-14.

LANG, P. L. M., WEIS, C. L., KERSTEN, S., LATORRE, S. M., NAGEL, S., NICKEL, B., MEYER, M. & BURBANO, H. A. 2020. Hybridization ddRAD-sequencing for population genomics of nonmodel plants using highly degraded historical specimen DNA. *Molecular Ecology Resources,* n/a.

LE ROEX, N., PAXTON, M., ADENDORFF, J., FERREIRA, S. & O'RIAIN, M. J. 2018. Starting small: long-term consequences in a managed large-mammal population. *Journal of Zoology,* 306**,** 95-100.

LENSTRA, J. A. (2006). Marker-assisted conservation of European cattle breeds: an evaluation. *Animal genetics, 37*(5), 475-481.

LI, S., LI, B., CHENG, C., XIONG, Z., LIU, Q., LAI, J., CAREY, H. V., ZHANG, Q., ZHENG, H., WEI, S., ZHANG, H., CHANG, L., LIU, S., ZHANG, S., YU, B.,

ZENG, X., HOU, Y., NIE, W., GUO, Y., CHEN, T., HAN, J., WANG, J., WANG, J., CHEN, C., LIU, J., STAMBROOK, P. J., XU, M., ZHANG, G., GILBERT, M. T. P., YANG, H., JARVIS, E. D., YU, J. & YAN, J. 2014. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biology,* 15**,** 557.

LI, H. 2011. Improving SNP discovery by base alignment quality. *Bioinformatics,* 27**,** 1157-1158.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics,* 25**,** 1754-1760.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LI, C., WEEKS, D., & CHAKRAVARTI, A. (1993). Similarity of DNA fingerprints due to chance and relatedness. *Human heredity, 43*(1), 45-52.

LINKLATER, W. L. 2007. Translocation reverses birth sex ratio bias depending on its timing during gestation: evidence for the action of two sex-allocation mechanisms. *Reproduction, Fertility and Development,* 19**,** 831-839.

LIU, S., ZHOU, Z., LU, J., SUN, F., WANG, S., LIU, H., JIANG, Y., KUCUKTAS, H., KALTENBOECK, L., PEATMAN, E. & LIU, Z. 2011. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics,* 12**,** 53.

LÓPEZ-DOMINGO, F. J., FLORIDO, J. P., RUEDA, A., DOPAZO, J. & SANTOYO-LOPEZ, J. 2014. ngsCAT: a tool to assess the efficiency of targeted enrichment sequencing. *Bioinformatics,* 30**,** 1767-1768.

LUDES, B., PFITZINGER, H. & MANGIN, P. 1993. DNA Fingerprinting from Tissues After Variable Postmortem Periods. *Journal of Forensic Sciences,* 38**,** 686-690.

LYNCH, M., & RITLAND, K. (1999). Estimation of Pairwise Relatedness With Molecular Markers. *Genetics, 152*(4), 1753-1766.

MABLE, B. K. 2019. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation Genetics,* 20**,** 89-100.

MACGREGOR, J. 1989. The paradoxes of wildlife conservation in Africa. *Africa Insight,* 19(4)**,** 201-212.

MALOMANE, D. K., REIMER, C., WEIGEND, S., WEIGEND, A., SHARIFI, A. R. & SIMIANER, H. 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. *BMC Genomics,* 19**,** 22.

MAMOOZADEH, N. R., GRAVES, J. E. & MCDOWELL, J. R. 2019. Genome-wide SNPs resolve spatiotemporal patterns of connectivity within striped marlin (Kajikia audax), a broadly distributed and highly migratory pelagic species. *Evolutionary Applications,* 00**,** 1-22.

MAROSO, F., HILLEN, J. E. J., PARDO, B. G., GKAGKAVOUZIS, K., COSCIA, I., HERMIDA, M., FRANCH, R., HELLEMANS, B., VAN HOUDT, J., SIMIONATI, B., TAGGART, J. B., NIELSEN, E. E., MAES, G., CIAVAGLIA, S. A., WEBSTER, L. M. I., VOLCKAERT, F. A. M., MARTINEZ, P., BARGELLONI, L. & OGDEN, R. 2018. Performance and precision of double digestion RAD (ddRAD) genotyping in large multiplexed datasets of marine fish species. *Marine Genomics,* 39**,** 64-72.

MARSHALL, T., SLATE, J., KRUUK, L. & PEMBERTON, J. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular ecology,* 7**,** 639-655.

MARTINEZ-CRUZ, B., GODOY, J. A. & NEGRO, J. J. 2007. Population fragmentation leads to spatial and temporal genetic structure in the endangered Spanish imperial eagle. *Molecular Ecology,* 16**,** 477-486.

MCCUE, M. E., BANNASCH, D. L., PETERSEN, J. L., GURR, J., BAILEY, E., BINNS, M. M., DISTL, O., GUÉRIN, G., HASEGAWA, T., HILL, E. W., LEEB, T., LINDGREN, G., PENEDO, M. C. T., RØED, K. H., RYDER, O. A., SWINBURNE, J. E., TOZAKI, T., VALBERG, S. J., VAUDIN, M., LINDBLAD-TOH, K., WADE, C. M. & MICKELSON, J. R. 2012. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLOS Genetics,* 8**,** e1002451.

MCGUIRE, J. A., COTORAS, D. D., O'CONNELL, B., LAWALATA, S. Z. S., WANG-CLAYPOOL, C. Y., STUBBS, A., HUANG, X., WOGAN, G. O. U., HYKIN, S. M. & REILLY, S. B. 2018. Squeezing water from a stone: high-throughput sequencing from a 145-year old holotype resolves (barely) a cryptic species problem in flying lizards. *PeerJ,* 6**,** e4470.

MCTAVISH, E. J. & HILLIS, D. M. 2015. How do SNP ascertainment schemes and population demographics affect inferences about population history? *BMC Genomics,* 16**,** 266.

MEAGHER, T. R. & THOMPSON, E. 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. *Theoretical Population Biology,* 29**,** 87-106.

MERENLENDER, A. M., WOODRUFF, D. S., RYDER, O. A., KOCK, R., & VAHALA, J. (1989). Allozyme variation and differentiation in African and Indian rhinoceroses. *J Hered, 80*(5), 377-382.

MEUWISSEN, T. H. E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science,* 75**,** 934-940.

MEYER, M. & KIRCHER, M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols,* 2010**,** pdb. prot5448.

MICKELBERG, J. L. 2011. *Understanding and managing isolation in a fragmented population of golden lion tamarins, Leontopithecus rosalia.* George Mason University.

MILLER, M. R., DUNHAM, J. P., AMORES, A., CRESKO, W. A. & JOHNSON, E. A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research,* 17**,** 240-248.

MILLER, S. M., DRUCE, D. J., DALTON, D. L., HARPER, C. K., KOTZE, A., PACKER, C., SLOTOW, R. & BLOOMER, P. 2019. Genetic rescue of an isolated African lion population. *Conservation Genetics*.

MILLER, P. S. 1995. Selective Breeding Programs for Rare Alleles: Examples from the Przewalski's Horse and California Condor Pedigrees. *Conservation Biology,* 9**,** 1262-1273.

MILLER, S. M., HARPER, C. K., BLOOMER, P., HOFMEYR, J. & FUNSTON, P. J. 2015. Fenced and Fragmented: Conservation Value of Managed Metapopulations. *PLOS ONE,* 10**,** e0144605.

MILLS, L. S., & ALLENDORF, F. W. (1996). The One-Migrant-per-Generation Rule in Conservation and Management. *Conservation Biology, 10*(6), 1509-1518.

MKUMBUKWA, A. R. 2008. The evolution of wildlife conservation policies in Tanzania during the colonial and post-independence periods. *Development Southern Africa,* 25**,** 589-600.

MOODLEY, Y., ISA-RITA, M. R., JAN, R., DESIRÉ, L. D., ANTOINETTE, K., STEVE, S., JAN, S., OLIVER, A. R., ROBERT, H., CHRIS, W. & MICHAEL, W. B. 2018. Contrasting evolutionary history, anthropogenic declines and genetic contact in the northern and southern white rhinoceros *(Ceratotherium simum)*. *Proceedings of the Royal Society B: Biological Sciences,* 285**,** 20181567.

MOODLEY, Y., RUSSO, I.-R. M., DALTON, D. L., KOTZÉ, A., MUYA, S., HAUBENSAK, P., BÁLINT, B., MUNIMANDA, G. K., DEIMEL, C., SETZER, A., DICKS, K., HERZIG-STRASCHIL, B., KALTHOFF, D. C., SIEGISMUND, H. R., ROBOVSKÝ, J., O'DONOGHUE, P. & BRUFORD, M. W. 2017. Extinctions, genetic erosion and conservation options for the black rhinoceros (Diceros bicornis). *Scientific Reports,* 7**,** 41417.

MORAES, A. M., RUIZ-MIRANDA, C. R., RIBEIRO, M. C., GRATIVOL, A. D., CARVALHO, C. D. S., DIETZ, J. M., KIERULFF, M. C. M., FREITAS, L. A. & GALETTI, P. M. 2017. Temporal genetic dynamics of reintroduced and translocated populations of the endangered golden lion tamarin (Leontopithecus rosalia). *Conservation Genetics,* 18**,** 995-1009.

MORAN, P., TEEL, D. J., LAHOOD, E. S., DRAKE, J. & KALINOWSKI, S. 2006. Standardising multi-laboratory microsatellite data in Pacific salmon: an historical view of the future. *Ecology of Freshwater Fish,* 15**,** 597-605.

MORRISSEY, M. B. & WILSON, A. J. 2010. pedantics: an r package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Molecular Ecology Resources,* 10**,** 711-719.

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**,** 2841-2847.

NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic acids research,* 44**,** D7-D19.

NEI, M. & TAKAHATA, N. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. *Journal of Molecular Evolution,* 37**,** 240-244.

NEI, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences,* 70**,** 3321-3323.

NGUYEN, T. T. T., HAYES, B. J., & INGRAM, B. A. (2014). Genetic parameters and response to selection in blue mussel (Mytilus galloprovincialis) using a SNP-based pedigree. *Aquaculture, 420-421*, 295-301.

NIELSEN, L., MEEHAN-MEOLA, D., KILBOURN, A., & ALCIVARl-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. *Conservation Genetics, 9*(1), 239-242.

NIELSEN, R. 2004. Population genetic analysis of ascertained SNP data. *Human Genomics,* 1**,** 218.

NIELSEN, R., KORNELIUSSEN, T., ALBRECHTSEN, A., LI, Y. & WANG, J. 2012. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. *PLOS ONE,* 7**,** e37558.

NIELSEN, R., MATTILA, D. K., CLAPHAM, P. J. & PALSBØLL, P. J. 2001. Statistical Approaches to Paternity Analysis in Natural Populations and Applications to the North Atlantic Humpback Whale. *Genetics,* 157**,** 1673-1682.

NUNNEY, L. 1999. THE EFFECTIVE SIZE OF A HIERARCHICALLY STRUCTURED POPULATION. *Evolution,* 53**,** 1-10.

OGDEN, R. & LINACRE, A. 2015. Wildlife forensic science: A review of genetic geographic origin assignment. *Forensic Science International: Genetics,* 18**,** 152-159.

O'GRADY, J. J., BROOK, B. W., REED, D. H., BALLOU, J. D., TONKYN, D. W., & FRANKHAM, R. (2006). Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation, 133*(1), 42-51.

OWEN-SMITH, N. 1977. On territoriality in ungulates and an evolutionary model. *The Quarterly Review of Biology,* 52**,** 1-38.

OWEN-SMITH, R. N. 1975. The social ethology of the white rhinoceros *Ceratotberium simum* (*Burchell 1817**). *Zeitschrift für Tierpsychologie,* 38**,** 337-384.

OZKAN, E., SOYSAL, M. I., OZDER, M., KOBAN, E., SAHIN, O. & TOGAN, İ. 2009. Evaluation of parentage testing in the Turkish Holstein population based on 12 microsatellite loci. *Livestock Science,* 124**,** 101-106.

PANETTO, J. C. D. C., MACHADO, M. A., DA SILVA, M. V. G. B., BARBOSA, R. S., DOS SANTOS, G. G., LEITE, R. D. M. H. & PEIXOTO, M. G. C. D. 2017. Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. *Livestock Science,* 204**,** 33-38.

PANG-CHING, J. M., PAXTON, K. L., PAXTON, E. H., PACK, A. A., & HART, P. J. (2018). The effect of isolation, fragmentation, and population bottlenecks on song structure of a Hawaiian honeycreeper. *Ecology and evolution, 8*(4), 2076-2087.

PARIS, J. R., STEVENS, J. R. & CATCHEN, J. M. 2017. Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution,* 8**,** 1360-1373.

PARRA, G. J., CAGNAZZI, D., JEDENSJÖ, M., ACKERMANN, C., FRERE, C., SEDDON, J., NIKOLIC, N. & KRÜTZEN, M. 2018. Low genetic diversity, limited gene flow and widespread genetic bottleneck effects in a threatened dolphin species, the Australian humpback dolphin. *Biological Conservation,* 220**,** 192-200.

PAVLOVA, A., BEHEREGARAY, L. B., COLEMAN, R., GILLIGAN, D., HARRISSON, K. A., INGRAM, B. A., KEARNS, J., LAMB, A. M., LINTERMANS, M., LYON, J., NGUYEN, T. T. T., SASAKI, M., TONKIN, Z., YEN, J. D. L. & SUNNUCKS, P. 2017. Severe consequences of habitat fragmentation on genetic diversity of an endangered Australian freshwater fish: A call for assisted gene flow. *Evolutionary Applications,* 10**,** 531-550.

PEMBERTON, J. M. 2008. Wild pedigrees: the way forward. *Proceedings of the Royal Society of London B: Biological Sciences,* 275**,** 613-621.

PEMBERTON, J. M., SLATE, J., BANCROFT, D. R. & BARRETT, J. A. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology,* 4**,** 249-252.

PETERSON, B. K., WEBER, J. N., KAY, E. H., FISHER, H. S. & HOEKSTRA, H. E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one,* 7**,** e37135.

