Association Analysis of Coat Colour in Blue Wildebeest (*Connochaetes taurinus taurinus*)

by

Riana van Deventer

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science at Stellenbosch University

Supervisor: Rouvay Roodt-Wilding, Ph.D.

Co-Supervisors: Clint Rhode, Ph.D., Pr. Sci. Nat.

Munro Marx, Ph.D., M.B.A.

Department of Genetics

December 2019

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Riana van Deventer

December 2019

ABSTRACT

The blue wildebeest (*Connochaetes taurinus taurinus*) is an economically important African antelope species that is regularly sold on game auctions and is widely utilised in commercial hunting in South Africa. A colour variant of the blue wildebeest, known as the golden wildebeest, has become one of the most common colour variant animals that wildlife ranchers breed with. Despite the economic importance of blue wildebeest, very few genomic resources are currently available for this species that can be used for management and research. This study aimed to use a genotyping-by-sequencing (GBS) approach to discover and genotype single nucleotide polymorphisms (SNPs) in blue wildebeest, and to use these SNPs in a case-control genome-wide association study (GWAS) to identify putative candidate genes involved in blue wildebeest pigmentation. The sample population consisted of 94 blue wildebeest, which included 35 blue wildebeest with the golden coat colour variation. The DArTseq genotyping platform was used for SNP discovery and a total of 20 563 SNPs, each located in a 69 bp marker sequence, were identified. The generated SNP markers were of high quality with a high average reproducibility (>99%) and a low percentage (~9.21%) missing data. In addition, the DArTseq platform was able to generate a large number of informative SNPs for blue wildebeest, exhibiting high heterozygosity and resolving power among the population samples. Furthermore, the *Bos taurus* genome was used for the *in silico* mapping of the marker sequences to estimate the distribution of the generated SNPs throughout the genome and to identify putative orthologous genes. A total of 6 020 (29.28%) SNP sequences were successfully mapped against the bovine genome of which 3 907 mapped to putative gene orthologues. The SNP sequences mapped to all of the bovine chromosomes establishing their genome-wide distribution. Finally, functional annotation of the marker sequences revealed a wide range of different putative functions. The GWAS identified 377 SNPs that were significantly associated with coat colour in blue wildebeest. Eight of these DArTseq SNP markers mapped to seven different genes that were considered as putative candidate genes for coat colour determination based on previous literature reports. Based on their reported biological function, the putative candidate genes for coat colour determination in blue wildebeest are *myosin VC* (*MYO5C*), *myosin VIIA* (*MYO7A*), *solute carrier family 6 member 3* (*SLC6A3*), *solute carrier family 28 member 2* (*SLC28A2*), *dopamine receptor D2* (*DRD2*), *frizzled class receptor 4* (*FZD4*) and *tyrosinase* (*TYR*). The established mode of inheritance for the majority of the SNPs located in these candidate genes were additive. This observation, together with the limited gene-gene interaction observed between the candidate SNPs, suggest that coat colour in blue wildebeest could be a quantitative threshold trait rather than a simple autosomal recessive trait as originally suspected. This study represents the first large genetic polymorphism discovery performed in an African antelope species that is farmed commercially. The generated DArTseq markers could be a useful resource for the genetic management of ranched blue wildebeest, and could also be used to facilitate research in this species. Furthermore, this study provides a foundation to further investigate the genetic underpinning of coat colour in blue wildebeest.

OPSOMMING

Die blou wildebees (*Connochaetes taurinus taurinus*) is ‘n wildsbokspesie van Afrika wat gereeld op wildsveilings verkoop word en van kommersiële belang is in die Suid-Afrikaanse jagbedryf. ‘n Kleurvariant van die blou wildebees, bekend as die goue wildebees, het onlangs een van die mees algemene kleurvariante geword waarmee wildsboere teël. Ten spyte van die ekonomiese belang van die blou wildebees, is die genomiese hulpbronne beskikbaar vir populasiebestuur en navorsing baie beperk vir hierdie spesie. Hierdie studie het gepoog om ‘n genotipering-deur-volgordebepaling (GBS) benadering te gebruik om enkel nukleotied polimorfismes (SNPs) te identifiseer en te genotipeer in die genoom van die blou wildebees. Verder, is hierdie SNPs gebruik in ‘n genoom-wye assosiasie studie (GWAS) om kandidaatgene te identifiseer wat betrokke is by pigmentasie in blou wildebeeste. Die steekproefpopulasie het bestaan uit 94 blou wildebeeste, wat 35 individue met die goue pelskleur ingesluit het. Die DArTseq platform is gebruik vir SNP identifikasie. ‘n Totaal van 20 563 SNPs, elkeen geleë in ‘n 69 bp merkervolgorde, is geïdentifiseer. Die geïdentifiseerde SNPs was van ‘n hoë kwaliteit, aangesien die SNPs hoogs reproduseerbaar was (>99%) en die persentasie ontbrekende data in die datastel laag was (~9.21%). Verder, was die DArTseq platform in staat om ‘n groot aantal informatiewe SNPs te genereer wat hoë heterosigositeit en diskriminasievermoë getoon het tussen die individue van die steekproefpopulasie. Die *Bos taurus* genoom is gebruik vir die *in silico* kartering van die merkervolgordes om, sodoende, die verspreiding van die SNPs deur die genoom te beraam en om ortoloë-gene te identifiseer. ‘n Totaal van 6 020 (29.28%) van die SNP merkervolgordes was suksesvol karteer teen die beesgenoom, waarvan 3 907 karteer is teen ortoloë-gene. Die merkervolgordes is karteer oor al die chromosome van die bees, wat die genoom-wye verspreiding van die SNPs bevestig het. Verder, het funksionele annotasie van die merkervolgordes ‘n wye verskeidenheid funksies onthul. Die GWAS het 377 SNPs identifiseer wat beduidend geassosieer was met pelskleur in die blou wildebees. Agt van hierdie DArTseq SNPs was gekarteer teen sewe verskillende gene wat oorweeg was as kandidaatgene vir die bepaling van pelskleur gebaseer op vorige literatuurverslae. Hierdie kandidaat gene is *miosien VC* (*MYO5C)*, *miosien VIIA* (*MYO7A*), *oplossingsdraer famile 6 lid 3* (*SLC6A3*), *oplossingsdraer 28 lid 2* (*SLC28A2*), *dopamien reseptor D2* (*DRD2*), *frizzel-klas reseptor 4* (*FZD4*) and *tirosinase* (*TYR*). Die meeste van die SNPs geleë in die kandidaatgene het ‘n additatiewe oorerwingsmodel getoon. Hierdie bevinding, tesame met die beperkte geen-geen interaksie wat waargeneem is tussen die kandidaat SNPs, stel voor dat pelskleur in die blou wildebees moontlik ‘n kwantitatiewe drumpel-eienskap is eerder as ‘n outosomaal resessiewe eienskap soos oorspronklik vermoed. Hierdie studie verteenwoordig die eerste grootskaalse polimorfisme identifikasie wat uitgevoer is in ‘n wildsbokspesie waarmee kommersieël geboer word. Hierdie nuwe SNP merkers kan ‘n nuttige hulpbron wees vir die genetiese bestuur van blou wildebees populasies en kan ook gebruik word om navorsing in hierdie spesie te fasiliteer. Laastens, dien hierdie studie as ‘n fondasie vir verdere ondersoek in verband met die onderliggende genetiese basis van pelskleur in die blou wildebees.

ACKNOWLEDGEMENTS

I wish to acknowledge the following institutions for their contributions to this study (in alphabetical order): Castle de Wildt (LBG Holdings), Unistel Medical Laboratories (Pty) Ltd, University of Stellenbosch. This research project would not have been possible without their aid and support.

I would like to express my sincere gratitude to my supervisor, Prof. Rouvay Roodt-Wilding, and co-supervisors, Dr Clint Rhode and Dr Munro Marx, for their academic guidance and support. Thank you for your valuable insights, patience and steadfast encouragement. I would also like to thank Dr Barbara Picone for assisting with the mapping of the marker sequences and Annesca Joubert for her insightful suggestions.

A very special thanks is due to my friends and colleagues for their non-academic support. Thank you for your prayers, encouragement and for taking an interest in my research project. I am especially grateful to all the members of the Unistel game team. Thank you for being so understanding and taking on extra work whenever I was off to a meeting or busy with my write-up. Lastly, I would like to thank my loving family for providing me with unfailing support and continuous encouragement. Thank you for always believing in me.

PROJECT CONTRIBUTIONS

Oral presentation – 37th International Society for Animal Genetics Conference, Lleida, Spain (7-12 July 2019)

Association Analysis of Coat Colour in Blue Wildebeest.

Poster presentation – 37th International Society for Animal Genetics Conference, Lleida, Spain (7-12 July 2019)

Association Analysis of Coat Colour in Blue Wildebeest.

Articles to be submitted for review – *Animal Genetics*

(i) *The Development of Genome-Wide Single Nucleotide Polymorphism (SNP) Markers in Blue Wildebeest using the DArTseq Platform.*

(ii) *The Elucidation of Coat Colour Genetics in Blue Wildebeest.*

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

APPENDICES

APPENDIX C

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

APPENDICES

APPENDIX A

APPENDIX B

APPENDIX D

LIST OF ABBREVIATIONS AND SYMBOLS

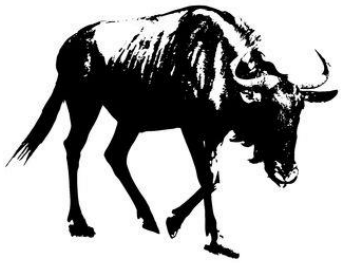
αMSH	α-Melanocyte stimulating hormone
~	Approximately
°	Degree
˚C	Degree Celsius
=	Equals
>	Greater than
<	Less than
≤	Less than or equal to
µl	Microliter
'	Minute
%	Percentage
±	Plus or minus
®	Registered trademark
×	Times
A	Adenine
ACTH	Adrenocorticotropic hormone
AFLP	Amplified fragment length polymorphism
AHCY	Adenosylhomocysteinase
AIC	Akaike Information Criterion
AIM-1	Absent in melanoma-1 protein
ASIP	Agouti signalling protein
ASO	Allele-specific oligonucleotide
BGD	Bovine Genome Database
BIC	Bayesian Information Criterion
Bp	Base pair
Blast	Basic local alignment search tool
Blastn	Nucleotide basic local alignment search tool (blast)
Blastx	Translated nucleotide query basic local alignment search tool (blast)
C	Cytosine
CAPS	Cleaved amplified polymorphic sequences
CI	Confidence interval
Contig	Contiguous sequence
D'	Lewontin standardised disequilibrium coefficient
DArT	Diversity Arrays Technology
DArTseq	Diversity Arrays Technology Sequencing

DCT	Dopachrome tautomerase
ddRAD	Double digest restriction associated DNA
DEA	Department of Environmental Affairs
DGGE	Denaturing gradient gel electrophoresis
DHI	5,6-Dihydroxyindole
DHICA	5,6-Dihydroxyindole (DHI) -2-carboxylic acid
DNA	Deoxyribonucleic acid
DOPA	3,4-Dihydroxyphenylalanine
DRD2	Dopamine receptor D2
E	East
E-value	Expect value
EDN3	Endothelin 3
EDNRB	Endothelin receptor type B
EDTA	Ethylenediamine tetraacetic acid
e.g.	*exempli gratia* (for example)
ER	Endoplasmic reticulum
ESPCR	European Society for Pigment Cell Research
EST	Expressed sequence tag
etc.	*et cetera* (and so forth)
et al.	*et alia* (and others)
EWT	Endangered Wildlife Trust
FASTQ	Fast alignment search tool-quality
FZD4	Frizzled class receptor 4
G	Guanine
Gbp	Gigabase pairs
GBS	Genotyping-by-sequencing
GDP	Gross domestic product
GGT	Gamma-glutamyl transpeptidase-encoding protein
GO	Gene ontology
gp75	Tyrosinase-related protein 1
gp100	Pre-melanosome protein
GWAS	Genome-wide association study
Hexp	Expected heterozygosity
Ho	Observed heterozygosity
HPS1	Hermansky-Pudlak syndrome 1
HPS6	Hermansky-Pudlak syndrome 6
HWE	Hardy-Weinberg equilibrium

ID	Identifier
i.e.	*id est* (that is)
IGFBP7	Insulin-like growth factor-binding protein 7
ING3	Inhibitor of growth protein 3
ITCH	Itchy E3 ubiquitin protein ligase
K	Thousand
Kbp	Kilobase pairs
KIT	Tyrosine kinase receptor
LD	Linkage disequilibrium
LRP	Low density lipoprotein receptor-related protein
LYST	Lysosomal trafficking regulator
MAF	Minor allele frequency
MART-1	Melanoma-associated antigen recognised by T cells (Melan-A)
MATP	Membrane-associated transporter protein
Mbp	Megabase pairs
MC1R	Melanocortin 1 receptor
MITF	Microphthalmia-associated transcription factor
MLPH	Melanophilin
MYO5	Myosin V
MYO5A	Myosin VA
MYO5C	Myosin VC
MYO7A	Myosin VIIA
NA	Not applicable
NAMC	National Agricultural Marketing Council
NCBI	National Centre for Biotechnology Information
NCC	Neural crest cell
ng/µl	Nanogram per microliter
NGS	Next-generation sequencing
Nr	Non-redundant
nsSNP	Non-synonymous single nucleotide polymorphism
OA1	Ocular albinism type 1
OCA2	Oculocutaneous albinism II
OR	Odds ratio
PANTHER	Protein Analysis Through Evolutionary Relationships
PAX3	Paired box 3
PCR	Polymerase chain reaction
PDGFRA	Platelet-derived growth factor receptor alpha

PIC	Polymorphic information content
PMEL	Pre-melanosome protein
POMC	Proopiomelanocortin
(Pty) Ltd	Proprietary limited
p-value	Probability value
Q-score	Quality score
r	Pearson correlation coefficient
r^2	Squared correlation coefficient between the alleles of two loci
R^2	Coefficient of determination
Rab27	Ras-related protein Rab-27
Rab27A	Ras-related protein Rab-27A
RADseq	Restriction-site associated DNA sequencing
RALY	RNA-binding protein Raly
RAPD	Random amplified polymorphic DNA
RE	Restriction enzyme
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RPE	Retinal pigment epithelium
RRL	Reduced representation library
RRS	Reduced representation sequencing
S	South
SCF	Stem cell factor
SE	Standard error
SD	Standard deviation
SILV	Silver
SLC	Solute carrier
SLC6A3	Solute carrier family 6 member 3
SLC7A11	Solute carrier family 7 member 11
SLC24A5	Solute carrier family 24 member 5
SLC28A2	Solute carrier family 28 member 2
SLC36A1	Solute carrier family 36 member 1
SLC45A2	Solute carrier family 45 member 2
SNP	Single nucleotide polymorphism
SOX10	Sex-determining region Y (SRY)-box 10
SRY	Sex-determining region Y
SSCP	Single strand conformational polymorphism
SSR	Simple sequence repeat

T	Thymine
TDT	Transmission disequilibrium test
TGN	Trans-Golgi network
TM	Trademark
TRP	Tyrosinase-related protein
TRP1	Tyrosinase-related protein 1
TRP2	Tyrosinase-related protein 2
Ts	Transition
Ts:Tv	Transition to Transversion ratio
Tv	Transversion
TYR	Tyrosinase
UCSC	University of California, Santa Cruz
UVR	Ultraviolet radiation
vs	*Versus*
Wnt	Wingless/Integrated
WNT16	Wingless/Integrated (Wnt) family member 16
ZAR	South African Rand (currency)



"A more whimsical compound than the Gnu could scarcely have been thrown together, or a monster imagined of a more fantastical and anomalous exterior"

- W. Cornwallis Harris, 1840 -

CHAPTER 1

Literature Review & Introduction

The game colour variant market has become a fundamental part of the South African wildlife industry over the past years. A coat colour variant of the blue wildebeest (*Connochaetes taurinus taurinus*), known as the golden wildebeest, especially has become a profitable game colour variant to breed with. Although the blue wildebeest and its golden coat colour variant is of economic importance, limited resources are available for their genetic management, and the genetic architecture underlying the aesthetic golden coat colour phenotype has also not been investigated.

The rapid progress and steadily lowering cost of sequencing technologies, together with the development of advanced bioinformatics tools, have greatly facilitated the large-scale discovery of genomic resources in non-model species. This revolution in genomic technologies has also opened-up new avenues for studying the genetic basis of animal colouration, surpassing the limitations of traditional approaches to a great extent. Research concerning the golden coat colour variant of blue wildebeest may thus benefit significantly from the application of high-throughput sequencing data.

1.1. The South African Wildlife Ranching Industry

1.1.1. A Brief Overview

In recent years, the wildlife ranching industry has become one of the fastest growing sectors of the South African economy. The industry has developed from simply being an alternative method of production on marginal land to a multi-billion South African Rand (ZAR) enterprise that contributes to biodiversity preservation, ecotourism and employment creation (Reilly *et al.*, 2003; Taylor *et al.*, 2016). The economic potential of wildlife along with the promulgation of the Game Theft Act in 1991, which enabled private ownership of game animals, were the greatest driving forces in the rapid development of the industry (Van Hoven, 2015; Taylor *et al.*, 2016).

Reports suggest that South Africa currently has approximately 9 000 wildlife ranches, which cover an area of 20 million hectares [National Agricultural Marketing Council (NAMC), 2006; Oberem, 2015; Taylor *et al.*, 2016]. These wildlife ranches harbour more than 16 million game animals – in comparison there are approximately 14 million head of cattle in South Africa (Jaja *et al.*, 2016; Taylor *et al.*, 2016). The majority of wildlife ranches (~80%) are

situated in the Eastern Cape, Northern Cape and Limpopo, with the remainder spread across the other provinces of South Africa (Taylor *et al.*, 2016).

The main contributors to the income generated by wildlife ranching can be divided into four subsectors, generally referred to as "pillars", namely (1) hunting, (2) game meat production, (3) ecotourism and (4) breeding and live sales (Oberem, 2015; Taylor *et al.*, 2016). Together these four subsectors annually contribute more than ZAR 20 billion to the gross domestic product (GDP) of the country (Oberem, 2015). Figure 1.1 illustrates the contribution of each subsector to the economy of the wildlife industry.

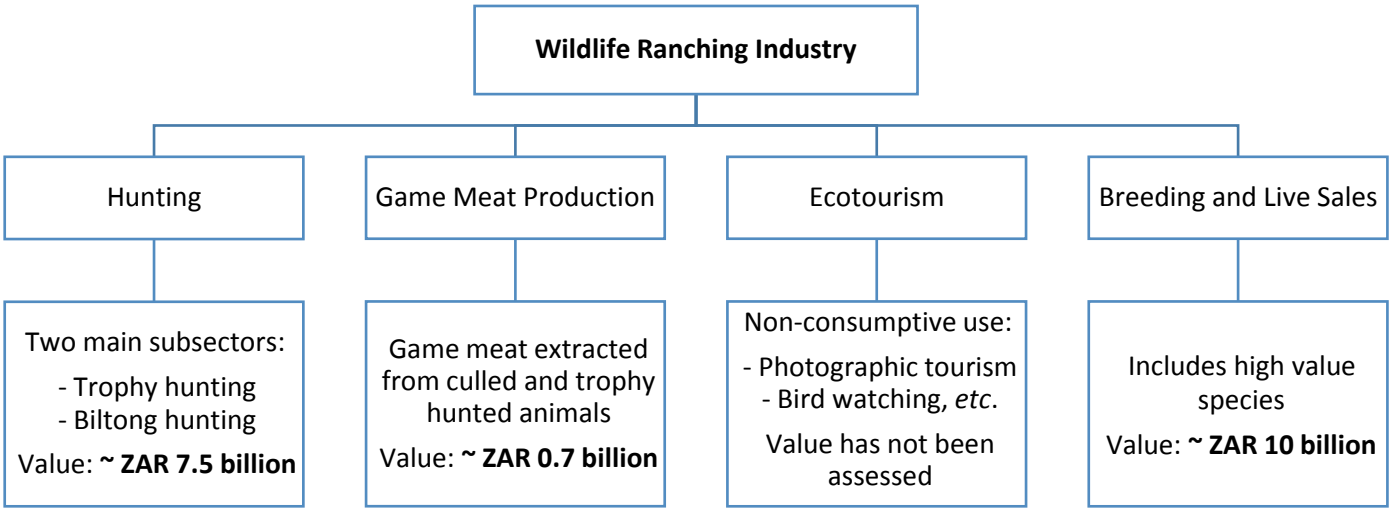


Figure 1.1. The four subsectors (pillars) of the South African wildlife ranching industry (Van der Merwe & Saayman, 2005; Oberem, 2015; Janovsky, 2016; Taylor *et al.*, 2016).

1.1.2. The Wildlife Breeding Industry

The breeding and live sales subsector of wildlife ranching has been responsible for the majority of the growth and development in the industry during the past few years (Cloete, 2015a). The first official wildlife auction in South Africa was held in 1975 and generated a revenue of ZAR 20 362 (Taylor *et al.*, 2016). The revenue generated from official game auctions has since increased from ZAR 9 million in 1991 to a historic high of ZAR 1.8 billion in 2014. With annual inflation rates taken into account, this represents an average annual increase of 18% over a period of 23 years (Taylor *et al.*, 2016). The growth in live auction sales has, however, became more stable in recent years, contributing ZAR 1.7 billion to the economy in 2016 [Department of Environmental Affairs (DEA), 2016;

2018]. Considering that only 20% of game is sold at formal auctions, the revenue generated by game breeding is currently estimated to be in excess of ZAR 10 billion per annum (Cloete, 2015b; Janovsky, 2016).

Over the years, wildlife ranching has shifted from breeding large numbers of common game, such as impala (*Aepyceros melampus*) and kudu (*Tragelaphus strepsiceros*), towards breeding fewer high-value animals under intensive conditions (Lindsey *et al.*, 2009; Pitman *et al.*, 2016). Due to this continued shift, the demand for high-value game animals has increased and, consequently, the profitability associated with the breeding of these animals has also significantly improved (Cloete, 2015b). High-value game can generally be divided into the following subcategories: rare game, trophy animals and colour variants (Taylor *et al.*, 2016). Rare game refers to game species that are not common and, therefore, are also valuable, while trophy animals refer to game animals with exceptional measurements for desirable traits, such as horn length (Strauss, 2015). Colour variants, also known as colour morphs, are game animals that express a rare colour phenotype. These naturally occurring colour variants are not considered to be a separate species, subspecies or hybrid, but are thought to be the result of genetic mutations that occur in wild populations (Taylor *et al.*, 2016). In recent years, colour variations of African antelope species have become a topic of great interest in the South African wildlife industry and, consequently, have received extensive media coverage (Bezuidenhout, 2012; Van Rooyen, 2012; Strauss, 2015; Heyneke, 2017).

1.1.2.1. The Colour Variant Market and the Associated Controversy

Game colour variants are not a new phenomenon and are known to have occurred naturally in wild populations in the past. The first sighting of a black springbuck (*Antidorcas marsupialis*) was documented in 1886 (Kruger *et al.*, 1979; Olivier, 2015). Furthermore, in the 1920s, golden wildebeest were discovered in the Limpopo River basin, adjacent to the Tuli block of Botswana, while golden oryx (*Oryx gazelle*) were historically widespread and was first recorded to be hunted in 1906 (Oberem, 2015; Olivier, 2015). Colour variants have since become significantly more prevalent on wildlife ranches and selective breeding has also resulted in the emergence of numerous new colour variants (Strauss, 2015; Taylor *et al.*, 2016). Currently, there are more than 40 distinct colour variations of African Bovidae (Table 1.1), of which the numerous springbuck colour forms (Figure 1.2) are presumably the best known example (Olivier, 2015). The main reason for the increased prevalence of colour variants has been financial gain as they have become sought after by wildlife ranchers (Taylor *et al.*, 2016). For example, in 2016, a common impala sold for an average auction price of ZAR 1 721, while a black impala and a white impala sold for an average auction price of ZAR 62 291 and ZAR 48 333, respectively (Cloete, 2017a).

Table 1.1. Bovidae colour variants commonly found on wildlife ranches in South Africa (Olivier, 2015; Cloete, 2017b).

Game Species	Common Colour Variations
Blesbuck (*Damaliscus pygargus phillipsi*)	apache, copper, painted, red, saddleback, white, yellow
Blue Wildebeest (*Connochaetes taurinus taurinus*)	golden, kings
Eland (*Tragelaphus oryx*)	Cape, Livingstone, white
Impala (*Aepyceros melampus*)	black, saddleback, white, white flanked
Kudu (*Tragelaphus strepsiceros*)	black, white
Oryx (*Oryx gazelle*)	golden, red, white
Springbuck (*Antidorcas marsupialis*)	black, blue, bont, coffee, copper, cream, ivory, white

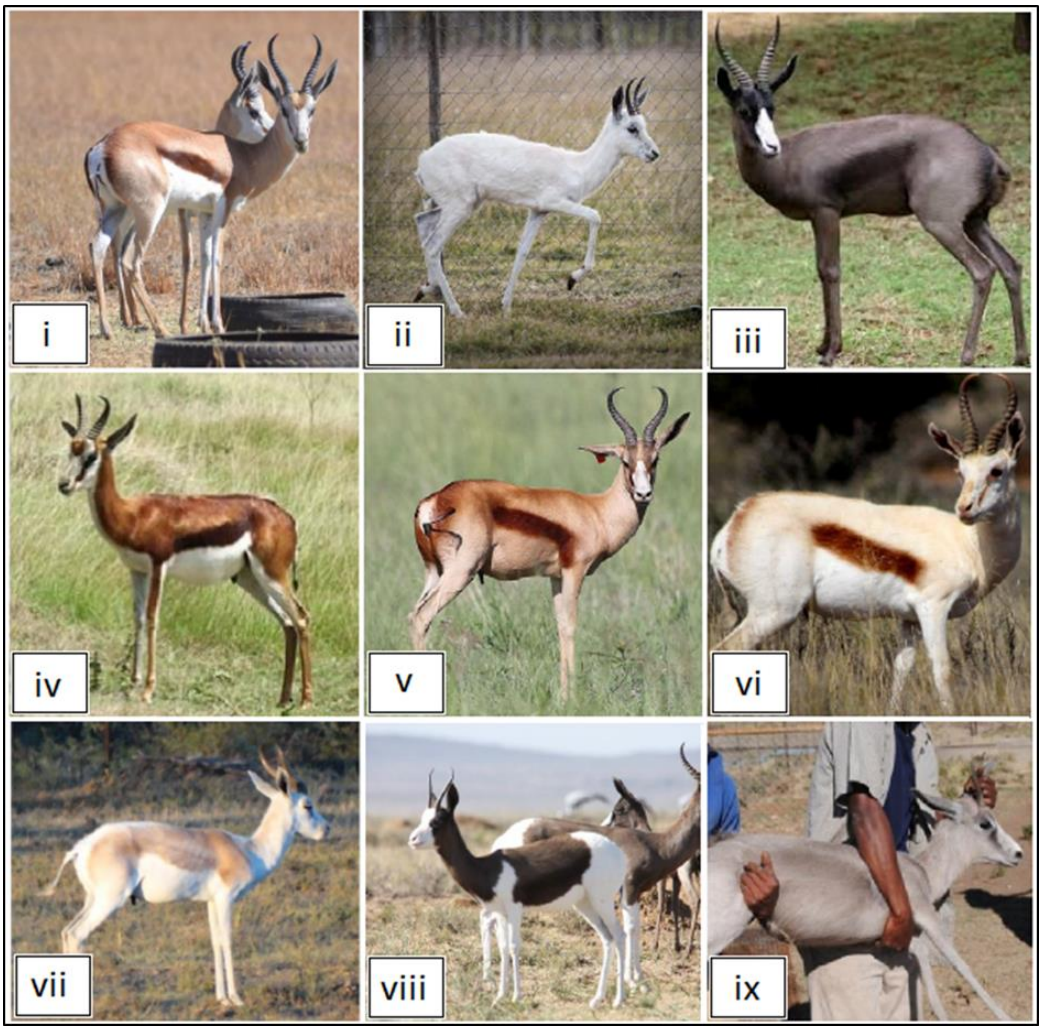


Figure 1.2. Colour variations of springbuck (*Antidorcas marsupialis*): (i) common springbuck, (ii) white, (iii) black, (iv) coffee, (v) copper, (vi) ivory, (vii) cream, (viii) bont, (ix) blue (figures taken from Olivier, 2015; Furstenburg, 2016).

The intensive breeding of colour variants has, however, become a controversial topic within the wildlife industry. The lack of published scientific data regarding colour variant game has sparked fierce debate between those who support the breeding of colour variants and those who are opposed to it, preferring minimal intervention in game breeding (Olivier, 2015; Van Hoven, 2015). The greatest dangers related to the ongoing selective breeding of colour variants are considered to the loss of genetic diversity due to inbreeding, genetic drift and the founder effect (Taylor *et al.*, 2016; Russo *et al.*, 2019). Selective breeding can, therefore, potentially result in the fixation of deleterious alleles that might be inherited together with coat colour. The reproductive potential of the breeding herd and the evolutionary potential of the species can also be affected by the loss of allelic diversity and heterozygosity (Russo *et al.*, 2019). In addition, mutations associated with coat colour often have pleiotropic effects including skin diseases, such as melanoma, and disorders of the immune system, reproductive tract and sensory organs. If colour variations in game species are the result of such deleterious mutations it could have severe consequences for animal welfare (Reissmann & Ludwig, 2013; Charon & Lipka, 2015).

Since colour variant game have become such an integral part of the South African wildlife industry, research regarding these animals has become increasingly important to facilitate breeding management and to assist in the regulation of colour variant Bovidae (Olivier, 2015). A well-founded breeding strategy must be supported by scientific knowledge and, therefore, there is an urgent need to investigate the genetic underpinning of the coat colours that are currently being selected for in game animals. The broad knowledge base on pigmentation genetics in model organisms and other animals will be able to assist in the elucidation of coat colour variation in game animals (Russo *et al.*, 2019).

1.2. The Biology and Genetics of Pigmentation

The study of pigmentation genetics in model organisms has a long and rich history. As a result, more than 150 genes have been identified that have an effect on animal colouration and patterning, either directly or indirectly (Cieslak *et al.*, 2011; D'Mello *et al.*, 2016). The majority of these genes were first discovered in laboratory mice (*Mus musculus*), although, the comparative mapping of numerous pigmentation loci has enabled the identification and characterisation of the corresponding orthologous genes in several mammalian species (Sturm, 2006; Nicoloso *et al.*, 2008; André *et al.*, 2017). The genes that play a role in pigmentation are generally involved in the development and survival of pigment producing melanocytes, the regulation of pigment synthesis or in pigment transport and transfer (Cieslak *et al.*, 2011; D'Mello *et al.*, 2016).

1.2.1. Melanocytes: Biology and Development

Melanocytes are responsible for the production of the melanin pigment in mammals, reptiles, birds and fish (Sturm, 2006; D'Mello *et al.*, 2016). Melanocytes are dendritic cells of the neural ectoderm. The unpigmented precursor cells of melanocytes, known as melanoblasts, are derived from neural crest cells (NCCs). During embryonic development, after the neural tube has closed, melanoblasts migrate to different regions of the body where they develop into melanocytes, as well as the choroid of the eye, bone and cartilage of the head skeleton and cells of the peripheral nervous system. Melanoblasts that differentiate into melanocytes are mainly located in hair follicles and the basal layer of the epidermis (D'Mello *et al.*, 2016).

The development, migration and survival of melanocytes depend on several transcription factors and signalling systems. Among the transcription factors crucial for melanocyte specification and development are paired box 3 (PAX3), sex-determining region Y (SRY)-box 10 (SOX10) and the microphthalmia-associated transcription factor (MITF), a basic helix-loop-helix leucine zipper protein regarded as the pigment cell master regulator (Hou *et al.*, 2000; Colombo *et al.*, 2011; D'Mello *et al.*, 2016). The major signalling pathways involved in melanocyte development include Wingless/Integrated (Wnt) signalling, endothelin 3 (EDN3)/endothelin receptor type B (EDNRB) signalling and stem cell factor (SCF)/tyrosine kinase receptor Kit signalling (Hou *et al.*, 2000; Colombo *et al.*, 2011). These transcription factors and signalling pathways instruct NCCs to migrate along the dorsolateral pathway, between the mesodermal and ectodermal layers, and acquire a melanogenic fate (Colombo *et al.*, 2011).

1.2.2. Melanin Synthesis: Melanosomes and Melanogenesis

Melanocytes contain melanosomes, which are specialised subcellular compartments that synthesise and package melanin. Melanosomes are related to conventional lysosomes, sharing many common features such as structure and the presence of lysosomal membrane proteins and hydrolases (Sturm, 2006). Two independent melanosome forms have been described, which differ in chemical composition, shape and size: eumelanosomes are opaque oval-shaped particles that synthesise brown-black eumelanin, while spheroid pheomelanosomes produce the yellowish-red pheomelanin (D'Mello *et al*., 2016).

A four-stage model, formulated through the ultrastructural study of melanin deposition in melanocytes, is used to describe the maturation of melanosomes (Figure 1.3). During stage I, pre-melanosomes bud from the endoplasmic reticulum (ER). The pre-melanosomes are empty vacuoles with an amorphous matrix which are not capable of melanin synthesis (Sturm, 2006). Enzymatic and structural proteins are trafficked from the trans-Golgi network (TGN) to the developing melanosome in stage II. Tyrosinase (TYR) and tyrosinase-related protein 2 (TRP2),

which determine the quality and quantity of melanin, are among the crucial enzymes, while melan-A (MART-1) and pre-melanosome protein (PMEL) are critical structural proteins required for the formation of the visible fibrillar matrix (Sturm, 2006; D'Mello *et al.*, 2016). During stage III, the onset of melanin production, catalysed by TYR, takes place and the pigment is deposited on the internal fibrils of the melanosome (Cichorek *et al.*, 2013a). In the last stage (IV), the melanin pigment fills the whole melanosome. These mature, heavily pigmented melanosomes lose their TYR activity and are transported to the neighbouring keratinocytes (Sturm, 2006; Cichorek *et al.*, 2013b). Each type of melanin is synthesised in a separate melanosome, but both melanosome types can occur in a single melanocyte (Cichorek *et al.*, 2013a).

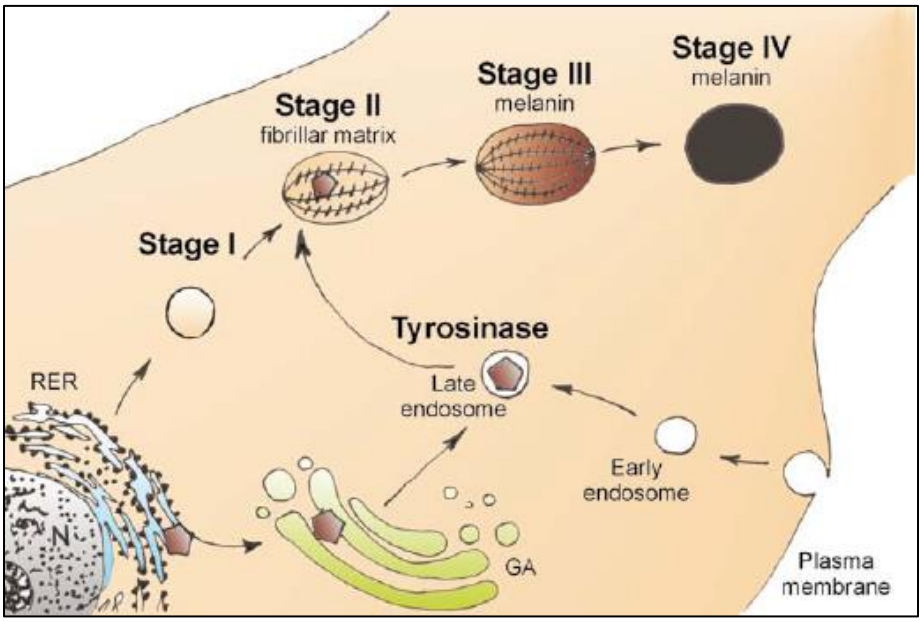


Figure 1.3. The four stages of melanosome development. (I) Formation of pre-melanosomes; (II) fibrillar matrix formed by glycoproteins; (III) onset of melanin production; (IV) melanin pigment fills the whole melanosome (figure taken from Cichorek *et al.*, 2013a).

Melanogenesis, also called the Raper-Mason pathway, is the biochemical pathway in melanosomes responsible for the synthesis of melanin (Cichorek *et al.*, 2013b; D'Mello *et al.*, 2016). The two melanin types, eumelanin and pheomelanin, differ in colour and also in the method of synthesis. In the first step, common for all melanin types, TYR carries out the hydroxylation of L-tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA) which is quickly oxidised to DOPAquinone. Thereafter, the function of the melanogenesis enzymes and substrate availability determine the type of melanin produced (Figure 1.4; Cichorek *et al.*, 2013a). In the presence of thiols, such as thioredoxin, glutathione and cysteine, DOPAquinone reacts with these substrates to create 3- or 5-cysteinylDOPA that then oxidises and polymerises to form the yellowish-red pheomelanin (Cichorek *et al.*, 2013b). In the absence of thiols,

brown-black eumelanin is produced when DOPAchrome is formed by the spontaneous non-enzymatic cyclisation and rearrangement of DOPAquinone (D'Mello *et al.*, 2016). DOPAchrome gives rise to eumelanins through two different pathways: non-enzymatic and enzymatic, via the use of tyrosine related proteins (TRPs; Cichorek *et al.*, 2013a; D'Mello *et al.*, 2016). The non-enzymatic pathway involves the spontaneous decarboxylation of DOPAchrome which produces 5,6-dihydroxyindole (DHI). Dark brown/black DHI-melanin is then formed through the rapid oxidation and polymerisation of DHI (Cichorek *et al.*, 2013a). In the enzymatic pathway, TRP2 tautomerises DOPAchrome to yield DHI-2-carboxylic acid (DHICA). Tyrosinase-related protein 1 (TRP1) catalyses the oxidation and polymerisation of DHICA to form DHICA-melanin, which is light brown in colour.

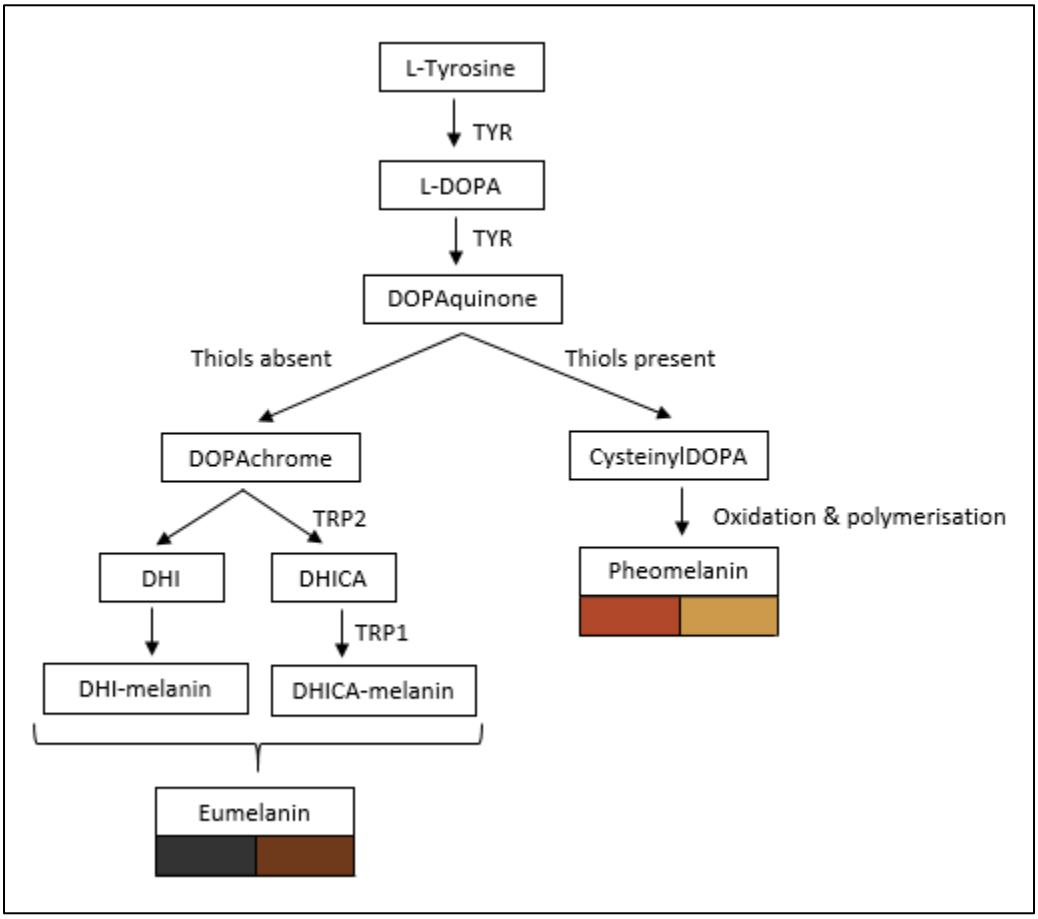


Figure 1.4. A simplified illustration of the melanogenesis pathway (adapted from Cichorek *et al.*, 2013b).

1.2.3. Melanin Transport and Transfer

Pigmentation depends on the transport of melanosomes from the centre of the cell, where melanosome biogenesis occurs, to the peripheral melanocyte dendrites, before transfer to the keratinocytes. Two major polarised cytoskeleton macromolecular polymers direct the intracellular transport of melanosomes: actin filaments and microtubules. Short-range movement occurs along actin filaments that consist of actin monomers, while microtubules, comprised of α- and β-tubulin dimers, facilitate fast long-range transport (Cieslak *et al.*, 2011; Colombo *et al.*, 2011; Van Gele & Lambert, 2011). In mammals, bi-directional transport of melanosomes along microtubules is accomplished by the motor proteins kinesin and dynein. Proteins of the kinesin superfamily facilitate anterograde movement towards the microtubule plus-ends at the cell periphery, whereas retrograde movement towards the microtubule minus-ends is powered by dynein and dynein-associated proteins (Barral & Seabra, 2004; Van Gele & Lambert, 2011). Once the mature melanosomes have reached the cell periphery, their transport and capture within the distal, actin-rich regions of the dendrites are achieved via a tripartite complex formed by Rab27a, melanophilin (MLPH) and an actin-based motor called myosin VA (Kondo & Hearing, 2011; Van Gele & Lambert, 2011).

Mature melanosomes captured in the peripheral dendrites are transferred to the surrounding keratinocytes. Since melanocytes have many dendrites, each melanocyte is able to make contact with multiple keratinocytes (Colombo *et al.*, 2011; D'Mello *et al.*, 2016). The cellular and molecular mechanisms involved in the transfer of melanosomes to keratinocytes have only been partly elucidated. However, it seems that membrane fusion between the plasma membrane of the melanocyte or keratinocyte and the melanosome is involved in the transfer process (Cieslak *et al.*, 2011; Colombo *et al.*, 2011).

1.2.4. Genetic Determination of Coat Colour

Basic coat colour in mammals are determined by the ratio and distribution of eumelanin and pheomelanin. Several genes have been identified that can influence basic coat colour by acting on the pathways responsible for the synthesis of either of the two pigment types or their distribution. Different mutation types in the coding or regulatory regions of these genes, including insertions, deletions, duplications and single nucleotide changes, have been shown to be responsible for variation in coat colour (Cieslak *et al.*, 2011).

Extension and *Agouti* are the major loci that control the relative amount of pheomelanin and eumelanin pigment produced by melanocytes (Fontanesi *et al.*, 2009). Molecular studies have revealed that the *Extension* locus encodes a G protein-coupled receptor called the melanocortin 1 receptor (MC1R; Yamaguchi *et al.*, 2007;

Fontanesi *et al.*, 2009). Melanocortin 1 receptor function is regulated by the agonists, α-melanocyte stimulating hormone (αMSH) and adrenocorticotropic hormone (ACTH). The agonists stimulate the expression of the melanogenic cascade through the activation of MC1R that results in the expression of eumelanin (Yamaguchi *et al.*, 2007). The agouti signalling protein (ASIP) is encoded by the *Agouti* locus and acts as the antagonist of MC1R by blocking αMSH-receptor interaction (Fontanesi *et al.*, 2009). Consequently, the pigment type produced is switched from eumelanin to pheomelanin (Hoekstra, 2006; Fontanesi *et al.*, 2009). In many species, the wild-type state involves the synthesis of both types of melanin. However, a mutation at either locus may commit the melanocyte to the exclusive synthesis of one pigment type. In several mammals, dominant alleles at the *Extension* locus produce a uniform black coat colour due to MC1R being constitutively active or hyper-responsive to melanocortins, while recessive alleles are the result of a loss-of-function mutation in *MC1R* and leads to the production of red/yellow pigment (Jackson, 1997; Pielberg, 2004). Conversely, dominant *Agouti* alleles result in the exclusive production of pheomelanin because MC1R is permanently blocked, while recessive loss-of-function alleles lead to a black/dark coat colour due to the accumulation of eumelanin (Fontanesi *et al.*, 2011). Furthermore, the *Extension* and *Agouti* loci are known to have epistatic interactions, since the expression of the *Agouti* allele usually requires the presence of a wild-type *Extension* allele (Fontanesi *et al.*, 2011). In addition to the *Extension* and *Agouti* loci, other genes can also influence the pigment type produced, for example, *TRP1* and *TRP2* regulate the eumelanin pathway, while genes such as *gamma-glutamyl transpeptidase-encoding protein* (*GGT*) and *solute carrier family 7 member 11* (*SLC7A11*) affect the production of pheomelanin (Chintala *et al.*, 2005; Hoekstra, 2006).

A number of genes have also been identified that can result in the dilution of the overall colouration (Nascimento *et al.*, 2003; Hoekstra, 2006). These genes are mainly involved in the rate of melanin synthesis and the distribution of melanosomes in melanocytes. For example, in mice, more than a 100 alleles have been reported and characterised for *TYR*, which encodes the rate-limiting enzyme in melanogenesis (Beermann *et al.*, 2004; Hoekstra, 2006; Challa *et al.*, 2016). These alleles range from null alleles, that result in the complete absence of pigmentation (albino), to alleles that limit melanin synthesis due to their reduced function (Hoekstra, 2006). Furthermore, mutations in genes involved in the transport and distribution of melanosomes in melanocytes, such as *MLPH*, *RAB27* and *MYO5*, can disrupt melanosome organisation that, in turn, result in diluted skin or coat colouration.

It is thus clear that many different molecular changes can influence the density, distribution or type of melanin pigment that can result in variation in coat colouration. Table 1.2 lists a few of the major genes known to influence pigmentation. However, variation in colouration is not necessarily the result of a mutation in one of these

candidate genes. Often changes in different genes can produce similar phenotypic results (Sturm & Frudakis, 2004).

Table 1.2. Major genes known to play a role in skin and hair colouration (adapted from Sturm & Frudakis, 2004).

Gene/Locus	Protein	Function
Melanosome Proteins		
TYR	Tyrosinase	Oxidation of L-tyrosine and L-DOPA
TRP1	TRP1/gp75	Oxidation and polymerisation of DHICA, TYR stabilisation
TRP2	TRP2/DCT	DOPAchrome tautomerase
PMEL/SILV	PMEL/gp100	Melanosomal striation, DHICA-polymerisation
MATP	MATP/AIM-1	Maturation of melanosome
OCA2	P-protein	Melanosomal pH and maturation of melanosome
Signal Proteins		
Extension	MC1R	G protein-coupled receptor
POMC	POMC, αMSH, ACTH	MC1R agonists
Agouti	ASIP	MC1R antagonist
MITF	MITF	Transcription factor that activates the transcription of TYR and TRP1
OA1	OA1 protein	Receptor for tyrosine, L-DOPA and dopamine; involved in intracellular signal transduction
Proteins involved in melanosome transport or keratinocyte uptake		
MYO5A	Myosin VA	Motor protein
RAB27A	Rab27A	Ras family protein involved in protein transport and small GTPase mediated signal transduction
HPS1	HPS1	Melanosome biogenesis and size
HPS6	HPS6	Melanosome biogenesis

Abbreviations: 3,4-Dihydroxyphenylalanine (DOPA); 5,6-dihydroxyindole-2-carboxylic acid (DHICA); α-melanocyte stimulating hormone (αMSH); absent in melanoma-1 protein (AIM-1); adrenocorticotropic hormone (ACTH); agouti signalling protein (ASIP); dopachrome tautomerase (DCT); Hermansky-Pudlak syndrome 1 (HPS1); Hermansky-Pudlak syndrome 6 (HPS6); melanocortin 1 receptor (MC1R); membrane-associated transporter protein (MATP); microphthalmia-associated transcription factor (MITF); myosin VA (MYO5A); ocular albinism type 1 (OA1); oculocutaneous albinism II (OCA2); pre-melanosome protein (PMEL/gp100); proopiomelanocortin (POMC); Ras-related protein Rab-27 (Rab27A); silver (SILV); tyrosinase (TYR); tyrosinase-related protein 1 (TRP1/gp75); tyrosinase-related protein 2 (TRP2).

1.2.4.1. Pleiotropic Effects Associated with Coat Colour Genes

Pleiotropy can be defined as a phenomenon in which one gene affects multiple, seemingly unrelated, phenotypic traits (Paaby & Rockman, 2013). Several pigmentation genes have been found to have pleiotropic effects and often mutations that result in coat colour variation are responsible for serious health disorders. Sensory organs and nerves are especially affected by disorders associated with coat colour due to the shared origin of neurocytes and melanocytes in the neural crest. These disorders are of particular concern for animal breeders, since they could result in financial losses due to the suffering or death of affected animals (Reissmann & Ludwig, 2013). Furthermore, studies have suggested that coat colour genes could also influence reproduction traits and, therefore, these genes could have an effect on the profitability of animal breeding (Charon & Lipka, 2015). A few examples of coat colour-associated genes in which mutations have pleiotropic effects are listed in Table 1.3. It is, however, important to note that the phenotypic expression of either coat colour or the associated pleiotropic effect can vary widely depending on the genetic state (homozygote or heterozygote) of the mutation, and the epistatic influence of other genes (Reissmann & Ludwig, 2013).

Table 1.3. Examples of coat colour-associated genes in which mutations have pleiotropic effects.

Gene	Species	Mutation(s)	Pleiotropic Effect	Effect on Coat Colour	Reference
KIT	Horse (*Equus caballus*)	2 Missense mutations, 2 frameshift mutations, 3 splice site mutations	Lethal *in utero* (homozygous state)	White (heterozygous state)	Haase *et al*., 2009
	Pig (*Sus scrofa*)	Duplication of *KIT*, splice mutation in one of the *KIT* copies	Mild macrocytic anaemia (homozygous state)	White (heterozygous or homozygous status)	Johansson *et al*., 2005
MITF	Cattle *(Bos taurus)*	Missense mutation	Bilateral deafness (heterozygous or homozygous state)	White (heterozygous or homozygous state)	Philipp *et al*., 2011
PMEL	Dog (*Canis lupus familiaris*)	Retrotransposon insertion	Auditory and ophthalmologic abnormalities (heterozygous or homozygous state), skeletal, reproductive and cardiac	Merle pattern (heterozygous or homozygous state)	Clark *et al*., 2006

Gene	Species	Mutation(s)	Pleiotropic Effect	Effect on Coat Colour	Reference
			abnormalities (homozygous state)		
PAX3	Horse (*Equus caballus*)	Missense mutation	Lethal *in utero* (homozygous state)	Splashed white (heterozygous state)	Hauswirth *et al*., 2012
LYST	Cattle (*Bos taurus*)	Missense mutation	Increased bleeding tendency (homozygous state)	Hypopigmentation (homozygous state)	Kunieda, 2005.

Abbreviations: Lysosomal trafficking regulator (LYST); microphthalmia-associated transcription factor (MITF); paired box 3 (PAX3); pre-melanosome protein 17 (PMEL); tyrosine kinase receptor (KIT).

In addition to the pleiotropic effects of coat colour genes, melanoma could also possibly result from mutations in these genes. Often mutations in coat colour genes result in diluted coat colour phenotypes. Melanomas are generally more prevalent in diluted coat colour phenotypes due to reduced ultraviolet radiation (UVR) protection (Reissmann & Ludwig, 2013).

1.3. Genomic Tools

Genomic resources are often used to investigate and unravel the genetic mechanisms underlying important phenotypic traits. Molecular genetic markers represent powerful tools that can be used for the analysis of genomes. In addition, molecular markers enable the association of heritable phenotypic traits with underlying variations in the genome (Duran *et al.*, 2009). However, the number of molecular markers available for game species are very limited at present (Taylor *et al.*, 2016). The rapid progress in next-generation sequencing (NGS) technologies, along with the continuous decline in sequencing cost have facilitated the large-scale discovery of molecular markers in numerous model and non-model species (Da Fonseca *et al.*, 2016; Morin *et al.*, 2018). These high-throughput genotyping technologies could, therefore, also assist in developing molecular markers for game species.

1.3.1. Molecular Markers

By definition, molecular markers are heritable DNA polymorphisms that can be readily detected and used to identify individuals, populations or species (Vignal *et al.*, 2002; Liu & Cordes, 2004; Davey *et al.*, 2011). Based on the method of analysis, molecular marker types can generally be classified in three major groups: (1) hybridisation-

based molecular markers, (2) polymerase chain reaction (PCR)-based markers and (3) sequencing-based molecular markers (Teneva, 2009; Garrido-Cardenas *et al.*, 2018). The first generation of molecular marker systems employed hybridisation techniques (Yang *et al.*, 2015). A classic example of a molecular marker based on DNA hybridisation is restriction fragment length polymorphisms (RFLPs), which involves the hybridisation of short DNA probes with genomic DNA that has been fragmented by restriction enzymes (REs; Park *et al.*, 2009). Molecular markers based on PCR generally represent the second generation of DNA markers and involve the *in vitro* amplification of particular DNA sequences or loci (Teneva, 2009). Examples of PCR-based markers include random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs) and microsatellites or simple sequence repeats (SSRs; Park *et al.*, 2009; Teneva, 2009). Sequencing-based molecular markers, such as single nucleotide polymorphisms (SNPs), represent the newest generation of molecular marker systems. Table 1.4 compares the characteristics of the marker systems that have been most commonly used for animal genome analysis up until now.

In modern genetic analysis, SSRs and SNPs predominate in most studies (Ozerov *et al.*, 2013; Vieira *et al.*, 2016; Yadav *et al.*, 2017). However, in recent years, SNPs have become the marker of choice due to their high abundance, genome-wide distribution and amenability to cost-effective, high-throughput genotyping (Von Thaden *et al.*, 2017). Furthermore, SNPs are more suitable for lineage-based analyses since they have lower mutation rates than SSRs and are, therefore, more stable and less likely to change over generations (Pontes *et al.*, 2015). Single nucleotide polymorphisms are also easy to score and do not require standardisation across different laboratories and thus the data can be readily compared (Von Thaden *et al.*, 2017). These characteristics make SNPs excellent molecular makers to study complex genetic traits and heredity (Duran *et al.*, 2009).

1.3.1.1. Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms are genetic variations, caused by a single point mutation, for which a short flanking sequence is known (Ben-Ari & Lavi, 2012). These point mutations result in different alleles that contain alternative nucleotide bases at a given position within a locus (Liu & Cordes, 2004; Jin *et al.*, 2016). Single nucleotide polymorphisms could in principle produce as many as four alleles, each containing one of the four bases at the particular SNP site: cytosine (C), guanine (G), adenine (A) and thymine (T) (Liu & Cordes, 2004). However, tri- and tetra-allelic SNPs are rare, hence SNPs are generally bi-allelic (Doveri *et al.*, 2008; Duran *et al.*, 2009). This disadvantage, when compared with highly polymorphic markers such as SSRs, is compensated for by the relative abundance and dense distribution of SNPs within a genome (Duran *et al.*, 2009; Hauser *et al.*, 2011).

Table 1.4. Comparison of commonly used markers systems for animal genome analysis (adapted from Liu & Cordes, 2004; Park *et al.*, 2009; Miah *et al.*, 2013; Nadeem *et al.*, 2018)**.**

Characteristic	RFLP	RAPD	AFLP	SSR	SNP
Type of Polymorphism	Single nucleotide change, insertion, deletion, inversion	Single nucleotide change, insertion, deletion, inversion	Single nucleotide change, insertion, deletion, inversion	Variation in repeat length	Single nucleotide change
Genomic Abundance	High	Very high	Very high	Moderate	Very high
Dominance	Co-dominant	Dominant	Dominant	Co-dominant	Co-dominant
Locus under Investigation	Single	Multiple	Multiple	Single	Single
Reproducibility	High	Low	Intermediate	High	High
Polymorphism/Power	Low	Low to moderate	High	High	High
DNA Quality	High	High	High	Moderate	Moderate
Cost	High	Low	Moderate	High initial cost	Variable
Prior DNA Sequence Information	Not required	Not required	Not required	Required	Required
Automation	Low	Moderate	Moderate	High	High
Status	Dated	Dated	Dated	Current	Current

Abbreviations: Amplified fragment length polymorphism (AFLP); random amplified polymorphic DNA (RAPD); restriction fragment length polymorphism (RFLP); single nucleotide polymorphism (SNP); simple sequence repeat (SSR).

For example, one SNP occurs approximately every 434 bp in *Bos taurus* and approximately every 104 bp in *Bos indicus* (Gurgul *et al.*, 2019). Furthermore, SNPs can be divided into two different categories: transitions (G/A or C/T base changes) and transversions (T/G, C/G, C/A and A/T base changes; Duran *et al.*, 2009).

Single nucleotide polymorphisms can often have important functional implications if they are located in coding DNA sequences (Gurgul *et al.*, 2019). When occurring in coding regions, these polymorphisms can either be synonymous or non-synonymous. Synonymous SNPs have different alleles that code for the same amino acid and thus the protein sequence remains unchanged (Murphy *et al.*, 2014). Conversely, non-synonymous SNPs (nsSNPs) alter the amino acid sequence of a protein (Zhang *et al.*, 2012; Murphy *et al.*, 2014). Missense mutations are non-synonymous base changes that cause a substitution of a single amino acid and can result in the alteration or loss of protein function (Jehan & Lakhanpaul, 2006; Zhang *et al.*, 2012). Alternatively, nonsense mutations are non-synonymous SNPs that result in a premature stop codon. Consequently, a truncated, and usually non-functional, protein product is produced (Zhang *et al.*, 2012; Murphy *et al.*, 2014). The type of SNP and the location of the SNP in the gene can determine the level of phenotypic effect that a SNP produces. Synonymous SNPs more often are neutral with no functional effect, while nonsense mutations are more likely to be deleterious. The effect of missense mutations depends on the nature of the substitution, as well as the location of the amino acid change (Murphy *et al.*, 2014).

Single nucleotide polymorphisms located in non-coding regions could also have functional consequences (Zhang & Lupski, 2015). The non-coding region between genes often contain *cis*-regulatory elements such as promotors, enhancers and silencers, and sequence variants in these regions could possibly alter transcription factor identification and binding (Guo & Jamison, 2005). Consequently, non-coding SNPs can affect gene regulation by a gain of function, resulting in gene overexpression, or a loss of function, resulting in reduced gene expression. (Spielmann & Mundlos, 2016).

1.3.2. SNP Genotyping Technologies

Significant efforts have been devoted to developing cost-effective, accurate and rapid technologies for SNP genotyping over the past several years (Kim & Misra, 2007; Yang *et al.*, 2013; Gamaniel & Gwaza, 2017). Early SNP genotyping technologies relied on direct sequencing, as well as simple hybridisation methods such as RFLP and the allele-specific oligonucleotide (ASO) technique, and gel-based methods such as cleaved amplified polymorphic sequences (CAPS), single strand conformational polymorphism (SSCP) assays and denaturing gradient gel electrophoresis (DGGE; Liu & Cordes, 2004; Thomson, 2014; Rasheed *et al.*, 2017). However, these

traditional methods share the common limitations of being costly, labour intensive and low in throughput (Rasheed *et al.*, 2017). High-throughput genotyping technologies are required to efficiently genotype a large number of SNPs in a cost-effective and accurate manner (Gurgul *et al.*, 2019). Early successes with high-throughput SNP genotyping relied on fixed SNP marker sets assayed using microarrays (Thomson, 2014). Microarrays are one of the most commonly used and powerful genotyping tools, since they are capable of analysing thousands of SNPs in parallel (Gurgul *et al.*, 2019). Other advantages of microarrays include robust allele calling with high call rates and low-cost per data point when genotyping a large number of SNPs (Thomson, 2014). However, prior knowledge on the characteristics of the SNPs and the DNA sequences surrounding the SNPs are needed during the microarray design process and should, consequently, be preceded by whole-genome sequencing or other efficient genotyping techniques (Trevino *et al*., 2007). Furthermore, the development of microarrays for initial studies or small experiments are often costly and unjustified (Gurgul *et al*., 2019).

Recently developed genotyping techniques based on NGS technology and the enrichment of genome fragments containing RE cut sites, offer a solution to the drawbacks associated with microarrays and other traditional SNP genotyping techniques. A RE-based enrichment strategy enables the sequencing of only a portion of the genome. This approach allows for higher sample multiplexing and also significantly reduces experiment costs when compared with whole genome sequencing (Elshire *et al.*, 2011). Other advantages of this approach include the genome-wide detection of polymorphisms and the large-scale genotyping of populations. The use of REs to reduce genome complexity is also relatively simple, specific and reproducible (Gurgul *et al.*, 2019). In addition, since repetitive elements are often heavily methylated, repetitive regions of the genome can be excluded and low copy regions can be targeted with higher efficiency when appropriate methylation sensitive REs are used (Gore *et al.*, 2007).

Several protocols involving the sequencing of restriction site-associated genomic DNA for SNP discovery and genotyping have been developed recently. Of these methods, the genotyping-by-sequencing (GBS) approach (Figure 1.5) seems to be one of the simplest methods that has found a wide application (He *et al.*, 2014; Gurgul *et al.*, 2019). The first step of the GBS approach involves the digestion of genomic DNA with one or two frequently cutting REs, followed by the ligation of cut-site specific adapters. These adapters contain a universal primer binding site, as well as a unique index. Ligation products of appropriate size are then amplified, multiplexed and sequenced (Gurgul *et al.*, 2019).

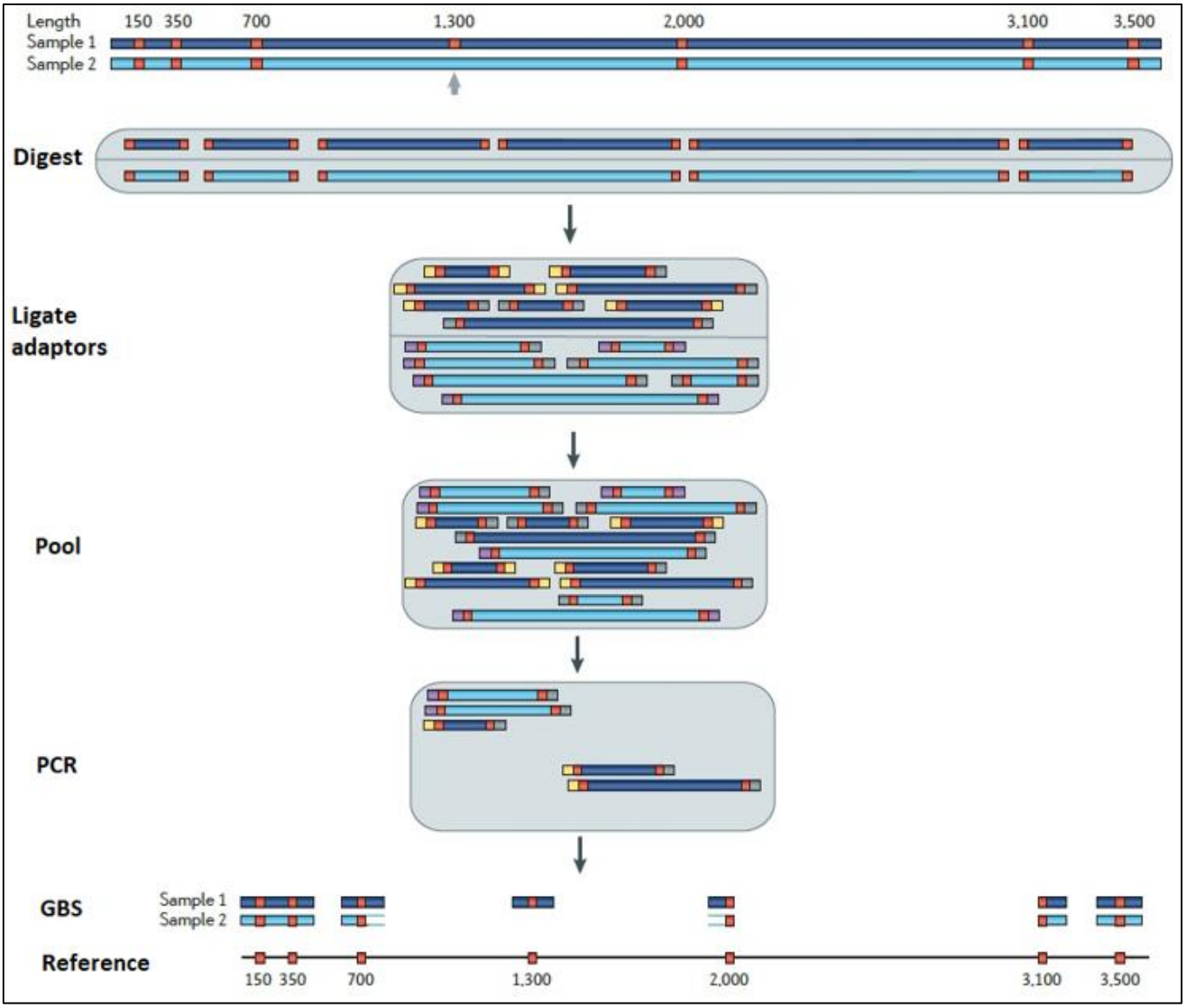


Figure 1.5. Method for Genotyping-by-Sequencing. An example genomic region contains restriction sites indicated in red. A DNA sample from two individuals is to be sequenced. The DNA from individual 1 is shown as dark blue and the DNA from individual 2 is shown as light blue. Sample 2 has a polymorphism in the restriction site at 1 300 bases (indicated by grey arrow) and, therefore, will not be cut at this site. Firstly, the DNA is digested with appropriate restriction enzymes. Thereafter, common adaptors (grey) and barcoded adaptors (yellow for sample 1 and purple for sample 2) are ligated to the digested fragments. The following adaptor combinations are produced: barcode+barcode, common+common and common+barcode. Samples are then pooled and amplified. Short fragments with a common+barcode adaptor combination are the only samples that will be amplified for sequencing. The light blue lines in the last step represent regions that will be filtered during amplification, but could be imputed later. Finally, the fragments will be sequenced on a NGS platform (figure adapted from Davey *et al.*, 2011).

Advantages of GBS include fewer DNA purification steps and DNA size selection is not required. Consequently, the time and cost of library preparation is reduced. Digestion of genomic DNA and adapter ligation are also carried out in a single well which results in reduced sample handling (Davey *et al.*, 2011; He *et al.*, 2014; Gurgul *et al.*, 2019). Furthermore, the generated sequence tags and their supporting reads can be used for large-scale SNP discovery and genotyping without mapping them to a reference genome sequence, thus enabling whole genome genotyping in practically any organism (Elshire *et al.*, 2011; Gurgul *et al.*, 2019). Therefore, GBS-based genotyping assays can result in tangible benefits in animal genomics, particularly in species for which limited or no commercial genotyping tools are available, such as game species (Gurgul *et al.*, 2019). The only point of caution in the GBS design procedure is the selection of REs. The REs that are selected for genome complexity reduction should not cut frequently in the major repetitive genome fragments and should also produce overhangs (De Donato *et al.*, 2013).

1.3.2.1. Diversity Array Technology Sequencing (DArTseq™)

The DArTseq platform, a variant of the GBS method developed by Diversity Arrays Technology (Pty) Ltd (Canberra, Australia; https://www.diversityarrays.com), has experienced an increasing interest worldwide as it has sufficiently met the requirements for cost-effective, high-throughput genotyping based on NGS sequencing results (Milczarski *et al.*, 2016). The DArT platform was first developed in 2000 and enabled the identification of DNA polymorphisms without the requirement for prior DNA sequence information (Jaccoud *et al.*, 2001). The technology was based on hybridisation and solid-state surfaces, instead of relying on electrophoretic separation to resolve DNA polymorphisms and, therefore, displayed improved accuracy and throughput compared to traditional PCR/electrophoresis based genotyping methods (Melville *et al.*, 2017). Today, classic DArT markers have been substituted by DArTseq markers that involve the use of DArT markers in combination with NGS (Baloch *et al.*, 2017; Melville *et al.*, 2017).

In brief, the DArTseq technology allows the selection of genome fractions that mainly correspond to active genes through the use of optimised combinations of REs that separate low copy sequences from repetitive fractions of the genome. Moreover, the technology is optimised for the specific organism by testing different combinations of REs and selecting the most suitable complexity reduction method (Melville *et al.*, 2017; Schultz *et al.*, 2018). These low copy sequences, which are informative for marker discovery, are then sequenced using Illumina short-read sequencing (Milczarski *et al.*, 2016). The sequences that are generated consist of both reference and polymorphic fragments and are generally 69 bp in length (Melville *et al.*, 2017). Two groups of markers can be detected in the generated sequences, namely SNP markers and *silico*DArT markers. The SNP markers contain

detailed data about the position of mutations and enables the detection of both homozygotes and heterozygotes (co-dominant markers). Conversely, the *silico*DArTs are dominant markers and only score for presence or absence of a sequence variant (Milczarski *et al.*, 2016). The DArTseq technology and its application in game species are discussed in detail in chapter 2 of this thesis.

1.4. Association Studies

The recent availability of genotyping platforms that are capable of generating large panels of SNPs in non-model species has given new momentum to the search for causal genes and mutations underlying phenotypic variation. Two methodological approaches, both based on genetic recombination, are commonly used to identify genetic variants that contribute to traits of interest: linkage studies and linkage disequilibrium (LD) mapping (Table 1.5). Linkage analysis was the first approach used to identify genomic regions related to a specific phenotype by testing whether a marker allele and a phenotype show correlated transmission within a pedigree (Borecki & Province, 2008; Fingerlin *et al.*, 2016). More recently, however, the focus has shifted to LD mapping, which depends on historic LD to link phenotypes to genotypes (Fingerlin *et al.*, 2016). The ability of LD mapping to exploit unobserved recombination events in past generations generally allows for finer mapping than linkage analysis, which are limited by the amount of observed recombination events in the pedigrees (Ott *et al.*, 2015). In addition, LD mapping also has more power to detect common genetic variants with modest effects (Geng *et al.*, 2017).

Table 1.5. A comparison between linkage analysis and linkage disequilibrium mapping (Ott *et al.*, 2011; Geng *et al.*, 2017).

Feature of Mapping Approach	Linkage Analysis	Linkage Disequilibrium Mapping
Biological basis of approach	Observes recombination events in pedigree data	Exploits unobserved recombination events in past generations
Detection of variants with modest effect	Not suitable	Suitable
Power to detect rare alleles	Yes	No
Pedigree required	Yes	No (unrelated and related individuals can be used)
Range of effect detected	Long (≤5 Mbp)	Short (≤100 Kbp)
Effect of population structure	No effect	Reduces power

Abbreviations: Kilobase pairs (Kbp); megabase pairs (Mbp).

1.4.1. Linkage Disequilibrium Mapping

Linkage disequilibrium is defined as the non-random association of alleles at two or more loci in a population (Bush & Moore, 2012). Linkage disequilibrium is a sensitive indicator of the population genetic forces that shape a genome (Slatkin, 2008). Factors creating LD in a population include mutation rate, recombination rate, non-random mating, genetic drift and population structure (Scherer & Christensen, 2016). The presence of LD creates two possible outcomes for a genetic association study. In the first outcome, the SNP that has an effect on a biological system, that ultimately determines the phenotype, is directly genotyped in the study and found to be statistically linked with the trait. This outcome is referred to as a "direct association", while the genotyped SNP is sometimes called the "functional SNP" or "causal SNP". The other outcome is that the functional SNP is not directly genotyped, but instead a tag SNP (representative SNP in a region of the genome) in high LD with the functional SNP is typed as statistically associated with the phenotype. This second possibility is referred to as "indirect association" (Hirschhorn & Daly, 2005; Bush & Moore, 2012). Due to these two possible outcomes, a significant SNP association identified in an association study should not be assumed to be the causal variant and additional studies may be required to map the exact location of the functional SNP.

1.4.1.1. Study Population: Case-Control Studies vs *Family-Based Studies*

The two basic types of study designs used for LD mapping are case-control study designs, sometimes also referred to as population-based designs, and family-based study designs (Fingerlin *et al.*, 2016). A case-control study aims to determine whether statistical association exists between the trait of interest and a genetic marker by comparing the allele or genotype frequencies at the marker locus in unrelated individuals with and without the trait of interest (Clarke *et al.*, 2011). Cases and controls should ideally be drawn from the same population. Failure to do so will often result in spurious associations due to biased selection and differences in genetic structure (Fingerlin *et al.*, 2016). Conversely, in family based designs, family pedigrees are used to study associations in which relatives of case individuals are used as control subjects. The transmission disequilibrium test (TDT) has become the most widely used family-based LD method for mapping genes related to a phenotypic trait of interest. This method compares the transmission *versus* the non-transmission of alleles to the offspring, therefore, assuming linkage between the marker and the trait (Soto-Cerda & Cloutier, 2012). The TDT design requires genotyping of a family trio: one affected offspring and two parents of which at least one should be heterozygous at the genetic marker of interest. The major advantage of family-based LD studies are that, unlike case-control studies, they are robust to population structure because the parental genotype is used as the control (Fingerlin *et al.*, 2016). However, the case-control study design is the most commonly used approach because it is simpler to

implement than family-based designs and the use of unrelated individuals also usually require less resources, money and time (Evangelou *et al.*, 2006; Laird & Lange, 2006; Fingerlin *et al.*, 2016).

1.4.1.2. Candidate Gene Studies vs *Genome-Wide Association Studies*

There are two approaches for the genetic dissection of phenotypic traits: the traditional candidate gene approach and genome-wide association studies (Amos *et al.*, 2011). The candidate gene approach involves the selection of putative candidates gene based on prior knowledge of the biochemical function and/or location of the specific genes (Gordish-Dressman & Devaney, 2011; Patnala *et al.*, 2013). Thereafter, the effects of variants in the putative genes are assessed in an association analysis (Zhu & Zhao, 2007; Patnala *et al.*, 2013). Conversely, genome-wide association studies (GWAS) first emerged about a decade ago due to the availability of large marker panels generated by high-throughput genotyping methods. This approach offers the unprecedented opportunity to study the genetic mechanisms underlying specific phenotypes by examining the entire genome of unrelated individuals without the prior need for a hypothesis relating to gene identity and functionality (Gordish-Dressman & Devaney, 2011).