PETRIE, M. & KEMPENAERS, B. 1998. Extra-pair paternity in birds: explaining variation between species and populations. *Trends in Ecology & Evolution,* 13**,** 52-58.

PÉREZ-RODRÍGUEZ, A., KHIMOUN, A., OLLIVIER, A., ERAUD, C., FAIVRE, B. & GARNIER, S. 2018. Habitat fragmentation, not habitat loss, drives the prevalence of blood parasites in a Caribbean passerine. *Ecography,* 41**,** 1835-1849.

PERTOLDI, C., WÓJCIK, J. M., TOKARSKA, M., KAWAŁKO, A., KRISTENSEN, T. N., LOESCHCKE, V., GREGERSEN, V. R., COLTMAN, D., WILSON, G. A., RANDI, E., HENRYON, M. & BENDIXEN, C. 2010. Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics,* 11**,** 627-634.

PIENAAR, U. D. V. 1970. The recolonisation history of the square-lipped (white) rhinoceros *Ceratotherium Simum Simum (Burchell)* in the Kruger National Park (October 1961 - November 1969). *Koedoe,* 13**,** 14.

PIENAAR, D. J., BOTHMA, J. d. P., & THERON, G. K. (1993). Landscape preference of the white rhinoceros in the central and northern Kruger National Park. *1993, 36*(1), 7.

PITMAN, R. T., FATTEBERT, J., WILLIAMS, S. T., WILLIAMS, K. S., HILL, R. A., HUNTER, L. T. B., SLOTOW, R. & BALME, G. A. 2017. The conservation costs of game ranching. *Conservation Letters,* 10**,** 403-413.

PLAYER, I. 1967. Translocation of white rhinoceros in South Africa. *Oryx,* 9**,** 137-150.

POPLIN, R., RUANO-RUBIO, V., DEPRISTO, M. A., FENNELL, T. J., CARNEIRO, M. O., VAN DER AUWERA, G. A., KLING, D. E., GAUTHIER, L. D., LEVY-MOONSHINE, A., ROAZEN, D., SHAKIR, K., THIBAULT, J., CHANDRAN, S., WHELAN, C., LEK, M., GABRIEL, S., DALY, M. J., NEALE, B., MACARTHUR, D. G. & BANKS, E. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv***,** 201178.

PORTIK, D. M., SMITH, L. L. & BI, K. 2016. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources,* 16**,** 1069-1083.

PROCTOR, M. F., MCLELLAN, B. N., STROBECK, C. & BARCLAY, R. M. R. 2005. Genetic analysis reveals demographic fragmentation of grizzly bears yielding vulnerably small populations. *Proceedings of the Royal Society B: Biological Sciences,* 272**,** 2409-2416.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. W., DALY, M. J. & SHAM, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics,* 81**,** 559-575.

PURISOTAYO, T., JONSSON, N. N., MABLE, B. K. & VERREYNNE, F. J. 2019. Combining molecular and incomplete observational data to inform management of southern white rhinoceros (*Ceratotherium simum simum*). *Conservation Genetics*, 20, 639-652.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

QUINN, C. B., ALDEN, P. B. & SACKS, B. N. 2019. Noninvasive Sampling Reveals Short-Term Genetic Rescue in an Insular Red Fox Population. *Journal of Heredity,* 110**,** 559-576.

R Core TEAM. 2013. R: A language and environment for statistical computing.

RACHLOW, J. L. & BERGER, J. 1998. Reproduction and population density: trade-offs for the conservation of rhinos in situ. *Animal Conservation,* 1**,** 101-106.

RACHLOW, J. L., BERKELEY, E. V. & BERGER, J. 1998. Correlates of male mating strategies in white rhinos (*Ceratotherium simum*). *Journal of Mammalogy,* 79**,** 1317-1324.

RALLS, K., BALLOU, J. D. & TEMPLETON, A. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology,* 2**,** 185-193.

RALLS, K. & BALLOU, J. D. 2004. Genetic Status and Management of California Condors. *The Condor,* 106**,** 215-228.

RALLS, K., BALLOU, J. D., DUDASH, M. R., ELDRIDGE, M. D. B., FENSTER, C. B., LACY, R. C., FRANKHAM, R. (2018). Call for a Paradigm Shift in the Genetic Management of Fragmented Populations. *Conservation Letters, 11*(2), e12412.

REY-IGLESIA, A., GOPALAKRISHAN, S., CARØE, C., ALQUEZAR-PLANAS, D. E., AHLMANN NIELSEN, A., RÖDER, T., BRUHN PEDERSEN, L., NÆSBORG-NIELSEN, C., SINDING, M.-H. S., FREDENSBORG RATH, M., LI, Z., PETERSEN, B., GILBERT, M. T. P., BUNCE, M., MOURIER, T. & HANSEN, A. J. 2019. MobiSeq: De novo SNP discovery in model and non-model species through sequencing the flanking region of transposable elements. *Molecular Ecology Resources,* 19**,** 512-525.

RAYMOND, M. & ROUSSET, F. 1995. An Exact Test for Population Differentiation. *Evolution,* 49**,** 1280-1283.

REID, C. E., VERSTEEGE, L., LUDWIG, A. & HERMES, R. 2012. Analysis of the European captive southern white rhinoceros *Ceratotherium simum simum* population using the International Studbook for the African White Rhinoceros (2001–2004) and integrating reproductive health findings. *International Zoo Yearbook,* 46**,** 209-220.

RIESTER, M., STADLER, P. F. & KLEMM, K. 2009. FRANz: reconstruction of wild multi-generation pedigrees. *Bioinformatics,* 25**,** 2134-2139.

ROCHETTE, N. C. & CATCHEN, J. M. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols,* 12**,** 2640.

RODRÍGUEZ, A. & DELIBES, M. 2003. Population fragmentation and extinction in the Iberian lynx. *Biological Conservation,* 109**,** 321-331.

RODRIGUEZ-RODRIGUEZ, M. A., GASCA-PINEDA, J., MEDELLIN, R. A. & EGUIARTE, L. E. 2015. Analysis of genetic diversity of bighorn sheep (*Ovis canadensis*) from Mexican populations. *Journal of Mammalogy,* 96**,** 473-480.

ROOKMAAKER, K. & ANTOINE, P.-O. 2012. New maps representing the historical and recent distribution of the African species of rhinoceros: *Diceros bicornis*, *Ceratotherium simum* and *Ceratotherium cottoni*. *Pachyderm,* 52**,** 91-96.

ROS-FREIXEDES, R., WHALEN, A., CHEN, C.-Y., GORJANC, G., HERRING, W. O., MILEHAM, A. J. & HICKEY, J. M. 2019. Accuracy of whole-genome sequence imputation using hybrid peeling in large pedigreed livestock populations. *bioRxiv***,** 771576.

ROSENBLUM, E. B. & NOVEMBRE, J. 2007. Ascertainment Bias in Spatially Structured Populations: A Case Study in the Eastern Fence Lizard. *Journal of Heredity,* 98**,** 331-336.

ROTH, T. L., & SWANSON, W. F. (2018). From petri dishes to politics – a multi-pronged approach is essential for saving endangered species. *Nature Communications, 9*(1), 2588.

ROTH, T. L., SCHOOK, M. W. & STOOPS, M. A. 2018. Monitoring and controlling ovarian function in the rhinoceros. *Theriogenology,* 109**,** 48-57.

ROUSSET, F. (2002). Inbreeding and relatedness coefficients: what do they measure? *Heredity, 88*(5), 371-380.

RYDER, O. A., FRIESE, C., GREELY, H. T., SANDLER, R., SARAGUSTY, J., DURRANT, B. S., & REDFORD, K. H. Exploring the limits of saving a subspecies: The ethics and social dynamics of restoring northern white rhinos (Ceratotherium simum cottoni). *Conservation Science and Practice*, e241.

SAMBROOK, J. & RUSSELL, D. W. 2006. Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harbor Protocols,* 2006**,** pdb. prot4455 1940-3402.

SÁNCHEZ BARREIRO, F., GOPALAKRISHNAN, S., WESTBURY, M. V., MARGARYAN, A., VIEIRA, F. G., KALTHOFF, D., TIMMONS, Z., SICHERITZ-PONTÉN, T., DALÉN, L. & RYDER, O. 2020. Historical Population Declines Prompted Significant Genomic Erosion in the Northern and Southern White Rhinoceros (Ceratotherium simum).

SARAGUSTY, J., DIECKE, S., DRUKKER, M., DURRANT, B., BEN-NUN, I. F., GALLI, C., HILDEBRANDT, T. B. (2016). Rewinding the Process of Mammalian Extinction. *Zoo Biology, 34*(4), 280-292. doi:https://doi.org/10.1002/zoo.21284

SARO, I., ROBLEDO-ARNUNCIO, J. J., GONZÁLEZ-PÉREZ, M. A. & SOSA, P. A. 2014. Patterns of pollen dispersal in a small population of the Canarian endemic palm (Phoenix canariensis). *Heredity,* 113**,** 215-223.

SAURA, M., FERNÁNDEZ, A., RODRÍGUEZ, M. C., TORO, M. A., BARRAGÁN, C., FERNÁNDEZ, A. I. & VILLANUEVA, B. 2013. Genome-Wide Estimates of Coancestry and Inbreeding in a Closed Herd of Ancient Iberian Pigs. *PLOS ONE,* 8**,** e78314.

SAWAYA, M. A., CLEVENGER, A. P. & SCHWARTZ, M. K. 2019. Demographic fragmentation of a protected wolverine population bisected by a major transportation corridor. *Biological Conservation,* 236**,** 616-625.

SCHIEGG, K., DANIELS, S. J., WALTERS, J. R., PRIDDY, J. A. & PASINELLI, G. 2006. Inbreeding in red-cockaded woodpeckers: Effects of natal dispersal distance and territory location. *Biological Conservation,* 131**,** 544-552.

SCHMIDT, D. A., CAMPBELL, N. R., GOVINDARAJULU, P., LARSEN, K. W. & RUSSELLO, M. A. 2020. Genotyping-in-Thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Molecular Ecology Resources,* 20**,** 114-124.

SCHLAEPFER, D. R., BRASCHLER, B., RUSTERHOLZ, H.-P. & BAUR, B. 2018. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. *Ecosphere,* 9**,** e02488.

SCOTT, C. A. 2008. *Microsatellite variability in four contemporary rhinoceros species: implications for conservation.* Master of Science, Queen's University.

SEEB, J. E., CARVALHO, G., HAUSER, L., NAISH, K., ROBERTS, S. & SEEB, L. W. 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol Ecol Resour,* 11 Suppl 1**,** 1-8.

SEKINO, M., SUGAYA, T., HARA, M. & TANIGUCHI, N. 2004. Relatedness inferred from microsatellite genotypes as a tool for broodstock management of Japanese flounder Paralichthys olivaceus. *Aquaculture,* 233**,** 163-172.

SELKOE, K. A., & TOONEN, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters, 9*(5), 615-629.

SHULTZ, A. 2018. *Whole-genome resequencing for population genomics (Fastqc to VCF)* [Online]. Harvard FAS Informatics. Available: https://informatics.fas.harvard.edu/whole-genome-resquencing-for-population-genomics-fastq-to-vcf.html#variantcalling [Accessed 14 April 2020 2020].

SIGG, D. P. 2006. Reduced genetic diversity and significant genetic differentiation after translocation: Comparison of the remnant and translocated populations of bridled nailtail wallabies (Onychogalea fraenata). *Conservation Genetics,* 7**,** 577-589.

SIMPSON, J. T., WONG, K., JACKMAN, S. D., SCHEIN, J. E., JONES, S. J. & BIROL, I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome research,* 19**,** 1117-1123.

SINNWELL, J. P., THERNEAU, T. M. & SCHAID, D. J. 2014. The kinship2 R package for pedigree data. *Human heredity,* 78**,** 91-93.

SMIT, A., HUBLEY, R. & GREEN, P. 2015. RepeatMasker Open-4.0. 2013–2015.

SMITH, C. T., TEMPLIN, W. D., SEEB, J. E. & SEEB, L. W. 2005. Single Nucleotide Polymorphisms Provide Rapid and Accurate Estimates of the Proportions of U.S. and Canadian Chinook Salmon Caught in Yukon River Fisheries. *North American Journal of Fisheries Management,* 25**,** 944-953.

SOMMER, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology,* 2**,** 16.

STÄDELE, V. & VIGILANT, L. 2016. Strategies for determining kinship in wild populations using genetic data. *Ecology and Evolution,* 6**,** 6107-6120.

STEVANOVIC, J., STANIMIROVIC, Z., DIMITRIJEVIC, V., STOJIC, V., FRATRIC, N., & LAZAREVIC, M. (2009). Microsatellite DNA polymorphism and its usefulness for pedigree verification in Simmental cattle from Serbia. *Acta Veterinaria, 59*(5-6), 621-631.

STRUCKEN, E. M., GUDEX, B., FERDOSI, M. H., LEE, H. K., SONG, K. D., GIBSON, J. P., KELLY, M., PIPER, E. K., PORTO-NETO, L. R., LEE, S. H. & GONDRO, C. 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. *Animal Genetics,* 45**,** 572-575.

STRUCKEN, E. M., LEE, S. H., LEE, H. K., SONG, K. D., GIBSON, J. P. & GONDRO, C. 2016. How many markers are enough? Factors influencing parentage testing in different livestock populations. *Journal of Animal Breeding and Genetics,* 133**,** 13-23.

SUCHAN, T., PITTELOUD, C., GERASIMOVA, N. S., KOSTIKOVA, A., SCHMID, S., ARRIGO, N., PAJKOVIC, M., RONIKIER, M. & ALVAREZ, N. 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PloS one,* 11**,** e0151651.

SYVÄNEN, A.-C. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics,* 2**,** 930-942.

TAUTZ, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids research, 17*(16), 6463-6471.

THERMO SCIENTIFIC, T. 2010. Nucleic Acid Thermo Scientific NanoDrop Spectrophotometers. Wilmington, DE, USA: Thermo Scientific.

THOMPSON, S., AVENT, T. & DOUGHTY, L. S. 2016. Range analysis and terrain preference of adult southern white rhinoceros (*Ceratotherium simum*) in a South African private game reserve: insights into carrying capacity and future management. *PLOS ONE,* 11**,** e0161724.