Each of these approaches has unique advantages and drawbacks. Since the candidate gene approach is a hypothesis-driven approach, it enables the targeted evaluation of selected alleles and, therefore, markers can be concentrated in the candidates genes. Putative functional markers could also be purposefully targeted. This approach may confer inferential advantages compared to GWAS, where coverage is spread across the whole genome and functional SNPs are not targeted specifically. This enhanced power is of great value when lower-frequency SNPs are studied or when the study population is small (Jorgensen *et al.*, 2009). Furthermore, candidate gene studies are also a more cost-efficient approach for direct gene discovery (Zhu & Zhao, 2007). However, this approach is largely limited by its reliance on existing biological knowledge on the trait of interest and is also incapable of discovering new genes or gene combinations that contribute to the trait of interest (Zhu & Zhao, 2007; Amos *et al.*, 2011). Therefore, when the genetic and biological underpinnings of a phenotype is completely unknown, as is the case for many animal species, candidate gene studies are not very efficient. In comparison, GWAS can scan the entire genome to identify loci possibly associated with the phenotypic outcome (Gordish-Dressman & Devaney, 2011). However, by exploring many SNP markers throughout the genome there is a high risk of finding false significant associations. Subsequently, a correction for multiple comparisons has to be employed to avoid this error (Riancho, 2012). Another limitation of GWAS is the inability to identify variants with small effect (Korte & Farlow, 2013).

1.4.2. Association Studies to Elucidate Coat Colour Genetics

Coat colour has traditionally been the target of many genetic studies due to the remarkable diversity observed in animal colouration (Andersson, 2001; San-Jose & Roulin, 2017). In the past, the study of the genetic basis of animal colouration has been mainly limited to model species, whereas research on non-model species has been either neglected or mostly been based on candidate gene studies. Nevertheless, studies that have made use of the candidate gene approach have been very useful to gain a better understanding of the genetic mechanisms underlying colour variation in a wide range of species, and have highlighted the recurrent role of certain genes, such as *MC1R*, *ASIP* and *TYR* (San-Jose & Roulin, 2017). The candidate gene approach has also been successful in identifying the causal variant of the black coat colour phenotype observed in impala (Miller *et al.*, 2016b). The study found a single base-pair deletion in the *ASIP* gene to be responsible for the black phenotype and has been the only study, up until now, to focus on the genetic basis of coat colour in African antelope.

Although the candidate gene approach has been beneficial in studies on animal colouration, the study of the genetic architecture of colour traits has been biased toward genes with well-known functions in pigmentation. Genome-wide association studies have proven to be a powerful tool to fill the gaps in our understanding of the biology of animal colouration (San-Jose & Roulin, 2017). Since variations in different genes can produce the same coat colour phenotype, GWAS offers a more robust approach to identify variants that influence coat colour (Cieslak *et al.*, 2011; San-Jose & Roulin, 2017). Examples of studies that successfully made use of GWAS to elucidate the genetic underpinnings of coat colour are presented in Table 1.6. Consequently, GWAS can also assist in identifying the genes that play a role in coat colour in other game species.

Table 1.6. Examples of genome-wide association studies in animals to elucidate coat colour genetics.

Species	Genotyping Platform	Coat Colour Variation	Major Candidate Genes Identified	Reference
Tibetan and Kele pigs (*Sus scrofa*)	Porcine 60K SNP array	Brown	*TRP1*	Ren *et al.*, 2011
Bengal tiger (*Panthera tigris tigris*)	Whole-genome sequencing and RADseq (509 220 SNPs)	White	*SLC45A2*	Xu *et al.*, 2013
Finnsheep (*Ovis aries*)	Ovine 50K SNP array	Non-white	*TRP1*, *ASIP*, *MITF*	Li *et al.*, 2014

Species	Genotyping Platform	Coat Colour Variation	Major Candidate Genes Identified	Reference
Holstein cattle (*Bos taurus*)	Bovine 50K SNP array	Proportion of black and teat colour	*KIT*, *IGFBP7*, *WNT16*, *MITF*, *PDGFRA*, *ING3*	Fan *et al.*, 2014
Saanen goat (*Capra falconeri*)	Caprine 50K SNP array	Pink and pink neck	*ASIP*	Martin *et al.*, 2016
Markhoz goat (*Capra falconeri*)	Caprine 50K SNP array	Black, brown and white	*ASIP*, *ITCH*, *RALY*, *AHCY* (black and brown) and *KIT*, *PDGFRA* (white)	Nazari-Ghadikolaei *et al.*, 2018

Abbreviations: adenosylhomocysteinase (AHCY); agouti signalling protein (ASIP); inhibitor of growth protein 3 (ING3); insulin-like growth factor-binding protein (IGFBP7); itchy E3 ubiquitin protein ligase (ITCH); microphthalmia-associated transcription factor (MITF); platelet-derived growth factor receptor alpha (PDGFRA); restriction-site associated DNA sequencing (RADseq); RNA-binding protein Raly (RALY); single nucleotide polymorphism (SNP); solute carrier family 45 member 2 (SLC45A2); tyrosinase-related protein 1 (TRP1); tyrosine kinase receptor (KIT); Wingless/Integrated (Wnt) family member 16 (WNT16).

1.5. Comparative Genomics

Numerous organisms have had their genomes completely or partially sequenced over the past few years (Wachi *et al*., 2017; Hu *et al*., 2019; Kersey, 2019). The availability of genome sequences, generated with NGS, has recently given rise to the field of comparative genomics in which the genome sequences of different species are compared. Comparative genomics provides a powerful tool to identify genes conserved among species, as well as genes that contribute to unique characteristics between species.

The functional annotation of marker sequences associated with a phenotype is an important task once these markers have been identified. Currently, comparative genomics is one of the major approaches used in the functional annotation of genomic sequences, and involves the identification of orthologous genes to facilitate annotation (Mazza *et al.*, 2009; Gabaldón & Koonin, 2013; König *et al.*, 2018). Orthologous genes are genes related by vertical descent from a common ancestor and are generally assumed to retain equivalent functions in different organisms. Therefore, genes will often be encoded within DNA conserved between the species due to functional constraints (Nobrega & Pennacchio, 2004). The core process in comparative genomics thus involves the alignment of DNA sequences, which entails the mapping of nucleotides in one sequence onto the nucleotides of the other sequence in order to identify conserved sequences between different species (Hardison, 2003; Nobrega & Pennacchio, 2004). Several bioinformatics tools for comparative sequence analysis are generally available. These provide integrative views of extensive annotation of genes or genomic regions, and are known to be

computationally demanding. Therefore, these bioinformatics tools are usually hosted on very powerful servers (Miller *et al.*, 2004; Negrisolo *et al.*, 2008). Three major browsers provide a wide variety of annotation functionalities: the Ensembl browser (https://www.ensembl.org), the University of California, Santa Cruz (UCSC) Genome Browser (https://genome.ucsc.edu) and Genome Data Viewer at the National Centre for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/; Miller *et al.*, 2004). Analyses performed on these browsers include full annotation of genomic sequences, identification of regulatory elements, prediction of gene boundaries and identification of syntenic regions among genomes (Negrisolo *et al.*, 2008).

1.5.1. Use of The Bovine Genome

The complete genome sequence of a domestic Hereford cow (*Bos taurus*) was published by the Bovine Genome Sequencing Consortium in April 2009 (Bovine Genome Sequencing and Analysis Consortium *et al.*, 2009). Annotation of the bovine genome identified approximately 22 000 genes. Since then, the bovine genome and its associated resources, such as the Bovine Genome Database (BGD), have been used in several other bovid studies to facilitate the identification of genes. For example, the bovine genome has been used to functionally annotate expressed sequence tags (ESTs) in water buffalo (*Bubalus bubalis*) due to their shared ancestry (Bajetha *et al.*, 2013). In addition, it has also been used to characterise differences in gene content, structure and regulation between cattle and water buffalo in order to identify genomic structural variations, which may contribute to the unique phenotypes of water buffalo (Li *et al.*, 2019). Furthermore, the cattle genome has been used as a reference genome to assemble the complete genome sequence of the African buffalo (*Syncerus caffer*) and was used for the functional annotation of the assembled genome (Glanzmann *et al.*, 2016). A study by Qiu *et al.* (2012) also made use of the cattle genome to gain a better understanding of how domestic yaks (*Bos grunniens*) adapted to living at high altitude. Genomic comparison between yak and cattle identified an enrichment of protein domains involved in sensing the extracellular environment and hypoxic stress, as well as an expansion in gene families related to energy metabolism and sensory perception. Consequently, the cattle genome could also be of great value as a genetic resource for studies in African antelope species.

1.6. Study Rational and Outline

1.6.1. Study Species: Blue Wildebeest (Gnu)

1.6.1.1. Overview of Biology, Ecology and Evolution

Wildebeest, genus *Connochaetes*, are large African antelope that form part of the Bovidae family. They belong to the Antilopinae subfamily and the Alcelaphini tribe (Røed *et al.*, 2011). The *Connochaetes* genus first evolved about 2.5 million years ago, and contains two extant species: black wildebeest (*Connochaetes gnou*) and blue wildebeest (*Connochaetes taurinus*; Hilton-Barber & Berger, 2004; Ackermann *et al.*, 2010). Fossil records and molecular analyses indicate that these two species diverged approximately one million years ago (Ackermann *et al.*, 2010). Today, C*. taurinus* consists of five subspecies that each inhabit different regions of the African continent (Figure 1.6): Cookson's wildebeest (*C. t. cooksoni*), Nyassa wildebeest (*C. t. johnstoni*), East African white-bearded wildebeest (*C. t. albojubatus*), western white-bearded wildebeest (*C. t. mearnsi*) and the southern brindled blue wildebeest, also known as brindled gnu or common wildebeest (*C. t. taurinus*; Ackermann *et al.*, 2010; Tambling *et al.*, 2016; Grobler *et al.*, 2018). The only blue wildebeest subspecies that is indigenous to South Africa is *C. t. taurinus* (Ackermann *et al.*, 2010; Tambling *et al.*, 2016).

The common name of the blue wildebeest is derived from the conspicuous silvery blue sheen of their coats. The coat colour of a fully mature blue wildebeest can, however, vary from a deep slate or bluish grey to a light grey or even a greyish-brown. Moreover, brindled stripes mark the neck and forequarters, which results in a perception of skin wrinkling (Grubb, 2005; Ackermann *et al.*, 2010). The blue wildebeest also has a long black horse-like tail, a dark mane that runs from the shoulders to the top of the neck, and a long beard down their chin and neck. The beard can vary in colour, depending on the subspecies of blue wildebeest (Ackermann *et al.*, 2010). The head of the animal is large with a distinct robust muzzle, and the face is convex in shape. Both sexes possess a pair of large horns shaped like parentheses, extending outward to the side and then curving up and inwards. Moreover, the shoulders of the blue wildebeest are markedly higher and more robust than the rump which gives the animal an ungainly appearance (Huffman, 2005).

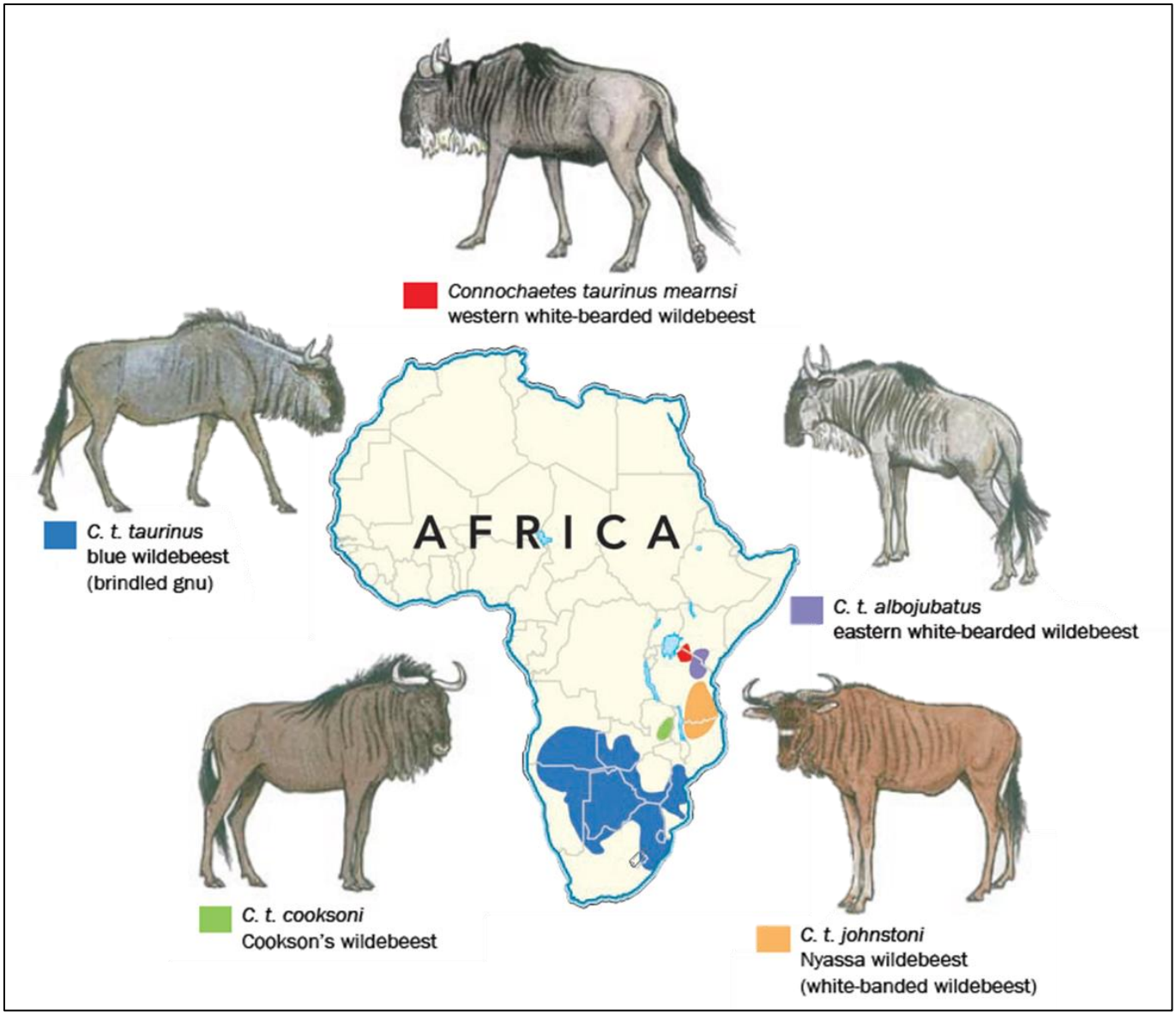


Figure 1.6. Blue wildebeest subspecies and their distribution (adapted from Estes & East, 2009).

Blue wildebeest inhabit a wide variety of habitats, such as bushlands, grasslands, open woodlands and floodplains. However, since they are herbivorous animals that primarily feed on short grasses, they have a special preference for Acacia savannahs, which is a mixed woodland-grassland biome. Blue wildebeest will also feed on foliage from trees and shrubs if grass becomes sparse during dry periods (De Haas, 2016). Furthermore, blue wildebeest are highly dependent on water and surface water for daily drinking is essential. Therefore, they readily migrate, following the seasonal rains (Furstenburg, 2013).

Bulls become sexually mature at the age of 36 months, whereas cows become sexually mature between the age of 18 and 27 months. However, bulls usually only start to breed from the age of 4.5 years, while cows start breeding from the age of 2.5 years (Furstenburg, 2013). Blue wildebeest are seasonal breeders. Rutting generally takes place from April to June (Tambling *et al.*, 2016). In wild populations, the mating ratio is generally one socially mature bull to two or three mature females (Furstenburg, 2013). Most calves are born between November and January. Usually a single calf is born after a gestation period of approximately 255 days (Furstenburg, 2013; Tambling *et al.*, 2016). The calving interval for cows is approximately 10 to 18 months (Furstenburg, 2013).

1.6.1.2. Golden Coat Colour Variant

The golden wildebeest is a colour variant of the common blue wildebeest (*C. t. taurinus*; Taylor *et al.*, 2016; Figure 1.7). The coat of the golden wildebeest is light brown to yellow-brown in colour (Adetunji *et al.*, 2018). As mentioned previously, the golden wildebeest originates from the Limpopo River basin, adjacent to the Tuli block of Botswana, and has long been known to occur in migratory herds. The earliest records of sporadic sightings of golden wildebeest date back to the 1920s and early farmers referred to them as "red wildebeest" or "Vos wildebeest" (Wildswinkel, 2015). The golden coat colour variation is thought to be an simple autosomal recessive Mendelian trait, although the pattern of inheritance requires confirmation as the molecular mechanisms underlying the golden coat colour phenotype has not been previously studied (Taylor *et al.*, 2016). The term "split" is often used to refer to animals that are normal in appearance, but are thought to be a carrier of the causal variant (Smith, 2013).

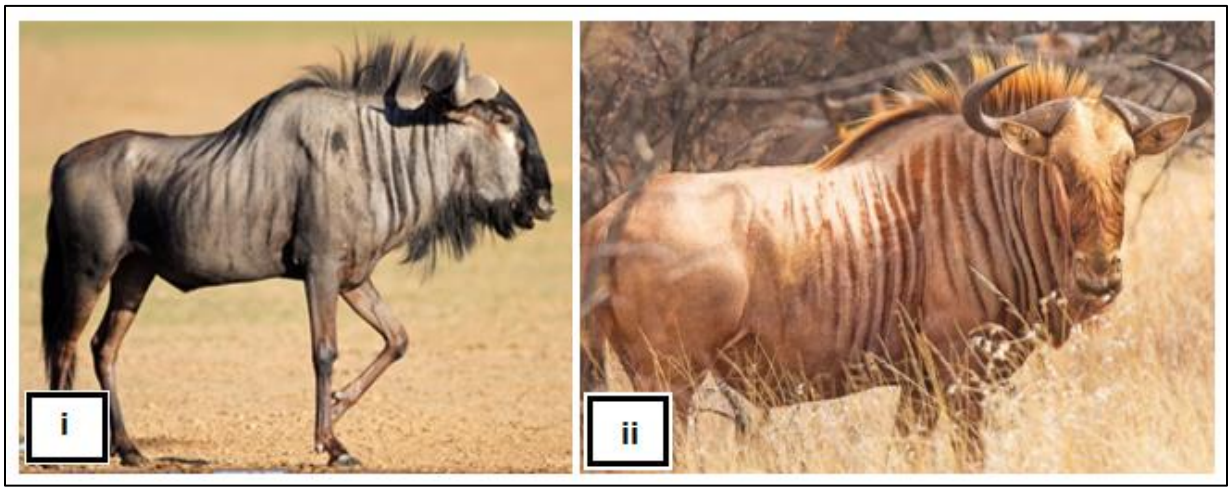


Figure 1.7. Colour variants of the blue wildebeest (*Connochaetes taurinus taurinus)*:
(i) common (figure taken from https://www.aboutanimals.com/mammal/blue-wildebeest/),
(ii) golden (figure taken from http://www.xtremewildlifegroup.co.za/wildebeest/gold/).

1.6.2. Ranching of Blue and Golden Wildebeest

Blue wildebeest are very tough, vigorous and disease resistant animals. Furthermore, they are known to be very fertile since they have a 90% conception rate, as well as a high birth rate (Gouws, 2015; Wildswinkel, 2015). The natural occurrence of a colour variant, therefore, make them among the most profitable and sustainable game to breed with (Wildswinkel, 2015). In 1991, only three golden wildebeest were in private ownership, but this number has grown to an estimated 1 500 animals in 2015 due to the high monetary value associated with these animals (Wildswinkel, 2015). In 2016, a blue wildebeest sold for an average auction price of ZAR 3 137 while a golden wildebeest sold for an average price of ZAR 418 090 (Cloete, 2017a). In addition, two particular standouts in the sale of golden wildebeest also occurred in 2016: a golden wildebeest bull sold for a new South African record price of ZAR 6 million while another golden wildebeest bull sold for ZAR 5.7 million (Wildswinkel, 2016; Cloete, 2017a). This indicates that game breeders have started to focus on methods that can combine the breeding of colour variants with the breeding of animals with excellent genetic quality to give them a competitive advantage in the market (Joubert, 2017).

The common herd structure among breeders is a single adult bull to 20 to 30 breeding females. Furthermore, three distinguishable wildlife production systems are used for blue- and golden wildebeest ranching, namely the extensive wildlife production system, the semi-extensive production system and the intensive production system (Van Hoving, 2011; Taylor *et al.*, 2016). Extensive wildlife ranching entails the managed production of free living wildebeest on a large, usually fenced, area with limited human interference in the form of supplemental food and water, veterinary health care and parasite control (Van Hoving, 2011). In semi-extensive breeding environments, habitat integrity is maintained by regular management interventions and animals are provided with supplemental food and water during dry periods. However, interference with animals through the provision of veterinary care is uncommon. Conversely, in intensive breeding environments wildebeest are confined in small to medium sized fenced enclosures where they are protected from predators and provided with water, food and veterinary health care (NAMC, 2006; Taylor *et al.*, 2016). Due to the high individual value of golden wildebeest, breeding is mostly conducted under intensive or semi-extensive conditions that reduce the risk of depredation and increase the production rate (Taylor *et al.*, 2016).

1.6.3. Study Rationale

Since the golden wildebeest has become such a common colour variant for wildlife ranchers to breed with in South Africa, research concerning these animals has become increasingly important to assist in their management and

regulation. This research includes the elucidation of the genetic mechanisms underlying coat colour variation in blue wildebeest in order to gain a better understanding of the coat colour phenotype that is currently being selected for. As mentioned previously, genomic resources, such as molecular markers, are often used to investigate and unravel the genetic basis of important phenotypic traits. However, the number of genomic resources available for blue wildebeest are very limited at present. The only molecular markers that have been identified for blue wildebeest is a set of 17 blue wildebeest-specific microsatellite loci published by Røed *et al.* (2011) and a small set of 11 microsatellite markers and 23 SNPs, identified in both blue and black wildebeest, published recently by Van Wyk e*t al*. (2018; 2019). A panel of microsatellite makers consisting of markers originally developed for cattle, sheep and goat have also been optimised by Miller *et al.* (2016a) for use in blue wildebeest.

In light of the numerous benefits inherent in having a set of genome-wide SNP markers available, the first aim of this study was, therefore, to develop a SNP marker panel for blue wildebeest using the DArTseq platform. For this purpose, the following objectives were set:

(i) Assess the efficacy and capability of the DArTseq platform to generate high-quality SNPs for blue wildebeest.

(ii) Estimate the genome-wide distribution of the DArTseq SNPs by performing *in silico* mapping of the DArTseq marker sequences to the contiguous sequences (contigs) of the bovine genome.

(iii) Assess the possible structural and functional effects of the DArTseq generated SNPs by aligning the SNP marker sequences to the bovine genome to identify orthologous genes, and by performing functional analysis in order to assign the SNPs to different gene ontology (GO) terms.

The second aim of this study was to identify genes that are associated with coat colour in blue wildebeest. The following objectives were set to achieve this aim:

(i) Apply the DArTseq generated SNPs in a case-control GWAS to identify SNPs associated with coat colour in blue wildebeest.

(ii) Identify SNPs significantly associated with coat colour in blue wildebeest that are located in genes that might play a functional role in pigmentation.

(iii) Perform a gene-gene interaction analysis to determine if a possible interaction occurs between any of the SNPs located in the pigmentation candidate genes.

(iv) Establish the mode of inheritance of the golden coat colour phenotype in blue wildebeest.

1.6.4. Study Outline

The objectives of the first aim will be addressed in the first experimental chapter (chapter 2) which will cover the development of the blue wildebeest DArTseq SNP markers, and will discuss the quality of the developed SNP markers compared to those generated by other similar platforms. The functional significance of the generated SNPs will also be discussed. The second experimental chapter (chapter 3) will address the objectives of the second aim, and will discuss the possible role of the significant SNPs, and their associated genes, in determining coat colour in blue wildebeest. The final chapter (chapter 4) will summarise the findings of the experimental chapters and discuss how the findings of the current study can be of use in the South African wildlife industry. In addition, the final chapter will also review the limitations of the study and provide suggestions for future research projects.

REFERENCES

Ackermann, R. R., Brink, J. S., Vrahimis, S., & De Klerk, B. (2010). Hybrid wildebeest (Artiodactyla: Bovidae) provide further evidence for shared signatures of admixture in mammalian crania. *South African Journal of Science*, *106*(11–12), 1–5.

Adetunji, S. A., Krecek, R. C., O'Dell, N., Prozesky, L., Steyl, J., & Arenas-Gamboa, A. M. (2018). Melanoma in golden and king wildebeests. *Journal of Zoo and Wildlife Medicine*, *49*(1), 134–142.

Amos, W., Driscoll, E., & Hoffman, J. I. (2011). Candidate genes versus genome-wide associations: Which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society B: Biological Sciences*, *278*(1709), 1183–1188.

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics*, *2*(2), 130–138.

André, C., Guaguère, E., Chaudieu, G., Genevois, J.-P., & Devauchelle, P. (2017). The importance of dogs for comparative pathology and genetics: Examples of shared resources and programmes. *Revue Vétérinaire Clinique*, *52*(3), 55–70.

Bajetha, G., Bhati, J., Jaiswal, S., Iquebal, M. A., Rai, A., Arora, V., & Kumar, D. (2013). Analysis and functional annotation of expressed sequence tags of water buffalo. *Animal Biotechnology*, *24*(1), 25–30.

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Çiftçi, V., Sáenz de Miera, L. E., Aasim, M., Nadeem, M. A., Aktaş, H., Özkan, H., & Hatipoğlu, R. (2017). A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. *PLoS ONE*, *12*(1), e0167821.

Barral, D. C., & Seabra, M. C. (2004). The melanosome as a model to study organelle motility in mammals. *Pigment Cell Research*, *17*(2), 111–118.

Beermann, F., Orlow, S. J., & Lamoreux, M. L. (2004). The *Tyr* (*albino*) locus of the laboratory mouse. *Mammalian Genome*, *15*(10), 749–758.

Ben-Ari, G., & Lavi, U. (2012). Marker-assisted selection in plant breeding. In A. Altman, & P. M. Hasegawa (Eds.), *Plant Biotechnology and Agriculture: Prospects for the 21st Century* (1st ed., pp. 163–184). London: Academic Press.

Bezuidenhout, R. (2012, July 20). Colour under investigation. *Farmer's Weekly*. Retrieved from https://www.farmersweekly.co.za [accessed 22 March 2018].

Borecki, I. B., & Province, M. A. (2008). Linkage and association: Basic concepts. *Advances in Genetics*, *60*, 51-74.

Bovine Genome Sequencing and Analysis Consortium, Elsik, C. G., Tellam, R. L., & Worley, K. C. (2009). The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science*, *324*(5926), 522–528.

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Computational Biology*, *8*(12), e1002822.

Challa, A. K., Boitet, E. R., Turner, A. N., Johnson, L. W., Kennedy, D., Downs, E. R., Hymel, K. M., Gross, A. K., & Kesterson, R. A. (2016). Novel hypomorphic alleles of the mouse tyrosinase gene induced by CRISPR-Cas9 nucleases cause non-albino pigmentation phenotypes. *PLoS ONE*, *11*(5), e0155812.

Charon, K. M., & Lipka, K. R. (2015). The effect of a coat colour-associated genes polymorphism on animal health - A review. *Annals of Animal Science*, *15*(1), 3–17.

Chintala, S., Li, W., Lamoreux, M. L., Ito, S., Wakamatsu, K., Sviderskaya, E. V., Bennett, D. C., Park, Y.-M., Gahl, W. A., Huizing, M., Spritz, R. A., Ben, S., Novak, E. K., Tan, J., & Swank, R. T. (2005). *Slc7a11* gene controls production of pheomelanin pigment and proliferation of cultured cells. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 10964–10969.

Cichorek, M., Wachulska, M., & Stasiewicz, A. (2013a). Heterogeneity of neural crest-derived melanocytes. *Central European Journal of Biology*, *8*(4), 315–330.

Cichorek, M., Wachulska, M., Stasiewicz, A., & Tymińska, A. (2013b). Skin melanocytes: Biology and development. *Postępy Dermatologii i Alergologii*, *30*(1), 30–41.

Cieslak, M., Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Colours of domestication. *Biological Reviews*, *86*(4), 885–899.

Clark, L. A., Wahl, J. M., Rees, C. A., & Murphy, K. E. (2006). From the cover: Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proceedings of the National Academy of Sciences of the United State of America*, *103*(5), 1376–1381.

Clarke, G. M., Anderson, C. A., Pettersson, F. H., Cardon, L. R., & Andrew, P. (2011). Basic statistical analysis in genetic case-control studies. *Nature Protocols*, *6*(2), 121–133.

Cloete, F. (2015a). *Growth Expectations for the South African Game Ranching Industry*. Retrieved from https://www.grainsa.co.za/growth-expectations-for-the-south-african-game-ranching-industry [accessed 11 September 2017].

Cloete, F. (2015b). Plains game, colour variants and strange species. *ABSA Agricultural Outlook*, 95–97.

Cloete, F. (2017a). Average prices: Wildlife auctions 2014-2016. *Wildlife Stud Services: Wildlife Breeders Journal*, 146–149.

Cloete, F. (2017b). Did the bubble in the game breeding industry burst? *Wildlife Stud Services: Wildlife Breeders Journal*, 12–13.

Colombo, S., Berlin, I., Delmas, V., & Larue, L. (2011). Classical and nonclassical melanocytes in vertebrates. In J. Borovanský, & P. A. Riley (Eds.), *Melanins and Melanosomes: Biosynthesis, Biogenesis, Physiological and*

Pathological Functions (1st ed., pp. 21–62). Weinheim: Wiley-Blackwell.

D'Mello, S. A. N., Finlay, G. J., Baguley, B. C., & Askarian-Amiri, M. E. (2016). Signaling pathways in melanogenesis. *International Journal of Molecular Sciences*, *17*(7), 1144.

Da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L., Zepeda-Mendoza, M. L., Campos, P. F., Heler, R., & Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*, 3–13.

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*(7), 499–510.

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. (2013). Genotyping-by-sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE*, *8*(5), e62137.

De Haas, A. R. (2016). *Monitoring adrenocortical function as a measure of stress in blue wildebeest (Connochaetes taurinus)* (Unpublished MSc thesis). University of Pretoria, Pretoria, South Africa.

Department of Environmental Affairs (DEA) (2016). *National biodiversity economy strategy (NBES)*. Retrieved from https://www.environment.gov.za/sites/default/files/reports/nationalbiodiversityeconomystrategy.pdf [accessed 6 February 2018].

Department of Environmental Affairs (DEA) (2018). *Unlocking the socio-economic potential of South Africa's biodiversity assets through sustainable use of wildlife resources* [PowerPoint presentation]. Retrieved from http://pmg-assets.s3-website-eu-west-1.amazonaws.com/180227DEA.pdf [accessed 6 February 2018].

Doveri, S., Lee, D., Maheswaran, M., & Powell, W. (2007). Molecular markers: History, features and application. In C. Kole, & A. G. Abbott (Eds.), *Principles and Practices of Plant Genomics, Volume 1* (1st ed., pp. 23–68). Enfield: Science Publishing Group.

Duran, C., Appleby, N., Edwards, D., & Batley, J. (2009). Molecular genetic markers: Discovery, applications, data storage and visualisation. *Current Bioinformatics*, *4*(1), 16–27.

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*(5), e19379.

Estes, R., & East, R. (2009). *Status of the wildebeest (*Connochaetes taurinus*) in the wild, 1967-2005* (WCS Working paper no 37). Retrieved from https://library.wcs.org/DesktopModules/Bring2mind/DMX/Download.aspx?EntryId=30992&PortalId=0&DownloadMethod=attachment [accessed 12 August 2018].

Evangelou, E., Trikalinos, T. A., Salanti, G., & Ioannidis, J. P. A. (2006). Family-based versus unrelated case-control designs for genetic associations. *PLoS Genetics*, *2*(8), 1147–1155.

Fan, Y., Wang, P., Fu, W., Dong, T., Qi, C., Liu, L., Guo, G., Li, C., Cui, X., Zhang, S., Zhang, Q., Zhang, Y., & Sun, D. (2014). Genome-wide association study for pigmentation traits in Chinese Holstein population. *Animal Genetics*, *45*(5), 740–744.

Fingerlin, T. E., Weiss, S. T., & Schwartz, D. A. (2016). Genetics of lung disease. In V. C. Broaddus, R. J. Mason, J. D. Ernst, T. E. King, S. C. Lazarus, J. F. Murray, J. A. Nadel, A. S. Slutsky, & M. B. Gotway (Eds.), *Murray and Nadel's Textbook of Respiratory Medicine* (6th ed., pp. 32–43). Philadelphia: Saunders.

Fontanesi, L., Beretti, F., Riggio, V., Dall'Olio, S., González, E. G., Finocchiaro, R., Davoli, R., Russo, V., & Portolano, B. (2009). Missense and nonsense mutations in melanocortin 1 receptor (*MC1R*) gene of different goat breeds: Association with red and black coat colour phenotypes but with unexpected evidences. *BMC Genetics*, *10*(47), 1–12.

Fontanesi, L., Dall'Olio, S., Beretti, F., Portolano, B., & Russo, V. (2011). Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (*ASIP*) and melanocortin 1 receptor (*MC1R*) genes. *Animal*, *5*(1), 8–17.

Furstenburg, D. (2013). *Focus on the Blue Wildebeest (*Connochaetes taurinus*)*. Retrieved from https://www.researchgate.net/publication/316154370_Focus_on_the_Blue_Wildebeest_Connochaetes_taurinus [accessed 12 August 2018].

Furstenburg, D. (2016). Springbok (*Antidorcas marsupialis*). In P. Oberem, & P. Oberem (Eds.), *The New Game Rancher* (1st ed., pp. 226–234). Pretoria: Briza Publications.

Gabaldón, T., & Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, *14*(5), 360–366.

Gamaniel, I. B., & Gwaza, D. S. (2017). Molecular characterization of animal genetics resources, its potential for use in developing countries. *Journal of Genetics and Genetic Engineering*, *1*(1), 43–57.

Garrido-Cardenas, J. A., Mesa-Valle, C., & Manzano-Agugliaro, F. (2018). Trends in plant research using molecular markers. *Planta*, *247*(3), 543–557.

Geng, X., Degui, Z., & Liu, Z. (2017). Genome-wide association studies of performance traits. In Z. Liu (Ed.), *Bioinformatics in Aquaculture: Principles and Methods* (1st ed., pp. 415–433). New York: John Wiley & Sons.

Glanzmann, B., Möller, M., Le Roex, N., Tromp, G., Hoal, E. G., & Van Helden, P. D. (2016). The complete genome sequence of the African buffalo (*Syncerus caffer*). *BMC Genomics*, *17*(1001), 1–7.

Gordish-Dressman, H., & Devaney, J. M. (2011). Statistical and methodological considerations in excercise genomics. In L. S. Pescatello, & S. M. Roth (Eds.), *Exercise Genomics* (1st ed., pp. 23–44). London: Humana Press.

Gore, M., Bradbury, P., Hogers, R., Kirst, M., Verstege, E., Van Oeveren, J., Peleman, J., Buckler, E., & Van Eijk, M.

(2007). Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. *Crop Science*, *47*(S2), S135–S148.

Gouws, A. (2015). How golden wildebeest saved a cattle farm. *Stockfarm*, *5*(9), 18–19.

Grobler, P., Van Wyk, A. M., Dalton, D. L., Van Vuuren, B. J., & Kotzé, A. (2018). Assessing introgressive hybridization between blue wildebeest (*Connochaetes taurinus*) and black wildebeest (*Connochaetes gnou*) from South Africa. *Conservation Genetics*, *19*(4), 981–993.

Grubb, P. (2005). Order Artiodactyla. In D. E. Wilson, & D. M. Reeder (Eds.), *Mammal Species of the World: A Taxonomic and Geographic Reference, Volume 2* (3rd ed., pp. 676). Baltimore: John Hopkins University Press.

Guo, Y., & Jamison, D. C. (2005). The distribution of SNPs in human gene regulatory regions. *BMC Genomics*, *6*(140), 1–11.

Gurgul, A., Miksza-Cybulska, A., Szmatoła, T., Jasielczuk, I., Piestrzyńska-Kajtoch, A., Fornal, A., Semik-Gurgul, E., & Bugno-Poniewierska, M. (2019). Genotyping-by-sequencing performance in selected livestock species. *Genomics*, *111*(2), 186–195.

Haase, B., Brooks, S. A., Tozaki, T., Burger, D., Poncet, P. A., Rieder, S., Hasegawa, T., Penedo, C., & Leeb, T. (2009). Seven novel *KIT* mutation in horses with white coat colour phenotypes. *Animal Genetics*, *40*(5), 623-629.

Hardison, R. C. (2003). Comparative genomics. *PLoS Biology*, *1*(2), 156–160.

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*(S1), 150–161.

Hauswirth, R., Haase, B., Blatter, M., Brooks, S. A., Burger, D., Drögemüller, C., Gerber, V., Henke, D., Janda, J., Jude, R., Gary Magdesian, K., Matthews, J. M., Poncet, P.-A., Svansson, V., Tozaki, T., Wilkinson-White, L., Penedo, C. T., Rieder, S., & Leeb, T. (2012). Mutations in *MITF* and *PAX3* cause "splashed white" and other white spotting phenotypes in horses. *PLoS Genetics*, *8*(4), e1002653.

He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, *5*, 484.

Heyneke, G. (2017). *The colourful game changer*. Retrieved from http://www.wrsa-ec.co.za/2017/02/03/1315/ [accessed 26 February 2018].