THORNTON, T., TANG, H., HOFFMANN, THOMAS J., OCHS-BALCOM, HEATHER M., CAAN, BETTE J. & RISCH, N. 2012. Estimating Kinship in Admixed Populations. *The American Journal of Human Genetics,* 91**,** 122-138.

TIENDEREN, K. V., HELSEN, P. & GALBUSERA, P. 2013. Improving the Golden-headed lion tamarin (Leontopithecus chrysomelas) ex-situ breeding schemes by establishing founder kinship through combined microsatellite maker and pedigree analysis. *The 9th International Conference on Behaviour, Physiology and Genetics of Wildlife.* Berlin.

TJIBAE, M. 2001. SADC Regional Programme for Rhino Conservation. SADC Rhino Range States and Consortium Meeting, 2001 Pilanesberg. 4 - 5.

TOCHER, M. D., FLETCHER, D., & BISHOP, P. J. (2006). A modelling approach to determine a translocation scenario for the endangered New Zealand frog Leiopelma hamiltoni. *The Herpetological Journal, 16*(2), 97-106.

TOKARSKA, M., MARSHALL, T., KOWALCZYK, R., WOJCIK, J., PERTOLDI, C., KRISTENSEN, T., LOESCHCKE, V., GREGERSEN, V. R. & BENDIXEN, C. 2009. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity,* 103**,** 326-332.

TORO, M. A., VILLANUEVA, B. & FERNÁNDEZ, J. 2014. Genomics applied to management strategies in conservation programmes. *Livestock Science,* 166**,** 48-53.

TOWNSEND, S. M. & JAMIESON, I. G. 2013. Molecular and pedigree measures of relatedness provide similar estimates of inbreeding depression in a bottlenecked population. *Journal of Evolutionary Biology,* 26**,** 889-899.

TUBBS, C. W., DURRANT, B. S. & MILNES, M. R. 2017. Reconsidering the use of soy and alfalfa in southern white rhinoceros diets. *Pachyderm***,** 135-139.

TUNSTALL, T., KOCK, R., VAHALA, J., DIEKHANS, M., FIDDES, I., ARMSTRONG, J., PATEN, B., RYDER, O. A. & STEINER, C. C. 2018. Evaluating recovery potential of the northern white rhinoceros from cryopreserved somatic cells. *Genome Research*.

VAN DER GOOT, A. C., MARTIN, G. B., MILLAR, R. P., PARIS, M. C. J. & GANSWINDT, A. 2015. Profiling patterns of fecal 20-oxopregnane concentrations during ovarian cycles in free-ranging southern white rhinoceros (*Ceratotherium simum simum*). *Animal Reproduction Science,* 161**,** 89-95.

VAN TASSELL, C. P., SMITH, T. P. L., MATUKUMALLI, L. K., TAYLOR, J. F., SCHNABEL, R. D., LAWLEY, C. T., HAUDENSCHILD, C. D., MOORE, S. S., WARREN, W. C. & SONSTEGARD, T. S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods,* 5**,** 247-252.

VANDEPUTTE, M. & HAFFRAY, P. 2014. Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in Genetics,* 5.

VANDEPUTTE, M., MAUGER, S. & DUPONT-NIVET, M. 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Molecular Ecology Notes,* 6**,** 265-267.

VELDHUIS, M. P., GOMMERS, M. I., OLFF, H., & BERG, M. P. (2018). Spatial redistribution of nutrients by large herbivores and dung beetles in a savanna ecosystem. *Journal of Ecology, 106*(1), 422-433.

VERREYNNE, F. J. 2012. *Update of the central database and studbook for all southern white rhinoceros (Cerathotherium simum simum) under private management in the*

Republic of Botswana). Master in the Philosophy in Wildlife Management, University of Pretoria.

VERVERS, C., VAN ZIJLL LANGHOUT, M., HOSTENS, M., OTTO, M., GOVAERE, J., DURRANT, B. & VAN SOOM, A. 2017. Reproductive performance parameters in a large population of game-ranched white rhinoceroses (Ceratotherium simum simum). *PloS one,* 12**,** 1932-6203.

VERVERS, C., VAN ZIJLL LANGHOUT, M., OTTO, M., GOVAERE, J. & VAN SOOM, A. 2018. *Breeding soundness and pregnancy evaluation by ultrasonography in female game-ranched southern white rhinoceroses (Ceratotherium simum simum).* Ghent University.

VIN RICHTER, W. 1973. Black and square-lipped rhinoceroses in Botswana. *Biological Conservation,* 5**,** 59-60.

WADE, T. G., RIITERS, K. H., WICKHAM, J. D., & JONES, K. B. (2003). Distribution and Causes of Global Forest Fragmentation. *Conservation Ecology, 7*(2), 7.

WALDRAM, M. S., BOND, W. J., & STOCK, W. D. (2008). Ecological Engineering by a Mega-Grazer: White Rhino Impacts on a South African Savanna. *Ecosystems, 11*(1), 101-112.

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.

WAN, Q.-H., ZHANG, P., NI, X.-W., WU, H.-L., CHEN, Y.-Y., KUANG, Y.-Y., GE, Y.-F. & FANG, S.-G. 2011. A novel HURRAH protocol reveals high numbers of monomorphic MHC class II loci and two asymmetric multi-locus haplotypes in the Père David's Deer. *PLOS ONE,* 6**,** e14518.

WANG, J. 2004. Application of the one-migrant-per-generation rule to conservation and management. *Conservation Biology,* 18**,** 332-343.

WANG, J. 2007. Parentage and sibship exclusions: higher statistical power with more family members. *Heredity,* 99**,** 205-217.

WANG, J. 2011. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources,* 11**,** 141-145.

WANG, D. G., FAN, J.-B., SIAO, C.-J., BERNO, A., YOUNG, P., SAPOLSKY, R., GHANDOUR, G., PERKINS, N., WINCHESTER, E., SPENCER, J., KRUGLYAK, L., STEIN, L., HSIE, L., TOPALOGLOU, T., HUBBELL, E., ROBINSON, E., MITTMANN, M., MORRIS, M. S., SHEN, N., KILBURN, D., RIOUX, J., NUSBAUM, C., ROZEN, S., HUDSON, T. J., LIPSHUTZ, R., CHEE, M. & LANDER, E. S. 1998. Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science,* 280**,** 1077-1082.

WANG, J. 2004a. Sibship Reconstruction From Genetic Data With Typing Errors. *Genetics,* 166**,** 1963-1979.

WANG, J. 2016. Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology,* 107**,** 4-13.

WANG, J. (2004b). Application of the one-migrant-per-generation rule to conservation and management. *Conservation Biology, 18*(2), 332-343.

WEBER, D. S., STEWART, B. S., SCHIENMAN, J. & LEHMAN, N. 2004. Major histocompatibility complex variation at three class II loci in the northern elephant seal. 13**,** 711-718.

WEINMAN, L. R., SOLOMON, J. W., & RUBENSTEIN, D. R. (2015). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular ecology resources, 15*(3), 502-511.

WEIR, B. S. & GOUDET, J. 2017. A unified characterization of population structure and relatedness. *Genetics,* 206**,** 2085-2103.

WEIR, B. S., ANDERSON, A. D. & HEPLER, A. B. 2006. Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics,* 7**,** 771-780.

WESTNEAT, D. F. & SHERMAN, P. W. 1997. Density and extra-pair fertilizations in birds: a comparative analysis. *Behavioral Ecology and Sociobiology,* 41**,** 205-215.

WHALEN, A., GORJANC, G. & HICKEY, J. M. 2019. Parentage assignment with genotyping-by-sequencing data. *Journal of Animal Breeding and Genetics,* 136**,** 102-112.

WHITLOCK, M. C. (2011). GST and D do not replace FST. *Molecular Ecology, 20*(6), 1083-1091.

WHITE, A. M., SWAISGOOD, R. R., & CZEKALA, N. (2007). Ranging patterns in white rhinoceros, Ceratotherium simum simum: implications for mating strategies. *Animal behaviour, 74*(2), 349-356. doi:https://doi.org/10.1016/j.anbehav.2006.12.011

WHITELEY, A. R., FITZPATRICK, S. W., FUNK, W. C., & TALLMON, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution, 30*(1), 42-49.

WILLOUGHBY, J. R., IVY, J. A., LACY, R. C., DOYLE, J. M. & DEWOODY, J. A. 2017. Inbreeding and selection shape genomic diversity in captive populations: Implications for the conservation of endangered species. *PloS one,* 12**,** e0175996.

WILLOUGHBY, J. R., FERNANDEZ, N. B., LAMB, M. C., IVY, J. A., LACY, R. C. & DEWOODY, J. A. 2015. The impacts of inbreeding, drift and selection on genetic diversity in captive breeding populations. *Molecular Ecology,* 24**,** 98-110.

WINGETT, S. & ANDREWS, S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control [version 2; peer review: 4 approved]. *F1000Research,* 7.

WRIGHT, S. 1922. Coefficients of inbreeding and relationship. *The American Naturalist,* 56**,** 330-338.

WRIGHT, S. 1950. Genetical structure of populations. *Nature,* 166**,** 247-249.

XUE, Y., PRADO-MARTINEZ, J., SUDMANT, P. H., NARASIMHAN, V., AYUB, Q., SZPAK, M., FRANDSEN, P., CHEN, Y., YNGVADOTTIR, B., COOPER, D. N.,

DE MANUEL, M., HERNANDEZ-RODRIGUEZ, J., LOBON, I., SIEGISMUND, H. R., PAGANI, L., QUAIL, M. A., HVILSOM, C., MUDAKIKWA, A., EICHLER, E. E., CRANFIELD, M. R., MARQUES-BONET, T., TYLER-SMITH, C. & SCALLY, A. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science,* 348**,** 242-245.

YU, G., TANG, Q., LONG, K., CHE, T., LI, M. & SHUAI, S. 2015. Effectiveness of microsatellite and single nucleotide polymorphism markers for parentage analysis in European domestic pigs. *Genetics and Molecular Research,* 14**,** 1362-1370.

ZANE, L., BARGELLONI, L. & PATARNELLO, T. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology,* 11**,** 1-16.

ZHANG, J., KOBERT, K., FLOURI, T. & STAMATAKIS, A. 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics,* 30**,** 614-620.

ZHAO, H., Li, C., HARGROVE, J. S., BOWEN, B. R., THONGDA, W., ZHANG, D., PEATMAN, E. (2018). SNP marker panels for parentage assignment and traceability in the Florida bass (Micropterus floridanus). *Aquaculture, 485*, 30-38.

ZSCHOKKE, S., GAUTSCHI, B., & BAUR, B. (2003). Polymorphic microsatellite loci in the endangered Indian rhinoceros, Rhinoceros unicornis. *Molecular Ecology Notes, 3*(2), 233-235.

ZSCHOKKE, S., STUDER, P. & BAUR, B. 1998. Past and future breeding of the Indian rhinoceros in captivity. *International Zoo News***,** 261-277.

Appendices

Appendix 1

Table 1-A1 Review of genetic parameters observed in populations of southern white rhinoceros (SWR) and black rhinoceros (BR). The parameters including number of observed haplotypes, numbers of alleles per locus (N_a), observed (H_o) and expected (H_e) heterozygosities were estimated using different marker systems: mitochondrial DNA control region (control region); microsatellite (MS); Single Nucleotide Polymorphism (SNP).

Species	Marker systems	Locations	Number of samples	H_o	H_e	N_a/Haplotypes	References
SWR	5 MS loci	Umfolozi Gamer Reserve, South Africa	30	0.597	0.578	N_a = 2.8	(Florescu et al., 2003)
SWR	13 MS loci DQA, DQB	South Africa, DR Congo, Zimbabwe, Namibia	163	0.440	0.450	- N_a = 2.6 - Monomorphic for both DQA and DQB	(Coutts, 2009)

	mtDNA control region					- 4 haplotypes	
SWR	11 MS loci	Namibia	31	0.450	0.450	2.7	(Guerier et al., 2012)
BR, SWR	22 MS loci	Kruger National Park, South Africa	BR = 33 SWR = 367	0.365 0.363	0.510 0.393	4.857 2.8	(Harper et al., 2013)
SWR	9 MS loci 33 SNP loci	Songimvelo Nature Reserve, South Africa	32	0.478 0.357	0.508 0.350		(Labuschagne et al., 2017)
BR	mtDNA control region	20 African countries	403			20 haplotypes	(Moodley et al., 2017)

Appendix 2

Table 2-A1 The results of parentage assignments using different approaches (see Figure 2-1): field-observed (Pedigree A); field-observed plus exclusion (Pedigree B); likelihood-based assignment (Pedigree C); and combination of all three approaches (Pedigree D). Results of assignments using Bayes' theorem are also provided; posterior probability of a parent-offspring pair that is lesser than 0.05 is considered a genuine parent-offspring pair. For possible field-observed parents that were excluded based on microsatellites, the number of locus mismatches are indicated in parentheses. Note that the unsampled animals are labelled as 999, 888, 889, and 127. NA = unassignable parent; * 95% confidence; (+) 80% confidence; (-) less than 80% confidence; ? = more possible parents. Please refer to the electronic version of this table at https://doi.org/10.1007/s10592-019-01166-4.

ID	Field-observed (Ped. A)		Field-observed plus exclusion (Ped. B)				Likelihood assignment (Ped. C)			Bayesian Probability (SOLOMON) (Ped. C)				Three-approach combination (Ped. D)	
	Sire	Dam	Sire	Excluded sire (no. mismatches)	Dam	Excluded dam (no. mismatches)	Sire	Dam	Confidence	Sire	Most likely sire posterior prob.	Dam	Most likely dam posterior prob.	Sire	Dam
111	999	132	999		NA	132 (5)	NA	NA	NA	NA	NA	NA	NA	999	NA
	130														
113	999	134	999		NA	134 (4)	130*	135*	*	NA	NA	NA	NA	999	135
	130	177				177 (1)									
114	999	134	999		NA	134 (4)	130*	151*	*	NA	NA	151	1	999	151

	130	177	130			177 (1)									
120	999	888	999		888		NA	NA	NA	NA	NA	NA	NA	999	132
		132			132										
121	130	888	130		888		NA	NA	NA	130	0.022406469	NA	NA	130	888
	124			124 (3)											
	133			133 (4)											
124	999	889	999		889		NA	NA	NA	NA	NA	NA	NA	999	889
125	999	151	999		NA	151 (3)	NA	NA	NA	NA	NA	NA	NA	999	NA
131	999	888	999		888		NA	NA	NA	NA	NA	NA	NA	999	888
136	124	131	NA	124 (1)	NA	131 (2)	124*	166*	*	NA	NA	NA	NA	124	166
	170	141		170 (4)		141 (5)									
	130			130 (5)											
140	130	134	130		151	134 (3)	NA	NA	NA	130	0.592487914	151	0.671811931	130	151
	124	177		124 (2)		177 (1)									
	133	151		133 (4)											
141	130	151	130		151		130*	151*	*	130	0.002061655	151	0.218006491	130	151
	124			124 (4)											
	133			133 (5)											
142	130	141	NA	130 (1)	NA	141 (1)	133(-)	151(-)	*	NA	NA	NA	NA	133	151
	124	131		124 (2)		131 (3)									
	133			133 (1)											

143	130	166	NA	130 (4)	132	166 (1)	133(+)	132*	NA	NA	NA	177	0.973935217	133	132
	124	132		124 (1)		131 (1)									
	133	131													
144	130	889	130		889		130*	177*	*	130	0.15470911	NA	NA	130	889
145	999	135	999		NA	135 (5)	NA	NA	NA	NA	NA	NA	NA	999	NA
146	124	141	124		131	141 (1)	124*	131*	*	124	0.118263674	131	0.067804187	124	131
	130	131		130 (1)		151 (1)									
	133	151		133 (4)											
147	130	134	130		151	134 (4)	130*	151*	*	NA	NA	151	1	130	151
	999	151													
153	130	166	130		NA	166 (1)	130*	177	*	130	0.448026338	135	0.437459604	130	NA
	124	132		124 (4)		132 (2)									
	133	131		133 (3)		131 (2)									
154	130	131	NA	130 (1)		131 (2)	NA	NA	NA	NA	NA	151	0.138705195	NA	NA
	124	141		124 (3)	141										
	133	151		133 (3)	151										
155	124	177	124		NA	177 (1)	124*	135*	*	124	1	135	1	124	135
	130	151		130 (2)		151 (2)									
	133	134		133 (1)		134 (2)									
		141				141 (2)									
156	130	131	NA	130 (1)	NA	131 (1)	130*	131*	*	NA	NA	NA	NA	130	131

	124	132		124 (4)		132 (6)									
	133	166		133 (6)		166 (3)									
167	130	166	NA	130 (3)	NA	166 (2)	133(+)	125*	*	NA	NA	NA	NA	133	125
	124	132		124 (3)		132 (2)									
	133	131		133 (1)		133 (1)									
171	130	132	NA	130 (3)	NA	132 (1)	120(-)	177*	(+)	NA	NA	177	1	NA	NA
	124			124 (1)											
	133			133 (2)											
172	124	166	124		177	166 (1)	124*	114*	*	124	0.021615277	141	0.03118103	124	177
	130	134		130 (1)		134 (3)									
	133	177		133 (2)											
173	130	132	NA	130 (1)	NA	132 (1)	NA	NA	NA	NA	NA	NA	NA	NA	NA
	124	135		124 (3)		135 (1)									
	133			133 (2)											
174	170	141	NA	170 (2)	141		124(-)	141(+)	(+)	NA	NA	NA	NA	NA	141
	124			124 (1)											
	133			133 (2)											
175	130	131	130		131		144(-)	131(-)	(+)	130	0.841185007	114	0.421149511	130	NA
	133	151		133 (3)	151										
	124	141		124 (2)	141										
176	170	155	NA	170 (2)	155		120*	155*	*	120	0.207076461	155	0.384960429	120	155

	133			133 (3)											
	?														
200	170	127	NA	170 (3)	127		144 (-)	131(-)	(+)	124	0.399730097	NA	NA	NA	127
	133			133 (2)											
	?														
201	124	141	124		NA	141 (3)	NA	NA	NA	124	0.831782588	166	0.753259843	124	NA
	130	131		130 (4)		131 (1)									
	133	151		133 (1)		151 (3)									
202	124	151	NA	124 (2)	NA	151 (5)	133(-)	125(+)	*	NA	NA	NA	NA	133	125
	133			133 (1)											
203	170	166	NA	170 (2)	166		120*	166*	*	120	0.771579354	131	0.607765565	120	166
	133			133 (2											
	?														
204	130	166	130		NA	166 (1)	130*	177*	*	130	0.512097086	177	0.99151543	130	177
	124	132		124 (3)		132 (2)									
	133			133 (3)											
205	170	125	NA	170 (2)	125		124(-)	125*	(+)	NA	NA	125	0.299762283	NA	125
	133			133 (3)											
	?														
206	170	135	170		135		171(-)	135*	(+)	170	0.762524048	135	0.549934573	170	135
	133			133 (3)											

	?														
207	170	177	NA	170 (3)	177		120*	177	*	120	0.132914772	135	0.407366141	120	177
	133			133 (4)											
	?														
210	170	151	133	170 (1)	NA	151 (1)	145*	147*	*	133	0.672630269	141	0.989264099	133	147
	133														
	?														
238	170	132	NA	170 (4)	NA	132 (2)	NA	NA	NA	NA	NA	NA	NA	NA	NA
	133			133 (1)											
	?														
242	133	132	NA	133 (2)	NA	132 (2)	124(-)	151*	(+)	120	0.999712433	151	0.416625998	NA	NA
	124			124 (1)											
	130			130 (3)											
255	170	131	NA	170 (2)	131		124(-)	131(+)	(+)	120	0.465465719	131	0.143738648	NA	131
	133			133 (3)											
	?														
271	170	272	NA	170 (2)	272		120*	272*	*	120	0.320952683	272	0.103836762	120	272
	133			133 (5)											
	?														
274	170	141	170		141		170*	141*	*	170	0.489314259	141	0.312170933	170	141
	133			133 (3)											

	?														
275	170	135	170		135		170*	135*	*	170	0.544272112	135	0.129976486	170	135
	133			133 (5)											
	?														
316	170	134	NA	170 (1)	NA	134 (3)	120(*)	125(-)	(+)	120	0.417143502	131	0.500228999	NA	NA
	133			133 (2)											
	?														
320	170	NA	NA	170 (3)	NA		124*	132(-)	*	124	0.03953796	NA	NA	124	132
	133			133 (2)											
	?														

Appendix 2-A2 Kinship matrix, showing the pairwise kinship coefficients between individuals (ID number) based on the final pedigree. Shaded diagonal illustrates the kinship coefficient between an individual and itself. Note that inbred individuals were defined as showing self-kinship greater than 0.5. Please refer to the electronic version of this table at https://doi.org/10.1007/s10592-019-01166-4.

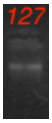
ID	111	113	114	120	121	124	125	131	136	140	141	142	143	144	145	146	147	153	155	156	167	172	174	175	176	200	201	202	203	204	205	206	207	210	255	271	274	275	320	272	177	170	166	151	135	133	132	130	999	888	889	127	ID-mk
111	0.500	0.125	0.125	0.125	0.000	0.125	0.125	0.125	0.063	0.000	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.063	0.063	0.063	0.063	0.000	0.000	0.094	0.000	0.063	0.063	0.063	0.000	0.063	0.000	0.063	0.000	0.063	0.063	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.051
113	0.125	0.500	0.125	0.125	0.000	0.125	0.125	0.125	0.063	0.000	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.188	0.063	0.063	0.063	0.000	0.000	0.156	0.000	0.063	0.063	0.063	0.000	0.063	0.125	0.063	0.000	0.063	0.063	0.000	0.125	0.063	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.064
114	0.125	0.125	0.500	0.125	0.000	0.125	0.125	0.125	0.063	0.125	0.125	0.125	0.000	0.000	0.125	0.125	0.125	0.000	0.063	0.063	0.063	0.063	0.063	0.000	0.094	0.000	0.063	0.063	0.063	0.000	0.063	0.000	0.063	0.063	0.063	0.063	0.063	0.000	0.063	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.069
120	0.125	0.125	0.125	0.500	0.000	0.125	0.125	0.125	0.063	0.000	0.000	0.000	0.125	0.000	0.125	0.125	0.000	0.000	0.063	0.063	0.063	0.063	0.000	0.000	0.281	0.000	0.063	0.063	0.250	0.000	0.063	0.000	0.250	0.000	0.063	0.250	0.000	0.000	0.188	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.250	0.000	0.000	0.000	0.075
121	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.125	0.000	0.125	0.125	0.000	0.000	0.125	0.000	0.063	0.125	0.125	0.000	0.188	0.000	0.000	0.063	0.125	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.063	0.063	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.250	0.000	0.000	0.048
124	0.125	0.125	0.125	0.125	0.000	0.500	0.125	0.125	0.250	0.000	0.000	0.000	0.000	0.125	0.125	0.313	0.000	0.000	0.250	0.063	0.063	0.250	0.000	0.000	0.188	0.000	0.250	0.063	0.063	0.000	0.063	0.000	0.063	0.000	0.063	0.063	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.250	0.000	0.082
125	0.125	0.125	0.125	0.125	0.000	0.125	0.500	0.125	0.063	0.000	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.063	0.063	0.250	0.063	0.000	0.000	0.094	0.000	0.063	0.250	0.063	0.000	0.250	0.000	0.063	0.000	0.063	0.063	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.062
131	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.500	0.063	0.000	0.000	0.000	0.000	0.000	0.125	0.313	0.000	0.000	0.063	0.250	0.063	0.063	0.000	0.000	0.094	0.000	0.063	0.063	0.063	0.000	0.063	0.000	0.063	0.000	0.250	0.063	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.250	0.000	0.000	0.069
136	0.063	0.063	0.063	0.063	0.000	0.250	0.063	0.063	0.500	0.000	0.000	0.000	0.000	0.063	0.063	0.156	0.000	0.000	0.125	0.031	0.031	0.125	0.000	0.000	0.094	0.000	0.125	0.031	0.156	0.000	0.031	0.000	0.031	0.000	0.031	0.031	0.000	0.000	0.125	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.055
140	0.000	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.000	0.500	0.250	0.125	0.000	0.125	0.000	0.000	0.250	0.125	0.000	0.125	0.000	0.000	0.125	0.125	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.055
141	0.000	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.000	0.250	0.500	0.125	0.000	0.125	0.000	0.000	0.250	0.125	0.000	0.125	0.000	0.000	0.250	0.125	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.060
142	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.125	0.500	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.063	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.188	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.042
143	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.063	0.000	0.000	0.125	0.063	0.000	0.000	0.000	0.063	0.125	0.000	0.063	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.250	0.000	0.000	0.000	0.000	0.000	0.038
144	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.063	0.125	0.125	0.000	0.000	0.500	0.000	0.063	0.125	0.125	0.063	0.125	0.000	0.063	0.063	0.125	0.031	0.000	0.063	0.000	0.000	0.125	0.000	0.000	0.000	0.063	0.000	0.000	0.063	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.250	0.000	0.052
145	0.125	0.125	0.125	0.125	0.000	0.125	0.125	0.125	0.063	0.000	0.000	0.000	0.000	0.000	0.500	0.125	0.000	0.000	0.063	0.063	0.063	0.063	0.000	0.000	0.094	0.000	0.063	0.063	0.063	0.000	0.063	0.000	0.063	0.000	0.063	0.063	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.051
146	0.125	0.125	0.125	0.125	0.063	0.313	0.125	0.313	0.156	0.000	0.000	0.000	0.000	0.063	0.125	0.563	0.000	0.000	0.156	0.156	0.063	0.156	0.000	0.000	0.141	0.000	0.156	0.063	0.063	0.000	0.063	0.000	0.063	0.000	0.156	0.063	0.000	0.000	0.156	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.125	0.125	0.000	0.080
147	0.000	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.000	0.250	0.250	0.125	0.000	0.125	0.000	0.000	0.500	0.125	0.000	0.125	0.000	0.000	0.125	0.125	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.250	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.058
153	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.125	0.000	0.000	0.125	0.500	0.000	0.125	0.000	0.000	0.063	0.125	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.063	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.037
155	0.063	0.188	0.063	0.063	0.000	0.250	0.063	0.063	0.125	0.000	0.000	0.000	0.000	0.063	0.063	0.156	0.000	0.000	0.500	0.031	0.031	0.125	0.000	0.000	0.281	0.000	0.125	0.031	0.031	0.000	0.031	0.125	0.031	0.000	0.031	0.031	0.000	0.125	0.125	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.064
156	0.063	0.063	0.063	0.063	0.188	0.063	0.063	0.250	0.031	0.125	0.125	0.000	0.000	0.125	0.063	0.156	0.125	0.125	0.031	0.500	0.031	0.031	0.063	0.125	0.047	0.000	0.031	0.031	0.031	0.125	0.031	0.000	0.031	0.063	0.125	0.031	0.063	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.125	0.125	0.000	0.000	0.069
167	0.063	0.063	0.063	0.063	0.000	0.063	0.250	0.063	0.031	0.000	0.000	0.125	0.125	0.000	0.063	0.063	0.000	0.000	0.031	0.031	0.500	0.031	0.000	0.000	0.047	0.000	0.031	0.250	0.031	0.000	0.125	0.000	0.031	0.125	0.031	0.031	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.125	0.000	0.000	0.000	0.053
172	0.063	0.063	0.063	0.063	0.000	0.250	0.063	0.063	0.125	0.000	0.000	0.000	0.000	0.063	0.063	0.156	0.000	0.000	0.125	0.031	0.031	0.500	0.000	0.000	0.094	0.000	0.125	0.031	0.031	0.125	0.031	0.000	0.156	0.000	0.031	0.031	0.000	0.000	0.125	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.058
174	0.000	0.000	0.063	0.000	0.063	0.000	0.000	0.000	0.000	0.125	0.250	0.063	0.000	0.063	0.000	0.000	0.125	0.063	0.000	0.063	0.000	0.000	0.500	0.063	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000	0.063	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.037
175	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.125	0.000	0.000	0.125	0.125	0.000	0.125	0.000	0.000	0.063	0.500	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.063	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.037
176	0.094	0.156	0.094	0.281	0.000	0.188	0.094	0.094	0.094	0.000	0.000	0.000	0.063	0.031	0.094	0.141	0.000	0.000	0.281	0.047	0.047	0.094	0.000	0.000	0.531	0.000	0.094	0.047	0.141	0.000	0.047	0.063	0.141	0.000	0.047	0.141	0.000	0.063	0.156	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.188	0.000	0.063	0.000	0.074
200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.014
201	0.063	0.063	0.063	0.063	0.000	0.250	0.063	0.063	0.125	0.000	0.000	0.000	0.000	0.063	0.063	0.156	0.000	0.000	0.125	0.031	0.031	0.125	0.000	0.000	0.094	0.000	0.500	0.031	0.031	0.000	0.031	0.000	0.031	0.000	0.031	0.031	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.048
202	0.063	0.063	0.063	0.063	0.000	0.063	0.250	0.063	0.031	0.000	0.000	0.125	0.125	0.000	0.063	0.063	0.000	0.000	0.031	0.031	0.250	0.031	0.000	0.000	0.047	0.000	0.031	0.500	0.031	0.000	0.125	0.000	0.031	0.125	0.031	0.031	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.125	0.000	0.000	0.000	0.053
203	0.063	0.063	0.063	0.250	0.000	0.063	0.063	0.063	0.156	0.000	0.000	0.000	0.063	0.000	0.063	0.063	0.000	0.000	0.031	0.031	0.031	0.031	0.000	0.000	0.141	0.000	0.031	0.031	0.500	0.000	0.031	0.000	0.125	0.000	0.031	0.125	0.000	0.000	0.094	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.052
204	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.125	0.000	0.000	0.125	0.125	0.000	0.125	0.000	0.125	0.063	0.125	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.125	0.063	0.000	0.000	0.063	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.047
205	0.063	0.063	0.063	0.063	0.000	0.063	0.250	0.063	0.031	0.000	0.000	0.000	0.000	0.000	0.063	0.063	0.000	0.000	0.031	0.031	0.125	0.031	0.000	0.000	0.047	0.000	0.031	0.125	0.031	0.000	0.500	0.000	0.031	0.000	0.031	0.031	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.038
206	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.125	0.250	0.000	0.000	0.000	0.250	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032
207	0.063	0.063	0.063	0.250	0.000	0.063	0.063	0.063	0.031	0.000	0.000	0.000	0.063	0.000	0.063	0.063	0.000	0.000	0.031	0.031	0.031	0.156	0.000	0.000	0.141	0.000	0.031	0.031	0.125	0.125	0.031	0.000	0.500	0.000	0.031	0.125	0.000	0.000	0.094	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.054
210	0.000	0.000	0.063	0.000	0.063	0.000	0.000	0.000	0.000	0.125	0.125	0.188	0.125	0.063	0.000	0.000	0.250	0.063	0.000	0.063	0.125	0.000	0.063	0.063	0.000	0.000	0.000	0.125	0.000	0.063	0.000	0.000	0.000	0.500	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.250	0.000	0.125	0.000	0.000	0.000	0.000	0.050
255	0.063	0.063	0.063	0.063	0.063	0.063	0.063	0.250	0.031	0.000	0.000	0.000	0.000	0.000	0.063	0.156	0.000	0.000	0.031	0.125	0.031	0.031	0.000	0.000	0.047	0.000	0.031	0.031	0.031	0.000	0.031	0.000	0.031	0.000	0.500	0.031	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.125	0.000	0.000	0.042
271	0.063	0.063	0.063	0.250	0.000	0.063	0.063	0.063	0.031	0.000	0.000	0.000	0.063	0.000	0.063	0.063	0.000	0.000	0.031	0.031	0.031	0.031	0.000	0.000	0.141	0.000	0.031	0.031	0.125	0.000	0.031	0.000	0.125	0.000	0.031	0.500	0.000	0.000	0.094	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.050
274	0.000	0.000	0.063	0.000	0.063	0.000	0.000	0.000	0.000	0.125	0.250	0.063	0.000	0.063	0.000	0.000	0.125	0.063	0.000	0.063	0.000	0.000	0.125	0.063	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.125	0.000	0.063	0.000	0.000	0.500	0.125	0.000	0.000	0.000	0.250	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.047
275	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.125	0.500	0.000	0.000	0.000	0.250	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.032
320	0.063	0.063	0.063	0.188	0.000	0.250	0.063	0.063	0.125	0.000	0.000	0.000	0.125	0.063	0.063	0.156	0.000	0.000	0.125	0.031	0.031	0.125	0.000	0.000	0.156	0.000	0.125	0.031	0.094	0.000	0.031	0.000	0.094	0.000	0.031	0.094	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.125	0.000	0.125	0.000	0.063
272	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014
177	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.024
170	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.250	0.250	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.024
166	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019
151	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.250	0.250	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.041
135	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031
133	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.034
132	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.125	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.034
130	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.250	0.250	0.000	0.000	0.250	0.000	0.000	0.250	0.250	0.000	0.250	0.000	0.000	0.125	0.250	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.125	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.060
999	0.250	0.250	0.250	0.250	0.000	0.250	0.250	0.250	0.125	0.000	0.000	0.000	0.000	0.000	0.250	0.250	0.000	0.000	0.125	0.125	0.125	0.125	0.000	0.000	0.188	0.000	0.125	0.125	0.125	0.000	0.125	0.000	0.125	0.000	0.125	0.125	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.088
888	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.026
889	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.250	0.000	0.125	0.000	0.000	0.125	0.000	0.000	0.125	0.000	0.000	0.063	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.035
127	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.014