Hilton-Barber, B., & Berger, L. R. (2004). *Field Guide to the Cradle of Humankind : Sterkfontein, Swartkrans, Kromdraai & Environs World Heritage Site* (2nd ed.). Cape Town: Struik.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, *6*(2), 95–108.

Hoekstra, H. E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*,

97(3), 222–234.

Hou, L., Panthier, J. J., & Arnheiter, H. (2000). Signaling and transcriptional regulation in the neural crest-derived melanocyte lineage: Interactions between KIT and MITF. *Development*, *127*(24), 5379–5389.

Hu, Z.-L., Park, C. A., & Reecy, J. M. (2019). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Research*, *47*(Database Issue), D701–D710.

Huffman, B. (2005). *Brindled gnu, blue wildebeest*. Retrieved from http://www.ultimateungulate.com/Artiodactyla/Connochaetes_taurinus.html [accessed 23 January 2018].

Jaccoud, D., Peng, K., Feinstein, D., & Kilian, A. (2001). Diversity Arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, *29*(4), e25.

Jackson, I. J. (1997). Homologous pigmentation mutations in human, mouse and other model organisms. *Human Molecular Genetics*, *6*(10), 1613–1624.

Jaja, I. F., Mushonga, B., Green, E., & Muchenje, V. (2016). Prevalence of lung lesions in slaughtered cattle in the Eastern Cape Province , South Africa. *Journal of the South African Veterinary Association*, *87*(1), 1362.

Janovsky, E. (2016). Wildlife industry expected to continue to grow. *Wildlife Stud Services: Wildlife Breeders Journal*, 12–14.

Jehan, T., & Lakhanpaul, S. (2006). Single nucleotide polymorphism (SNP) - Methods and applications in plant genetics: A review. *Indian Journal of Biotechnology*, *5*(4), 435–459.

Jin, Y., Liu, S., Yuan, Z., Yang, Y., Tan, S., & Liu, Z. (2016). Catfish genomic studies: Progress and perspectives. In S. MacKenzie, & S. Jentoft (Eds.), *Genomics in Aquaculture* (1st ed., pp. 73–104). London: Academic Press.

Johansson, A., Pielberg, G., Andersson, L., & Edfors-Lilja, I. (2005). Polymorphisms at the porcine *Dominant white*/*KIT* locus influence coat colour and peripheral blood cell measures. *Animal Genetics*, *36*(4), 288–296.

Jorgensen, T. J., Ruczinski, I., Kessing, B., Smith, M. W., Shugart, Y. Y., & Alberg, A. J. (2009). Hypothesis-driven candidate gene association studies: Practical design and analytical considerations. *American Journal of Epidemiology*, *170*(8), 986–993.

Joubert, A. (2017). Microsatellite markers: Understanding the basics. *Wildlife Stud Services: Wildlife Breeders Journal*, 44–46.

Kersey, P. J. (2019). Plant genome sequences: Past, present, future. *Current Opinion in Plant Biology*, *48*, 1–8.

Kim, S., & Misra, A. (2007). SNP genotyping: Technologies and biomedical applications. *Annual Review of Biomedical Engineering*, *9*(1), 289–320.

Kondo, T., & Hearing, V. J. (2011). Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Review of Dermatology*, *6*(1), 97–108.

König, S., Romoth, L., & Stanke, M. (2018). Comparative genome annotation. In J. Setuba, J. Stoye, & P. Stadler (Eds.), *Comparative Genomics: Methods and Protocols, Volume 1704* (1st ed., pp. 189–212). New York: Humana Press.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*(29), 1–9.

Kruger, J. C., Skinner, J. D., & Robinson, T. J. (1979). On the taxonomic status of the black and white springbok, *Antidorcas marsupialis*. *South African Journal of Science*, *75*(99), 411–412.

Kunieda, T. (2005). Identification of genes responsible for hereditary diseases in Japanese beef cattle. *Animal Science Journal*, 76(6), 525–533.

Laird, N. M., & Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics*, *7*(5), 385–394.

Li, M. H., Tiirikka, T., & Kantanen, J. (2014). A genome-wide scan study identifies a single nucleotide substitution in *ASIP* associated with white versus non-white coat-colour variation in sheep (*Ovis aries*). *Heredity*, *112*(2), 122–131.

Li, W., Bickhart, D. M., Ramunno, L., Iamartino, D., Williams, J. L., & Liu, G. E. (2019). Comparative sequence alignment reveals River Buffalo genomic structural differences compared with cattle. *Genomics*, *111*(3), 418–425.

Lindsey, P. A., Romañach, S. S., & Davies-Mostert, H. T. (2009). The importance of conservancies for enhancing the value of game ranch land for large mammal conservation in southern Africa. *Journal of Zoology*, *277*(2009), 99–105.

Liu, Z. J., & Cordes, J. F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, *238*(1–4), 1–37.

Martin, P. M., Palhière, I., Ricard, A., Tosser-Klopp, G., & Rupp, R. (2016). Genome wide association study identifies new loci associated with undesired coat color phenotypes in Saanen goats. *PLoS ONE*, *11*(3), e0186029.

Mazza, R., Strozzi, F., Caprera, A., Ajmone-Marsan, P., & Williams, J. L. (2009). The other side of comparative genomics: Genes with no orthologs between the cow and other mammalian species. *BMC Genomics*, *10*(604), 1–15.

Melville, J., Haines, M. L., Boysen, K., Hodkinson, L., Kilian, A., Smith Date, K. L., Potvin, D. A., & Parris, K. M. (2017). Identifying hybridization and admixture using SNPs: Application of the DArTseq platform in phylogeographic research on vertebrates. *Royal Society Open Science*, *4*(7), 161061.

Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. A., Islam, K. N., & Latif, M. A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance.

International Journal of Molecular Sciences, *14*(11), 22499–22528.

Milczarski, P., Hanek, M., Tyrka, M., & Stojałowski, S. (2016). The application of GBS markers for extending the dense genetic map of rye (*Secale cereale* L.) and the localization of the *Rfc1* gene restoring male fertility in plants with the C source of sterility-inducing cytoplasm. *Journal of Applied Genetics*, *57*(4), 439–451.

Miller, S. M., Clarke, A. B., Bloomer, P., Guthrie, A. J., & Harper, C. K. (2016a). Evaluation of microsatellites for common ungulates in the South African wildlife industry. *Conservation Genetics Resources*, *8*(3), 329–341.

Miller, S. M., Guthrie, A. J., & Harper, C. K. (2016b). Single base-pair deletion in *ASIP* exon 3 associated with recessive black phenotype in impala (*Aepyceros melampus*). *Animal Genetics*, *47*(4), 511–512.

Miller, W., Makova, K. D., Nekrutenko, A., & Hardison, R. C. (2004). Comparative genomics. *Annual Review of Genomics and Human Genetics*, *5*(1), 15–56.

Morin, P. A., Foote, A. D., Hill, C. M., Simon-Bouhet, B., Lang, A. R., & Louis, M. (2018). SNP discovery from single and multiplex genome assemblies of non-model organisms. In S. R. Head, P. Ordoukhanian, & D. R. Salomon (Eds.), *Next Generation Sequencing: Methods and Protocols* (1st ed., pp. 113–144). New York: Springer.

Murphy, K. M., Cooper, A., & Tobias, E. S. (2014). The human genome, gene regulation and genomic variation. In S. Padmanabhan (Ed.), *Handbook of Pharmacogenomics and Stratified Medicine* (1st ed., pp. 41–56). London: Academic Press.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnology and Biotechnological Equipment*, *32*(2), 261–285.

Nascimento, A. A., Roland, J. T., & Gelfand, V. I. (2003). Pigment cells: A model for the study of organelle transport. *Annual Review of Cell and Developmental Biology*, *19*(1), 469–491.

National Agricultural Marketing Council (NAMC) (2006). *Report on the Investigation to Identify Problems for Sustainable Growth and Development in South African Wildlife Ranching*. Retrieved from https://www.namc.co.za/wp-content/uploads/2017/09/Report-on-Wildlife-ranching.pdf [accessed 17 February 2018].

Nazari-Ghadikolaei, A., Mehrabani-Yeganeh, H., Miarei-Aashtiani, S. R., Staiger, E. A., Rashidi, A., & Huson, H. J. (2018). Genome-wide association studies identify candidate genes for coat color and mohair traits in the Iranian Markhoz Goat. *Frontiers in Genetics*, *9*, 105.

Negrisolo, E., Bargelloni, L., Patarnello, T., Ozouf-Costaz, C., Pisano, E., Di Prisco, G., & Verde, C. (2008). Comparative and evolutionary genomics of globin genes in fish. In R. Eckenhoff, & I. Dmochowski (Eds.), *Methods in Enzymology, Volume 602: Chemical and Biochemical Approaches for the Study of Anesthetic Function, Part A* (1st ed., pp. 511–537). Cambridge: Academic Press.

Nicoloso, L., Negrini, R., Milanesi, E., & Crepaldi, P. (2008). Identification of polymorphism in the *SCL24A5* gene of cattle. *Italian Journal of Animal Science*, *7*(4), 505–512.

Nobrega, M. A., & Pennacchio, L. A. (2004). Comparative genomic analysis as a tool for biological discovery. *The Journal of Physiology*, *554*(1), 31–39.

Oberem, P. (2015, October 22). Wildlife ranching in South Africa [Blog post]. Retrieved from https:// africageographic.com/blog/wildlife-ranching-in-south-africa/ [accessed 28 March 2018].

Olivier, P. A. (2015). *Colour variation of African Bovidae: Consequences for conservation and the wildlife ranching industry* (Unpublished MSc thesis). University of the Free State, Bloemfontein, South Africa.

Ott, J., Kamatani, Y., & Lathrop, M. (2011). Family-based designs for genome-wide association studies. *Nature Reviews Genetics*, *12*(7), 465–474.

Ott, J., Wang, J., & Leal, S. M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. *Nature Reviews Genetics*, *16*(5), 275–284.

Ozerov, M., Vasemägi, A., Wennevik, V., Diaz-Fernandez, R., Kent, M., Gilbey, J., Prusov, S., Niemelä, E., & Vähä, J. P. (2013). Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. *PLoS ONE*, *8*(12), e82434.

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. *Trends in Genetics*, *29*(2): 66–73.

Park, Y.-J., Lee, J. K., & Kim, N.-S. (2009). Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. *Molecules*, *14*(11), 4546–4569.

Patnala, R., Clements, J., & Batra, J. (2013). Candidate gene association studies: A comprehensive guide to useful *in silico* tools. *BMC Genetics*, *14*(39), 1–11.

Philipp, U., Lupp, B., Mömke, S., Stein, V., Tipold, A., Eule, J. C., Rehage, J., & Distl, O. (2011). A *MITF* mutation associated with a dominant white phenotype and bilateral deafness in German Fleckvieh cattle. *PLoS ONE*, *6*(12), e28857.

Pielberg, G. (2004). *Molecular coat colour genetics: Dominant white in pigs and greying with age in horses* (Unpublished PhD thesis). Swedish University of Agricultural Sciences, Uppsala, Sweden.

Pitman, R. T., Fattebert, J., Williams, S. T., Williams, K. S., Hill, R. A., Hunter, L. T. B., Slotow, R., & Balme, G. A. (2016). The conservation costs of game ranching. *Conservation Letters*, *10*(4), 402–412.

Pontes, M. L., Fondevila, M., Laréu, M. V., & Medeiros, R. (2015). SNP markers as additional information to resolve complex kinship cases. *Transfusion Medicine and Hemotherapy*, *42*(6), 385–388.

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D. M., Auvil, L., Capitanu, B., Ma, J., Lewin, H. A., Qian, X., Lang, Y., Zhou, R., Wang, L., Wang, K., Xia, J., Liao, S., Pan, S., Lu, X., Hou, H., Wang, Y., Zang, X., Yin, Y., Ma, H., Zhang, J., Wang, Z., Zhang, Y., Zhang, D., Yonezawa, T., Hasegawa, M.,

Zhong, Y., Liu, W., Zhang, Y., Huang, Z., Zhang, S., Long, R., Yang, H., Wang, J., Lenstra, J. A., Cooper, D. N., Wu, Y., Wang, J., Shi, P., Wang, J., & Liu, J. (2012). The yak genome and adaptation to life at high altitude. *Nature Genetics*, *44*(8), 946–949.

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., & He, Z. (2017). Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. *Molecular Plant*, *10*(8), 1047–1064.

Reilly, B. K., Sutherland, E. A., & Harley, V. (2003). The nature and extent of wildlife ranching in Gauteng Province, South Africa. *South African Journal of Wildlife Research*, *33*(2), 141–144.

Reissmann, M., & Ludwig, A. (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Seminars in Cell & Developmental Biology*, *24*(6), 576–586.

Ren, J., Mao, H., Zhang, Z., Xiao, S., Ding, N., & Huang, L. (2011). A 6-bp deletion in the *TYRP1* gene causes the brown colouration phenotype in Chinese indigenous pigs. *Heredity*, *106*(5), 862–868.

Riancho, J. A. (2012). Genome-wide association studies (GWAS) in complex diseases: Advantages and limitations. *Reumatología Clínica (English Edition)*, *8*(2), 56–57.

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

Russo, I.-R. M., Hoban, S., Bloomer, P., Kotzé, A., Segelbacher, G., Rushworth, I., Birss, C., & Bruford, M. W. (2019). 'Intentional genetic manipulation' as a conservation threat. *Conservation Genetics Resources*, *11*(2), 237–247.

San-Jose, L. M., & Roulin, A. (2017). Genomics of coloration in natural animal populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1724), 20160337.

Scherer, A., & Christensen, G. B. (2016). Concepts and relevance of genome-wide association studies. *Science Progress*, *99*(1), 59–67.

Schultz, A. J., Cristescu, R. H., Littleford-Colquhoun, B. L., Jaccoud, D., & Frère, C. H. (2018). Fresh is best: Accurate SNP genotyping from koala scats. *Ecology and Evolution*, *8*(6), 3139–3151.

Slatkin, M. (2008). Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), 477–485.

Smith, N. (2013, July 29). The York family & breeding golden wildebeest. *Farmer's Weekly*. Retrieved from https://www. farmersweekly.co.za [accessed 25 June 2017].

Soto-Cerda, B. J., & Cloutier, S. (2012). Association mapping in plant populations. In M. Çalişkan (Ed.), *Genetic Diversity in Plants* (1st ed., pp. 29–54). Rijeka: IntechOpen.

Spielmann, M., & Mundlos, S. (2016). Looking beyond the genes: The role of non-coding variants in human disease.

Human Molecular Genetics, *25*(R2), R157-R165.

Strauss, J. S. (2015). *The long term financial sustainability of the high value game breeding sector in South Africa* (Unpublished MBA thesis). University of the Free State, Bloemfontein, South Africa.

Sturm, R. A. (2006). A golden age of human pigmentation genetics. *Trends in Genetics*, *22*(9), 464–468.

Sturm, R. A., & Frudakis, T. N. (2004). Eye colour: Portals into pigmentation genes and ancestry. *Trends in Genetics*, *20*(8), 327–332.

Tambling, C., Viljoen, P., & Kotze, A. (2016). A conservation assessment of *Connochaetes taurinus*. In M. F. Child, L. Roxburgh, E. Do Lihn San, D. Raimondo, & H. T. Davies-Mostert (Eds.), *The Red List of Mammals of South Africa, Swaziland and Lesotho*. South African National Biodiversity Institute and Endangered Wildlife Trust, South Africa.

Taylor, W. A., Lindsey, P. A., & Davies-Mostert, H. (2016). *An assessment of the economic, social and conservation value of the wildlife ranching industry and its potential to support the green economy in South Africa*. Endangered Wildlife Trust, South Africa.

Teneva, A. (2009). Molecular markers in animal genome analysis. *Biotechnology in Animal Husbandry*, *25*(5–6), 1267–1284.

Thomson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breeding and Biotechnology*, *2*(3), 195–212.

Trevino, V., Falciani, F., & Hugo, A. B.-S. (2007). DNA microarrays: A powerful genomic tool for biomedical and clinical research. *Molecular Medicine*, *13*(9–10), 527–541.

Van der Merwe, P., & Saayman, M. (2005). Game farms as sustainable ecotourist attractions. *Koedoe*, *48*(2), 1–10.

Van Gele, M., & Lambert, J. (2011). Transport and distribution of melanosomes. In J. Borovanský, & P. A. Riley (Eds.), *Melanins and Melanosomes: Biosynthesis, Biogenesis, Physiological and Pathological Functions* (1st ed., pp. 295–322). Weinheim: Wiley-Blackwell.

Van Hoven, W. (2015). Private game reserves in southern Africa. In R. Van der Duim, M. Lamers, & J. Van Wyk (Eds.), *Institutional Arrangements for Conservation, Development and Tourism in Eastern and Southern Africa: A Dynamic Perspective* (1st ed., pp. 101–118). Dordrecht: Springer.

Van Hoving, S. C. P. (2011). *An evaluation of different extensive wildlife production systems in the Western Cape Province of South Africa* (Unpublished MSc thesis). University of Stellenbosch, Stellenbosch, South Africa.

Van Rooyen, L. (2012, July 2). Colour variant game naturally profitable. *Farmer's Weekly*. Retrieved from https://www. farmersweekly.co.za [accessed 5 October 2017].

Van Wyk, A. M., Kotzé, A., Grobler, J. P., Janse van Vuuren, B., Barrow, L. N., & Dalton, D. L. (2018). Isolation and

characterization of species-specific microsatellite markers for blue and black wildebeest (*Connochaetes taurinus and C. gnou*). *Journal of Genetics*, *97*(Online Resources), e101–e109.

Van Wyk, A. M., Labuschagne, C., Kropff, A. S., Kotzé, A., Grobler, J. P., Jansen van Vuuren, B., & Dalton, D. L. (2019). A targeted gene approach to SNP discovery in the blue (*Connochaetes taurinus*) and black wildebeest (*C. gnou*). *Conservation Genetic Resources*, *11*(1), 35–38.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, *39*(3), 312–328.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution*, *34*(3), 275–305.

Von Thaden, A., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlidis, A. A., Tiesmeyer, A., Nowak, C., & Muñoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports*, *7*(1), 10768.

Wachi, N., Matsubayashi, K. W., & Maeto, K. (2017). Application of next-generation sequencing to the study of non-model insects. *Entomological Science*, *21*(1), 3–11.

Wildswinkel (2015). *Golden wildebeest are regal animals*. Retrieved from https://wildswinkel.co.za/golden-wildebeest-is-royal-game/ [accessed 9 October 2017].

Wildswinkel (2016). *Piet du Toit Wildbedryf Auction 2016 - Final Auction Results*. Retrieved from https://wildswinkel.co.za/wp-content/uploads/2018/09/Wildswinkel-Piet-du-Toit-Auction-2016-Results-Final-150216.pdf [accessed 15 November 2018].

Xu, X., Dong, G. X., Hu, X. S., Miao, L., Zhang, X. L., Zhang, D. L., Yang, H. D., Zhang, T. Y., Zou, Z. T., Zhang, T. T., Zhuang, Y., Bhak, J., Cho, Y. S., Dai, W. T., Jiang, T. J., Xie, C., Li, R., & Luo, S. J. (2013). The genetic basis of white tigers. *Current Biology*, *23*(11), 1031–1035.

Yadav, A. K., Tomar, S. S., Jha, A. K., & Singh, J. (2017). Importance of molecular markers in livestock improvement: A review. *International Journal of Agriculture Innovations and Research*, *5*(4), 614–621.

Yamaguchi, Y., Brenner, M., & Hearing, V. J. (2007). The regulation of skin pigmentation. *Journal of Biological Chemistry*, *282*(38), 27557–27561.

Yang, H.-B., Kang, W.-H., Nahm, S.-H., & Kang, B.-C. (2015). Methods for developing molecular markers. In H.-J. Koh, S.-Y. Kwon, & M. Thomson (Eds.), *Current Technologies in Plant Molecular Breeding: A Guide Book of Plant Molecular Breeding for Researchers* (1st ed., pp. 15–50). Dordrecht: Springer.

Yang, W., Kang, X., Yang, Q., Lin, Y., & Fang, M. (2013). Review on the development of genotyping methods for assessing farm animal diversity. *Journal of Animal Science and Biotechnology*, *4*(2), 1–6.

Zhang, F., & Lupski, J. R. (2015). Non-coding genetic variants in human disease. *Human Molecular Genetics*, *24*(R1),

R102-R110.

Zhang, Z., Miteva, M. A., Wang, L., & Alexov, E. (2012). Analyzing effects of naturally occurring missense mutations. *Computational and Mathematical Methods in Medicine*, *2012*, 805827.

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: Progress and challenges. *International Journal of Biological Sciences*, *3*(7), 420–427.

CHAPTER 2

The Development of Genome-Wide Single Nucleotide Polymorphism (SNP) Markers in Blue Wildebeest using the DArTseq Platform

Abstract

Blue wildebeest (*Connochaetes taurinus taurinus*) and its golden coat colour variant are economically important antelope species that are widely utilised in the South African wildlife industry. However, very few genomic resources are available for the blue wildebeest that can assist in breeding management and facilitate research. This study, therefore, aimed to develop a set of genome-wide SNP markers for blue wildebeest. The DArTseq genotyping platform was selected for SNP discovery. A total of 20 563 SNPs, each located within a 69 bp sequence, were generated using the DArTseq platform. The developed SNP markers had a high average scoring reproducibility (>99%) and a low percentage missing data (~9.21%) compared to other reduced representation sequencing approaches. Furthermore, the number of candidate SNPs per nucleotide position decreased toward the 3' end of sequence reads. In addition, the ratio of transitions (Ts) to transversions (Tv) remained similar for each read position, indicating that there was no read position bias, such as the identification of false SNPs due to low sequencing quality towards the tail-end of sequencing reads. The DArTseq platform was also successful in identifying a large number of informative SNPs with desirable polymorphism parameters such as a high minor allele frequency (MAF). The *Bos taurus* genome was used for the *in silico* mapping of the marker sequences and a total of 6 020 (29.28%) sequences were successfully mapped against the bovine genome. The marker sequences mapped to all of the bovine chromosomes establishing the genome-wide distribution of the SNPs. Moreover, the high observed Ts:Tv ratio (2.84:1) indicate that the DArTseq platform targeted gene-rich regions of the blue wildebeest genome. Finally, functional annotation of the marker sequences revealed a wide range of different putative functions indicating that these SNP markers can be useful in functional gene studies. The DArTseq platform, therefore, represents a high-throughput, robust and cost-effective genotyping platform, which may find adoption in several other African antelope species.

2.1. Introduction

The South African wildlife industry is a multi-billion Rand (ZAR) enterprise centred on wildlife ranching activities that include the stocking, breeding, trading, and hunting of game animals (Cloete & Rosouw, 2014). During the past decade, the industry has shifted from mainly focusing on hunting and ecotourism to the breeding of high-value colour and morphological variants (Miller *et al.*, 2016; Joubert, 2017). The practice generally involves the intensive breeding of game animals in small to medium sized fenced enclosures where breeding conditions may be manipulated to select for animals with desirable traits [Endangered Wildlife Trust (EWT), 2016; Taylor *et al.*, 2016]. Recently, the profitability and ecological sustainability of intensive wildlife breeding has become a highly debated topic among conservationists (Russo *et al.*, 2019). Specific concerns regarding the intensive breeding of wildlife have been raised, which include the distortion of the natural process of evolution, the loss of genetic diversity, the fixation of deleterious alleles, weakened resilience to environmental changes and reduced reproductive fitness of captive stock (Chardonnet & Mallon, 2015; Miller *et al.*, 2016; Taylor *et al.*, 2016).

It is becoming increasingly evident that genetics should be considered when developing management strategies and conservation policies for game species (Kotzé *et al.*, 2002; Miller *et al.*, 2016). Genetic markers can provide a powerful tool to address wildlife management concerns such as effective population size, levels of gene flow and population viability (DeYoung & Honeycutt, 2005; Russo *et al.*, 2019). The novel insights genetic markers are able to provide into the genetic diversity and structure of a population can be effectively integrated into wildlife management and monitoring programmes (Sarre & Georges, 2009). Furthermore, the data can also be used to assess genotype-phenotype associations in order to develop a deeper understanding of the phenotypic traits that are currently being selected for in the industry (Peterson *et al.*, 2012; Russo *et al.*, 2019). In recent years, SNPs have been advocated as an excellent marker choice for genetic analysis due to the numerous advantages associated with these markers, such as low mutation rates, cost efficiency, processing efficiency and ease of scoring (Ozerov *et al.*, 2013; Gurgul *et al.*, 2019). In addition, compared to traditional microsatellite markers, SNPs have a denser and more uniform distribution across genomes, which make these polymorphisms ideal molecular markers for population and mapping studies (Xing e*t al.*, 2005).

Despite the many advantages associated with the availability of a molecular marker panel, the number of genetic resources available to assist with decisions in the wildlife industry are very limited at present (Blanchong *et al.*, 2016; Miller *et al.*, 2016; Taylor *et al.*, 2016). Consequently, an urgent need exists to identify a low-cost, high-throughput, robust system for the development of genomic resources, such as SNPs, in game species. The advent of next-generation sequencing (NGS) has greatly facilitated the development of several methods that enable the

simultaneous discovery and genotyping of thousands of SNPs, even in species for which little or no genetic information is available (Ray & Satya, 2014; Torkamaneh *et al.*, 2016). This revolution in the discovery of genetic markers has greatly facilitated advances in molecular breeding and population management (Torkamaneh *et al.*, 2016). The most commonly used NGS-based genotyping approaches make use of restriction enzymes (REs) to capture a reduced representation of a genome, followed by sequencing of multiplexed samples. This family of reduced-representation genotyping approaches are collectively generically referred to as genotyping-by-sequencing (GBS; Davey *et al.*, 2011; Torkamaneh *et al.*, 2016).

The DArTseq platform, developed by Diversity Arrays Technology (https://www. diversityarrays.com), is a variant of the GBS approach that has experienced an increasing interest worldwide. This technology is able to generate thousands of high quality molecular markers in a timely and cost-effective manner (approximately ZAR 350-500/sample at the time of this report; Sansaloni *et al.*, 2011; Nguyen *et al.*, 2018). The DArTseq approach implements complexity reduction methods, initially optimised on the DArT microarray platform (Jaccoud *et al.*, 2001), that effectively targets low-copy sequences of the genome (Melville *et al.*, 2017). This usually involves the combined use of a rare cutting RE and a more frequently cutting RE. Furthermore, this process is optimised for each organism and application, by testing different combinations of REs and selecting the combination that is most effective in reducing genome complexity (Nguyen *et al.*, 2018). The DArTseq technology generates two independent molecular marker types – single nucleotide polymorphisms (SNPs) and *silico*DArTs (presence-absence variants) – that are identified from restriction site associated fragments recovered from the sequence data (Morse *et al.*, 2018). The DArTseq platform has traditionally been utilised mostly in polyploid plant species, such as rice (Courtois *et al.*, 2013), barley (Dracatos *et al.*, 2014) and maize (Dos Santos *et al.*, 2016), because SNP detection is mediated by high fidelity REs, rather than relying on the annealing of primers to genomic targets in the presence of homologous annealing sequences (Akbari *et al.*, 2006). More recently, however, DArTseq has also become increasingly more popular in animal systems, including frog (Lambert *et al.*, 2016; Melville *et al.*, 2017), koala (Schultz *et al.*, 2018), and several aquatic species (Lind *et al.*, 2017; Morse *et al.*, 2018; Nguyen *et al.*, 2018).

The blue wildebeest (*Connochaetes taurinus taurinus*; family: Bovidae) is an economically important African antelope species that is regularly sold on game auctions and is widely utilised in commercial hunting in South Africa (Bothma *et al.*, 2010). Furthermore, a colour variant of blue wildebeest, known as golden wildebeest, has become one of the most common colour variants that wildlife ranchers intensively breed with (Tambling *et al.*, 2016; Taylor *et al.*, 2016). Despite the economic importance of blue wildebeest, the only available molecular markers for this species is a set of 17 blue wildebeest-specific microsatellite loci (Røed *et al.*, 2011), a small set of 11 microsatellite markers and 23 SNPs identified in blue and black wildebeest (Van Wyk *et al.*, 2018; 2019), as well

as a microsatellite panel consisting of markers originally developed for cattle, sheep and goat (Miller *et al.*, 2016). Moreover, a study by Miller *et al.* (2016) found the discriminatory power of some of these markers to be low. There is thus an urgent need to develop informative molecular markers for this species that will be able to assist in the genetic management of blue wildebeest. The DArTseq platform provides a cost-effective *de novo* approach that will allow the rapid, high-throughput discovery of SNP markers in blue wildebeest. The current study, therefore, aimed to (1) assess the efficacy and practicality of the DArTseq platform to generate high-quality, genome-wide SNPs for game species by using the blue wildebeest as a model and (2) characterise the DArTseq developed SNP markers for blue wildebeest, which includes performing functional analysis of the generated SNPs.

2.2. Materials and Methods

2.2.1. Sampling and DNA Extraction

Ninety-four blue wildebeest (*C. t. taurinus*) originating from a private game ranch (S24° 45.429' E28° 27.162') located in Limpopo Province, South Africa were selected for DArTseq genotyping, with written and informed consent from the rancher. The selected blue wildebeest included 35 golden wildebeest and 35 split wildebeest (phenotypically normal blue wildebeest, but presumed carriers of the causal variant for golden coat colour based on their parentage records). Samples consisted of either hair (with roots) or whole blood stored in EDTA, which were originally submitted to Unistel Medical Laboratories (Pty) Ltd (Cape Town, South Africa) for routine microsatellite genotyping. Genomic DNA was isolated from each blue wildebeest sample using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's protocol. The extracted DNA was quantified using a Nanodrop 1000 spectrophotometer and adjusted to 50 ng/µl in a final volume of 20 µl. All DNA samples were sent to Diversity Arrays Technology (Pty) Ltd in Canberra, Australia for full RE-digestion, library preparation and GBS data generation.

2.2.2. DArTseq Assay

Four enzyme combinations (*SbfI*/*MseI*, *SbfI*/*HpaII*, *PstI*/*HpaII*, *PstI*/*SphI*) were tested in a pilot study to identify the combination most suitable for genome complexity reduction in blue wildebeest. When the RE combinations for genome complexity reduction were compared, those showing skewed size ranges or increased percentages of repetitive elements were avoided. In this study, the combination of *PstI* and *SphI* performed better in genome complexity reduction and polymorphism detection efficiency.

DNA samples were processed in digestion/ligation reactions (Kilian *et al.*, 2012), ligating two different adaptors corresponding to the *PstI* and *SphI* RE overhangs. The *PstI*-compatible adaptor included the Illumina flow cell attachment sequence, sequencing primer and barcode. The designed barcodes differed in length (between 4 and 8 bp) in order to stagger the sequencing start position, similar to the method reported by Elshire *et al.* (2011). The reverse adaptor contained the Illumina flow cell attachment region and the *SphI*-compatible overhang sequence. The *PstI* –*SphI* ligated fragments were amplified by adapter-mediated polymerase chain reaction (PCR) as follows: initial denaturation at 94˚C for 1 minute, followed by 30 cycles of denaturation (94˚C for 20 seconds), annealing (58˚C for 30 seconds) and extension (72˚C for 45 seconds) and an additional final extension step at 72˚C for 7 minutes. After PCR amplification, equimolar amounts of amplification product for each sample were pooled before 77 cycles of single-read sequencing on an Illumina HiSeq2500 (https://www.illumina.com/). Since the *PstI*-compatible adapter included a sequencing primer, the sequence tags were always read from the *PstI* sites.

Following sequencing, the raw reads obtained were processed using Illumina CASAVA version 1.8.2 software to generate fast alignment search tool-quality (FASTQ) files (DiBattista *et al.*, 2018). Thereafter, the FASTQ files were processed using a proprietary DArT analytical pipeline (DArTtoolbox), which performed filtering, variant calling and generated final genotypes in sequential primary and secondary workflows (Cruz *et al.*, 2013; Lal *et al.*, 2017). The primary analysis, performed with the *DArTsoft14* algorithm, removed reads with a quality score (Q-score/ Phred score) <25 from further processing and applied stringent selection criteria to the barcode region of all sequences in order to increase confidence in genomic region recovery (Lal *et al.*, 2017). Next, individual samples were de-multiplexed by barcode and trimmed at 69 bp, which included 5 bp of the sequenced *PstI* recognition site. Subsequently, the sequences were compared to catalogued sequences in both DArTdb custom databases and GenBank archives to screen out contaminant sequences, such as viral and bacterial fragments (Lal *et al.*, 2017; Al-Breiki *et al.*, 2018; DiBattista *et al.*, 2018).

The secondary workflow analysis involved the *DArTsoft14* algorithm and the KDCompute framework (http://www.kddart.org/kdcompute.html) developed by DArT, as well as the DArTdb database. Approximately 2 420 000 sequences per sample were used in SNP marker calling during the secondary workflow. Single nucleotide polymorphisms were identified by aligning reads to create clusters across all individuals sequenced (Lal *et al.*, 2017). A threshold distance of three was used for clustering of sequence tags and, therefore, sequences could only be grouped in the same cluster if they did not differ for more than three bases (Lambert *et al.*, 2016). Thereafter, clusters were added to the DArTdb database and matched against each other to generate reduced-representation loci, based on their level of similarity and size (Lal *et al.*, 2017). Reference and SNP allele loci were identified within clusters and assigned arbitrarily, however, in most cases the reference allele refers to the allele

that was most frequent across all samples for that specific locus. To ensure robust variant calling, SNP loci had to be present in both homozygous and heterozygous states, monomorphic clusters were removed and technical sample replicates were performed to ensure scoring consistency. In addition, Mendelian distribution of alleles were tested to separate true allelic variants from paralogous sequences and clusters containing overrepresented or tri-allelic sequences were also excluded (Lal *et al.*, 2017; DiBattista *et al.*, 2018). Finally, once the SNP markers had been confidently identified, each locus was examined using KDCompute to determine several quality and information content parameters for each marker. These included call rate (percentage of valid scores compared to all possible scores for a marker), reproducibility (percentage of technical sample replicates for which the marker score was consistent), read depth (total number of aligned sequence reads that contain the SNP/reference allele), polymorphic information content (PIC) and heterozygosity (Baloch *et al.*, 2017; Melville *et al.*, 2017; Nadeem *et al.*, 2018; Schultz *et al.*, 2018). The SNP dataset was, however, not filtered based on these parameters, because the cut-off threshold required can vary depending on the application. Final genotype scores were exported to Microsoft Excel.

Although not relevant to this particular study, *silico*DArT markers were also identified in addition to the genome-wide SNP loci. The *silico*DArT markers were called based on a minimum reproducibility of 95% (Lal, 2016). Once the *silico*DArT markers were identified, the KDCompute framework was also used to determine the relevant quality and information content parameters (call rate, reproducibility, read depth and PIC), before genotype scores were exported to Microsoft Excel.

2.2.3. In Silico *Mapping of SNP Marker Sequences*

The genome of a domestic Hereford cow (*Bos taurus*), has been sequenced and annotated (Zimin *et al.*, 2009; Elsik *et al.*, 2016). The cattle genome has a total estimated size of 2.87 Gbp and consists of 29 autosomes and the sex chromosomes (Liu *et al.*, 2009). Approximately 91% of the assembled *B. taurus* genome has been anchored onto the chromosomes. The *B. taurus* genome assembly, therefore, provides a high-quality resource for ongoing studies in bovid species and is often used to address important questions in ruminant biology and evolution. In addition, the *B. taurus* annotated genome can also be used to identify orthologous genes in related species.

Consequently, the DArTseq SNP sequences were mapped to the contiguous sequences (contigs) of the *B. taurus* genome assembly, based on established synteny between the *B. taurus* and blue wildebeest chromosomes (Steiner *et al.*, 2014), to estimate the genomic distribution and relative density of the DArTseq SNPs. Genomic regions flanking the DArTseq SNPs were aligned to the *B. taurus* genome sequence assembly [UMD_3.1.1; Bovine

Genome Database (http://bovinegenome.org/)] using blastn with default parameter settings. Shorter alignments inherently have higher E-values and, therefore a less stringent E-value threshold is usually required. Since the SNP marker sequences are only 69 bp in length, an E-value cut-off of E ≤ 1.0E-01 was used for mapping. Furthermore, when a DArTseq SNP marker sequence mapped to an annotated *B. taurus* region, the name of the gene was retrieved as a putative orthologous gene.

In addition, a Pearson correlation analysis was performed to determine whether a possible relationship exists between DArTseq SNP density and the size of a chromosome and/or SNP density and gene density. The size of each *B. taurus* chromosome and the number of genes per chromosome were obtained from the bovine assembly UMD_3.1.1 (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context=genome&acc=GCF_000003055.6). Thereafter, the software RStudio (version 1.1.456, https://www.rstudio.com) was used to calculate the Pearson correlation coefficient (r) and the coefficient of determination (R^2) between SNP density (SNPs/Mbp) and chromosome size (Mbp) and SNP density (SNPs/Mbp) and gene density (number of genes per chromosome/Mbp).

2.2.4. Functional Annotation

The SNP marker sequences were functionally annotated using the Blast2GO pipeline (Blast2GO version 5.1.1; Götz *et al.*, 2008). Blast2GO was used to conduct a blastx search of the DArTseq generated SNP sequences against the non-redundant (nr) National Center for Biotechnology Information (NCBI) sequence database prior to mapping for Gene Ontology (GO) terms. Default blastx parameters were used, however, a mammalian taxonomic filter was applied to make the blastx search more time-efficient. Furthermore, similar as for the *in silico* mapping, blast matches were considered significant with E-values ≤ 1.0E-1. Thereafter, mapping and annotation at GO level 2 were performed with default Blast2GO settings in order to retrieve molecular function, biological process and cellular component terms.