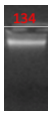
Appendix 3

Table 3-A1

Table 3-A1 List of total Botswana samples. Table shows the sample ID, the provenance of samples (Botswana1, Botswana2, Botswana3), the quality of DNA extractions based on 1% agarose gel electrophoresis (high = a tight band of high molecular weight DNA at > 1000 bp position; moderate = a smeared band but relatively denser at > 1,000 bp position; low = a completely smeared band, no noticeable band), and the quantify of DNA based on nanodrop readings. Please note that the size standard DNA ladders used to judge DNA quality cannot be seen in the gel pictures provided.

	Sample ID	Provenance	DNA quality Classification [a]	Gel picture	Quantity (ng/µl)
1	108	Botswana2	Moderate		28.8
2	118	Botswana2	Low		52.0
3	120	Botswana1	Low		162.0
4	121	Botswana1	Low		190.0
5	123	Botswana3	High		21.5

6	124	Botswana2	Low		27.3
7	125	Botswana1	High		39.9
8	126	Botswana3	Low	NA	29.6
9	127	Botswana1	High		13.5
10	128	Botswana2	Low	NA	9.8
11	130	Botswana1	Low		9.6
12	131	Botswana1	Moderate		34.7
13	132	Botswana1	Moderate		36.3
14	133	Botswana1	Moderate		35.3

15	134	Botswana1	High		45.1
16	135	Botswana1	Low		22.1
17	137	Botswana1	High		15.0
18	140	Botswana2	Low		29.3
19	141	Botswana1	Moderate		53.2
20	142	Botswana3	Low		140.0
21	146	Botswana1	Low		8.3
22	148	Botswana1	Low		132.0
23	150	Botswana3	Low	NA	99.3

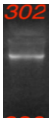
24	151	Botswana1	Low		43.27
25	154	Botswana1	Low		12.2
26	155	Botswana1	Moderate		18.3
27	156	Botswana1	Low		92.7
28	158	Botswana2	Moderate		141.3
29	166	Botswana1	Low		18.93
30	168	Botswana2	Low		138.0
31	171	Botswana1	Low		31.6
32	172	Botswana1	Low		38.13

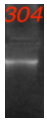
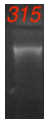
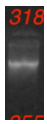
33	176	Botswana1	Low		49.4
34	177	Botswana1	High		14.5
35	178	Botswana2	Low		121.3
36	180	Botswana2	Low		78.7
37	181	Botswana1	Low		102.0
38	182	Botswana2	Low		10.96
39	183	Botswana2	Low		148.0
40	184	Botswana2	High		19.1
41	185	Botswana2	Low		64.3

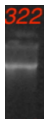
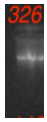
42	187	Botswana1	Low		101.0
43	188	Botswana1	Low		79.3
44	200	Botswana1	Low		88.0
45	201	Botswana1	High		14.2
46	202	Botswana1	Low		210.0
47	203	Botswana1	Low		280.0
48	204	Botswana1	Low		40.9
49	205	Botswana1	Low		303.3
50	206	Botswana1	Low		18.1

51	207	Botswana1	Low		156.0
52	208	Botswana3	Low		24.9
53	210	Botswana1	Moderate		148.0
54	211	Botswana2	Low		29.6
55	212	Botswana2	Moderate		30.3
56	213	Botswana2	Moderate		61.7
57	237	Botswana2	Moderate		20.4
58	238	Botswana1	Low		154.3
59	242	Botswana1	High		16.9

60	246	Botswana1	High		28.7
61	248	Botswana2	Low		42.1
62	251	Botswana3	Low	NA	6.7
63	255	Botswana1	Low		138.0
64	262	Botswana2	Low		176.0
65	263	Botswana2	Low	NA	15.1
66	266	Botswana2	Low		46.6
67	268	Botswana2	Low		14.17
68	271	Botswana1	Low		322.7

69	272	Botswana1	Moderate		19.7
70	274	Botswana1	Low		118
71	275	Botswana1	Low		260.0
72	278	Botswana2	Low		22.1
73	280	Botswana2	Moderate		87.3
74	286	Botswana3	Low	NA	6.1
75	300	Botswana3	Low	NA	9.8
76	301	Botswana2	Low		115.3
77	302	Botswana2	High		21.9

78	303	Botswana2	Low		138.0
79	304	Botswana2	High		19.1
80	308	Botswana3	Low	NA	7.7
81	311	Botswana2	Low		101.3
82	312	Botswana2	Moderate		38.3
83	314	Botswana2	Low		49.3
84	315	Botswana2	High		12.5
85	316	Botswana1	Low		17.6
86	318	Botswana1	High		60.8

87	320	Botswana1	Low		146.7
88	322	Botswana1	High		47.6
89	323	Botswana1	High		18.2
90	326	Botswana1	High		30.3
91	327	Botswana3	Low	NA	9.4
92	328	Botswana2	High		15.8
93	330	Botswana2	High		13.7
94	331	Botswana3	Low	NA	8.5
95	332	Botswana2	Moderate		66.6
96	333	Botswana2	Low		49.3

97	335	Botswana2	Low		150.7
98	343	Botswana2	Moderate		38.9
99	344	Botswana2	Low		87.3
100	347	Botswana1	Low		51.8
101	348	Botswana1	Low		22.9
102	352	Botswana1	High		27.4
103	353	Botswana3	Low	NA	5.7
104	355	Botswana1	High		17.6
105	360	Botswana2	Low		10.9

106	361	Botswana2	Low		14.9
107	362	Botswana2	Low		11.7
108	363	Botswana2	Low		17.7
109	366	Botswana2	Low		12.8
110	368	Botswana2	Low		96.7

Figure 3-A2

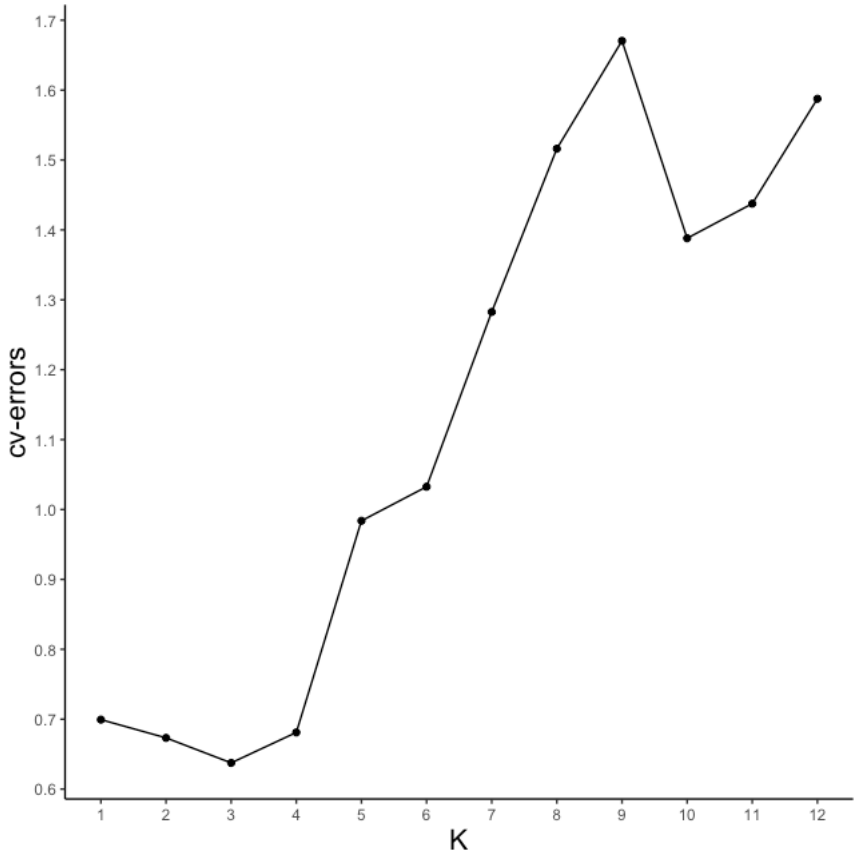


Figure 3-A2 10-fold cross validation errors (cv-errors) for different *K* values. To determine an appropriate *K* value (number of genetic clusters) to be included in the likelihood model implemented in ADMIXTURE (Alexander et al., 2015), *K* was varied from 1-12. The lowest cv-errors of 0.63751 was reported for *K* = 3; thus, it was employed in the model to determine genetic clusters for samples used in the initial screening for variation based on ddRAD data.