2.3. Results

2.3.1. DArTseq Generated Markers

The DArTseq genotyping platform identified 20 563 SNPs and 15 500 *silico*DArTs in the blue wildebeest genome. Summary statistics and quality parameters for both marker types are presented in Table 2.1. However, as the aim of the study was to develop SNP markers for the blue wildebeest, the remainder of this chapter will only focus on the DArTseq generated SNP markers. As mentioned previously, reference and SNP alleles were assigned arbitrarily

since the blue wildebeest is a non-model species. Generally, the reference allele refers to the allele that was most frequently sequenced across all samples for a specific locus.

Table 2.1. Summary statistics for DArTseq identified genomic markers in blue wildebeest.

	Marker Types	
	SNP	***Silico*DArT**
Number of markers	20 563	15 500
Number of unique sequence fragments	17 021	15 500
Length of fragments (bp)	69	69
Call rate (%)[a] ± SE	90.78 ± 1.03E-03	95.10 ± 3.68E-04
Average reference read depth[b] ± SE	18.31 ± 9.20E-02	19.95 ± 0.14
Average SNP read depth[b] ± SE	12.92 ± 6.46E-02	NA
Reproducibility (%)[c] ± SE	99.67 ± 6.75E-05	99.91 ± 5.08E-05
Average number of missing alleles per individual (%) ± SE	9.21 ± 0.19	4.90 ± 0.53

Abbreviations: Base pairs (bp); not applicable (NA); single nucleotide polymorphism (SNP); standard error (SE).

[a] Percentage of valid scores compared to all possible scores for a marker (Nadeem *et al.*, 2018).

[b] Total number of aligned sequence reads that contain the SNP/reference allele (Schultz *et al*., 2018).

[c] The proportion of technical sample replicates for which the marker score was consistent (Baloch *et al.*, 2017; Melville *et al.*, 2017).

2.3.2. Assessment of Genotyping Errors

2.3.2.1. Read Position of Generated SNPs

Previous studies have found that sequencing errors are more abundant towards the end of next-generation sequencing reads and are thought to result in the identification of more false SNPs (Dohm *et al.*, 2008; Kraus *et al.*, 2011; Smitz *et al.*, 2016). To validate whether the DArTseq identified SNPs are indeed true SNPs, the distribution of the SNPs was plotted over the 69 base pair positions of the sequence reads (Figure 2.1). No SNPs were identified in the first four base pair positions or at the last read position. Furthermore, a decrease in SNP frequency was observed towards the end of sequence reads.

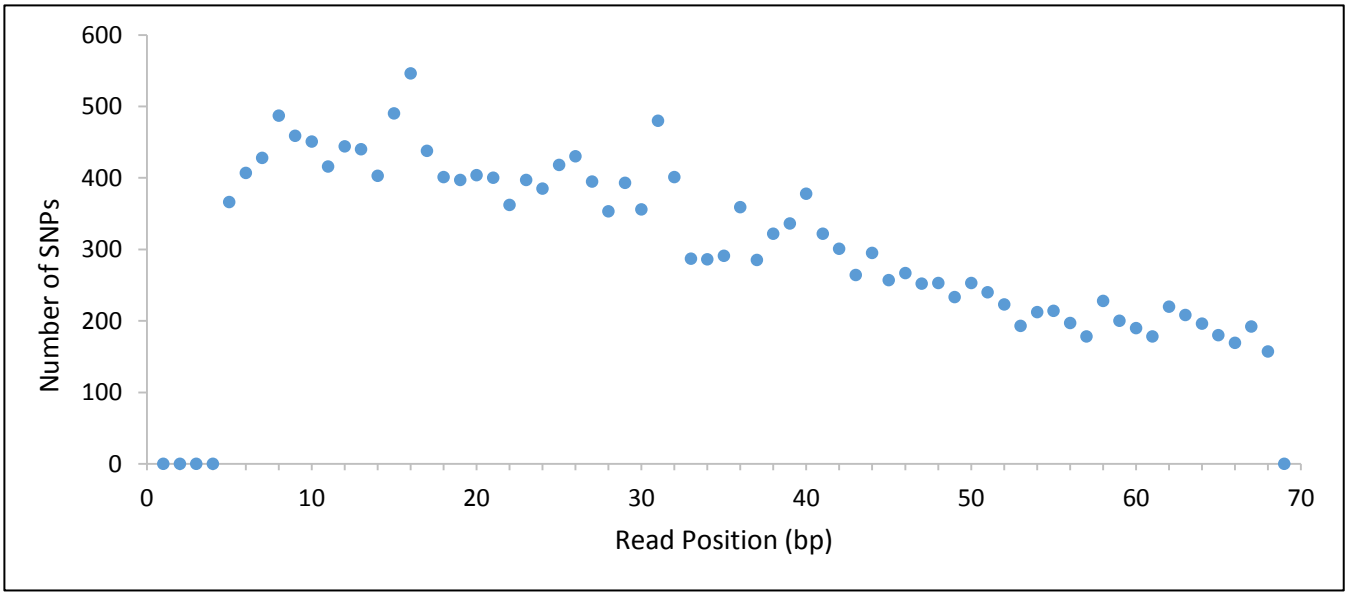


Figure 2.1. Total number of identified single nucleotide polymorphisms (SNPs) at each sequence read position.

2.3.2.2. Transition vs *Transversion Ratios*

The ratio of transitions (Ts; *i.e.* A/G or T/C substitutions) *versus* transversions (Tv; *i.e.* A/T, A/C, T/G or C/G substitutions) observed in the SNP dataset was 2.84:1 (15 204 transitions *versus* 5 359 transversions), which is higher than the expected empirical ratio of 2:1 (Kraus *et al.*, 2011; Smitz *et al.*, 2016). The number of A/G and T/C polymorphisms occurred at similar frequencies (7 679 A/G and 7 525 T/C). Although the four Tv nucleotide substitutions occurred at comparable frequencies, the number of A/T substitutions was slightly lower than the other Tv substitution types (Figure 2.2). In addition, the Ts:Tv ratio mainly ranged between 1.6 and 2.4 when plotted per read position, with the exception of a few outliers (Figure 2.3).

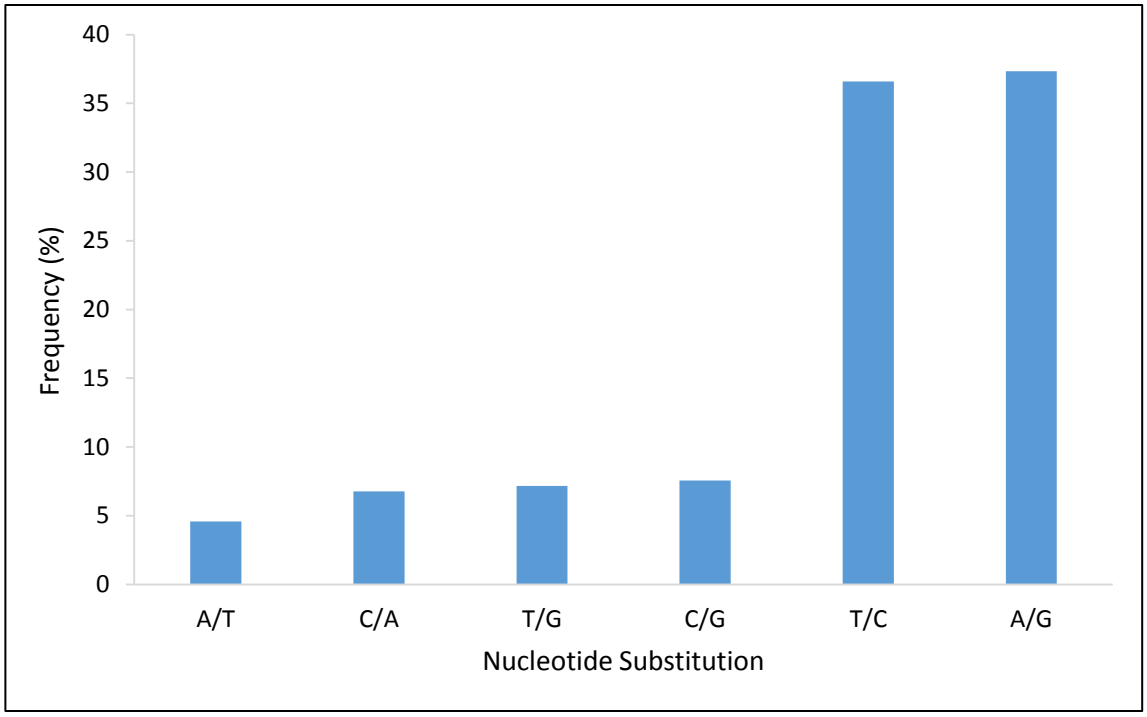


Figure 2.2. Distribution of the six types of nucleotide substitutions.

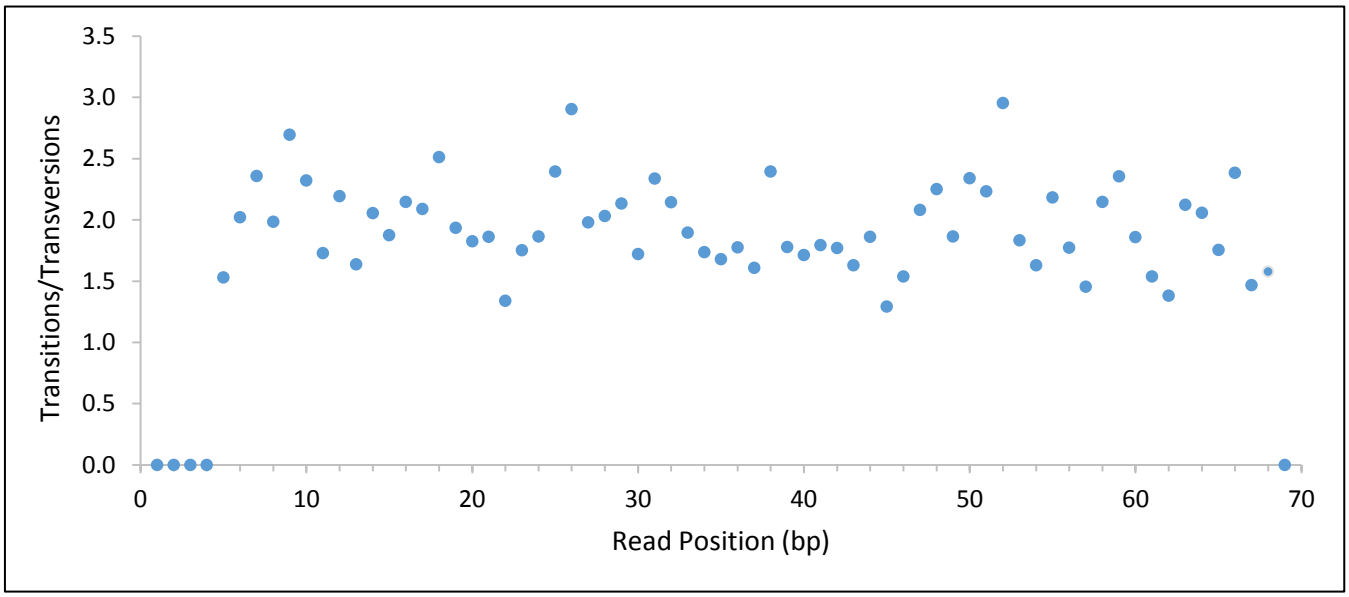


Figure 2.3. Transition/transversion ratio at each read position.

2.3.3. Information Content of the SNP Markers

The allelic frequencies and PIC were used as measures for the informativeness of the generated SNP markers. The distribution of the minor allele frequencies (MAF) indicated that the SNPs with MAF < 0.20 were overrepresented at more than half of the SNPs (60%). Furthermore, SNPs were nearly evenly distributed in high MAF bins that ranged from 0.25 to 0.50 (31%; Figure 2.4). The average MAF across the 20 563 generated SNPs was 0.19. In addition, the observed frequency of heterozygotes ranged from 0.00 to 0.70 (Figure 2.5). The average heterozygote frequency across all the SNP markers was 0.19. The generated SNPs had PIC values that ranged from 0.02 to 0.50, with an average of 0.24. The distribution of the SNP PIC values are presented in Figure 2.6.

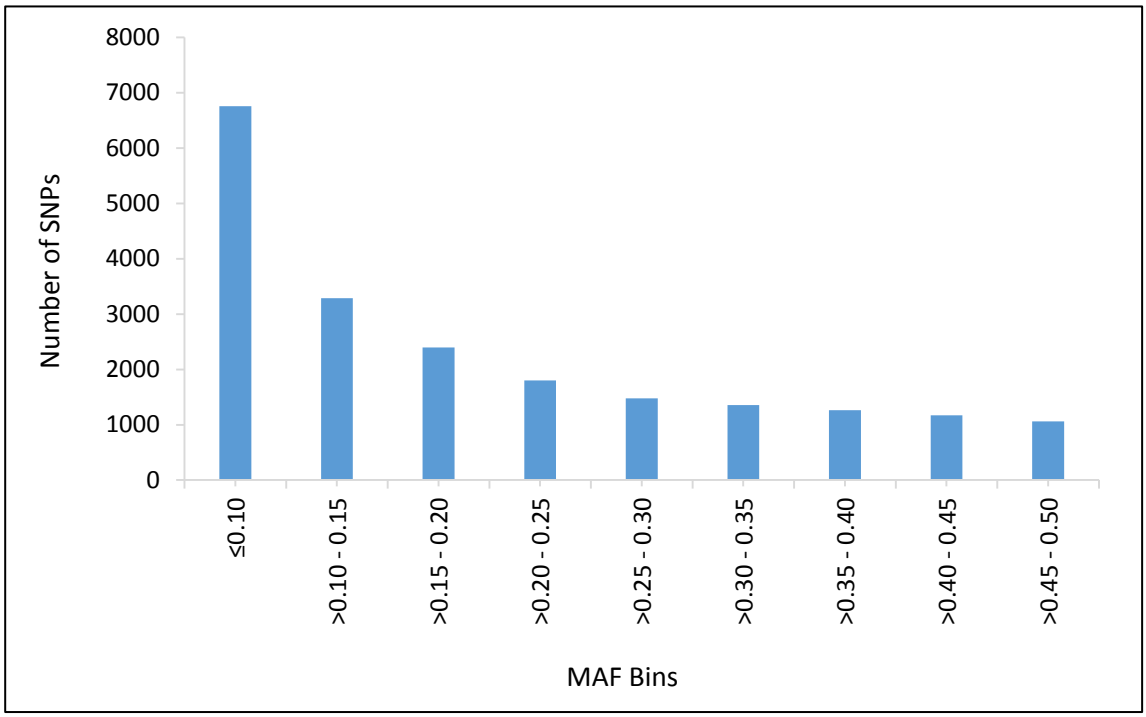


Figure 2.4. Distribution of the DArTseq generated single nucleotide polymorphisms (SNPs) across the minor allele frequency (MAF) bins.

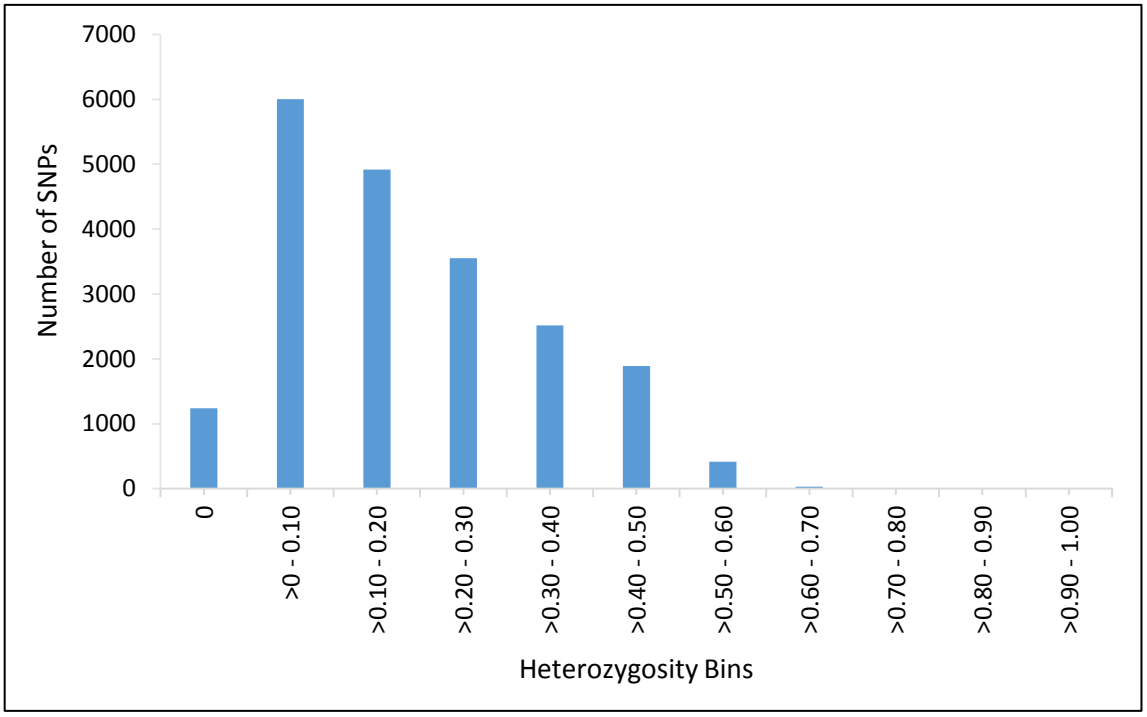


Figure 2.5. Frequency of the DArTseq generated single nucleotide polymorphisms (SNPs) across the heterozygosity bins.

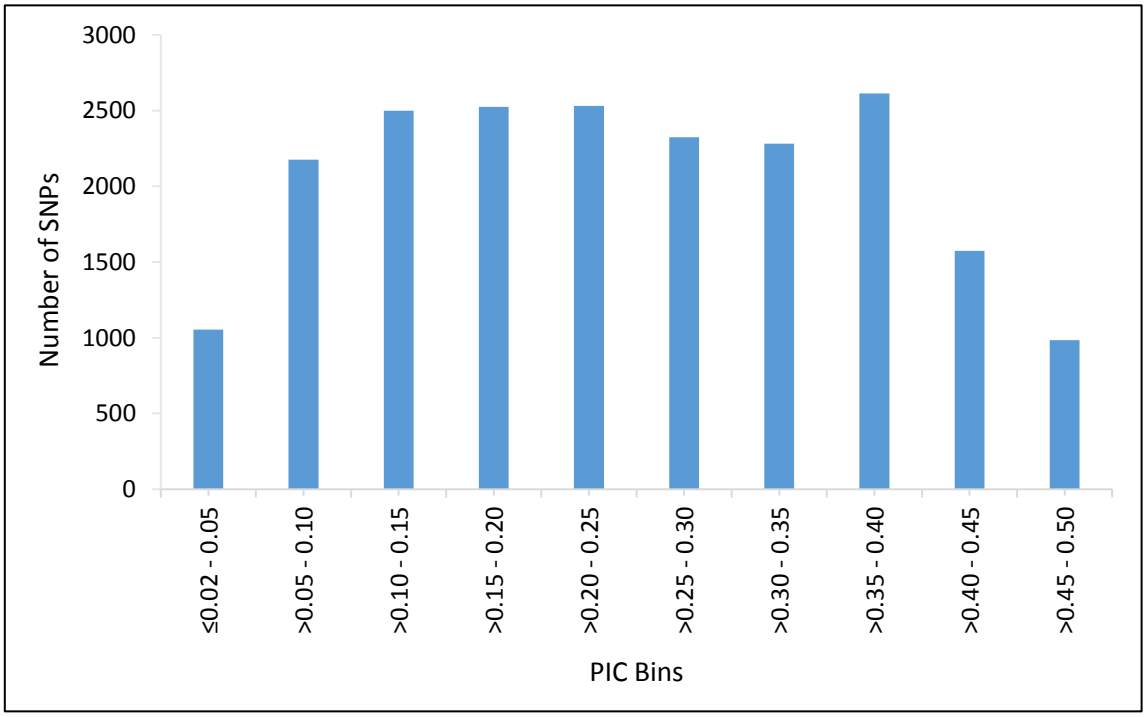


Figure 2.6. Distribution of the DArTseq generated single nucleotide polymorphisms (SNPs) across the polymorphic information content (PIC) bins.

2.3.4. In Silico *Mapping of SNP Marker Sequences*

Alignment of the DArTseq SNP sequences to the bovine genome assembly produced 6 020 significant hits. Of these DArTseq SNPs, 6 018 could be mapped to one of the 30 *B. taurus* chromosomes, while only two SNPs were assigned to unmapped contigs. The distribution of SNPs was not even across the chromosomes (Figure 2.7). The highest number of SNPs (899) mapped to bovine chromosome 5, whereas the lowest number of SNPs (20) mapped to chromosome 24. No significant correlation was observed between either the density of the DArTseq SNPs and the size of the *B. taurus* chromosomes (Figure 2.8; r = 0.207, R^2 = 0.043) or between SNP density and the gene density of the chromosome (Figure 2.9; r = 0.097, R^2 = 0.010).

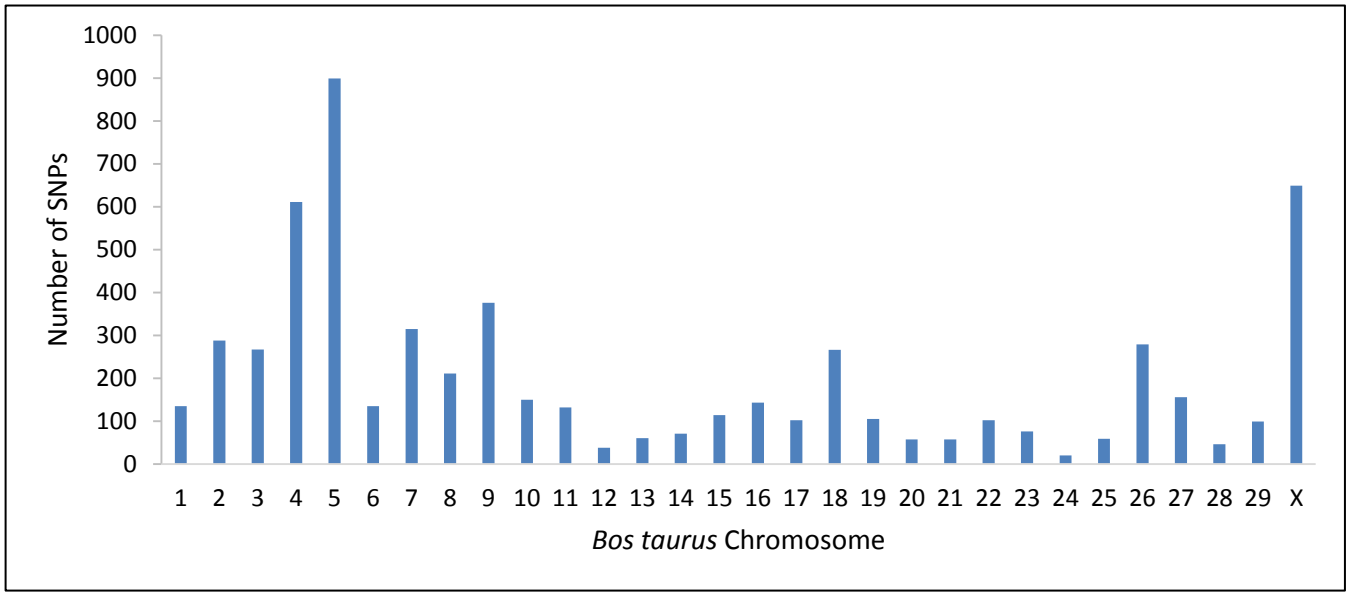


Figure 2.7. Distribution of the blue wildebeest DArTseq single nucleotide polymorphisms (SNPs) across the *Bos taurus* chromosomes.

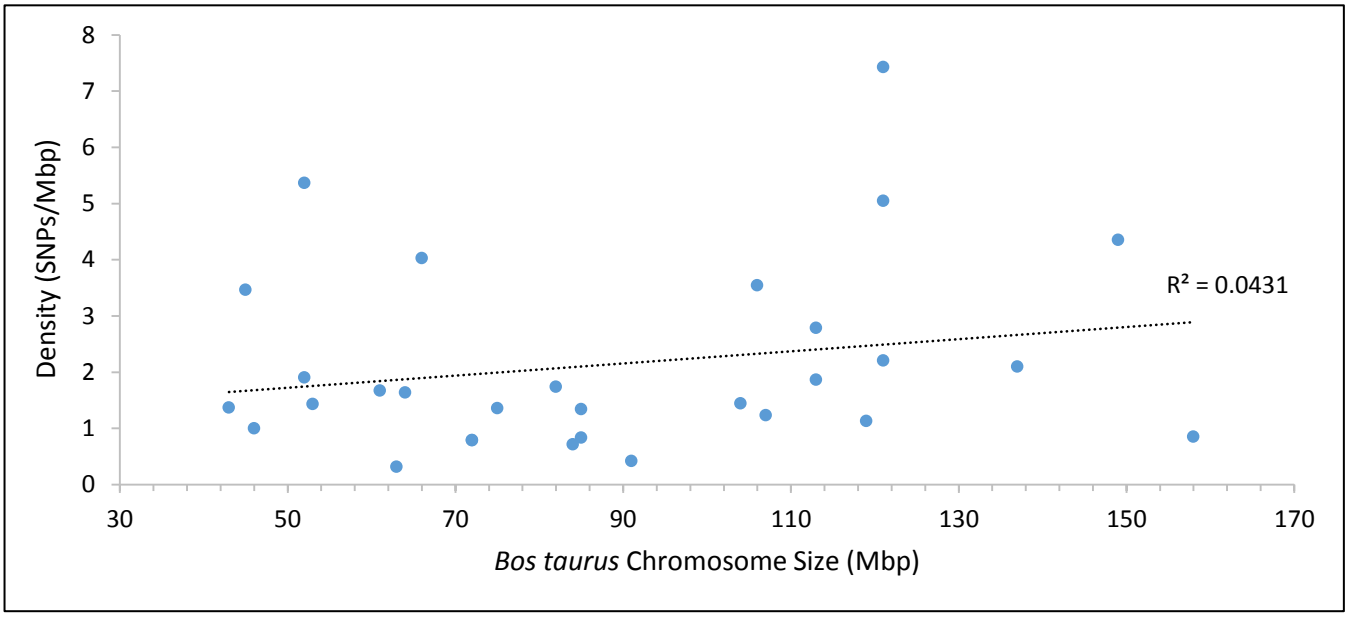


Figure 2.8. Average single nucleotide polymorphism (SNP) density per *Bos taurus* chromosome related to its size. The coefficient of determination is denoted as R^2. *Bos taurus* chromosome sizes were obtained from the bovine assembly UMD_3.1.1 (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context=genome&acc=GCF_000003055.6).

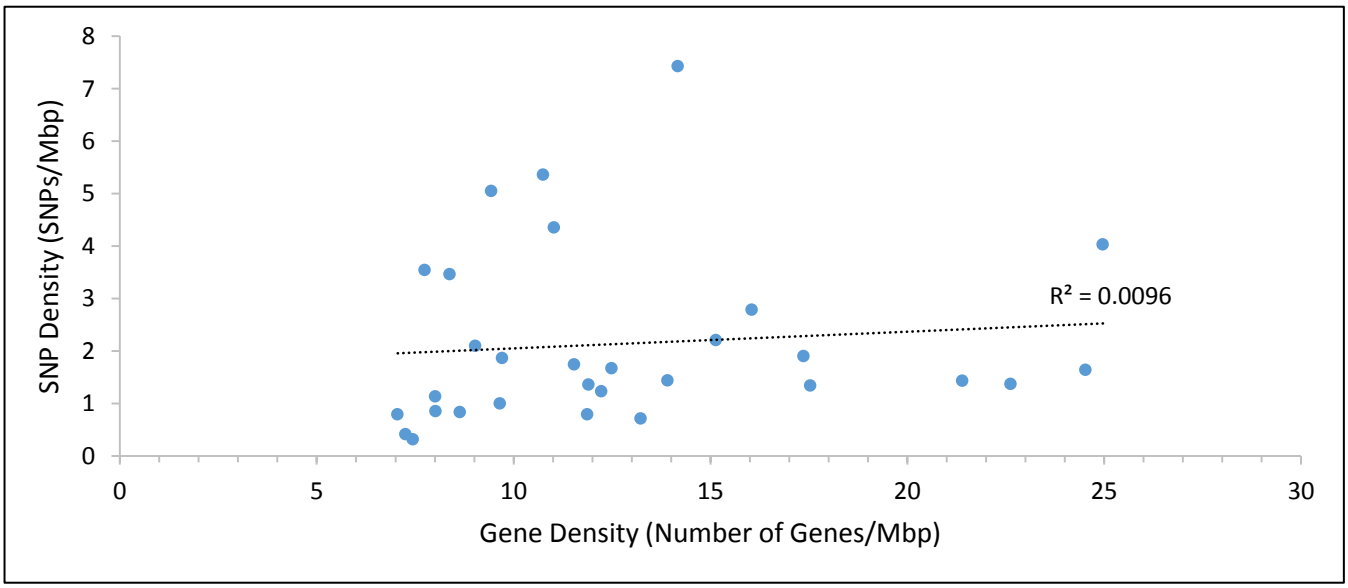


Figure 2.9. Average single nucleotide polymorphism (SNP) density compared to the average gene density of the *Bos taurus* chromosomes. The coefficient of determination is denoted as R^2. *Bos taurus* chromosome sizes and number of genes per chromosome were obtained from the bovine assembly UMD_3.1.1 (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context=genome&acc=GCF_000003055.6).

Among the 6 020 SNP sequences that could be mapped to the *B. taurus* genome, 3 907 (64.90%) had putative gene orthologues identified by similarity searches against the Bovine Genome Database. The remaining 2 113 SNP sequences (35.10%) aligned to *B. taurus* genome sequences that code for uncharacterised proteins, cattle genes or regulatory elements that have not been annotated yet or non-functional DNA sequences.

2.3.5. Functional Annotation

The blastx search performed using Blast2GO obtained significant hits for a total of 3 689 (17.94%) of the DArTseq SNP marker sequences. Figure 2.10 illustrates a listing of the species to which most of the sequences aligned during the blast search. The species most represented is *Bos taurus,* with nearly 10 000 significant blast hits, followed by *Ovis aries musimon* (European mouflon), *Bos mutus* (domestic yak), *Ovis aries* (domestic sheep) and *Bubalus bubalis* (water buffalo).

One or more GO terms were attributed to 1 733 of the sequences. “Binding” (807 sequences, 34%), “transporter activity” (614 sequences, 26%), and “catalytic activity” (542 sequences, 23%) were among the most abundant molecular function subcategories. In the biological processes category, several subcategories were represented, among which “cellular process” (971 sequences, 19%) and “biological regulation” (908 sequences, 18%) were most common, followed by “metabolic process” (875 sequences, 17%). Furthermore, the SNP sequences were assigned to a wide range of cellular component terms of which “membrane” (1 162 sequences, 21%), “membrane part” (1 074 sequences, 20%) and ‘cell part’ (868 sequences, 16%) were the most represented. In addition to these marker sequence that were assigned GO terms, 366 sequences were reported as predicted proteins, 266 sequences were reported as hypothetical proteins, 63 sequences were mapped but not annotated, 29 were assigned to uncharacterised proteins, and five were reported as unnamed protein products.

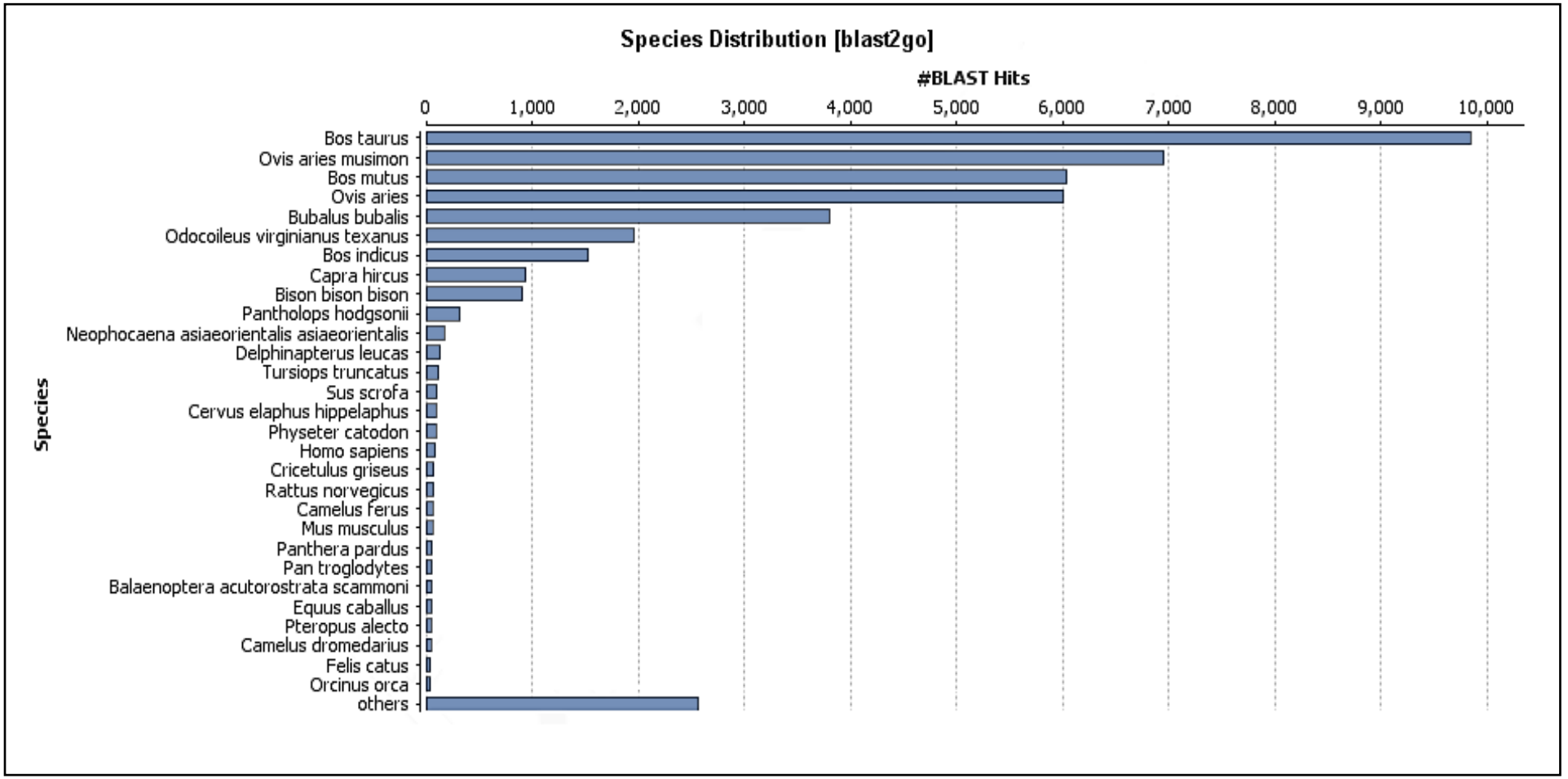


Figure 2.10. Species distribution for all blast hits obtained by performing a Blast2GO blastx search with the DArTseq single nucleotide polymorphism (SNP) marker sequences.

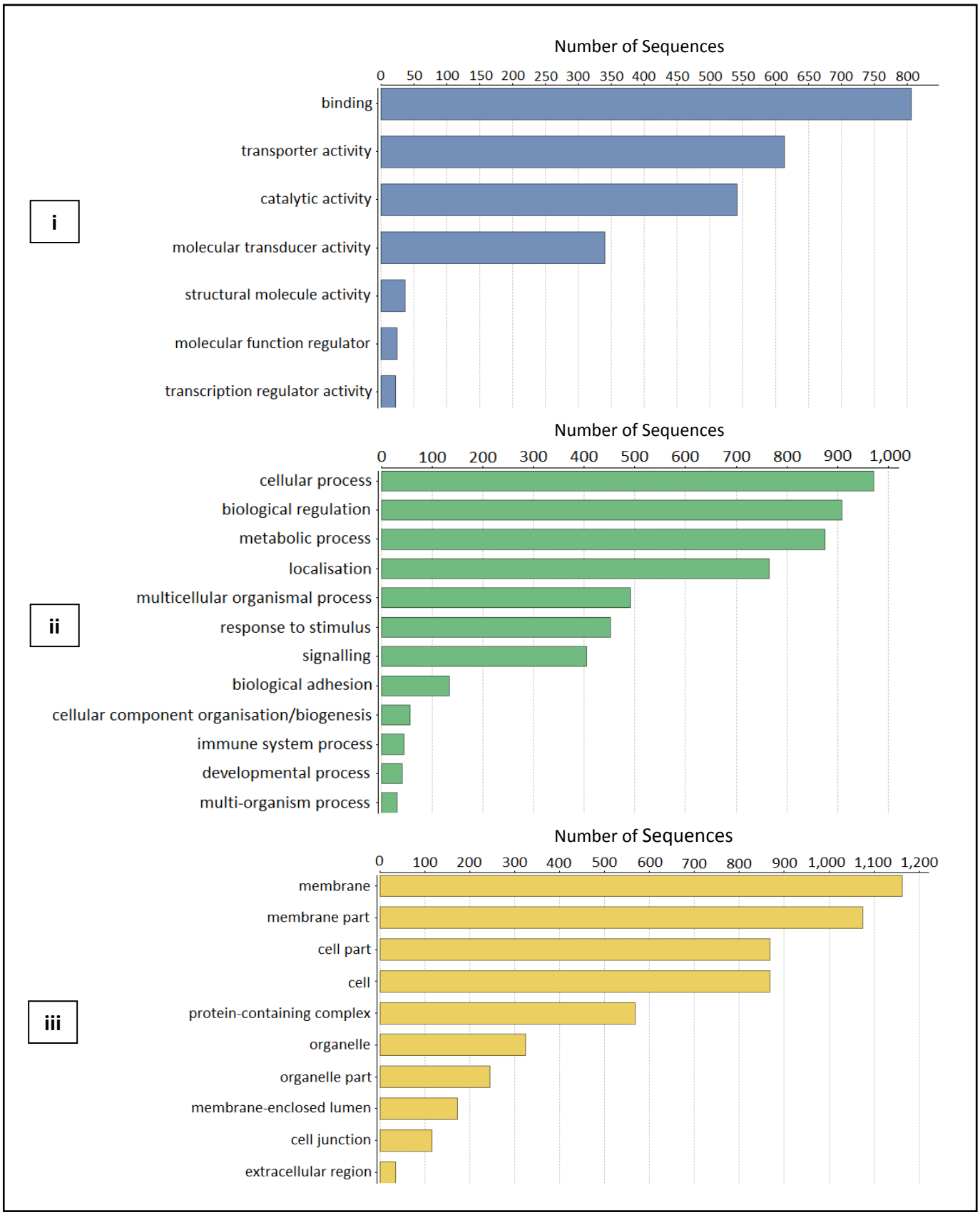


Figure 2.11. Gene ontology distribution of single nucleotide polymorphisms (SNPs) in blue wildebeest. The results are summarised as follows: (i) molecular functions, (ii) biological processes, (iii) cellular components.

2.4. Discussion

Single nucleotide polymorphisms (SNPs) are valuable tools for genetic analyses due to their high abundance and genome-wide distribution. The DArTseq genotyping platform provides an efficient and cost-effective approach for SNP discovery, enabling the identification of thousands of SNPs, even in the absence of an available reference genome (Melville *et al.*, 2017). Using the DArTseq approach, the present study enabled the identification of 20 563 SNP markers for blue wildebeest. This SNP study is the first large genetic polymorphism discovery performed in blue wildebeest, and is also the first study to identify SNPs, using a high-throughput platform, for an African antelope species that is farmed commercially.

2.4.1. General Overview of DArTseq Application in Blue Wildebeest

The DArTseq genotyping approach employs a combination of genome complexity reduction methods, initially optimised on the DArT microarray platform, and NGS (Melville *et al.*, 2017). This more recent application of the DArT complexity reduction method on a NGS platform is conceptually similar to double digest restriction associated DNA (ddRAD) sequencing, which is perhaps a more widely known implementation of GBS (Peterson *et al.*, 2012; Couch *et al.*, 2016). Nevertheless, the DArTseq platform was also able to discover and genotype a large number of SNPs (20 563) in blue wildebeest, as in the case of other studies using the ddRAD platform for SNP discovery in even-toed ungulates (Table 2.2). In addition, the number of DArTseq SNPs identified in blue wildebeest is also higher than the number of SNPs identified in many other studies making use of alternative reduced representation sequencing (RRS) approaches in even-toed ungulates, such as hog deer (*Axis porcinus*) and African buffalo (*Synerus caffer*; Table 2.2). It is, however, important to note that the number of SNPs identified with the DArTseq platform in blue wildebeest and other animals cannot be directly compared to the number of SNPs identified by other RRS platforms. This is due to the fact that SNP discovery can be influenced by various factors, such as genome size, SNP density in the genome, the size of the sample population and the choice of REs (Shirasawa *et al.,* 2016).

Although the DArTseq platform enables the discovery and genotyping of a large numbers of SNPs, the marker sequences that are generated are relatively short. While ddRAD and other RRS platforms generally yield SNP marker sequences that are 100 bp in length or longer (*e.g.*, Kess *et al.*, 2016; Yang *et al.*, 2016; Ba *et al.*, 2017; Souza *et al.*, 2017), DArTseq's stringent filtering criteria mostly produce SNP sequences approximately 69 bp in length (Melville *et al.*, 2017). The length of the DArTseq marker sequences can, however, range between 20 bp and 69 bp depending on the quality of the generated sequences (Schultz *et al.*, 2018). All the SNP marker

sequences generated for the blue wildebeest were 69 bp in length. Although these final DArTseq SNP marker sequences are relatively short, previous studies employing DArTseq have demonstrated that, depending on the read position of the SNP, they still provide sufficient sequence context to allow for the subsequent development of SNP genotyping assays (*e.g.* Sieber *et al.*, 2016; Barilli *et al.*, 2018; Okoń *et al.*, 2018).

Table 2.2. A comparison between the number of single nucleotide polymorphisms (SNPs) identified in the current study and the number identified in other reduced representation sequencing (RRS) studies.

Specie	Approach	Enzymes used	Number of SNPs Identified[a]	Fragment Length	Reference
Blue wildebeest (*Connochaetes taurinus taurinus*)	DArTseq	*PstI*/*SphI*	20 563	69 bp	Current study
Ring-tailed dragon (*Ctenophorus caudicinctus*)	DArTseq	*PstI*/*HpaII*	28 960	~69 bp	Melville *et al.*, 2017
Nile tilapia (*Oreochromis niloticus*)	DArTseq	*PstI*/*HpaII*	13 215	20-69 bp	Lind *et al.*, 2017
Sika deer (*Cervus nippon*)	ddRAD	*PstI*/*MseI*	96 188	~100 bp	Ba *et al.*, 2017
Addax (*Addax nasomaculatus*)	ddRAD	*SpeI*/*Sau3AI*	15 240	~300 bp	Ivy *et al.*, 2016
Indigenous Chinese cattle (*Bos taurus*)	RADseq	*EcoRI*	238 725	150 bp	Wang *et al.*, 2018
Hog deer (*Axis porcinus*)	RADseq	*EcoRI*	11 155	125 bp	Wang *et al.*, 2017
African buffalo (*Synerus caffer*)	RRL	*AluI*	2 534	~76 bp	Smitz *et al.*, 2016
White-tailed deer (*Odocoileus virginianus*)	RRL	*AluI*	10 476	~338 bp	Seabury *et al.*, 2011

[a]Number of SNPs retained after quality filtering.

Abbreviations: Base pairs (bp); Diversity Arrays Technology sequencing (DArTseq); double digest restriction associated DNA sequencing (ddRAD); restriction-site associated DNA sequencing (RADseq); reduced representation library (RRL).

The developed blue wildebeest SNP markers showed high reproducibility, with an average scoring reproducibility of >99% and a minimum reproducibility of 93.06%. A high average scoring reproducibility (>99%) has also generally been observed in previous studies making use of DArTseq genotyping (Melville *et al.*, 2017). Conversely, low reproducibility have often been reported by other restriction digest derived methods such as restriction-site associated DNA sequencing (RADseq; Davey *et al.*, 2013; Scheben *et al.*, 2017; Souza *et al.*, 2017). Therefore, the high reproducibility of the DArTseq approach can be seen as an evident advantage of this platform (Kilian *et al.*,

2016; Lind *et al.*, 2017; Melville *et al.*, 2017). In addition, the DArTseq platform results in considerably less missing data (20% and lower; Chen *et al.*, 2016; Lind *et al.*, 2017) compared to other reduced representation approaches. The percentage missing data in the blue wildebeest DArTseq dataset (~9.21%) was markedly less than that observed in some studies making use of the conceptually similar ddRAD approach. For example, the percentage missing data ranged between 17% and 22% for different butterfly specimen groups (*Speyeria* genus) in a ddRAD study by Campbell *et al.* (2017), while the percentage missing data for two different species of spiny rock lobster (*Jasus edwardsii* and *Sagmariasus verreauxi*) was 24% and 31%, respectively, in a study by Souza *et al.* (2017). Moreover, large amounts of missing data is especially a common problem encountered when using the RADseq platform (Moore & Benestan, 2018).

In the blue wildebeest DArTseq SNP dataset, the reference alleles had an average read depth greater than 18, while the SNP alleles had an average read depth greater than 12. The read depth obtained in this study is comparable, and in some cases, exceeds the read depth obtained in other similar studies. For example, a study by Smitz *et al.* (2016), making use of a reduced representation library (RRL) to develop SNPs in the African buffalo, had an average read depth of 5× for the SNP allele and 13× for the reference allele. In addition, a similar average depth range for SNP discovery using a RRS approach was also reported in wild mallard (*Anas platyrhynchos*; ~16×; Kraus *et al.*, 2011), barnacle goose (*Branta leucopsis*; 9.9×; Jonker *et al.*, 2012), domestic cattle (7.4×; Eck *et al.*, 2009) and sika deer (~23×; Ba *et al.*, 2017). The read depth obtained in this study is thus an acceptable depth range for SNP discovery compared to other reduced representation approaches.

2.4.2. Assessment of Genotyping Errors

In order to remove low quality reads, the DArTseq sequence reads were filtered based on their Q-scores, which are per-base quality scores that predict the probability of an error in base calling (Lal *et al.*, 2017). However, these Q-scores can sometimes be unreliable, since it is calibrated with instrument-dependent variables and can, therefore, co-vary with features such as the chemistry or software of the sequencing platform, machine cycle or sequence context (DePristo *et al*., 2011; Manley *et al*., 2016). Consequently, the reliability, and thus also the quality, of the DArTseq generated blue wildebeest SNPs were assessed based on the data characteristics of next-generation sequencing technology. The different estimators used to evaluate the risk of false positives in the dataset included the read position of the generated SNPs and the ratio of the substitution types.

2.4.2.1. Read Position of the Generated SNPs

More sequencing errors are often found at the tail-end of reads when using Illumina sequencing technology (Dohm *et al.*, 2008; Smitz *et al.*, 2016). Therefore, if a substantial number of SNPs in the dataset were the result of sequencing error, an increase in the number of SNPs toward the 3' end of sequencing reads is expected (Ba *et al.*, 2017). However, the results for the DArTseq dataset actually showed that the number of SNPs identified decreased from the 5' to the 3' end of sequencing reads. This observation can most likely be attributed to the filtering and trimming of sequencing reads, based on quality scores, during the first DArTseq analytical pipeline. Therefore, the identified SNPs are likely to be true polymorphisms and not the result of sequencing errors. Decreasing numbers of candidate SNPs per nucleotide position toward the 3' end of sequence reads have also been reported in other genome complexity reduction studies making use of single-end sequencing (Kerstens *et al.*, 2009; Van Bers *et al.*, 2010; Smitz *et al.*, 2016).

2.4.2.2. Transition vs *Transversion Ratios*

The Ts:Tv ratio is commonly computed in high-throughput sequencing studies as a quality measurement for variant calling (*e.g.* Liu *et al*., 2012; Wang *et al*., 2014; Smitz *et al*., 2016; Ba *et al.*, 2017). Thymine (T) and cytosine (C) are one-ring pyrimidine-based nucleotides, while guanine (G) and adenine (A) are two-ring purine-derived nucleotides. Transitions refer to the interchange of pyrimidine-based nucleotides (C/T) or purine-based nucleotides (A/G), while transversions are defined as the interchange between one-ring pyrimidine nucleobases and two-ring purine bases (A/T, A/C, T/G or C/G; Wang *et al*., 2015a). A Ts:Tv ratio of 0.5 would be expected if transition and transversion mutations occurred at random (*i.e.* without any biological influence), since there are two possible transition mutations and four possible transversions (Kraus *et al*., 2011; Wang *et al*., 2015a). However, in actuality, transversions occur less than transitions, because the interchange of a two-ring and a one-ring chemical structure requires more energy than nucleotide changes that do not involve a change in ring structure (Wang *et al*., 2015a). In addition, mutational mechanisms, such as oxidative deamination, favour transitions over transversions (Kraus *et al*., 2011). Therefore, the observed Ts:Tv ratio is often greater than 0.5. Generally, a ratio of 2.1:1 is observed in mammals but, the observed Ts:Tv ratio can vary depending on the genome region and functionality (DePristo *et al.*, 2011; Smitz *et al.*, 2016). The observed Ts:Tv ratio for SNPs tends to be the highest in coding regions due to the hyper-mutability of CpG islands (Jiang *et al.*, 2008; Hendre *et al.*, 2012; Wang *et al.*, 2015a; Ba *et al.*, 2017). For example, coding regions in the human genome typically have Ts:Tv ratios

of approximately 3.5:1 (Le & Durbin, 2011). Conversely, false sequence variants that arise due to sequencing error often have a Ts:Tv ratio closer to the expected ratio of 0.5 (Kraus *et al.*, 2011; Smitz *et al.*, 2016).

An average Ts:Tv ratio of 2.84:1 was observed in the blue wildebeest DArTseq SNP dataset. This ratio is similar to the higher Ts:Tv ratios observed in other studies making use of RRS for SNP discovery [*e.g.* Jonker *et al.*, 2012 (2.7:1) and Smitz *et al.*, 2016 (2.4:1)]. The high Ts:Tv ratio observed in the dataset is, therefore, not unexpected and indicates that most SNP calls were not random. The generated SNPs most likely represent true nucleotide polymorphisms. The high observed Ts:Tv ratio could also be an indication that the DArTseq approach targeted coding regions within the blue wildebeest genome. This preferential sampling of gene-rich regions is a well-known attribute of the DArTseq platform (Courtois *et al.*, 2013; Sánchez-Sevilla *et al.*, 2015; Garavito *et al.*, 2016; Gawroński *et al.*, 2016). In addition, the Ts:Tv ratio remained similar for each read position which is also a good indication that read position bias, such as the identification of false SNPs due to low sequencing quality towards the tail-end of reads, did not occur (Jonker *et al.*, 2012). Furthermore, although the number of A/T substitutions was slightly less than the other substitution types, it could be attributed to the fact that read coverage of sequenced regions are biased towards regions with elevated GC content (Dohm *et al.*, 2008).

2.4.3. Information Content of the SNP Markers

Population genetic studies are often required in order to formulate effective strategies for conservation and population management. Individual identification, parentage and relatedness determination, and population structure or hybridisation are some of the major applications in population genetic studies for which SNP loci has been selected (*e.g.* Ba *et al*., 2017; Von Thaden e*t al*., 2017; Blåhed *et al*., 2018; Fan *et al*., 2019). For such applications, it is advantageous for a SNP set to exhibit high heterozygosity and resolving power (*i.e.* high MAF) among population samples (Yousefi *et al*., 2018).

As expected from NGS technologies (Siu *et al.*, 2011; Lo *et al*., 2018), a wide variation in SNP MAF and heterozygosity was detected by the DArTseq genotyping platform in blue wildebeest. The large number of DArTseq SNPs included both rare (MAF < 10%) and common variants (MAF > 40%), and showed slightly higher levels of genomic variation compared to the small existing set of 23 SNPs identified in blue and black wildebeest (Van Wyk *et al*., 2019). The MAF of the DArTseq SNP set ranged between 0.01 and 0.50, while the MAF of the small SNP set ranged between 0.08 and 0.40. In addition, the distribution of MAF values for the DArTseq SNPs is characteristic for SNPs detected with NGS, with a large percentage of the SNPs representing rare variants (Gurgul e*t al*., 2019). Furthermore, the frequency of heterozygotes ranged between 0.00 and 0.70 for the DArTseq SNPs,

whereas the frequency of heterozygotes varied between 0.00 and 0.40 in the small SNP set. It is, however, important to note that the level of polymorphism of the small SNPs set identified by Van Wyk *et al*. (2019) was assessed in a population of blue and black wildebeest consisting of only 30 individuals. This SNP set might thus show greater variability when assessed in a larger population of individuals.

Furthermore, the average SNP MAF obtained in the current study (0.19) is comparable to the average MAF obtained in other similar GBS studies, such as 0.20 in sika deer (Ba *et al.*, 2017), 0.21 in African forest elephant (*Loxodonta cyclotis*; Bourgeois *et al*., 2017), and 0.21 in cattle (De Donato *et al*., 2013). However, the average MAF obtained in blue wildebeest is somewhat lower than the MAF generally observed in array-based platforms. For example, the average MAF ranged between 0.24 and 0.30 in a study by Matukumalli *et al.* (2009), that tested the Illumina BovineSNP50 microarray assay using a panel of 21 cattle breeds. Additionally, a study by Li & Kim (2015) in Korean native cattle (Hanwoo) found a MAF of 0.27 using the same BovineSNP50 microarray. A commercial buffalo SNP chip array (Affymetrix Axiom® Buffalo Genotyping Array 90K) has also recently become available and obtained a MAF of 0.30 in Egyptian water buffalo (*Bubalus bubalis*; El-Halawany *et al*., 2017). However, although array-based platforms generally result in higher MAFs than GBS platforms, the SNPs included in arrays are preselected and, thus, may not be geographically representative. As a result, these SNPs often have higher frequencies than SNPs that have been randomly sampled. Therefore, population genetic parameters derived from these SNPs (*e.g.* diversity, recombination estimates, population subdivision) may be biased. Conversely, SNPs are identified as they are sequenced in GBS platforms, such as DArTseq and thus they can be used in any population without ascertainment bias (De Donato *et al.*, 2013; Gurgul *et al.*, 2019).

The PIC index, first described by Botstein *et al.* (1980), is also frequently used in genetics as a measure for the level of polymorphism, and thus also the usefulness, of a marker (Yi *et al.*, 2013). According to the index, markers with a higher PIC value will be more informative. The PIC index is based on the number of detectable alleles and their frequencies, and is calculated as the sum of the square of all allele frequencies subtracted from one. An increase in the number of alleles will thus result in an increase in PIC. In addition, for a given number of alleles, the PIC value will also be greater the more equal the allele frequencies are (Liu & Cordes, 2004). Since SNPs are bi-allelic, they can have a maximum PIC value of 0.5 (Singh *et al.*, 2013). As a standard feature, PIC values are calculated for all DArTseq generated SNP markers and reported along with the genotyping results. The DArTseq generated blue wildebeest SNPs had an average PIC value of 0.24. This average PIC value is slightly higher than the average PIC value reported by other DArTseq studies (*e.g.* 0.18 and 0.19, Melville *et al.*, 2017; 0.22, Lind *et al.*, 2017). Furthermore, 1 337 of the blue wildebeest DArTseq SNPs had a highly informative PIC value ranging between 0.45 and 0.50.

The DArTseq platform was thus able to generate a large number of informative SNPs for blue wildebeest. A subset of these genome-wide, population-informative SNPs can be used to develop custom assays in order to allow genotyping in local laboratories that do not have access to sequencing technologies. The development of such panels will be able to greatly facilitate the genetic management of wild and ranched populations of blue wildebeest, and potential useful applications include analysis of population structure, parentage analysis and individual identification. However, once these assays have been developed, the extent of genetic variability at these new SNPs will have to be further investigated in other blue wildebeest populations (Bourgeois *et al*., 2017).

2.4.4. In Silico *Mapping of SNP Marker Sequences*

It is important to know the genomic position of SNPs as genetic markers in order to ensure that SNPs forming part of any marker panel are distributed throughout the whole genome (Kraus *et al.*, 2011). Since no reference genome was available for the blue wildebeest, the genome sequence of *Bos taurus*, the closest available relative with a well-annotated genome (MacEachern *et al.*, 2009), was used to estimate the physical distribution of the identified SNPs. The *in silico* mapping of the DArTseq marker sequences to the bovine genome was also used as an opportunity to identify orthologous genes in the blue wildebeest genome. The efficacy of this approach has been demonstrated in the past. For example, the chicken (*Gallus gallus domesticus*) reference genome has been used to map sequencing reads of turkey (*Meleagris gallopavo*; Kerstens *et al.*, 2009), and sequence reads of the great tit (*Parus major*) were mapped to the reference genome of the zebra finch (*Taeniopygia guttata*) in order to identify novel SNPs (Van Bers *et al.*, 2010). In addition, short sequence reads of the African Buffalo have also been mapped to the cattle genome to identify SNPs (Smitz e*t al*., 2016).

A total of 6 020 (29.28%) of the blue wildebeest sequences were successfully mapped against the bovine genome. Of these, 3 907 SNP marker sequences mapped to putative orthologous genes with known functions in cattle. Although the percentage of mapped reads is low, this is similar to those reported by Kerstens *et al.* (2009), where approximately 30% of the turkey sequence reads mapped to the chicken reference genome. In the study by Van Bers *et al.* (2010), the great tit sequence reads from two pools mapped successfully to the zebra finch genome for 26% and 32% of the reads, respectively, while in the study by Smitz *et al*. (2016) approximately 30% of the sequence reads mapped to the bovine genome. Increasing the length of the SNP marker sequences, now available from some Illumina sequencing platforms, might improve the percentage mapping in future studies.

The DArTseq SNP marker sequences could be mapped to all of the *B. taurus* chromosomes, which confirmed the genome-wide distribution of the SNPs. However, the SNPs were not evenly distributed among all of the *B. taurus* chromosomes. No significant correlation was observed between the density of SNPs on a chromosome and the

size of the chromosome, indicating that larger chromosomes did not necessarily result in the identification of more DArTseq SNPs. In addition, no significant correlation was observed between SNP density and gene density on a chromosome. A GBS study by De Donato *et al*. (2013), that utilised *PstI* for genome complexity reduction in bovine, reported a high correlation between SNP density and gene density. This could be related to the fact that *PstI* is a well-known methylation-sensitive enzyme, which is biased towards the incorporation of under-methylated, gene-rich regions of the genome (Pootakham *et al*., 2016). Since *PstI* was also used for genome complexity reduction in the current study, a similar correlation between DArTseq SNP density and gene density was expected. Nevertheless, the majority of the DArTseq SNPs that could be mapped onto the bovine genome mapped to chromosome 5, which is the third most gene-rich bovine chromosome (contains 1 715 genes; https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context=genome&acc=GCF_000003055.6). Conversely, the lowest number of blue wildebeest DArTseq SNPs mapped to chromosome 24, which is the bovine chromosome that contains the third least amount of genes (469 genes). This observation, together with the high Ts:Tv of the dataset, suggests that the DArTseq platform could have indeed targeted gene-rich regions. The difference in the correlation results between SNP density and gene density in bovine and blue wildebeest might reflect the consequences of genome reshuffling during the differentiation and history of adaption of these species (He e*t al*., 2018).

2.4.5. Functional Annotation

Gene ontology is frequently used to characterise gene functional annotation and classification. Gene ontology terms describe gene function using controlled vocabulary and hierarchy, which subsequently allows gene function to be grouped in one of the following main GO categories: molecular function, biological processes and cellular components (Sathiyamoorthy *et al.*, 2010). The functional annotation of the blue wildebeest DArTseq SNP marker sequences were carried out using Blast2GO software after a blastx search was performed. As expected, the majority of the significant blast hits were obtained against species of the Bovidae family, while *B. taurus* was the species most represented among the significant hits. Therefore, this also served as confirmation that the *B. taurus* genome was the most suitable reference genome to use for the *in silico* mapping of the marker sequences.

A total of 1 733 (8.43%) DArTseq marker sequences could be functionally annotated based on GO terms. This frequency is considerably lower than that typically observed in exome and transcriptome studies [*e.g.* 36.27% of domestic sheep (*Ovis aries*) unigenes annotated in a study by Yue e*t al*. (2015); 46.49% of goat (*Capra hircus*) transcriptome sequences annotated in a study by Wang *et al*. (2015b)]. Exome and RNA sequencing, as represented by the above-mentioned studies, are alternative reduced-representation approaches that allow more

selective sequencing, enabling focus on protein coding regions. Therefore, the sequences generated by these studies are likely to contain a high number of functional variants. Conversely, although the DArTseq platform is known to target gene rich regions, the majority of SNPs generated using GBS approaches are still located in non-coding regions and are often difficult to annotate (Scheben *et al*., 2017; Feng *et al*., 2018). Nevertheless, the high cost and prior requirements associated with exome sequencing and the biases, related to gene length and nucleotide composition, in RNA sequencing have resulted in GBS approaches becoming an increasingly popular genotyping method (Zheng *et al*., 2011; Scheben *et al*., 2017). The percentage of DArTseq sequences that could be functionally annotated in the current study is, however, also less than what would be typically expected for reduced representation approaches making use of REs. For example, in a study by Karam e*t al*. (2015), 10.66% of the conifer (*Cedrus atlantica*) RADseq sequences could be functionally annotated, while in a study by Sharma *et al*. (2012), on the Bornean elephant (*Elephas maximus borneensis*), 23% of the RADseq sequences could be associated with one or more GO terms. Functional annotation of a white-tailed deer RRL also resulted in only 9% of the sequences producing putative functional hits (Seabury *et al*., 2011). The low functional annotation success rate in the current study could thus be the result of poor antelope and deer sequence representation in existing sequence databases. Consequently, due to the short DArTseq query sequences, more distant but significant sequence similarities may be overlooked. Furthermore, in addition to the DArTseq sequences that were assigned putative function, another 63 sequences could be mapped, but could not be annotated. The reason for this is most likely that the default annotation threshold (55) was too high. However, this default value has been chosen by Blast2GO as it provides a good balance between quality and quantity of annotation (Blast2GO, n.d.). Consequently, the annotation threshold was not adjusted in order to avoid permissive annotation and to ensure accuracy of the results.

A wide range of different GO terms could be assigned to the DArTseq SNP marker sequences. The functional annotation results thus preliminarily indicated that the blue wildebeest DArTseq sequences represent various functional genes and these sequences could, therefore, be useful in functional gene studies (Feng e*t al*., 2018). Moreover, the GO terms most represented in each of the three main categories (molecular function, biological processes and cellular components) are very similar to those that have been reported by other studies in even-toed ungulates. A transcriptome study in forest musk deer (*Moschus berezovskii*) also reported “binding” and “catalytic activity” as the most common GO terms related to molecular function (Xu *et al*., 2017), while functional annotation of a white-tailed deer RRL also found binding activities to be among the most common molecular function GO terms (Seabury *et al*., 2011). In the biological processes category, all of the GO terms associated with the DArTseq sequences in the current study were also found to be associated with the blood transcriptome of

captive forest musk deer in a study by Sun *et al*. (2018). In addition, several studies in even-toed ungulates have reported detection of or response to stimuli to be among the most prevalent GO terms related to biological processes (*e.g.* Das *et al*., 2015; Da Silva *et al*., 2015; Sun *et al*., 2018). In the animal kingdom, the detection of and response to stimuli, such as smell and sound, plays a vital role in survival and reproduction (Da Silva e*t al*., 2015). Therefore, it is not surprising that this GO term is commonly associated with the genome sequences of even-toed ungulates. Lastly, the GO terms related to cellular components covered many of the important constituent parts of the cell and were also similar to those that have been reported in other ungulates (*e.g.* Sun e*t al*., 2018). The complete GO profile of blue wildebeest, obtained using the DArTseq platform, is thus very similar to the profiles that have been obtained for other even-toed ungulates using genotyping approaches such as RRLs and transcriptome sequencing. This suggests that the protocol differences between the DArTseq platform and other genotyping approaches did not introduce a bias in the distribution of GO terms. The level of identity between the GO profiles of different ungulate species thus reflect a shared molecular signature in their genome beyond the sampling effect of different genotyping platforms (Karam *et al*., 2015).

2.5. Conclusion

Based on comparison of conventional RRS approaches and the DArTseq platform, the present study concludes that DArTseq can be successfully utilised in blue wildebeest. The markers that were generated are of high quality, display desirable polymorphism parameters and are scattered across the whole genome. Furthermore, the high observed Ts:Tv ratio indicate that the DArTseq platform targeted gene-rich regions of the blue wildebeest genome, which is a well-known advantage of the platform. The identified DArTseq SNP markers represents a new valuable genomic resource for blue wildebeest that will be able to facilitate population management. These markers may also aid research efforts devoted to identifying the genetic basis of profitable traits in blue wildebeest, such as horn length and coat colour. The DArTseq platform, thus, represents a robust, cost-effective alternative to existing genotyping platforms, which may find adoption in numerous wildlife species with limited genetic resources.

REFERENCES

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., Hayden, M. J., Howes, N., Sharp, P., Vaughan, P., Rathmell, B., Huttner, E., & Kilian, A. (2006). Diversity Arrays Technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theoretical and Applied Genetics*, *113*(8), 1409–1420.

Al-Breiki, R. D., Kjeldsen, S. R., Afzal, H., Al Hinai, M. S., Zenger, K. R., Jerry, D. R., Al-Abri, M. A., & Delghandi, M. (2018). Genome-wide SNP analyses reveal high gene flow and signatures of local adaptation among the scalloped spiny lobster (*Panulirus homarus*) along the Omani coastline. *BMC Genomics*, *19*(690), 1–13.

Ba, H., Jia, B., Wang, G., Yang, Y., Kedem, G., & Li, C. (2017). Genome-wide SNP discovery and analysis of genetic diversity in farmed sika deer (*Cervus nippon*) in Northeast China using double-digest restriction site-associated DNA sequencing. *G3: Genes|Genomes|Genetics*, *7*(9), 3169–3176.

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Çiftçi, V., Sáenz de Miera, L. E., Aasim, M., Nadeem, M. A., Aktaş, H., Özkan, H., & Hatipoğlu, R. (2017). A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. *PLoS ONE*, *12*(1), e0167821.

Barilli, E., Cobos, M. J., Carrillo, E., Kilian, A., Carling, J., & Rubiales, D. (2018). A high-density integrated DArTseq SNP-based genetic map of *Pisum fulvum* and identification of QTLs controlling rust resistance. *Frontiers in Plant Science*, *9*, 167.

Blåhed, I.-D., Königsson, H., Ericsson, G., & Spong, G. (2018). Discovery of SNPs for individual identification by reduced representation sequencing of moose (*Alces alces*). *PLoS ONE*, *13*(5), e0197364.

Blanchong, J. A., Robinson, S. J., Samuel, M. D., & Foster, J. T. (2016). Application of genetics and genomics to wildlife epidemiology. *Journal of Wildlife Management*, *80*(4), 593–608.

Blast2GO (n.d.). *Frequently Asked Questions*. Retrieved from https://www.blast2go.com/support/faq#faqnoanchor [accessed 17 April 2019].

Bothma, J., Du Toit, J. T., & Van Rooyen, J. (2010). Buying and selling wild animals. In J. Bothma (Ed.), *Game Ranch Management* (5th ed., pp. 619–639). Pretoria: Van Schaik Publishers.

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, *32*(3), 314–331.

Bourgeois, S., Senn, H., Kaden, J., Taggart, J. B., Ogden, R., Jeffery, K. J., Bunnefeld, N., Abernethy, K., & McEwing, R. (2017). Single-nucleotide polymorphism discovery and panel characterization in the African forest elephant. *Ecology and Evolution*, *8*(4), 2207–2217.

Campbell, E. O., Davis, C. S., Dupuis, J. R., Muirhead, K., & Sperling, F. A. H. (2017). Cross-platform compatibility of

de novo-aligned SNPs in a nonmodel butterfly genus. *Molecular Ecology Resources*, *17*(6), e84–e93.

Chardonnet, P., & Mallon, D. (2015). *IUCN SSC Antelope Specialist Group position statement on the intentional genetic manipulation of antelopes.* Retrieved from http://cmsdata.iucn.org/downloads/asg_igm _posnsment _2015_final_19may_2015.pdf [accessed 17 June 2018].

Chen, J., Zavala, C., Ortega, N., Petroli, C., Franco, J., Burgueño, J., Costich, D. E., & Hearne, S. J. (2016). The development of quality control genotyping approaches: A case study using elite maize lines. *PLoS ONE*, *11*(6), e0157236.

Cloete, P. C., & Rosouw, R. (2014). The South African wildlife ranching sector: A social accounting matrix Leontief multiplier analysis. *Acta Commercii*, *14*(1), 1–10.

Couch, A. J., Unmack, P. J., Dyer, F. J., & Lintermans, M. (2016). Who's your mama? Riverine hybridisation of threatened freshwater Trout Cod and Murray Cod. *PeerJ*, *4*, e2593.

Courtois, B., Audebert, A., Dardou, A., Roques, S., Ghneim-Herrera, T., Droc, G., Frouin, J., Rouan, L., Gozé, E., Kilian, A., Ahmadi, N., & Dingkuhn, M. (2013). Genome-wide association mapping of root traits in a Japonica rice panel. *PLoS ONE*, *8*(11), e78037.

Cruz, V. M. V., Kilian, A., & Dierig, D. A. (2013). Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop lesquerella and related species. *PLoS ONE*, *8*(5), e64062.

Das, A., Panitz, F., Gregersen, V. R., Bendixen, C., & Holm, L.-E. (2015). Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes. *BMC Genomics*, *16*(1043), 1–12.

Da Silva, J. M., Giachetto, P. F., Da Silva, L. O. C., Cintra, L. C., Paiva, S. R., Caetano, A. R., & Yamagishi, M. E. B. (2015). Genomic variants revealed by invariably missing genotypes in Nelore cattle. *PLoS ONE*, *10*(8), 1–18.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD sequencing data: Implications for genotyping. *Molecular Ecology*, *22*(11), 3151–3164.

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*(7), 499–510.

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. (2013). Genotyping-by-sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE*, *8*(5), e62137.

DePristo, M. A., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel,

S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498.

DeYoung, R. W., & Honeycutt, R. L. (2005). The molecular toolbox : Genetic techniques in wildlife ecology and management. *Journal of Wildlife Management*, *69*(4), 1362–1384.

DiBattista, J. D., Wakefield, C. B., Moore, G. I., Bunce, M., Williams, A. J., O'Malley, J. M., Humphreys, R. L., Tuikolongahau, H., Williams, A., Green, M. A., Graham, K., Tucker, S. J., Cruz, E., & Newman, S. J. (2018). Genomic and life-history discontinuity reveals a precinctive lineage for a deep-water grouper with gene flow from tropical to temperate waters on the west coast of Australia. *Ecological Genetics and Genomics*, *9*, 23–33.

Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, *36*(16), e105.

Dos Santos, J. P. R., Pires, L. P. M., De castro Vasconcellos, R. C., Pereira, G. S., Von Pinho, R. G., & Balestre, M. (2016). Genomic selection to resistance to *Stenocarpella maydis* in maize lines using DArTseq markers. *BMC Genetics*, *17*(86), 1–10.

Dracatos, P. M., Khatkar, M. S., Singh, D., & Park, R. F. (2014). Genetic mapping of a new race specific resistance allele effective to *Puccinia hordei* at the *Rph9*/*Rph12* locus on chromosome 5HL in barley. *BMC Plant Biology*, *14*(1598), 1–12.

Eck, S. H., Benet-Pagès, A., Flisikowski, K., Meitinger, T., Fries, R., & Strom, T. M. (2009). Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome Biology*, *10*(8), R82.

El-Halawany, N., Abdel-Shafy, H., Shawky, A.-E.-M. A., Abdel-Latif, M. A., Al-Tohamy, A. F. M., & Abd El-Moneim, O. M. (2017). Genome-wide association study for milk production in Egyptian buffalo. *Livestock Science*, *198*, 10–16.

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*(5), e19379.

Elsik, C. G., Unni, D. R., Diesh, C. M., Tayal, A., Emery, M. L., Nguyen, H. N., & Hagen, D. E. (2016). Bovine Genome Database: New tools for gleaning function from the *Bos taurus* genome. *Nucleic Acids Research*, *44*(D1), D834–D839.

Endangered Wildlife Trust (EWT) (2016). *Endangered Wildlife Trust position statement on the intensive breeding of wildlife species with particular reference to selective breeding for colour variance*. Retrieved from https://www.ewt.org.za/scientific%20publications/position%20statements/EWT%20Wildlife%20intensive%20breeding%20and%20colour%20variation%20position%20statement.pdf [accessed 17 June 2018].

Fan, J., Zheng, X., Wang, H., Qi, H., Jiang, B., Qiao, M., Zhou, J., & Bu, S. (2019). Analysis of genetic diversity and

population structure in three forest musk deer captive populations with different origins. *G3: Genes|Genomes|Genetics*, *9*(4), 1037–1044.

Feng, J. Y., Li, M., Zhao, S., Zhang, C., Yang, S. T., Qiao, S., Tan, W. F., Qu, H. J., Wang, D. Y., & Pu, Z. G. (2018). Analysis of evolution and genetic diversity of sweetpotato and its related different polyploidy wild species *I. trifida* using RAD-seq. *BMC Plant Biology*, *18*(181), 1–12.

Garavito, A., Montagnon, C., Guyot, R., & Bertrand, B. (2016). Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. *BMC Plant Biology*, *16*(242), 1–12.

Gawroński, P., Pawełkowicz, M., Tofil, K., Uszyński, G., Sharifova, S., Ahluwalia, S., Tyrka, M., Wędzony, M., Kilian, A., & Bolibok-Brągoszewska, H. (2016). DArT markers effectively target gene space in the rye genome. *Frontiers in Plant Science*, *7*, 1600.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435.

Gurgul, A., Miksza-Cybulska, A., Szmatoła, T., Jasielczuk, I., Piestrzyńska-Kajtoch, A., Fornal, A., Semik-Gurgul, E., & Bugno-Poniewierska, M. (2019). Genotyping-by-sequencing performance in selected livestock species. *Genomics*, *111*(2), 186–195.

He, J., Guo, Y., Xu, J., Li, H., Fuller, A., Tait, R. G., Jr., Wu, X.-L., & Bauck, S. (2018). Comparing SNP panels and statistical methods for estimating genomic breed composition of individual animals in ten cattle breeds. *BMC Genetics*, *19*(1), 56.

Hendre, P. S., Kamalakannan, R., & Varghese, M. (2012). High-throughput and parallel SNP discovery in selected candidate genes in *Eucalyptus camaldulensis* using Illumina NGS platform. *Plant Biotechnology Journal*, *10*(6), 646–656.