Command 3-A3: Command lines used to determine effects of DNA quality on performance of RADcapture protocol Determine total number of reads:

- "*expr $(samtools view -f 4 $sample.against.assembly.bam -c) + $(samtools view -F 2308 $sample.against.assembly.bam -c)*"

- Determine mean read length:

  - "*samtools stats $sample.against.assembly.bam | grep ^SN | grep "average length"*"

- Estimate mean depth of coverage:

    - "*samtools idxstats $sample.sorted.bam | awk '{print $1, $3, $mean_read_length*$3/$2}' | awk '{ sum += $3 } END { if (NR > 0) print sum / NR }' > $sample.depth.text*"

- Determine percentage of duplications:

    - "*samtools view -f 1024 $sample.against.assembly.bam -c*" (this command provided the number of reads flagged as duplicates, then percent duplicates (number of reads flagged as duplicates/total number of reads)*100)

Command 3-A4: Command lines used to determine %sensitivity and %specificity

- Measure sensitivity of the bait set and RADcapture protocol:

    - "*bedtools genomecov -ibam $sample.against.bait.seq.bam -bga | awk '$4==0' | bedtools intersect -a all.regions.bed -b - > $sample.zero.coverage.regions.bed*" (this command generated a BED file containing regions that lacked coverage)

    - "*cat $sample.zero.coverage.regions.bed | awk -F'\t' 'BEGIN{SUM=0}{ SUM+=$3-$2+1 }END{print SUM}'*" (this command counted number of bases in a $sample.zero.coverage.regions.bed file)

- Measure specificity of the bait set and RADcapture protocol

- "*samtools view -f 4 $sample.against.bait.seq.bam -c*" (this command determined exact number of reads that were not mapped to the bait reference for each sample)

- "*expr $(samtools view -f 4 $sample.against.bait.seq.bam -c) + $(samtools view -F 2308 $sample.against.bait.seq.bam -c)*" (this command gave exact number of total reads for each sample where -F 2308 excluded reads flagged with 2308 that were either unmapped, supplementary alignments, or not primarily alignments)

Table 3-A5

Table 3-A5 Results of screening for exogenous DNA contamination. Approximately 2,000,000 reads were randomly chosen from each sample and were mapped against the genome of northern white rhinoceros and some common sources of exogenous DNA contamination: human, mouse, and *E.coli* genomes. Numbers and percentage of reads mapped uniquely to each genome are provided.

Sample IDs	Populations	Numbers and percentages of reads mapped uniquely to genomes of							
		NWR	% NWR	Human	% Human	Mouse	% Mouse	*E.coli*	% *E.coli*
131	Botswana1	1,393,476	71.02	350	0.02	105	0.01	4	0
132	Botswana1	1,418,118	70.92	314	0.02	66	0.00	0	0
134	Botswana1	1,224,195	62.36	280	0.01	103	0.01	0	0
135	Botswana1	1,429,792	66.73	381	0.02	120	0.01	0	0
151	Botswana1	1,261,533	70.01	288	0.02	115	0.01	0	0
154	Botswana1	1,354,278	66.07	308	0.02	70	0.00	0	0
155	Botswana1	1,426,824	70.41	323	0.02	99	0.00	0	0
166	Botswana1	1,369,067	71.05	324	0.02	94	0.00	0	0
171	Botswana1	1,421,571	71.46	352	0.02	93	0.00	0	0
172	Botswana1	1,295,503	65.32	292	0.01	80	0.00	0	0
177	Botswana1	1,266,645	62.18	324	0.02	110	0.01	1	0
184	Botswana2	1,163,999	57.23	285	0.01	116	0.01	0	0
185	Botswana2	1,338,185	67.96	320	0.02	119	0.01	0	0
187	Botswana1	1,226,958	62.38	303	0.02	80	0.00	0	0

188	Botswana1	1,225,717	62.23	312	0.02	68	0.00	0	0
201	Botswana1	1,178,298	57.43	330	0.02	137	0.01	1	0
205	Botswana1	1,195,095	57.95	327	0.02	92	0.00	0	0
207	Botswana1	1,139,736	58.59	312	0.02	100	0.01	0	0
238	Botswana1	1,401,154	68.68	373	0.02	127	0.01	0	0
242	Botswana1	1,415,797	70.75	374	0.02	125	0.01	0	0
278	Botswana2	1,252,888	62.38	329	0.02	94	0.00	4	0
302	Botswana2	1,416,852	70.78	466	0.02	122	0.01	0	0
304	Botswana2	1,436,662	70.86	366	0.02	95	0.00	0	0
312	Botswana2	1,303,851	66.42	334	0.02	108	0.01	0	0
315	Botswana2	1,272,517	63.17	358	0.02	125	0.01	0	0
326	Botswana1	1,427,257	68.99	388	0.02	146	0.01	1	0
328	Botswana2	1,068,104	55.14	268	0.01	89	0.00	1	0
330	Botswana2	1,284,210	65.59	297	0.02	126	0.01	0	0
347	Botswana1	1,057,918	55.83	247	0.01	88	0.00	0	0
348	Botswana1	1,217,293	59.42	259	0.01	109	0.01	0	0
352	Botswana1	1,319,564	66.59	375	0.02	122	0.01	396	0.02
355	Botswana1	1,266,924	65.08	331	0.02	133	0.01	0	0

Table 3-A6

Table 3-A6 Results of RADcapture, SNP calling, and efficiency of the bait set. Capture reaction IDs provide information about which samples were included in a 4-sample pooled reaction. Quality of DNA was judged immediately after DNA extraction based on 1% agarose gel electrophoresis: high quality = a tight band of high molecular weight DNA (> 1000 bp position); moderate quality = a smeared band but relatively dense at > 1000 bp position; and low quality = a completely smeared band (no noticeable band).

Capture reaction IDs	Sample IDs	DNA quality	Origin of samples	Performance parameters					Efficiency of bait set	
				Cleaned read numbers	Read lengths	Mapping coverage (×)	Percent duplicates (%)	Number of SNPs	Percent sensitivity (%)	Percent specificity (%)
1	355	High	Botswana1	27,255,148	138	7.54	82.64	3,760	60	8.72
1	184	High	Botswana2	36,609,269	137	12.17	85.12	4,861	62	11.18
1	201	High	Botswana1	34,877,428	129	9.97	91.04	3,614	60	8.47
1	177	High	Botswana1	40,740,570	131	13.1	88.05	5,212	57	6.50
2	135	Low	Botswana1	12,856,446	142	1.95	55.11	2,874	57	5.70
2	154	Low	Botswana1	38,948,276	142	7.26	69.43	6,206	57	4.26
2	185	Low	Botswana2	21,658,805	143	3.39	57.51	4,838	62	5.21
2	238	Low	Botswana1	18,360,541	142	3.89	68.68	3,438	60	5.10
3	172	Low	Botswana1	57,515,669	140	14.73	81.93	6,421	60	5.70
3	205	Low	Botswana1	22,686,495	141	6.46	76.62	4,506	58	10.20
3	207	Low	Botswana1	29,180,103	136	9.34	84.70	4,675	60	8.75
3	347	Low	Botswana1	17,052,961	142	4.69	71.80	3,867	56	11.07

4	242	High	Botswana1	28,014,956	142	4.18	59.54	5,569	64	4.42
4	312	Moderate	Botswana2	33,372,323	143	5.55	61.67	6,103	64	5.30
4	326	High	Botswana1	24,826,351	143	3.96	59.07	5,105	64	5.28
4	352	High	Botswana1	23,779,998	142	4.08	61.15	5,180	64	6.49
5	131	Moderate	Botswana1	19,622,097	143	2.81	56.88	3,863	53	2.79
5	132	Moderate	Botswana1	29,992,687	144	5.11	59.02	5,326	58	2.66
5	302	High	Botswana2	18,015,799	142	2.65	55.87	3,567	60	4.15
5	304	High	Botswana2	20,273,741	142	3.26	60.02	3,916	62	4.19
6	151	Low	Botswana1	9,009,770	142	1.16	45.83	1,356	51	3.89
6	155	Low	Botswana1	22,290,150	142	3.65	61.51	4,587	57	3.61
6	166	Low	Botswana1	21,195,925	142	3.55	62.99	4,115	55	2.86
6	171	Low	Botswana1	27,851,517	143	5.22	64.78	5,040	57	2.75
7	187	Low	Botswana1	53,109,231	141	13.1	78.93	6,418	62	6.55
7	188	Low	Botswana1	45,303,334	142	11.56	76.27	6,357	60	6.99
7	278	Low	Botswana2	18,075,025	141	4.03	67.65	4,094	59	8.50
7	348	Low	Botswana1	32,778,904	141	1.59	79.05	5,655	63	9.32
8	134	High	Botswana1	23,558,455	143	6.1	73.08	4,515	60	8.95
8	315	High	Botswana2	46,331,588	140	13.13	82.28	5,485	65	9.49
8	328	High	Botswana2	27,120,582	141	7.1	77.72	4,561	61	11.98
8	330	High	Botswana2	33,283,557	141	4.97	76.62	5,136	63	8.16

Mapping coverage = depth of reads mapped to the entire reference assembly

Appendix 4

Table 4-A1

Table 4-A1 Adaptor combinations and library concentration for 96 samples included in library preparation for RADcapture sequencing (both in Chapter 3 and 4). Each DNA sample was fragmented and ligated to a specific combination of i5 and i7 adaptors, followed by eight cycles of PCR amplification. Samples for which final library concentrations were < 20 ng/µl were not included in bait hybrid capture (denoted as NA in the first column). A total of 22 libraries were retained, which were divided into pools of 4 for subsequent capture by the baits.

Bait capture library reactions	Sample ID	Populations	i5 adaptor sequences	i7 adaptor sequences	Final library concentrations (ng/µl)
1	177	Botswana1	CCTATCCT	ACGAATTC	52.0
1	184	Botswana2	AGGCGAAG	GCGCATTA	75.9
1	201	Botswana1	GGCTCTGA	CGAGTAAT	56.2

1	355	Botswana1	AGGCGAAG	AGCTTCAG	87.3
2	135	Botswana1	ATAGAGGC	TCTCCGGA	29.7
2	154	Botswana1	CCTATCCT	CGAGTAAT	53.0
2	185	Botswana2	TAATCTTA	CGAGTAAT	84.0
2	238	Botswana1	GGCTCTGA	AATGAGCG	85.2
3	172	Botswana1	CCTATCCT	TTCTGAAT	69.3
3	205	Botswana1	GTACTGAC	AGCTTCAG	66.5
3	207	Botswana1	GGCTCTGA	TCTCCGGA	58.0

3	347	Botswana1	AGGCGAAG	GGAATCTC	69.2
4	242	Botswana1	GGCTCTGA	TTCTGAAT	79.3
4	312	Botswana2	TAATCTTA	ACGAATTC	72.8
4	326	Botswana1	AGGCGAAG	AATGAGCG	75.6
4	352	Botswana1	AGGCGAAG	ACGAATTC	95.4
5	131	Botswana1	TATAGCCT	AATGAGCG	23.1
5	132	Botswana1	TATAGCCT	GGAATCTC	22.7
5	302	Botswana2	TAATCTTA	GGAATCTC	80.6

5	304	Botswana2	TAATCTTA	TTCTGAAT	69.3
6	151	Botswana1	ATAGAGGC	GGAATCTC	22.2
6	155	Botswana1	CCTATCCT	TCTCCGGA	63.3
6	166	Botswana1	CCTATCCT	AATGAGCG	58.6
6	171	Botswana1	CCTATCCT	GGAATCTC	45.1
7	187	Botswana1	CCTATCCT	AGCTTCAG	84.7
7	188	Botswana1	CCTATCCT	GCGCATTA	85.1
7	278	Botswana2	TAATCTTA	TCTCCGGA	88.6

7	348	Botswana1	AGGCGAAG	TTCTGAAT	83.7
8	134	Botswana1	CAGGACGT	TTCTGAAT	60.6
8	315	Botswana2	TAATCTTA	AGCTTCAG	79.7
8	328	Botswana2	TAATCTTA	GCGCATTA	69.3
8	330	Botswana2	CAGGACGT	CGAGTAAT	78.3
9	158	Botswana2	ATAGAGGC	CATAGCCG	70.1
9	266	Botswana2	TAATCTTA	TTCGCGGA	58.8
9	271	Botswana1	TATAGCCT	ACGAATTC	56.7

9	344	Botswana2	CAGGACGT	TTCGCGGA	59.9
10	211	Botswana2	CCTATCCT	GCGCGAGA	58.5
10	262	Botswana2	GGCTCTGA	TATCGCT	70.0
10	274	Botswana1	TATAGCCT	AGCTTCAG	63.7
10	316	Botswana1	ATAGAGGC	TTCTGAAT	62.0
11	237	Botswana2	GGCTCTGA	TTCGCGGA	61.1
11	275	Botswana1	TATAGCCT	GCGCATTA	53.6
11	303	Botswana2	AGGCGAAG	CATAGCCG	62.2

11	368	Botswana2	CCTATCCT	CATAGCCG	60.7
12	183	Botswana2	CCTATCCT	TTCGCGGA	53.2
12	255	Botswana1	TATAGCCT	TTCTGAAT	57.9
12	320	Botswana1	ATAGAGGC	ACGAATTC	68.4
12	335	Botswana2	AGGCGAAG	CTATCGCT	49.2
13	108	Botswana2	ATAGAGGC	AGCTTCAG	44.0
13	212	Botswana2	CCTATCCT	CTATCGCT	53.5
13	248	Botswana2	GGCTCTGA	GCGCGAGA	48.3

13	311	Botswana2	AGGCGAAG	TTCGCGGA	51.2
14	123	Botswana3	TAATCTTA	CATAGCCG	51.1
14	178	Botswana2	ATAGAGGC	GCGCGAGA	36.6
14	343	Botswana2	CAGGACGT	CATAGCCG	62.7
14	363	Botswana2	GTACTGAC	TTCGCGGA	63.8
15	118	Botswana2	ATAGAGGC	GCGCATTA	29.2
15	124	Botswana2	TATAGCCT	CATAGCCG	24.5
15	126	Botswana3	TAATCTTA	GCGCGAGA	22.3

15	150	Botswana3	TATAGCCT	TTCGCGGA	61.2
16	168	Botswana2	ATAGAGGC	TTCGCGGA	18.8
16	213	Botswana2	GGCTCTGA	CATAGCCG	47.0
16	301	Botswana2	TAATCTTA	CTATCGCT	53.6
16	314	Botswana2	AGGCGAAG	GCGCGAGA	71.0
17	202	Botswana1	GTACTGAC	GGAATCTC	62.5
17	272	Botswana1	GGCTCTGA	AGCTTCAG	67.5
17	280	Botswana2	TAATCTTA	AATGAGCG	79.4

17	322	Botswana1	AGGCGAAG	CGAGTAAT	61.4
18	121	Botswana1	CAGGACGT	GGAATCTC	53.8
18	127	Botswana1	TATAGCCT	TCTCCGGA	27.6
18	204	Botswana1	GTACTGAC	ACGAATTC	60.4
18	323	Botswana1	GGCGAAG	TCTCCGGA	82.2
19	120	Botswana1	CAGGACGT	AATGAGCG	56.5
19	148	Botswana1	CAGGACGT	AGCTTCAG	78.5
19	176	Botswana1	GTACTGAC	CGAGTAAT	66.7

19	181	Botswana1	GTACTGAC	TCTCCGGA	58.0
20	141	Botswana1	CAGGACGT	ACGAATTC	72.0
20	156	Botswana1	CAGGACGT	GCGCATTA	78.0
20	200	Botswana1	GTACTGAC	AATGAGCG	68.9
20	203	Botswana1	GTACTGAC	TTCTGAAT	63.6
21	206	Botswana1	GTACTGAC	GCGCATTA	64.1
21	210	Botswana1	GGCTCTGA	GGAATCTC	72.2
21	246	Botswana1	GGCTCTGA	ACGAATTC	68.3

21	318	Botswana1	GGCTCTGA	GCGCATTA	65.3
22	133	Botswana1	ATAGAGGC	CGAGTAAT	23.0
22	208	Botswana3	GTACTGAC	CTATCGCT	60.0
22	332	Botswana2	CAGGACGT	TCTCCGGA	74.1
22	366	Botswana2	GTACTGAC	GCGCGAGA	68.8
NA	125	Botswana1	TATAGCCT	CGAGTAAT	17.9
NA	137	Botswana1	ATAGAGGC	AATGAGCG	10.7
NA	140	Botswana2	TATAGCCT	GCGCGAGA	9.3

NA	142	Botswana3	TATAGCCT	CTATCGCT	11.9
NA	180	Botswana2	ATAGAGGC	CTATCGCT	17.6
NA	360	Botswana2	CAGGACGT	GCGCGAGA	2.6
NA	361	Botswana2	CAGGACGT	CTATCGCT	2.0
NA	362	Botswana2	GTACTGAC	CATAGCCG	2.9

Table 4-A2 Per-site summary statistics of 302 SNPs genotyped in 88 individuals including: nucleotide diversity (π); observed (H_o) and expected heterozygosity (H_e); significance of deviation from Hardy-Weinberg Equilibrium (*HWE*), *NS* = not significant, *p* = statistic *p-value*; inbreeding coefficient (F_{is}); tests for heterozygote excess or deficit (Excess/deficit), *NS* = not significant, *excess* = significant for excess, *deficit* = significant for deficiency.

SNPs	π	H_o	H_e	*HWE*	F_{is}	*Excess/deficit*
1	0.0976	0.0795	0.0970	*NS*	0.1858	*NS*
2	0.1847	0.1818	0.1836	*NS*	0.0156	*NS*
3	0.2114	0.1932	0.2101	*NS*	0.0865	*NS*
4	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
5	0.1376	0.1477	0.1368	*NS*	-0.0741	*NS*
6	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
7	0.3532	0.4318	0.3512	*p < 0.05*	-0.2240	*excess*
8	0.1755	0.1932	0.1745	*NS*	-0.1013	*NS*
9	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
10	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*
11	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
12	0.3405	0.3864	0.3386	*NS*	-0.1355	*NS*
13	0.4140	0.5795	0.4116	*p < 0.001*	-0.4032	*excess*
14	0.4321	0.4659	0.4297	*NS*	-0.0786	*NS*
15	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
16	0.1662	0.1818	0.1653	*NS*	-0.0943	*NS*
17	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
18	0.1662	0.1591	0.1653	*NS*	0.0432	*NS*
19	0.2691	0.2500	0.2676	*NS*	0.0713	*NS*
20	0.5029	1.0000	0.5000	*p < 0.01*	-1.0000	*excess*
21	0.5029	1.0000	0.5000	*p < 0.01*	-1.0000	*excess*
22	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
23	0.1078	0.0227	0.1072	*p < 0.001*	0.7901	*deficit*
24	0.2369	0.2273	0.2355	*NS*	0.0408	*NS*
25	0.1662	0.1818	0.1653	*NS*	-0.0943	*NS*
26	0.2285	0.2159	0.2272	*NS*	0.0554	*NS*
27	0.1473	0.1591	0.1465	*NS*	0.0807	*NS*
28	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
29	0.1376	0.1477	0.1368	*NS*	-0.0741	*NS*
30	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
31	0.4997	0.8523	0.4968	*p < 0.001*	-0.7126	*excess*
32	0.2691	0.3182	0.2676	*NS*	-0.1837	*NS*

33	0.1278	0.1136	0.1270	*NS*	0.1113	*NS*
34	0.2285	0.2159	0.2272	*NS*	0.0554	*NS*
35	0.0976	0.0795	0.0970	*NS*	0.1858	*NS*
36	0.4321	0.6250	0.4297	*p < 0.001*	-0.4500	*excess*
37	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
38	0.1937	0.2159	0.1926	*NS*	-0.1154	*NS*
39	0.1179	0.1023	0.1172	*NS*	0.1329	*NS*
40	0.1568	0.1250	0.1559	*NS*	0.2038	*NS*
41	0.1278	0.1364	0.1270	*NS*	-0.0675	*NS*
42	0.1278	0.1136	0.1270	*NS*	0.1113	*NS*
43	0.2451	0.2386	0.2437	*NS*	0.0266	*NS*
44	0.1278	0.1136	0.1270	*NS*	0.1113	*NS*
45	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
46	0.3532	0.4545	0.3512	*p < 0.05*	-0.2889	*excess*
47	0.2612	0.2614	0.2598	*NS*	-0.0005	*NS*
48	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*
49	0.1278	0.1364	0.1270	*NS*	-0.0675	*NS*
50	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
51	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
52	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
53	0.1179	0.0341	0.1172	*p < 0.001*	0.7119	*deficit*
54	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
55	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
56	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
57	0.2200	0.2273	0.2187	*NS*	-0.0333	*NS*
58	0.3828	0.5114	0.3806	*p < 0.001*	-0.3385	*excess*
59	0.2285	0.2614	0.2272	*NS*	-0.1447	*NS*
60	0.2200	0.2500	0.2187	*NS*	-0.1373	*NS*
61	0.1078	0.0455	0.1072	*p < 0.001*	0.5797	*deficit*
62	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
63	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
64	0.1568	0.1477	0.1559	*NS*	0.0583	*NS*
65	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
66	0.2114	0.2159	0.2101	*NS*	-0.0216	*NS*
67	0.1376	0.1477	0.1368	*NS*	-0.0741	*NS*
68	0.3937	0.3295	0.3914	*NS*	0.1637	*NS*
69	0.4742	0.4886	0.4715	*NS*	-0.0306	*NS*
70	0.3135	0.3864	0.3117	*p < 0.05*	-0.2340	*excess*
71	0.3205	0.3977	0.3186	*p < 0.05*	-0.2429	*excess*
72	0.3273	0.4091	0.3255	*p < 0.05*	-0.2518	*excess*
73	0.3135	0.3636	0.3117	*NS*	-0.1610	*NS*
74	0.3714	0.4659	0.3692	*p < 0.05*	-0.2564	*excess*
75	0.3771	0.4773	0.3750	*p < 0.05*	-0.2674	*excess*
76	0.3205	0.3750	0.3186	*NS*	-0.1714	*NS*

77	0.0555	0.0568	0.0552	*NS*	-0.0235	*NS*
78	0.2919	0.3295	0.2902	*NS*	-0.1299	*NS*
79	0.0555	0.0568	0.0552	*NS*	-0.0235	*NS*
80	0.0555	0.0568	0.0552	*NS*	-0.0235	*NS*
81	0.0447	0.0455	0.0444	*NS*	-0.0175	*NS*
82	0.4321	0.6023	0.4297	*p < 0.001*	-0.3968	*excess*
83	0.3828	0.5114	0.3806	*p < 0.001*	-0.3385	*excess*
84	0.3883	0.5227	0.3861	*p < 0.001*	-0.3488	*excess*
85	0.3655	0.4773	0.3634	*p < 0.05*	-0.3083	*excess*
86	0.3205	0.3977	0.3186	*p < 0.05*	-0.2429	*excess*
87	0.3532	0.4545	0.3512	*p < 0.05*	-0.2889	*excess*
88	0.3714	0.4886	0.3692	*p < 0.05*	-0.3182	*excess*
89	0.3532	0.4545	0.3512	*p < 0.05*	-0.2889	*excess*
90	0.3469	0.4432	0.3450	*p < 0.001*	-0.2794	*excess*
91	0.3883	0.5227	0.3861	*p < 0.001*	-0.3488	*excess*
92	0.3771	0.5000	0.3750	*p < 0.001*	-0.3282	*excess*
93	0.1755	0.1023	0.1745	*p < 0.05*	0.4187	*deficit*
94	0.2114	0.1477	0.2101	*p < 0.05*	0.3023	*deficit*
95	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
96	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
97	0.2114	0.2386	0.2101	*NS*	-0.1299	*NS*
98	0.1278	0.1136	0.1270	*NS*	0.1113	*NS*
99	0.2369	0.2500	0.2355	*NS*	-0.0557	*NS*
100	0.1078	0.0682	0.1072	*p < 0.05*	0.3688	*deficit*
101	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*
102	0.1179	0.0795	0.1172	*p < 0.05*	0.3263	*deficit*
103	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
104	0.2532	0.2500	0.2518	*NS*	0.0129	*NS*
105	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
106	0.4950	0.7841	0.4922	*p < 0.001*	-0.5894	*excess*
107	0.1278	0.0909	0.1270	*p < 0.05*	0.2898	*deficit*
108	0.2451	0.2386	0.2437	*NS*	0.0266	*NS*
109	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
110	0.1568	0.1023	0.1559	*p < 0.05*	0.3491	*deficit*
111	0.1568	0.1023	0.1559	*p < 0.05*	0.3491	*deficit*
112	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
113	0.3205	0.3523	0.3186	*NS*	-0.0999	*NS*
114	0.2612	0.2841	0.2598	*NS*	-0.0880	*NS*
115	0.1376	0.1023	0.1368	*NS*	0.2578	*NS*
116	0.1179	0.0795	0.1172	*p < 0.05*	0.3263	*deficit*
117	0.1937	0.2159	0.1926	*NS*	-0.1154	*NS*
118	0.2285	0.2614	0.2272	*NS*	-0.1447	*NS*
119	0.1937	0.2159	0.1926	*NS*	-0.1154	*NS*
120	0.2114	0.1932	0.2101	*NS*	0.0865	*NS*

121	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
122	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
123	0.1179	0.1023	0.1172	*NS*	0.1329	*NS*
124	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*
125	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
126	0.1376	0.1477	0.1368	*NS*	-0.0741	*NS*
127	0.1568	0.1705	0.1559	*NS*	-0.0875	*NS*
128	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
129	0.0976	0.0795	0.0970	*NS*	0.1858	*NS*
130	0.1847	0.1591	0.1836	*NS*	0.1392	*NS*
131	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
132	0.1278	0.0909	0.1270	*p < 0.05*	0.2898	*deficit*
133	0.1278	0.0909	0.1270	*p < 0.05*	0.2898	*deficit*
134	0.2844	0.3409	0.2828	*NS*	-0.2000	*NS*
135	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*
136	0.2451	0.2386	0.2437	*NS*	0.0266	*NS*
137	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
138	0.0976	0.0568	0.0970	*p < 0.05*	0.4192	*deficit*
139	0.1078	0.0682	0.1072	*p < 0.05*	0.3688	*deficit*
140	0.0976	0.0341	0.0970	*p < 0.001*	0.6520	*deficit*
141	0.0976	0.0341	0.0970	*p < 0.001*	0.6520	*deficit*
142	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
143	0.1179	0.1023	0.1172	*NS*	0.1329	*NS*
144	0.2200	0.2273	0.2187	*NS*	-0.0333	*NS*
145	0.1847	0.1364	0.1836	*p < 0.05*	0.2627	*deficit*
146	0.1376	0.1250	0.1368	*NS*	0.0920	*NS*
147	0.1278	0.0682	0.1270	*p < 0.05*	0.4679	*deficit*
148	0.1662	0.1818	0.1653	*NS*	-0.0943	*NS*
149	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
150	0.1376	0.1477	0.1368	*NS*	-0.0741	*NS*
151	0.4590	0.5455	0.4564	*NS*	-0.1897	*NS*
152	0.3205	0.3977	0.3186	*p < 0.05*	-0.2429	*excess*
153	0.4882	0.8295	0.4855	*p < 0.001*	-0.7059	*excess*
154	0.4882	0.8295	0.4855	*p < 0.001*	-0.7059	*excess*
155	0.0976	0.0568	0.0970	*p < 0.05*	0.4192	*deficit*
156	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
157	0.2114	0.2386	0.2101	*NS*	-0.1299	*NS*
158	0.1662	0.1818	0.1653	*NS*	-0.0943	*NS*
159	0.3469	0.4432	0.3450	*p < 0.05*	-0.2794	*excess*
160	0.3714	0.4886	0.3692	*p < 0.05*	-0.3182	*excess*
161	0.1847	0.1364	0.1836	*p < 0.05*	0.2627	*deficit*
162	0.0976	0.0795	0.0970	*NS*	0.1858	*NS*
163	0.1078	0.0455	0.1072	*p < 0.001*	0.5797	*deficit*
164	0.1278	0.1364	0.1270	*NS*	-0.0675	*NS*