Ivy, J. A., Putnam, A. S., Navarro, A. Y., Gurr, J., & Ryder, O. A. (2016). Applying SNP-derived molecular coancestry estimates to captive breeding programs. *Journal of Heredity*, *107*(5), 403–412.

Jaccoud, D., Peng, K., Feinstein, D., & Kilian, A. (2001). Diversity Arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, *29*(4), E25.

Jiang, Z., Wu, X. L., Zhang, M., Michal, J. J., & Wright, R. W. (2008). The complementary neighborhood patterns and methylation-to-mutation likelihood structures of 15,110 single-nucleotide polymorphisms in the bovine genome. *Genetics*, *180*(1), 639–647.

Jonker, R. M., Zhang, Q., Van Hooft, P., Loonen, M. J. J. E., Van der Jeugd, H. P., Crooijmans, R. P. M. A., Groenen, M. A. M., Prins, H. H. T., & Kraus, R. H. S. (2012). The development of a genome wide SNP set for the Barnacle Goose *Branta leucopsis*. *PLoS ONE*, *7*(7), e38412.

Joubert, A. (2017). Microsatellite markers: Understanding the basics. *Wildlife Stud Services: Wildlife Breeders Journal*, 44–46.

Karam, M.-J., Lefèvre, F., Bou Dagher-Kharrat, M., Pinosio, S., & Vendramin, G. G. (2015). Genomic exploration and molecular marker development in a large and complex conifer genome using RADseq and mRNAseq. *Molecular Ecology Resources*, *15*(3), 601–612.

Kerstens, H. H. D., Crooijmans, R. P. M. A., Veenendaal, A., Dibbits, B. W., Chin-A-Woeng, T. F. C., Den Dunnen, J. T., & Groenen, M. A. M. (2009). Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: Applied to turkey. *BMC Genomics*, *10*(479), 1–11.

Kess, T., Gross, J., Harper, F., & Boulding, E. G. (2016). Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. *Journal of Molluscan Studies*, *82*(1), 104–109.

Kilian, A., Sanewski, G., & Ko, L. (2016). The application of DArTseq technology to pineapple. *Acta Horticulturae*, *1111*, 181–188.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity Arrays Technology: A generic genome profiling technology on open platforms. In F. Pompanon, & A. Bonin (Eds.), *Data Production and Analysis in Population Genomics: Methods and Protocols* (1st ed., pp. 67–89). New York: Humana Press.

Kotzé, A., Grobler, J. P., & Buduram, P. (2002). *The application of genetic markers in game farming in South Africa*. Paper presented at the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, 19–23 August 2002. Retrieved from http://www.wcgalp.org/system/files/proceedings/2002/application-genetic-markers-game-farming-south-africa.pdf [accessed 4 February 2018].

Kraus, R. H. S., Kerstens, H. H. D., Van Hooft, P., Crooijmans, R. P. M. A., Van der Poel, J. J., Elmberg, J., Vignal, A., Huang, Y., Li, N., Prins, H. H. T., & Groenen, M. A. M. (2011). Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC Genomics*, *12*(150), 1–11.

Lal, M. M. (2016). *Genetic structure, population connectivity and taxonomic identity of the black-lip pearl oyster* Pinctada margaritifera *(Bivalvia: Pteriidae), across its Indo-Pacific distribution* (Unpublished PhD thesis). James Cook University, Queensland, Australia.

Lal, M. M., Southgate, P. C., Jerry, D. R., Bosserelle, C., & Zenger, K. R. (2017). Swept away: Ocean currents and seascape features influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *BMC Genomics*, *18*(66), 1–21.

Lambert, M. R., Skelly, D. K., & Ezaz, T. (2016). Sex-linked markers in the North American green frog (*Rana*

clamitans) developed using DArTseq provide early insight into sex chromosome evolution. *BMC Genomics*, *17*(844), 1–13.

Le, S. Q., & Durbin, R. (2011). SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. *Genome Research*, *21*(6), 952–960.

Li, Y., & Kim, J. J. (2015). Effective population size and signatures of selection using bovine 50K SNP chips in Korean native cattle (Hanwoo). *Evolutionary Bioinformatics*, *11*(11), 143–153.

Lind, C. E., Kilian, A., & Benzie, J. A. H. (2017). Development of Diversity Arrays Technology markers as a tool for rapid genomic assessment in Nile tilapia, *Oreochromis niloticus*. *Animal Genetics*, *48*(3), 362–364.

Liu, Q., Guo, Y., Li, J., Long, J., Zhang, B., & Shyr, Y. (2012). Steps to ensure accuracy in genotype and SNP calling from Illumina sequence data. *BMC Genomics*, *13*(S8), S8.

Liu, Y., Qin, X., Song, X.-Z., Jiang, H., Shen, Y., Durbin, K. J., Lien, S., Kent, M. P., Sodeland, M., Ren, Y., Zhang, L., Sodergren, E., Havlak, P., Worley, K. C., Weinstock, G. M., & Gibbs, R. A. (2009). *Bos taurus* genome assembly. *BMC Genomics*, *10*(180), 1–11.

Liu, Z. J., & Cordes, J. F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, *238*(1–4), 1–37.

Lo, E., Bonizzoni, M., Hemming-Schroeder, E., Ford, A., Janies, D. A., James, A. A., Afrane, Y., Etemesi, H., Zhou, G., Githeko, A., & Yan, G. (2018). Selection and utility of single nucleotide polymorphism markers to reveal fine-scale population structure in human malaria parasite *Plasmodium falciparum*. *Frontiers in Ecology and Evolution*, *6*, 145.

MacEachern, S., McEwan, J., & Goddard, M. (2009). Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae). *BMC Genomics*, *10*(177), 1–17.

Manley, L. J., Ma, D., & Levine, S. S. (2016). Monitoring error rates in Illumina sequencing. *Journal of Biomolecular Techniques*, *27*(4), 125–128.

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P. L., Sonstegard, T. S., & Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE*, *4*(4), e5350.

Melville, J., Haines, M. L., Boysen, K., Hodkinson, L., Kilian, A., Smith Date, K. L., Potvin, D. A., & Parris, K. M. (2017). Identifying hybridization and admixture using SNPs: Application of the DArTseq platform in phylogeographic research on vertebrates. *Royal Society Open Science*, *4*(7), 161061.

Miller, S. M., Clarke, A. B., Bloomer, P., Guthrie, A. J., & Harper, C. K. (2016). Evaluation of microsatellites for common ungulates in the South African wildlife industry. *Conservation Genetics Resources*, *8*(3), 329–341.

Moore, J.-S., & Benestan, L. (2018). Let's talk about sex : A rigorous statistical framework to assign the sex of

individuals from reduced-representation sequencing data. *Molecular Ecology Resources*, *18*(2), 191–193.

Morse, P., Kjeldsen, S. R., Meekan, M. G., Mccormick, M. I., Finn, J. K., Huffard, C. L., & Zenger, K. R. (2018). Genome-wide comparisons reveal a clinal species pattern within a holobenthic octopod—the Australian Southern blue-ringed octopus, *Hapalochlaena maculosa* (Cephalopoda: Octopodidae). *Ecology and Evolution*, *8*(4), 2253–2267.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., & Baloch, F. S. (2018). DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnology and Biotechnological Equipment*, *32*(2), 261–285.

Nguyen, N. H., Premachandra, H. K. A., Kilian, A., & Knibb, W. (2018). Genomic prediction using DArT-Seq technology for yellowtail kingfish *Seriola lalandi*. *BMC Genomics*, *19*(107), 1–9.

Okoń, S., Ociepa, T., & Nucia, A. (2018). Molecular identification of *Pm4* powdery mildew resistant gene in oat. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, *46*(2), 350–355.

Ozerov, M., Vasemägi, A., Wennevik, V., Diaz-Fernandez, R., Kent, M., Gilbey, J., Prusov, S., Niemelä, E., & Vähä, J. P. (2013). Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. *PLoS ONE*, *8*(12), e82434.

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*(5), e37135.

Pootakham, W., Sonthirod, C., Naktang, C., Jomchai, N., Sangsrakru, D., & Tangphatsornruang, S. (2016). Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: A case study in oil palm (*Elaeis guineensis*). *Molecular Breeding*, *36*(11), 154.

Ray, S., & Satya, P. (2014). Next generation sequencing technologies for next generation plant breeding. *Frontiers in Plant Science*, *5*, 367.

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

Russo, I.-R. M., Hoban, S., Bloomer, P., Kotzé, A., Segelbacher, G., Rushworth, I., Birss, C., & Bruford, M. W. (2019). 'Intentional genetic manipulation' as a conservation threat. *Conservation Genetics Resources*, *11*(2), 237–247.

Sánchez-Sevilla, J. F., Horvath, A., Botella, M. A., Gaston, A., Folta, K., Kilian, A., Denoyes, B., & Amaya, I. (2015).

Diversity arrays technology (DArT) marker platforms for diversity analysis and linkage mapping in a complex crop, the octoploid cultivated strawberry (*Fragaria* × *ananassa*). *PLoS ONE*, *10*(12), e0144960.

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proceedings*, *5*(S7), P54.

Sarre, S. D., & Georges, A. (2009). Genetics in conservation and wildlife management: A revolution since Caughley. *Wildlife Research*, *36*(1), 70–80.

Sathiyamoorthy, S., In, J.-G., Gayathri, S., Kim, Y.-J., & Yang, D. C. (2010). Generation and gene ontology based analysis of expressed sequence tags (EST) from a *Panax ginseng* C. A. Meyer roots. *Molecular Biology Reports*, *37*(7), 3465–3472.

Scheben, A., Batley, J., & Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: Choosing the right tool for the right application. *Plant Biotechnology Journal*, *15*(2), 149–161.

Schultz, A. J., Cristescu, R. H., Littleford-Colquhoun, B. L., Jaccoud, D., & Frère, C. H. (2018). Fresh is best: Accurate SNP genotyping from koala scats. *Ecology and Evolution*, *8*, 3139–3151.

Seabury, C. M., Bhattarai, E. K., Taylor, J. F., Viswanathan, G. G., Cooper, S. M., Davis, D. S., Dowd, S. E., Lockwood, M. L., & Seabury, P. M. (2011). Genome-wide polymorphism and comparative analyses in the white-tailed deer (*Odocoileus virginianus*): A model for conservation genomics. *PLoS ONE*, *6*(1), e15811.

Sharma, R., Goossens, B., Kun-Rodrigues, C., Teixeira, T., Othman, N., Boone, J. Q., Jue, N. K., Obergfell, C., O'Neill, R. J., & Chikhi, L. (2012). Two different high throughput sequencing approaches identify thousands of *de novo* genomic markers for the genetically depleted Bornean elephant. *PLoS ONE*, *7*(11), e49533.

Shirasawa, K., Hirakawa, H., & Isobe, S. (2016). Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and *in silico* optimization in tomato. *DNA Research*, *23*(2), 145–153.

Sieber, A. N., Longin, C. F. H., Leiser, W. L., & Würschum, T. (2016). Copy number variation of *CBF-A14* at the *Fr-A2* locus determines frost tolerance in winter durum wheat. *Theoretical and Applied Genetics*, *129*(6), 1087–1097.

Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., Singh, N. K., & Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE*, *8*(12), e84136.

Siu, H., Zhu, Y., Jin, L., & Xiong, M. (2011). Implication of next-generation sequencing on association studies. *BMC Genomics*, *12*(322), 1–11.

Smitz, N., Van Hooft, P., Heller, R., Cornélis, D., Chardonnet, P., Kraus, R., Greyling, B., Crooijmans, R., Groenen, M., & Michaux, J. (2016). Genome-wide single nucleotide polymorphism (SNP) identification and

characterization in a non-model organism, the African buffalo (*Syncerus caffer*), using next generation sequencing. *Mammalian Biology*, *81*(6), 595–603.

Souza, C. A., Murphy, N., Villacorta-Rath, C., Woodings, L. N., Ilyushkina, I., Hernandez, C. E., Green, B. S., Bell, J. J., & Strugnell, J. M. (2017). Efficiency of ddRAD target enriched sequencing across spiny rock lobster species (Palinuridae: *Jasus*). *Scientific Reports*, *7*(6781), 1–14.

Steiner, C. C., Charter, S. J., Houck, M. L., & Ryder, O. A. (2014). Molecular phylogeny and chromosomal evolution of Alcelaphini (Antilopinae). *Journal of Heredity*, *105*(3), 324–333.

Sun, X., Cai, R., Jin, X., Shafer, A. B. A., Hu, X., Yang, S., Li, Y., Qi, L., Liu, S., & Hu, D. (2018). Blood transcriptomics of captive forest musk deer (*Moschus berezovskii*) and possible associations with the immune response to abscesses. *Scientific Reports*, *8*(1), 599.

Tambling, C., Viljoen, P., & Kotze, A. (2016). A conservation assessment of *Connochaetes taurinus*. In M. F. Child, L. Roxburgh, E. Do Lihn San, D. Raimondo, & H. T. Davies-Mostert (Eds.), *The Red List of Mammals of South Africa, Swaziland and Lesotho*. South African National Biodiversity Institute and Endangered Wildlife Trust, South Africa.

Taylor, W. A., Lindsey, P. A., & Davies-Mostert, H. (2016). *An assessment of the economic, social and conservation value of the wildlife ranching industry and its potential to support the green economy in South Africa*. Endangered Wildlife Trust, South Africa.

Torkamaneh, D., Laroche, J., & Belzile, F. (2016). Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLoS ONE*, *11*(8), e0161333.

Van Bers, N. E. M., Van Oers, K., Kerstens, H. H. D., Dibbits, B. W., Crooijmans, R. P. M. A., Visser, M. E., & Groenen, M. A. M. (2010). Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. *Molecular Ecology*, *19*(S1), 89–99.

Van Wyk, A. M., Kotzé, A., Grobler, J. P., Janse van Vuuren, B., Barrow, L. N., & Dalton, D. L. (2018). Isolation and characterization of species-specific microsatellite markers for blue and black wildebeest (*Connochaetes taurinus and C. gnou*). *Journal of Genetics*, *97*(Online Resources), e101–e109.

Van Wyk, A. M., Labuschagne, C., Kropff, A. S., Kotzé, A., Grobler, J. P., Jansen van Vuuren, B., & Dalton, D. (2019). A targeted gene approach to SNP discovery in the blue (*Connochaetes taurinus*) and black wildebeest (*C. gnou*). *Conservation Genetic Resources*, *11*(1), 35–38.

Von Thaden, A., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlidis, A. A., Tiesmeyer, A., Nowak, C., & Muñoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports*, *7*(1), 10768.

Wang, G. T., Peng, B., & Leal, S. M. (2014). Variant association tools for quality control and analysis of large-scale

sequence and genotyping array data. *American Journal of Human Genetics*, *94*(5), 770–783.

Wang, J., Raskin, L., Samuels, D. C., Shyr, Y., & Guo, Y. (2015a). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics*, *31*(3), 318–323.

Wang, L., Zhang, Y., Zhao, M., Wang, R., Su, R., & Li, J. (2015b). SNP discovery from transcriptome of cashmere goat skin. *Asian-Australasian Journal of Animal Sciences*, *28*(9), 1235–1243.

Wang, W., Gan, J., Fang, D., Tang, H., Wang, H., Yi, J., & Fu, M. (2018). Genome-wide SNP discovery and evaluation of genetic diversity among six Chinese indigenous cattle breeds in Sichuan. *PLoS ONE*, *13*(8), e0201534.

Wang, W., Yan, H., Yu, J., Yi, J., Qu, Y., Fu, M., Chen, A., Tang, H., & Niu, L. (2017). Discovery of genome-wide SNPs by RADseq and the genetic diversity of captive hog deer (*Axis porcinus)*. *PLoS ONE*, *12*(3), e0174299.

Xing, C., Schumacher, F. R., Xing, G., Lu, Q., Wang, T., & Elston, R. C. (2005). Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis. *BMC Genetics*, *6*(S1), S29.

Xu, Z., Jie, H., Chen, B., Guar, U., Wu, N., Gao, J., Li, P., Zhao, G., Zeng, D., Yang, M., & Li, D. (2017). Illumina-based *de novo* transcriptome sequencing and analysis of Chinese forest musk deer. *Journal of Genetics*, *96*(6), 1033–1040.

Yang, G.-Q., Chen, Y.-M., Wang, J.-P., Guo, C., Zhao, L., Wang, X.-Y., Guo, Y., Li, L., Li, D.-Z., & Guo, Z.-H. (2016). Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. *Plant Methods*, *12*(39), 1–17.

Yi, T., Sun, J., Liang, X., He, S., Li, L., Wen, Z., & Shen, D. (2013). Effects of polymorphisms in *pepsinogen* (*PEP*), *amylase* (*AMY*) and *trypsin* (*TRY)* genes on food habit domestication traits in mandarin fish. *International Journal of Molecular Sciences*, *14*(11), 21504–21512.

Yousefi, S., Abbassi-Daloii, T., Kraaijenbrink, T., Vermaat, M., Mei, H., Van 't Hof, P., Van Iterson, M., Zhernakova, D. V., Claringbould, A., Franke, L., 't Hart, L. M., Slieker, R. C., Van der Heijden, A., De Knijff, P., BIOS Consortium, & 't Hoen, P. A. C. (2018). A SNP panel for identification of DNA and RNA specimens. *BMC Genomics*, *19*(90), 1–12.

Yue, Y. J., Liu, J. B., Yang, M., Han, J. L., Guo, T. T., Guo, J., Feng, R. L., & Yang, B. H. (2015). *De novo* assembly and characterization of skin transcriptome using RNAseq in sheep (*Ovis aries*). *Genetics and Molecular Research*, *14*(1), 1371–1384.

Zheng, W., Chung, L., & Zhao, H. (2011). Bias detection and correction in RNA-sequencing data. *BMC Bioinformatics*, *12*(290), 1–14.

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassel, C. P., Sonstegard, T. S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J. A., & Salzberg, S. L. (2009). A whole-

genome assembly of the domestic cow, Bos *taurus*. *Genome Biology*, *10*(4), R42.

CHAPTER 3

Elucidation of Coat Colour Genetics in Blue Wildebeest

Abstract

In blue wildebeest, coat colour is an important trait due to the high economic value associated with the golden coat colour variant. Based on parentage records, the prevailing hypothesis is that the golden coat colour is a recessive phenotype caused by the homozygous state of a single major mutation. However, the genetic basis of the golden coat colour phenotype has not been investigated. Therefore, the aim of the study was to elucidate the genetic underpinning and mode of inheritance of the golden coat colour phenotype. A genome-wide association study was performed with a set of 14 624 DArTseq single nucleotide polymorphisms (SNPs) to identify putative candidate genes involved in blue wildebeest pigmentation. A total of 374 SNPs were significantly associated with coat colour in blue wildebeest (p-value ≤ 0.001). Five of these SNPs mapped to four different *Bos taurus* orthologous genes that could possibly be involved in pigmentation based on previous literature reports. An additional three SNPs with an association p-value ≤ 0.05 mapped to well-known pigmentation genes and were also considered. Based on the reported biological function of the identified putative candidate genes, the *myosin VC* (*MYO5C*), *myosin VIIA* (*MYO7A*), *solute carrier family 6 member 3* (*SLC6A3*), *solute carrier family 28 member 2* (*SLC28A2*), *dopamine receptor D2* (*DRD2*), *frizzled class receptor 4* (*FZD4*) and *tyrosinase* (*TYR*) genes are promising candidate genes for coat colour determination in blue wildebeest. Of the SNPs located in the putative candidate genes, SNP 100076096|F|0--31:A>G located in *MYO7A* [OR: A/G= 2.63E+09 (1.69E+09 – 4.10E+09); G/G= 5.12E+10 (3.35E+10 – 7.83E+10)], SNP 100025667|F|0--39:T>C located in *SLC6A3* [OR: C/T= 2.62E+07 (2.62E+07 – 2.62E+07); T/T= 2.10E+17 (2.10E+17 – 2.10E+17)] and SNP 100048574|F|0--35:G>A located in *SLC28A2* [OR A/G: 1.71E+08 (9.86E+07 – 2.97E+08); G/G= 8.21E+08 (4.73+08 – 1.42E+09)] had the largest genetic size effects and are thus expected to have the greatest influence on coat colour determination. Moreover, based on the gene-gene interaction analysis performed between the SNPs located in the candidate genes, and their determined mode of inheritance, coat colour in blue wildebeest could possibly be a quantitative threshold trait rather than a simple autosomal recessive trait. This study provides a basis for further investigation on the genetic mechanisms of pigmentation in blue wildebeest.

3.1. Introduction

Pelt or plumage colour is an important trait in animals since it often has a specific function, such as camouflage, photoprotection, mate selection, communication or regulation of physiological processes (*e.g.* thermoregulation). Therefore, in natural populations, wild animals generally possess limited coat colour variability due to its prominent role in adaption (Linderholm & Larson, 2013). Conversely, in farmed populations, novel coat colours are often actively selected for due to human penchant, thus allowing the proliferation of mutations in coat colour genes (Koseniuk *et al.*, 2018). In recent years, coat colour variations of African antelope species have also become a topic of great interest in the South African wildlife ranching industry due to their novelty and rarity. The significant increase in the demand for these animals has also resulted in an increase in their monetary value (Taylor *et al.*, 2016). Consequently, colour variant game have become significantly more prevalent on wildlife ranches and numerous new colour variants have also become known through selective breeding (Strauss, 2015; Taylor *et al.*, 2016). Currently, there are more than 40 distinct colour variations of African Bovidae (Olivier, 2015). Common examples include golden wildebeest (*Connochaetes taurinus taurinus*), golden and white oryx (*Oryx gazelle*), black and saddleback impala (*Aepyceros melampus*) and copper, yellow and saddleback blesbuck (*Damaliscus pygargus phillipsi*; Olivier, 2015; Cloete, 2017).

The emergence of selective breeding and the intensive management of colour variant antelope has, however, raised concerns within the broader hunting and conservation fraternity, and particularly the conservation agencies in South Africa (Taylor *et al.,* 2016; Blackmore, 2017). The greatest dangers regarding the ongoing selective breeding of colour variants are considered to be inbreeding and the subsequent loss of genetic diversity, weakened resilience to environmental changes, reduced reproductive fitness of captive stock, and the fixation of deleterious alleles (Taylor *et al.*, 2016; Russo *et al.*, 2019). In addition, mutations influencing coat colour often have other pleiotropic effects including skin diseases, such as melanoma, and disorders of the immune system, reproductive tract and sensory organs. If colour variations in game species are the result of such deleterious mutations it could have severe consequences for the fitness of the affected animal (Reissmann & Ludwig, 2013; Charon & Lipka, 2015). Published scientific data concerning game colour variants are, however, very limited and most of the available information is in the form of websites, newsletters and articles in popular game ranching magazines (Olivier, 2015; Taylor *et al.*, 2016). Due to this lack in published scientific data, nature conservation authorities have not been able to set guidelines or assess the validity of the concerns regarding the breeding of colour variants (Taylor *et al.*, 2016). Since colour variant game have become such an integral part of the South African wildlife industry, research regarding the genetic basis of these coat colour variations has become

increasingly important to facilitate breeding management and to assist in the regulation of colour variant Bovidae (Olivier, 2015).

Through the study of model organisms, great insights have been gained into the molecular mechanisms underlying pigmentation, specifically melanin-based pigmentation. Furthermore, with advancing technology, it is now possible to study the genetic basis of pigmentation in non-model systems. Indeed, these studies have demonstrated that pigmentation biology and genetics are highly conserved across mammals (Hubbard *et al*., 2010). In mammals, a small population of specialised cells, called melanocytes, is responsible for pigmentation (Yamaguchi *et al*., 2007). Melanocytes express a limited number of specific proteins that act in a cascade, termed melanogenesis, to produce the melanin biopolymer. In melanocytes, melanin pigments are synthesised and stored in melanosomes, which are lysosome-related organelles, before distribution to the surrounding keratinocytes (D'Mello *et al*., 2016). Mammalian melanocytes can produce two distinct types of melanin: the brownish-black eumelanin and the reddish-yellow pheomelanin. Coat colour variability is mainly determined by the ratio of eumelanin to pheomelanin, which is regulated by the melanocortin 1 receptor (MC1R) ligand system. Coat colour can, however, be modified by several other genes involved in pigmentation (Cieslak *et al.*, 2011). Currently, more than a 150 genes are known to regulate pigmentation in mammals, either directly or indirectly, with updated lists of these genes being actively maintained by the European Society for Pigment Cell Research (ESPCR; http://www.espcr.org/).

Since pigmentation biology is greatly conserved across mammals, the candidate gene approach has been very successful in identifying the molecular basis underlying coat colour variation in the past. However, the candidate gene approach provides no guarantees as nearly identical coat colour phenotypes in a species or among species can arise from changes in different genes (Hoekstra, 2006; Cieslak *et al.*, 2011; Reissmann & Ludwig, 2013). Consequently, the genome-wide association study (GWAS) approach has become a more widely used approach since it is able to pinpoint novel casual variants not previously suspected to play a role in a specific phenotypic trait (Amos *et al.*, 2011). Rather than focusing on specific candidate genes, a large genome-wide set of genetic variants is tested in a GWAS in order to identify genes or genomic regions that are associated with a phenotypic trait (Lewis & Knight, 2012). The GWAS approach has been successfully used to elucidate coat colour variation in many different species, including porcine (Ren *et al.*, 2011), bovine (Fan *et al.*, 2014), ovine (Li *et al.*, 2014) and caprine (Nazari-Ghadikolaei *et al.*, 2018). Genome-wide association studies thus represent a promising approach for elucidating the genetic basis of coat colour variation in African antelope species.

Blue wildebeest, also known as brindled gnu, are even-toed ungulates of the Bovidae family that originate from southern and eastern Africa (Adetunji *et al.*, 2018; Grobler *et al.*, 2018). The blue wildebeest (*C. t. taurinus*) is an economically important species in the South African wildlife industry, due to its utilisation in hunting and live animal sales (Tambling *et al.*, 2016). Furthermore, the golden wildebeest, a colour morph of the blue wildebeest, has become one of the most common game colour variants that wildlife ranchers breed with (Taylor *et al.*, 2016). Unlike the silvery-blue coat colour of the blue wildebeest, the coat of the golden wildebeest is light brown to yellow-brown in colour and is expected to occur due to increased pheomelanin synthesis (Adetunji *et al.*, 2018; Grobler *et al.*, 2018). Furthermore, based on breeding records, the golden coat colour variation is believed to be the result of a simple autosomal recessive mutation (Taylor *et al.*, 2016). However, the genetic basis of the golden coat colour phenotype has not been studied previously and, consequently, this mode of inheritance has not been confirmed. In addition, it is also not known if the causal variant of the golden coat colour is linked to other pleiotropic effects (Olivier, 2015). If the casual variant is linked to such detrimental effects, it may pose significant economic drawbacks in terms of mortality, aesthetics and management costs (Adetunji *et al.*, 2018). Consequently, there is an urgent need to elucidate the genetic underpinnings of the golden coat colour. Genetic research concerning these animals will be able to facilitate breeding management and will also enable game ranchers to make informed decisions regarding the viability of golden wildebeest ranching (Russo *et al.*, 2019). This study, therefore, aimed to (1) identify single nucleotide polymorphisms (SNPs) that are significantly associated with coat colour in blue wildebeest by performing a case-control GWAS, (2) identify whether any of the marker sequences of the significant SNP mapped to genes with known functional roles in pigmentation, (3) elucidate the mode of inheritance of the golden coat colour variation, and (4) identify gene combinations that could play a role in determining coat colour in blue wildebeest.

3.2. Materials and Methods

3.2.1. Study Population: Pedigree Relations and Genetic Resources

Ninety-four blue wildebeest (*C. t. taurinus*) originating from a private game ranch (S24° 45.429' E28° 27.162') located in Limpopo Province, South Africa were selected for this association study after written informed consent was obtained from the rancher. The selected blue wildebeest included 35 golden wildebeest and 35 split wildebeest, which are phenotypically normal blue wildebeest that are presumed carriers of the golden causal variant based on parentage records. The majority of the selected blue wildebeest could be grouped into one of 15 families based on shared sires and, therefore, some of the individuals in the population were half-sibs.

Furthermore, five of the selected blue wildebeest were also the dams of offspring in other families in the study population (Appendix A).

The selected blue wildebeest were genotyped using the DArTseq platform, as described in chapter 2, and a set of 20 563 genome-wide SNP marker were available for this study population. The flanking sequences of these SNPs (69 bp in length) were aligned to the *Bos taurus* genome sequence assembly [UMD_3.1.1; Bovine Genome Database (http://bovinegenome.org/)]. Of these sequences, 3 907 mapped to putative bovine orthologues genes.

3.2.2. Association Analysis

The 20 563 SNPs were assessed for quality before the association analysis was conducted and removed if they had a minor allele frequency (MAF) less than 0.05 and a call rate less than 0.80. Thereafter, for the remaining SNPs (14 624), Hardy-Weinberg equilibrium (HWE), allele frequencies and genotype frequencies were calculated for the sample population, as well as for the case and control groups separately, using the software GenALEx version 6.503 (Peakall & Smouse, 2006; 2012). Association between the DArTseq generated SNPs and coat colour was assessed using the software UNPHASED version 3.1.7 (Dudbridge, 2008). Since coat colour in blue wildebeest is a binary trait, a case-control design was implemented for the association study. Individuals with the golden coat colour phenotype represented 'affected' individuals (cases), whereas phenotypically normal blue wildebeest represented 'unaffected' individuals (controls). To prevent confounding and to strengthen the power of the analysis, family relationships were taken into account for the association analysis where possible. The default "Full" analysis model for genotypic tests was used for the association analysis. Statistical significance was evaluated by permutation testing, performing 1000 random permutations. Single nucleotide polymorphisms were considered to be significantly associated with coat colour in blue wildebeest for p-value ≤ 0.001. In addition, SNPs that mapped to *B. taurus* genes with a known role in pigmentation (*i.e.* occurs in the ESPCR pigment gene database) were considered statistically significant for p-value ≤ 0.05. In order to estimate the genetic effect size of the statistically significant SNPs, the odds ratio (OR) for each genotype and upper and lower bound 95% confidence intervals (95% CI) were also calculated by the UNPHASED software.

The statistically significant SNPs were also analysed using the software SNPstats (Solé *et al.*, 2006; https://www.snpstats.net/start.htm), in order to confirm the association. In addition, SNPstats assessed multiple inheritance models (dominant, over-dominant, co-dominant, recessive and additive) to evaluate the associations between the statistically significant SNPs and coat colour in blue wildebeest. Each inheritance model was assessed by performing an unconditional logistic regression analysis, and the best-fitting inheritance model for each

polymorphism was identified based on the lowest Akaike Information Criterion (AIC) and/or Bayesian Information Criterion (BIC).

3.2.3. Assessment of Linkage Disequilibrium

The extent of linkage disequilibrium (LD), defined as the non-random association of alleles at different loci, was assessed between the SNPs found to be significantly associated with coat colour (Bush & Moore, 2012). The standard descriptive LD parameters, D' and r^2, were estimated for all pairwise SNP combinations using the LD plot function in the software Haploview version 4.2 (Barrett *et al*., 2005). Using LD categories defined by Espigolan *et al*. (2013), pairwise SNP combinations were classified as exhibiting low LD ($r^2 \leq 0.16$), moderate LD ($0.16 < r^2 \leq 0.70$) or high LD ($r^2 > 0.70$). In addition, LD for all possible pairs of loci were also tested with probability tests in Genepop version 4.2 using default Markov chain parameters (Raymond & Rousset, 1995; Rousset, 2008). Linkage disequilibrium was considered to be significant for p-values ≤ 0.05.

3.2.4. Functional Annotation and Identification of Candidate Loci

The marker sequences of all the statistically significant SNPs were functionally annotated using the Blast2GO pipeline (Blast2GO version 5.5.1; Götz *et al.*, 2008) in order to gain an overview of the functional processes the genes, in which these SNPs are situated, are involved in. The flanking sequence available for each of these SNPs was used to conduct a blastx search against the non-redundant (nr) National Center for Biotechnology Information (NCBI) sequence database prior to mapping for Gene Ontology (GO) terms. Default blastx parameters were used, however, a mammalian taxonomic filter was applied to make the blast search more time-efficient. Blast matches were considered statistically significant with E-values ≤ 1.0E-1 due to the short length of the SNP flanking sequences. Mapping and annotation of GO terms at level 2 were performed with default Blast2GO settings in order to retrieve molecular function, biological process and cellular component terms.

The marker sequences of the SNPs found to be statistically significant ($p \leq 0.001$) in the association study were then matched, where possible, to the *B. taurus* orthologous genes identified in blue wildebeest. Each of these genes were further investigated to determine if they could be involved in pigmentation. Genes that could possibly play a role in pigmentation, based on previous literature reports, represented putative candidate genes for the determination of coat colour in blue wildebeest. In addition, *B. taurus* orthologous genes that occurred in the ESPCR database were also considered as putative candidate genes for coat colour determination if the corresponding DArTseq SNP had a p-value ≤ 0.05 for the association study. The orthologous genes that were

identified as putative candidate genes for coat colour determination were then assigned GO terms using the PANTHER Classification System (Mi *et al*., 2019). The analyses were conducted using the Gene List Analysis tools available on the PANTHER website (http://www.pantherdb.org) and the GO-slim component terms for molecular function, biological processes and cellular components were retrieved for each of these putative candidate genes.

3.2.5. Gene-Gene Interaction Analysis of Candidate Loci

All possible combinations of the putative pigment-related candidate genes were assessed to determine whether a possible interaction exists between any of these genes. The software UNPHASED version 3.1.7 (Dudbridge, 2008) was used to test these combinations by performing gene-gene interaction analyses using the genotype tests option. A 1000 permutations were run to allow for multiple testing. Gene interactions were considered to be significant for p-value ≤ 0.05.

3.3. Results

3.3.1. Association Analysis

A total of 374 SNPs were significantly associated ($p \leq 0.001$) with coat colour in blue wildebeest (Appendix B). In addition three SNPs with an association p-value ≤ 0.05 mapped to genes occurring in the ESPCR pigment gene database and were, therefore, also considered to be significantly associated with coat colour in blue wildebeest. The descriptive statistics for the 377 statistically significant SNP are summarised in Appendix C, while the descriptive statistics for the SNPs located in the individually identified putative candidate genes are presented in Table 3.1.

3.3.2. Assessment of Linkage Disequilibrium

Both the D' and r^2 value for all possible pairwise SNP combinations varied between 0 and 1. Based on the LD categories defined by Espigolan *et al*. (2013), 69 219 (97.66%) pairwise SNP combinations were classified as exhibiting low LD, 1 642 (2.32%) were classified as exhibiting moderate LD and 15 (0.02%) were classified as exhibiting high LD. The average D' estimated for all pairwise SNP combinations was 0.39 [standard deviation (SD) = 0.27)], while the average r^2 was 0.04 (SD = 0.05). Furthermore, based on the results of the probability tests, 34.16% of the 70 876 possible SNP combinations showed significant LD ($p \leq 0.05$).

3.3.3. Functional Annotation and Identification of Candidate Loci

Before identification of the individual candidate genes, the marker sequences of all the statistically significant SNPs were functionally annotated using the Blast2GO pipeline to gain an overview of the functional processes the genes, in which these SNPs are situated, are involved in. In total, only 37 SNP sequences of the 377 statistically significant SNPs could be annotated and grouped into one or more of the three main Blast2GO Gene Ontology categories: (i) molecular functions, (ii) biological processes and (iii) cellular components (Figure 3.1). Among molecular functions, “binding” represented the most abundant category (22 sequences, 39%), followed by “catalytic activity” (16 sequences, 28%) and “transporter activity” (10 sequences, 18%). Among the biological processes, 19% (23 sequences) were annotated in both the “cellular process” and “metabolic process” categories and 15% (18 sequences) in the “regulation of biological process” and “biological regulation” categories. Single nucleotide polymorphism sequences associated with the “membrane” (22 sequences, 26%) and “membrane part” (17 sequences, 20%) represented the most dominant groups of the cellular component terms

Table 3.1. Summary statistics of the case-control association analysis for the SNPs associated with putative coat colour candidate genes.