165	0.1568	0.0795	0.1559	*p < 0.001*	0.4942	*deficit*
166	0.2026	0.2045	0.2015	*NS*	-0.0097	*NS*
167	0.1662	0.1818	0.1653	*NS*	-0.0943	*NS*
168	0.2612	0.3068	0.2598	*NS*	-0.1757	*NS*
169	0.2369	0.2727	0.2355	*NS*	-0.1523	*NS*
170	0.2114	0.2386	0.2101	*NS*	-0.1299	*NS*
171	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
172	0.2026	0.2045	0.2015	*NS*	-0.0097	*NS*
173	0.4590	0.6136	0.4564	*p < 0.05*	-0.3396	*excess*
174	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
175	0.2844	0.3409	0.2828	*NS*	-0.2000	*NS*
176	0.2992	0.2955	0.2975	*NS*	0.0127	*NS*
177	0.2844	0.2727	0.2828	*NS*	0.0413	*NS*
178	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
179	0.2114	0.1932	0.2101	*NS*	0.0865	*NS*
180	0.2612	0.2614	0.2598	*NS*	-0.0005	*NS*
181	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
182	0.2200	0.2500	0.2187	*NS*	-0.1373	*NS*
183	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
184	0.2200	0.2273	0.2187	*NS*	-0.0333	*NS*
185	0.1847	0.2045	0.1836	*NS*	-0.1083	*NS*
186	0.1376	0.0795	0.1368	*p < 0.05*	0.4233	*deficit*
187	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
188	0.0976	0.0568	0.0970	*p < 0.05*	0.4192	*deficit*
189	0.0976	0.0341	0.0970	*p < 0.001*	0.6520	*deficit*
190	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*
191	0.3937	0.3750	0.3914	*NS*	0.0478	*NS*
192	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
193	0.3273	0.4091	0.3255	*p < 0.05*	-0.2518	*excess*
194	0.3135	0.3864	0.3117	*p < 0.05*	-0.2340	*excess*
195	0.4964	0.8864	0.4936	*p < 0.001*	-0.7938	*excess*
196	0.1278	0.1364	0.1270	*NS*	-0.0675	*NS*
197	0.1278	0.1364	0.1270	*NS*	-0.0675	*NS*
198	0.4841	0.8068	0.4814	*p < 0.001*	-0.6731	*excess*
199	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
200	0.2114	0.2386	0.2101	*NS*	-0.1299	*NS*
201	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
202	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
203	0.2200	0.2273	0.2187	*NS*	-0.0333	*NS*
204	0.2285	0.2386	0.2272	*NS*	-0.0446	*NS*
205	0.3273	0.3864	0.3255	*NS*	-0.1818	*NS*
206	0.1662	0.1818	0.1653	*NS*	-0.0943	*NS*
207	0.2114	0.2386	0.2101	*NS*	-0.1299	*NS*
208	0.2114	0.2159	0.2101	*NS*	-0.0216	*NS*

209	0.2114	0.1932	0.2101	*NS*	0.0865	*NS*
210	0.2026	0.2045	0.2015	*NS*	-0.0097	*NS*
211	0.2369	0.2273	0.2355	*NS*	0.0408	*NS*
212	0.2532	0.2500	0.2518	*NS*	0.0129	*NS*
213	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
214	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
215	0.2026	0.2273	0.2015	*NS*	-0.1226	*NS*
216	0.3883	0.4773	0.3861	*p < 0.05*	-0.2307	*excess*
217	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
218	0.1179	0.1023	0.1172	*NS*	0.1329	*NS*
219	0.2768	0.3295	0.2752	*NS*	-0.1918	*NS*
220	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
221	0.1755	0.1705	0.1745	*NS*	0.0290	*NS*
222	0.1376	0.1250	0.1368	*NS*	0.0920	*NS*
223	0.1376	0.1250	0.1368	*NS*	0.0920	*NS*
224	0.3205	0.3295	0.3186	*NS*	-0.0285	*NS*
225	0.2114	0.2386	0.2101	*NS*	-0.1299	*NS*
226	0.1755	0.1932	0.1745	*NS*	-0.1013	*NS*
227	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
228	0.1755	0.1705	0.1745	*NS*	0.0290	*NS*
229	0.2026	0.2273	0.2015	*NS*	-0.1226	*NS*
230	0.0447	0.0455	0.0444	*NS*	-0.0175	*NS*
231	0.1278	0.1364	0.1270	*NS*	-0.0675	*NS*
232	0.1568	0.1477	0.1559	*NS*	0.0583	*NS*
233	0.1078	0.0682	0.1072	*p < 0.05*	0.3688	*deficit*
234	0.2691	0.2727	0.2676	*NS*	-0.0136	*NS*
235	0.4769	0.5227	0.4742	*NS*	-0.0967	*NS*
236	0.1568	0.1023	0.1559	*p < 0.05*	0.3491	*deficit*
237	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
238	0.3064	0.3750	0.3047	*p < 0.05*	-0.2254	*excess*
239	0.2285	0.2614	0.2272	*NS*	-0.1447	*NS*
240	0.0976	0.0795	0.0970	*NS*	0.1858	*NS*
241	0.1376	0.1477	0.1368	*NS*	-0.0741	*NS*
242	0.3883	0.4318	0.3861	*NS*	-0.1128	*NS*
243	0.3532	0.3864	0.3512	*NS*	-0.0943	*NS*
244	0.4901	0.4773	0.4873	*NS*	0.0264	*NS*
245	0.3532	0.3636	0.3512	*NS*	-0.0296	*NS*
246	0.4919	0.4659	0.4891	*NS*	0.0531	*NS*
247	0.4935	0.8636	0.4907	*p < 0.001*	-0.7576	*excess*
248	0.2532	0.2955	0.2518	*NS*	-0.1678	*NS*
249	0.4405	0.6477	0.4380	*p < 0.001*	-0.4746	*excess*
250	0.2451	0.2841	0.2437	*NS*	-0.1600	*NS*
251	0.5005	0.9091	0.4977	*p < 0.001*	-0.8249	*excess*
252	0.3714	0.4886	0.3692	*p < 0.05*	-0.3182	*excess*

253	0.4278	0.6136	0.4254	*p < 0.001*	-0.4380	*excess*
254	0.5018	0.9545	0.4989	*p < 0.001*	-0.9121	*excess*
255	0.2768	0.3295	0.2752	*NS*	-0.1918	*NS*
256	0.5012	0.9432	0.4984	*p < 0.001*	-0.8913	*excess*
257	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
258	0.5023	0.9659	0.4994	*p < 0.001*	-0.9333	*excess*
259	0.5026	0.9773	0.4998	*p < 0.001*	-0.9551	*excess*
260	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
261	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
262	0.2992	0.3636	0.2975	*NS*	-0.2168	*NS*
263	0.3405	0.4318	0.3386	*p < 0.05*	-0.2701	*excess*
264	0.2451	0.2841	0.2437	*NS*	-0.1600	*NS*
265	0.1755	0.1932	0.1745	*NS*	-0.1013	*NS*
266	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
267	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
268	0.1847	0.2045	0.1836	*NS*	-0.1083	*NS*
269	0.1179	0.1023	0.1172	*NS*	0.1329	*NS*
270	0.4187	0.5682	0.4164	*p < 0.001*	-0.3598	*excess*
271	0.4187	0.5682	0.4164	*p < 0.001*	-0.3598	*excess*
272	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
273	0.2451	0.2841	0.2437	*NS*	-0.1600	*NS*
274	0.5029	1.0000	0.5000	*p < 0.001*	-1.0000	*excess*
275	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
276	0.0976	0.0795	0.0970	*NS*	0.1858	*NS*
277	0.0976	0.0795	0.0970	*NS*	0.1858	*NS*
278	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
279	0.1568	0.1705	0.1559	*NS*	-0.0875	*NS*
280	0.2919	0.3523	0.2902	*NS*	-0.2083	*NS*
281	0.2200	0.1818	0.2187	*NS*	0.1744	*NS*
282	0.1473	0.1591	0.1465	*NS*	-0.0807	*NS*
283	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
284	0.4841	0.8068	0.4814	*p < 0.001*	-0.6731	*excess*
285	0.4655	0.7273	0.4628	*p < 0.001*	-0.5676	*excess*
286	0.4882	0.8295	0.4855	*p < 0.001*	-0.7059	*excess*
287	0.4590	0.7045	0.4564	*p < 0.001*	-0.5398	*excess*
288	0.2532	0.2727	0.2518	*NS*	-0.0774	*NS*
289	0.1179	0.1023	0.1172	*NS*	0.1329	*NS*
290	0.1179	0.0568	0.1172	*p < 0.05*	0.5193	*deficit*
291	0.1179	0.0568	0.1172	*p < 0.05*	0.5193	*deficit*
292	0.3532	0.3636	0.3512	*NS*	-0.0296	*NS*
293	0.1179	0.1250	0.1172	*NS*	-0.0610	*NS*
294	0.1376	0.1250	0.1368	*NS*	0.0920	*NS*
295	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*
296	0.1078	0.0909	0.1072	*NS*	0.1574	*NS*

297	0.1278	0.1136	0.1270	*NS*	0.1113	*NS*
298	0.4590	0.6136	0.4564	$p < 0.05$	-0.3396	*excess*
299	0.0976	0.1023	0.0970	*NS*	-0.0482	*NS*
300	0.1662	0.1818	0.1653	*NS*	-0.0943	*NS*
301	0.1078	0.1136	0.1072	*NS*	-0.0545	*NS*
302	0.1473	0.1136	0.1465	*NS*	0.2294	*NS*

Command 4-A3

Command 4-A3: Command lines of hard filtering criteria applied to SNPs called using GATK tools

 # create the VCT file containing only SNP using SelectVariants

java -jar gatk-package-local.jar SelectVariants -R concatenated_350Ns_assembly.fasta -V final.88.vcf --select-type-to-include SNP -O raw_SNPs_final.88.vcf

#apply criteria for hard filtering that output a new VCF file named "*filtered.final.88.vcf***" containing all variant sites with the annotations applied accordingly to the filter names, i.e.** *"badseq", "badMap"*

java -jar gatk-package-local.jar VariantFiltration -R concatenated_350Ns_assembly

-V raw_SNPs_final.88.vcf

*--filter-expression "(vc.isSNP() && (vc.hasAttribute('ReadPosRankSum') && ReadPosRankSum < -8.0)) || (vc.hasAttribute('QD') && QD < 2.0) " --filter-name "badSeq" *

*--filter-expression "(vc.isSNP() && ((vc.hasAttribute('MQ') && MQ < 40.0) " --filter-name "badMap" *

-O filtered.final.88.vcf

#create VCF file containing only the filter-passed SNPs

java -jar gatk-package-local.jar SelectVariants -R concatenated_350Ns_assembly -V filtered.final.88.vcf --exclude-filtered true -O filter_passed_variant_sites.vcf

Command 4-A4

Command 4-A4: Command lines used to create VCF file for parentage analysis

#filter out SNPs that were present in less than <100% of individuals

vcftools --vcf filter_passed_variant_sites.vcf --max-missing 1.0 --recode --recode-INFO-all --out filter_passed_variant_sites_100pct.vcf

#filter out SNPs with minor allele count < 3

vcftools --vcf filter_passes_variant_sites_100pct.vcf.recode.vcf --mac 3 --recode --recode-INFO-all --out filter_passed_variant_sites_100pct_mac3.vcf

#filter out SNPs that showed *MAF* > 0.05

vcftools --vcf out filter_passed_variant_sites_100pct_mac3.vcf --out filtered_100pct_mac3_nonref005.vcf --non-ref-af 0.05 --recode --recode-INFO-all

Command 4-A5

Command 4-A5: Command lines used to estimate kinship coefficient

#estimate population allele frequencies

angsd -b bam_list.txt -ref -R concatenated_350Ns_assembly -gl 2 -domajorminor 1 -snp_pval 1e-6 -domaf 1 -minmaf 0.05 -doglf 3 -remove_bads 1 -baq 1 -minMapQ 20 -minQ 20 -minInd 44 -doCounts 1 -minIndDepth 15 -setMinDepth 660 -setMaxDepth 6600 -doDepth 1 -dumpCounts 1 -out maf_ngsrelate

#extract only the column containing allele frequencies, remove the header

#this is the input format required by NGSrelate

zcat maf_ngsrelate_44_15.mafs.gz | cut -f 5 | sed 1d > freq_ngsrelate

#estimate mean kinship coefficient and other summary statistics

/ngsRelate -g maf_ngsrelate.glf -n 88 -f freq_ngsrelate > out_ngsrelate

Table 4-A6 Pedigree-based individual kinships of SWRs from Botswana1 estimated in Chapter 4. [a] Candidates for translocation identified using pedigree-based $MK_{Botswana1}$ + 2SD, [b] Candidates for breeding identified using pedigree-based $MK_{Botswana1}$ - 2SD

Rank	ID	MK-ID	Rank	ID	MK-ID
1	999 [a]	0.09040179	29	140	0.05133929
1	124 [a]	0.09040179	30	200	0.0499442
3	146 [a]	0.08398438	30	271	0.0499442
3	255 [a]	0.08398438	32	210	0.04910714
5	120 [a]	0.07756696	33	121	0.04464286
5	176 [a]	0.07756696	34	204	0.04352679
7	114	0.06863839	34	274	0.04352679
7	131	0.06863839	36	142	0.04129464
7	155	0.06863839	37	889	0.0390625
10	136	0.06752232	37	143	0.0390625
10	171	0.06752232	37	154	0.0390625

10	201	0.06752232	40	151	0.03794643
10	242	0.06752232	41	205	0.03766741
14	156	0.06668527	42	166	0.03571429
15	113	0.06640625	42	133	0.03571429
15	320	0.06640625	44	153	0.03459821
17	125	0.06194196	44	175	0.03459821
18	203	0.06110491	44	174	0.03459821
19	172	0.06082589	47	135	0.03348214
20	130	0.05580357	47	132	0.03348214
20	141	0.05580357	49	206 [b]	0.03236607
22	207	0.05440848	49	275 [b]	0.03236607
23	147	0.05357143	51	888 [b]	0.02455357
24	167	0.05329241	51	316 [b]	0.02455357
24	202	0.05329241	53	177 [b]	0.02232143

26	111	0.05189732	53	170 [b]	0.02232143
26	144	0.05189732	55	272 [b]	0.01339286
26	145	0.05189732	55	127 [b]	0.01339286