SNP ID	Alleles (Major/Minor Allele)	Blast E-Value	Gene	MAF			Heterozygosity						HWE p-value			Association p-value	OR (95% CI)		Best-Fitting Model of Inheritance
				Case	Control	Total Cohort	Observed (Case)	Expected (Case)	Observed (Control)	Expected (Control)	Observed (Total Cohort)	Expected (Total Cohort)	Case	Control	Total Cohort				
100028101|F|0--43:C>A	C/A	2.63E-03	*MYO5C*	0.31	0.42	0.48	0.40	0.43	0.54	0.49	0.49	0.50	0.67	0.40	0.85	5.46E-04**	A/C: 2.48 (0.62 – 10.02)	C/C: 11.33 (2.62 – 49.05)	Additive
100076096|F|0--31:A>G	A/G	3.20E-02	*MYO7A*	0.41	0.00	0.16	0.24	0.48	0.00	0.00	0.09	0.27	2.71E-03*	1.00	7.68E-10**	2.94E-08**	A/G: 2.63E+09 (1.69E+09 – 4.10E+09)	G/G: 5.12E+10 (3.35E+10 – 7.83E+10)	Dominant/ Additive
100025667|F|0--39:T>C	T/C	4.16E-07	*SLC6A3*	0.00	0.43	0.27	0.00	0.00	0.59	0.49	0.37	0.40	1.00	0.19	0.57	1.65E-12**	C/T: 2.62E+07 (2.62E+07 – 2.62E+07)	T/T: 2.10E+17 (2.10E+17 – 2.10E+17)	Dominant/ Additive
100071798|F|0--48:A>G	A/G	4.16E-07		0.50	0.28	0.36	0.59	0.50	0.24	0.40	0.37	0.46	0.30	2.57E-03*	0.06	1.85E-04**	A/G: 8.75 (2.82 – 27.17)	G/G: 5.10 (1.35 – 19.37)	Dominant
100048574|F|0--35:G>A	G/A	3.20E-02	*SLC28A2*	0.09	0.32	0.23	0.17	0.16	0.44	0.44	0.34	0.36	0.58	0.94	0.62	9.86E-04**	A/G: 1.71E+08 (9.86E+07 – 2.97E+08)	G/G: 8.21E+08 (4.73+08 – 1.42E+09)	Additive

SNP ID	Alleles (Major/Minor Allele)	Blast E-Value	Gene	MAF			Heterozygosity						HWE p-value			Association p-value	OR (95% CI)		Best-Fitting Model of Inheritance
				Case	Control	Total Cohort	Observed (Case)	Expected (Case)	Observed (Control)	Expected (Control)	Observed (Total Cohort)	Expected (Total Cohort)	Case	Control	Total Cohort				
100075122|F|0--28:G>A	G/A	2.63E-03	*DRD2*	0.33	0.17	0.23	0.31	0.44	0.26	0.28	0.28	0.35	0.09	0.69	0.06	3.64E-02*	A/G: 0.24 (0.04 – 1.45)	G/G: 0.14 (0.02 – 0.75)	Additive
100028737|F|0--18:G>A	G/A	3.20E-02	*FZD4*	0.47	0.24	0.32	0.49	0.50	0.37	0.36	0.42	0.44	0.88	0.82	0.60	4.03E-03*	A/G: 0.28 (0.06 – 1.24)	G/G: 0.10 (0.02 – 0.47)	Additive
100070025|F|0--27:C>T	C/T	7.52E-04	*TYR*	0.35	0.13	0.21	0.21	0.45	0.14	0.22	0.17	0.33	2.00E-03*	0.01*	4.11E-06**	9.64E-03*	C/T: 2.45 (0.76 – 7.92)	T/T: 7.47 (1.75 – 31.90)	Additive

*statistical significance at p-value cut-off of 0.05; ** statistical significance at p-value cut-off of 0.001.

Abbreviations: Confidence interval (CI); dopamine receptor D2 (DRD2); frizzled class receptor 4 (FZD4); Hardy-Weinberg equilibrium (HWE); minor allele frequency (MAF); myosin VC (MYO5C); myosin VIIA (MYO7A); odds ratio (OR); single nucleotide polymorphism (SNP); solute carrier family 6 member 3 (SLC6A3); solute carrier family 28 member 2 (SLC28A2); tyrosinase (TYR).

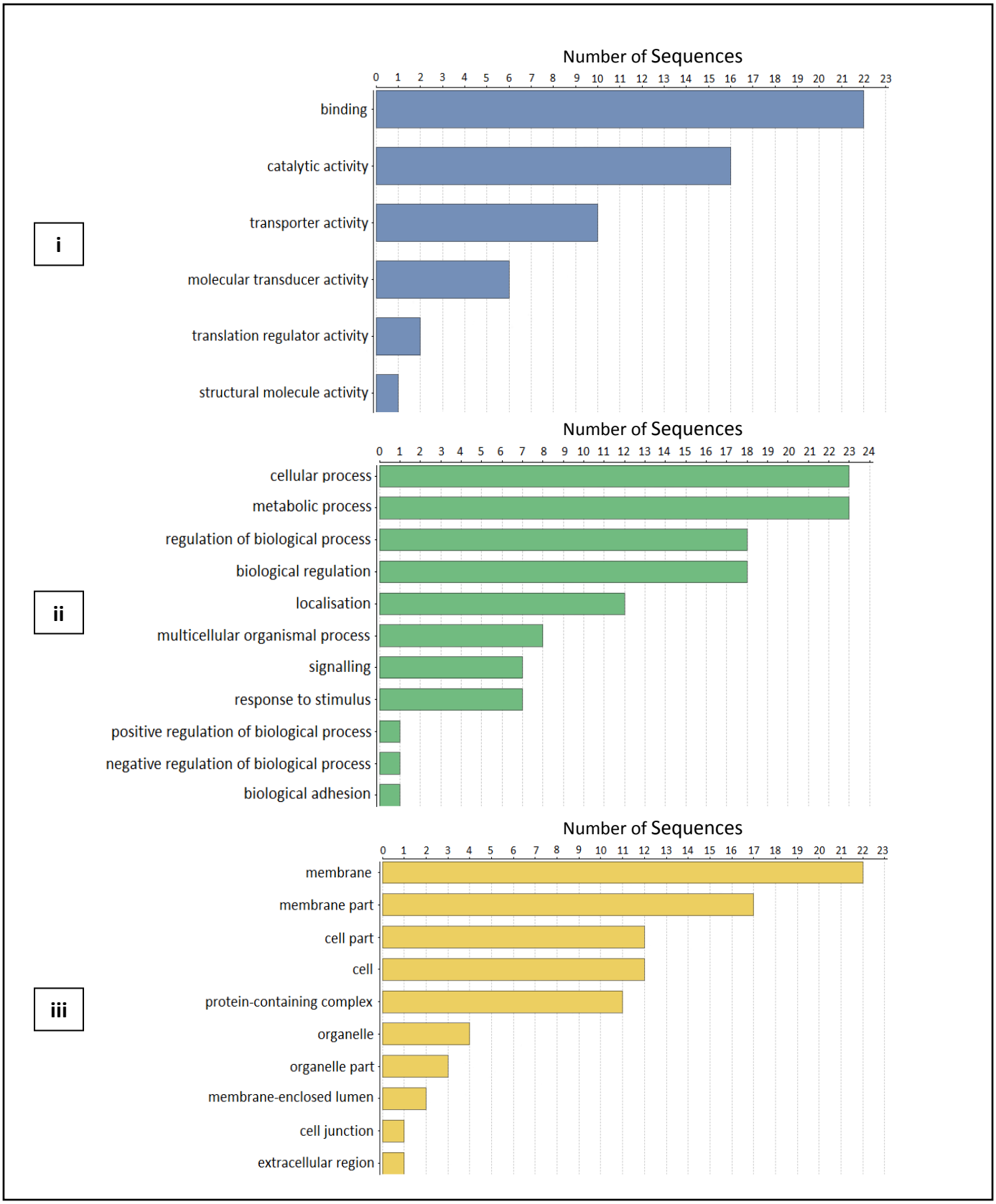


Figure 3.1. Gene ontology distribution of all single nucleotide polymorphisms (SNPs) significantly associated with coat colour in blue wildebeest. The results are summarised as follows: (i) molecular functions, (ii) biological processes, (iii) cellular components.

The marker sequences of the statistically significant SNPs ($p \leq 0.001$) were then matched, where possible, to the *B. taurus* orthologous genes previously identified in blue wildebeest. Of these significantly associated SNPs, 150 could be mapped the bovine genome. Bovine chromosomes 4, 5 and X were the chromosomes to which the most SNPs mapped. Of the 150 SNPs that mapped to the bovine genome, only 101 mapped to genes with known protein products.

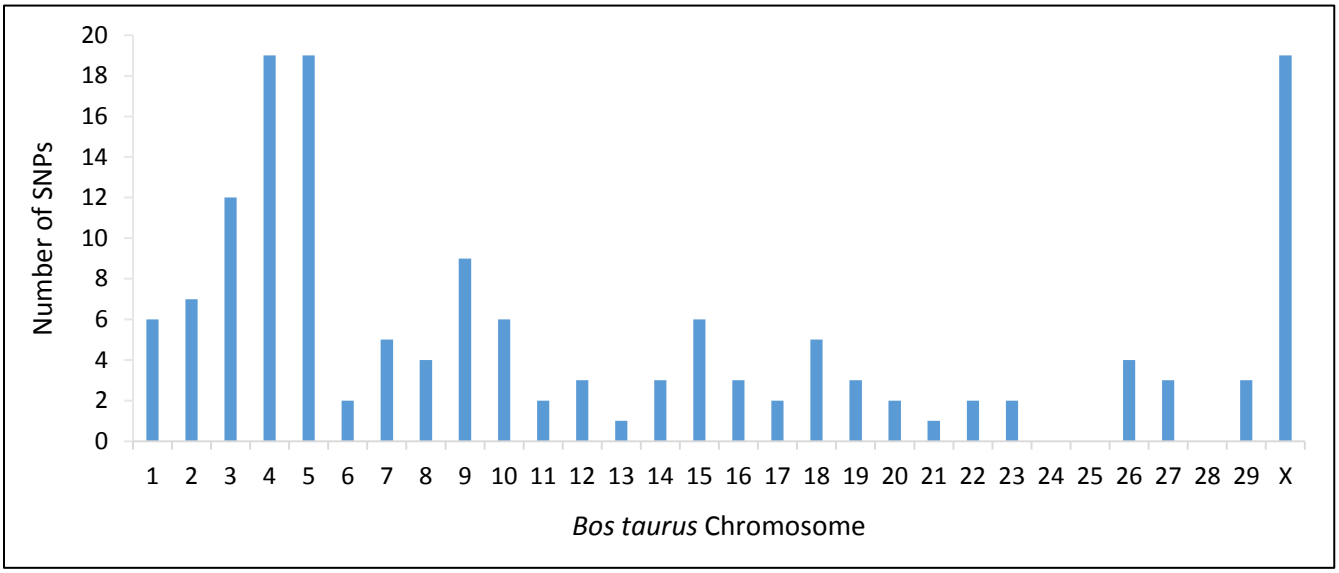


Figure 3.2. Distribution of the single nucleotide polymorphisms (SNPs), significantly associated with coat colour in blue wildebeest (p-value ≤ 0.001), across the *Bos taurus* chromosomes.

Each of these orthologous genes were further investigated to determine if they could be involved in pigmentation, and a total of four genes [*myosin VC* (*MYO5C*), *myosin VIIA* (*MYO7A*), *solute carrier family 6 member 3* (*SLC6A3*) and *solute carrier family 28 member 2* (*SLC28A2*)] were identified as putative candidate loci for coat colour determination in blue wildebeest. An additional three orthologous genes [*dopamine receptor D2* (*DRD2*), *frizzled class receptor 4* (*FZD4*) and *tyrosinase* (*TYR*)] previously identified in blue wildebeest occurred in the ESPCR database of pigmentation genes and were also considered as putative candidate genes for coat colour determination, since their corresponding SNPs had a p-value ≤ 0.05 for the association study. A total of seven genes were, thus, identified as putative candidate genes that could determine coat colour in blue wildebeest. The eight SNPs located in these genes are presented in Table 3.1. Except for SNP 100071798|F|0--48:A>G which has a dominant mode of inheritance according to the AIC and BIC values, and SNP 100025667|F|0--39:T>C and 100076096|F|0--31:A>G which could either have a dominant or additive mode of inheritance, all other SNPs located in the putative coat colour candidate loci fitted an additive model of inheritance best.

The GO component terms, retrieved using the PANTHER classification system, for the individually identified putative coat colour candidate genes (SNP with $p \leq 0.001$) are presented in Table 3.2. These identified candidate genes are mainly involved in myosin-dependent transport and the transmembrane transport of ions. Table 3.3 presents the GO terms retrieved for the putative candidate genes for which the associated SNP had an association p-value ≤ 0.05. These genes were identified as candidate genes with the prerequisite that they had to occur in the ESPCR's pigment gene database. The involvement of these genes in pigmentation has thus been studied previously.

3.3.4. Gene-Gene Interaction Analysis of Candidate Loci

Coat colour pigmentation is a polygenic trait and many of the coat colour-associated genes and their alleles often have epistatic interactions (Sturm *et al.*, 2001). In order to elucidate some of the gene interactions that influence coat colour variability in blue wildebeest, the SNPs located in all of the identified coat colour candidate loci were tested against each other for a possible gene-gene interaction. However, none of these interactions were found to be statistically significant, since all of the interaction p-values were > 0.05.

Table 3.2. Gene ontology (GO) terms of the individually identified putative coat colour candidate genes for which the associated SNP had a p-value of ≤ 0.001 for the association study.

SNP ID	Orthologous *Bos taurus* Gene	*Bos taurus* Protein Product	GO Classification			Possible Role in Pigmentation
			Molecular Function	Biological Process	Cellular Component	
100028101|F|0--43:C>A	*MYO5C*	Myosin VC	Enzyme regulator activity, Motor activity, Protein binding, Structural constituent of cytoskeleton	Vesicle-mediated transport, Cellular component movement, Cellular component morphogenesis, Intracellular protein transport, Intracellular signal transduction, Cytokinesis, Mitosis, Mesoderm development, Muscle contraction, Sensory perception	Actin cytoskeleton, Cell junction, Intracellular, Plasma membrane	Myosin VC is involved in the actin-dependant trafficking of integral membrane proteins to melanosomes (Bultema *et al.*, 2014).
100076096|F|0--31:G>A	*MYO7A*	Myosin VIIA	Enzyme regulator activity, Motor activity, Protein binding, Structural constituent of cytoskeleton	Anatomical structure morphogenesis, Cell cycle, Intracellular protein transport, Intracellular signal transduction, Mesoderm development, Muscle contraction, Sensory perception, Vesicle-mediated transport	Intracellular, Plasma membrane	The myosin VIIA protein is responsible for the transport of melanosomes in retinal pigment epithelium (RPE; Williams & Lopes, 2011).

SNP ID	Orthologous *Bos taurus* Gene	*Bos taurus* Protein Product	GO Classification			Possible Role in Pigmentation
			Molecular Function	Biological Process	Cellular Component	
100025667|F|0--39:T>C 100071798|F|0--48:A>G	*SLC6A3*	Transporter (sodium/ chloride transporter)	Cation transmembrane transporter activity, Neurotransmitter transporter activity	Biological regulation, Neuron-neuron synaptic transmission, Neurological system process	Integral component of plasma membrane, Neuronal cell body, Cell surface, Axon, Flotillin complex	The family of solute carriers is predicted to play an important role in the regulation of melanosomal pH (Kondo & Hearing, 2011). Melanosomal pH regulates the activity of tyrosinase and the melanin production rate (Ancans *et al.*, 2001).
100048574|F|0--35:G>A	*SLC28A2*	Sodium/ nucleoside cotransporter	Cation transmembrane transporter activity	Cellular process, Nucleobase-containing compound transport	Integral to membrane, Plasma membrane	

Table 3.3. Gene ontology (GO) terms of the coat colour candidate genes, with a known functional role in pigmentation, for which the associated SNP had a p-value of ≤ 0.05 for the association study.

SNP ID	Orthologous *Bos taurus* Gene	*Bos taurus* Protein Product	GO Classification			Role in Pigmentation
			Molecular Function	Biological Process	Cellular component	
Melanocyte and Keratinocyte Development						
100028737|F|0--18:G>A	*FZD4*	Frizzled (FZD) class receptor 4	Signal transducer activity, Protein binding, G-protein coupled receptor activity	Response to stimulus, Regulation of biological process	Integral to membrane	Wnt signalling is mediated by the family of Frizzled receptors and their co-receptors, low density lipoprotein receptor-related proteins (LRP) 5/6 (Yamada *et al.*, 2013).
Components of Melanosomes and their Precursors						
100070025|F|0--27:C>T	*TYR*	Tyrosinase	Metal ion binding, Oxidoreductase activity, Protein homo-dimerisation activity, Protein hetero-dimerisation activity, Monophenol monooxygenase activity	Cell proliferation, Pigmentation, Melanin biosynthetic process, Thymus development, Oxidation-reduction process	Melanosome membrane, Intracellular membrane-bond organelle, Perinuclear region of cytoplasm, Integral component of membrane	The tyrosinase protein hydroxylates tyrosine in the first step of melanin synthesis. This step is crucial for the formation of melanosomes (D'Mello *et al.*, 2016)

SNP ID	Orthologous *Bos taurus* Gene	*Bos taurus* Protein Product	GO Classification			Role in Pigmentation
			Molecular Function	Biological Process	Cellular component	
Melanin Synthesis and Switching						
100075122|F|0--28:G>A	*DRD2*	Dopamine receptor D2	Adenylate cyclase activity, G protein-coupled receptor activity, Signal transducer activity	Cell-cell signalling, Behaviour, Biosynthetic process, Intracellular signal transduction, Cyclic nucleotide metabolic process, Nitrogen compound metabolic process, Regulation of catalytic activity, Regulation of phosphate metabolic process, Regulation of nucleobase-containing compound metabolic process	Plasma membrane, Integral to membrane	The D2 dopamine receptor plays a critical role in the transcription of proopiomelanocortin (POMC), the precursor of α-melanocyte stimulating hormone (αMSH; Yamaguchi *et al.*, 1996; Lao *et al.*, 2007).

3.4. Discussion

In blue wildebeest, coat colour is a trait of great economic importance. The golden coat colour variant of the blue wildebeest is a very sought after game colour morph, often selling for a higher price than its wild-type counterpart (Taylor *et al.*, 2016). However, concerns have been raised regarding the breeding of game colour variants, since certain mutations in genes influencing coat colour in mammals have also been shown to have direct pleiotropic effects related to health (Reissmann & Ludwig, 2013; Charon & Lipka, 2015). Therefore, to elucidate the genetic underpinnings of coat colour in blue wildebeest, a GWAS was conducted with 14 624 SNPs. A total of 377 of these SNPs were found to be significantly associated with coat colour in blue wildebeest.

3.4.1. The Genetic Underpinning of Coat Colour in Blue Wildebeest

Only 37 of the significant SNP marker sequences could be functionally annotated. The SNP sequences that could not be functionally annotated could either be located in genes that have not been fully annotated yet or could be located in non-coding regions. Studies have shown that SNPs located in non-coding regions could also have functional consequences by altering the binding site of transcriptional machinery (Zhang & Lupski, 2015; Spielmann & Mundlos, 2016). Therefore, some of the SNPs that were not annotated could still influence coat colour in blue wildebeest by affecting the regulation of pigment gene expression. Moreover, marker or tag SNPs in high LD with the causal variant could also show statistical association with coat colour if the actual functional variant has not been directly genotyped (Bush & Moore, 2012). Wildlife ranchers often make use of line breeding (*i.e.* the selective breeding of animals within a closely related line in order to obtain a specific desired phenotype) to increase the number of colour variant offspring in their captive stock (Van Rooyen, 2012; Maiwashe, 2015). The natural decay of LD occurs at a much slower rate in inbreeding systems, such as line breeding, because there is a severe decrease in effective recombination and, consequently, genetic variants remain correlated over larger physical distances (Caldwell, 2016). In order to determine if the large number of SNPs that could not be annotated could indeed be tag SNPs of the causal variant, the level of LD (measured by D' and r^2) were estimated between all significant SNP combinations. Generally, the pattern of LD between neighbouring markers are high and decreases with an increase in marker distance (Berihulay *et al.*, 2019). However, since the positions of the significant SNPs relative to each other were not known in the current study, the LD between these markers could not be measured as a function of distance. As a result, the average D' and r^2 were estimated over all possible pairwise SNP combinations. Therefore, although the average D' and r^2 values seem low, it should be compared to

the average LD observed across all chromosomes in other species. In the Brazilian Santa Inês sheep breed (*Ovis aries*), the average D' and r^2 between all pairwise SNPs were 0.225 and 0.018, respectively (Alvarenga *et al*., 2018). Furthermore, 0.01% of all SNPs combinations were reported as being in high LD and 0.85% were reported as being in moderate LD (compared to 0.02% in high LD and 2.32% in moderate LD in the current study). In addition, a study by Al-Mamun *et al*. (2015) in domestic sheep breeds reported an average D' of 0.17 (Merino), 0.29 (Border Leicester) and 0.27 (Poll Dorset), and an average r^2 of 0.01 (Merino and Border Leicester) and 0.02 (Poll Dorset). Furthermore, in Nellore beef cattle, the average D' and r^2 over all marker pairs were 0.52 and 0.17, respectively (Espigolan *et al*., 2013), while in Sahiwal dairy cattle the average D' and r^2 were 0.55 and 0.18, respectively (Mustafa *et al*., 2018). The LD observed in the current study is thus higher than that generally observed in farmed domestic sheep, but lower than that observed in cattle. Based on these comparisons, the number of tag SNPs in the current study is expected to be slightly higher than the number of tag SNPs in sheep association studies, but less than the number in cattle association studies. The percentage pairwise SNP found to be in significant LD (34.16%), based on the probability test p-values, was also significantly more than the 14.58% reported in water buffalo (*Bubalus bubalis;* Nagarajan *et al*., 2009). In addition, although the significant SNPs were distributed throughout the *B. taurus* genome, more than a third (37.25%) of the mapped SNPs mapped to bovine chromosome 4, 5 and X, suggesting that these SNPs could possibly be physically linked.

Despite the large number of SNP sequences that could not be functionally annotated, the Blast2GO functional analysis is able to give an overview of the molecular functions, biological processes and cellular components involved in blue wildebeest pigmentation. Interestingly, the most prevalent cellular component GO term was “membrane”, while binding and transporter activities also seem to play an important role in blue wildebeest pigmentation. Consequently, binding proteins and membrane transport are expected to be involved in the determination of coat colour in blue wildebeest. Many previous studies have identified genes coding for ion transport proteins as key regulators of melanin synthesis (*e.g.* Smith *et al*., 2004; Bellono *et al*., 2016; Chao e*t al*., 2017). Ion transport across membranes has been suggested to serve as ionic messengers that could modulate pigmentation, while the subtle changes in membrane voltage and luminal pH are believed to regulate enzyme activity and organellar biogenesis (Bellono & Oancea, 2014).

Furthermore, a total of seven different genes were identified as putative candidate genes for coat colour determination in blue wildebeest. These identified candidate genes have a wide range of functions related to

pigmentation, including the regulation of melanogenesis, the transport and transfer of structural proteins and melanin, as well as a functional role in Wnt signalling.

3.4.1.1. Regulation of Melanogenesis

3.4.1.1.1. Tyrosinase Activity

Tyrosinase is the rate-limiting enzyme in melanognesis (D'Mello *et al*., 2016). The SNP that mapped to the *B. taurus TYR* gene (100070025|F|0--27:C>T) could thus be involved in the determination of coat colour in blue wildebeest. Previous studies have reported that the required level of TYR activity is higher for eumelanin synthesis than for pheomelanin synthesis (Burchill *et al.*, 1993). As mentioned previously, the coat colour observed in golden wildebeest is most likely the result of increased yellow/red pheomelanin synthesis. If the SNP in the *TYR* gene results in reduced TYR activity, it might shift melanogenesis towards pheomelanin synthesis.

The SNP that mapped to *TYR* deviated from Hardy-Weinberg equilibrium (HWE) in the golden wildebeest case group, as well as in the control group. There are several reasons why a population might deviate from HWE for a specific marker such as inbreeding, population substructure, purifying selection, genotyping error or copy number variation (Chen *et al*., 2017). In this case, the deviation from HWE is likely due to the selection of case and control individuals based on phenotype, since the minor T-allele of the SNP was significantly increased in the case group (0.35) compared to the control group (0.13). Furthermore, the genotype distribution of the SNP fitted an additive model of inheritance best, with each addition of the minor *T*-allele increasing the probability of the resulting coat colour phenotype being golden (C/T genotype: OR = 2.45; 95% CI: 0.76 – 7.92; T/T genotype: OR = 7.47; 95% CI: 1.75 – 31.90).

Mutations in the *TYR* gene has been associated with a paler coat colour, due to reduced melanin production, and pleiotropic effects such as retinal functional abnormalities, nystagmus and behavioural changes (Reissmann & Ludwig, 2013). Such pleiotropic effects have not been reported in golden wildbeest, however, further investigation is needed to ensure that such negative effects do not occur in animals with the golden coat colour phenotype.

3.4.1.1.2. Melanosomal pH

Three SNPs that mapped to members of the *solute carrier* (*SLC*) superfamily were found to be associated with coat colour in blue wildebeest (Table 3.2). Members of this gene family encode membrane-bound transporters that facilitate the movement of specific substrates, either against or with its concentration gradient (He *et al.*, 2009). A number of studies have found mutations in *SLC* genes to affect pigmentation. For example, a missense mutation in *solute carrier family 36 member 1* (*SLC36A1*) is responsible for the champagne dilution phenotype in horses, whereas a mutation in *solute carrier family 45 member 2* (*SLC45A2*) is responsible for the cream dilution phenotype (Cook *et al.*, 2008; Cieslak *et al.*, 2011). Furthermore, mutations in *SLC45A2* cause plumage colour variation in Japanese quail and chicken (Gunnarsson *et al.*, 2007). In humans, *SLC45A2* has been associated with skin colour variation (Graf *et al.*, 2005), while a similar missense mutation in *solute carrier family 24 member 5* (*SLC24A5*) has resulted in decreased melanin content among humans of European ancestry (Lamason *et al.*, 2005). S*olute carrier family 7 member 11* (*SLC7A11*) has also been found to control the production of pheomelanin directly in cultured cells (Chintala *et al.*, 2005).

As mentioned previously, ion exchange might play a pivotal role in the regulation of pigmentation through the regulation of melanosomal pH (Cheli *et al.*, 2009). A study by Ancans *et al.* (2001) has shown that melanosomal pH is an essential factor that regulates melanin synthesis. The activity of tyrosinase is optimal at a neutral pH (Ancans *et al.*, 2001; Cheli *et al.*, 2009). However, melanosomes, which are lysosome-related organelles, have an acidic pH. Therefore, melanogenesis is generally stimulated by an increase in melanosomal pH, while the activity of tyrosinase is gradually lost with decreasing pH (Cheli *et al.*, 2009). Melanosomal pH is thus able to control the rate of melanogenesis. Based on these findings, a SNP in one of the *SLC* candidate genes could possibly result in a change in melanosomal pH. This change in melanosomal pH could, in turn, result in a change in tyrosinase activity. Although, the gene-gene interaction analysis did not identify a significant interaction between the SNPs located in the *SLC* genes and the SNP that mapped to *TYR* (SNP 100070025|F|0--27:C>T), it does not undoubtedly prove that a *SLC* gene is not involved in the regulation of tyrosinase in blue wildebeest. Previous studies have shown that melanin synthesis does not always correlate with the level of *TYR* expression, and that tyrosinase is also regulated post-transcriptionally. The subsequent processing of tyrosinase to produce a mature, functional enzyme has been found to depend on the neutralisation of pH in the Golgi (Watabe *et al.*, 2004). A change in pH caused by a SLC protein could, thus, possibly affect the maturation or functionality of tyrosinase after translation, which in turn, could affect the rate of melanogenesis. Furthermore, an increase in melanosomal pH has also shown a preferential increase in the production of eumelanin (Ancans *et al.*, 2001), while a slightly acidic pH promotes

pheomelanogenesis (Wakamatsu *et al.*, 2017). A change in melanosomal pH could, thus, also alter the ratio of eumelanin to pheomelanin in wildebeest. Studies by Ito *et al.* (2013) and Wakamatsu *et al.* (2017) found that an slightly acidic pH chemically shifts mixed melanogenesis to more pheomelanic states by suppressing the late stages of eumelanogenesis, that occur after the stages catalysed by tyrosinase. Consequently, a polymorphism in a *SLC* gene could result in a slightly acid pH, which in turn could contribute to the shift in melanogenesis towards the synthesis of pheomelanin observed in golden wildebeest.

For SNP 100048574|F|0--35:G>A (*SLC28A2*), each addition of the G-allele increased the odds of the resulting coat colour phenotype being golden (A/G genotype: OR = 1.71E+08; CI: 9.86E+07 – 2.97E+08; G/G genotype: OR = 8.21E+08; CI: 4.73E+08 – 1.42E+09). Furthermore, for SNP 100025667|F|0--39:T>C (*SLC6A3*), none of the golden wildebeest case individuals were carriers of the minor C-allele, and each addition of the T-allele also increased the odds of the coat colour being golden (C/T genotype: OR = 2.62E+07; 95% CI: 2.62E+07 – 2.62E+07; T/T genotype: OR = 2.10E+17; CI: 2.10E+17 – 2.10E+17). In addition, both of these SNPs had very large genetic size effects and are thus expected to greatly influence coat colour in blue wildebeest. It is, however, important to note that the size effects of these SNPs could possibly be overestimated due to the small study population. Furthermore, both of these SNPs thus fit an additive model of inheritance, however, for SNP 100025667|F|0--39:T>C the logistic regression analysis indicated that either an additive or dominant model of inheritance is possible. Furthermore, one of the SNPs that mapped to *SLC6A3*, 100071798|F|0--48:A>G, significantly deviated from HWE in the control group, but not for the case group. This deviation in the control group is likely the result of the lower frequency of the minor G-allele in the control group (0.28) compared to the total cohort (0.36) and the case group (0.50). The major A-allele of this SNP could thus possibly have a “protective” effect against the golden coat colour phenotype. If this SNP results in an increase in melanosomal pH, it could shift melanogenesis towards the synthesis of eumelanin, thereby reducing the odds of the coat colour phenotype being golden.

3.4.1.1.3. Dopamine Receptor D2

The D2 dopamine receptor plays several roles in the pituitary, particularly in the inhibitory control of hormone secretion. The synthesis of αMSH, an agonist of MC1R, takes place in the pituitary where it is derived from POMC (Yamaguchi *et al.*, 1996). Therefore, the D2 dopamine receptor is also involved in the inhibitory regulation of αMSH. By binding to MC1R, αMSH controls the activation of TYR as well as the ratio of pheomelanin to eumelanin (D’Mello *et al.*, 2016). The D2 dopamine receptor could thus be indirectly involved in the regulation of TYR by

regulating αMSH levels. However, the genetic size effect was relatively small (A/G genotype: OR = 0.24; 95% CI: 0.04 – 1.45; G/G genotype: OR = 0.14; 95% CI: 0.02 – 0.75) and is thus not expected to greatly influence coat colour in blue wildebeest.

3.4.1.2. Transport and Transfer of Proteins and Melanin

Melanosomes are produced by the endosomal transport systems around the nucleus of melanocytes (Ohbayashi & Fukuda, 2012). Numerous proteins are involved in the sorting and trafficking of specific enzymes and structural proteins to melanosomes during melanin synthesis (Yamaguchi *et al.*, 2007). As the constituent proteins of melanosomes are delivered and the melanosomes become mature, the organelles themselves become cargos transported by various molecular motors from the perinuclear area to the cell periphery (Coudrier, 2007; Yamaguchi *et al.*, 2007). Thereafter, mature melanosomes are transferred to neighbouring keratinocytes (Yamaguchi *et al.*, 2007). All of these intra- and intercellular transport systems are required for efficient pigmentation in mammals. Mutations in the genes coding for proteins involved in the pigmentation transport system will typically result in a hypopigmentation.

The actin cytoskeleton is essential to the proper distribution of, and communication between, cellular compartments. Some actin-based myosin motors have been involved in this process. Two SNPs (100028101|F|0--43:C>A; 100076096|F|0--31:G>A) that mapped to genes coding for myosin motors, namely myosin VC and myosin VIIA, were found to be associated with coat colour in blue wildebeest in the current study. The frequency of the C/C-homozygous genotype for the myosin VC SNP (100028101|F|0--43:C>A) was significantly increased in the case group (0.53) compared to the control group (0.16). Carriers of the C-allele were also found to be significantly associated with a higher probability to result in the golden coat colour phenotype compared to the A/A homozygotes (OR = 2.48; 95% CI: 0.62 - 10.02). The Myosin Vs are well known for their role in the pigmentation of skin and hair. Myosin VC, specifically, has been found to function in the trafficking of integral proteins, such as TRP1 and TRRP2, to melanosomes (Bultema *et al.*, 2014). Tyrosinase-related protein 1 and TRP2 play crucial roles in catalysing eumelanin-producing reactions. Therefore, if the C-allele of the SNP associated with *MYO5C* results in a transport protein with reduced functional activity compared to the wild-type protein, it may disrupt the synthesis of eumelanin in blue wildebeest.

The SNP that mapped to myosin VII (100076096|F|0--31:A>G) significantly deviated from HWE in the case group, but not in the control group. Typically, HWE does not need to hold for the case group since the group consists of

non-randomly selected individuals based on phenotype (Namipashaki *et al*., 2015). The deviation observed in the case group is likely the result of the increased frequency of the minor G-allele (0.41) compared to the control group which is homozygous for the major A-allele. The G-allele of this SNP could thus also possibly disrupt the synthesis or distribution of eumelanin. Studies have found myosin VII to be involved in the transport of retinal melanosomes and, therefore, also in the determination of eye pigmentation (El-Amraoui *et al*., 2002; Williams & Lopes, 2011). Although the role of myosin VII has not been directly established in the determination of coat colour, the possibility of myosin VII playing a role in coat colour determination cannot be rejected. Additional studies are needed to determine the functional role of myosin VII in determining coat colour.

3.4.1.3. Wnt Signalling

The process of pigment cell development plays a crucial role in the determination of mammalian coat colouration. Previous studies have shown that Wnt signalling plays an important role in melanocyte development (Guo *et al.*, 2012; D'Mello *et al.*, 2016). The *FZD4* candidate gene identified in the current study can possibly contribute to melanocyte development by acting in Wnt signalling. Wnt signalling is mediated by the binding of Wnt molecules to FZD4 and its co-receptors, LRP 5 and 6 (Guo *et al.*, 2012), which induces various cellular events, such as differentiation, proliferation, migration and adhesion (Yamada *et al.*, 2013). Besides playing a critical role in melanocyte development, studies have shown that the Wnt signalling pathway is also involved in regulating TYR activity. Positive regulation of Wnt/β-catenin signalling results in increased TYR activity, while studies have reported that inhibition of Wnt/β-catenin signalling greatly inhibits melanin synthesis (Yamaguchi *et al.*, 2008; Guo *et al.*, 2012). The SNP located in the *FZD4* gene (100028737|F|0--18:G>A) is, however, not expected to have a large effect on coat colour determination in blue wildebeest since the genetic effect size of the genotypes were relatively small (A/G genotype: OR = 0.28; 95% CI: 0.06 - 1.24; G/G genotype: OR = 0.10; 95% CI: 0.02 - 0.47).

3.4.2. Mode of Inheritance

As mentioned previously, the prevailing hypothesis, based on parentage records, is that the golden coat colour phenotype in blue wildebeest is a simple autosomal recessive Mendelian trait. However, based on the results of the current study, coat colour in blue wildebeest has a complex mode of inheritance with multiple genes, with different size effects, influencing the coat colour phenotype. The quantitative threshold model is commonly used to explain how multiple genes can influence the expression of binary traits (Moorad & Promislow, 2011). Under

the threshold model, the expressed phenotype for a dichotomous trait is wholly determined by an underlying, unobserved continuous trait referred to as "liability". Each of the loci involved in the phenotypic expression of the trait act additively on the scale of liability. Individuals with liability that exceeds a fixed threshold will express one phenotypic character state (affected), while individuals with liability below the threshold will express the alternate state (unaffected; Moorad & Linksvayer, 2008; Revell, 2014). Based on the results of the current study, coat colour in blue wildebeest could thus be a quantitative threshold trait mainly determined by *SLC6A3*, *SLC28A2* and *MYO7A*, since the SNPs located in these genes had the largest genetic size effects. In addition, a few minor genes with smaller size effects (*MYO5C*, *DRD2*, *FZD4*, *TYR*) are also expected to contribute to the trait. This particular inheritance model for coat colour in blue wildebeest is also supported by the gene-gene interaction analysis results which indicated that the trait is more likely to be additive across loci, with limited epistatic gene interactions. In addition, the majority of the putative candidate loci also fitted an additive mode of inheritance best, providing further support for the threshold inheritance model for coat colour in blue wildebeest.

This particular mode of inheritance for coat colour in a mammalian species has not been previously reported in published literature. It is possible that previous studies regarding coat colour inheritance in mammals may have overlooked this specific inheritance model due to study design or the use of a limited number of molecular markers. Nevertheless, due to the novelty of the findings in this preliminary study of coat colour in blue wildebeest, additional studies are required to confirm the mode of inheritance.

3.5. Conclusion

The current study represented the first genetic study to identify genes and SNPs associated with coat colour in blue wildebeest. Eight DArTseq SNP markers were identified as putative candidate loci for coat colour determination in blue wildebeest. These SNPs mapped to seven different candidate genes: *TYR*, *SLC6A3*, *SLC28A2* *MYO5C*, *MYO7A*, *DRD2* and *FZD4*. Of these identified candidate genes, the *TYR* gene is the only gene, known from literature, to be associated with pleiotropic effects. However, none the negative effects associated with mutations in the *TYR* gene have been reported in golden wildebeest. Of the putative candidate SNPs, 100076096|F|0--31:A>G (*MYO7A)*, 100025667|F|0--39:T>C (*SLC6A3*) and 100048574|F|0--35:G>A (*SLC28A2*) had the largest genetic size effects. Furthermore, based on the gene-gene interaction analysis results and the mode of inheritance of the SNPs located in the putative candidate genes, coat colour in blue wildebeest could possibly be a quantitative threshold trait rather than a simple autosomal recessive trait as originally suspected. The findings of this

preliminary study, therfore, provides a valuable theoretical basis for further research regarding the genetic mechanisms of pigmentation in blue wildebeest.

REFERENCES

Adetunji, S. A., Krecek, R. C., O'Dell, N., Prozesky, L., Steyl, J., & Arenas-Gamboa, A. M. (2018). Melanoma in golden and king wildebeests. *Journal of Zoo and Wildlife Medicine*, *49*(1), 134–142.

Al-Mamun, H. A., Clark, S. A., Kwan, P., & Gondro, C. (2015). Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genetics Selection Evolution*, *47*, 90.

Alvarenga, A. B., Rovadoscki, G. A., Petrini, J., Coutinho, L. L., Morota, G., Spangler, M. L., Pinto, L. F. B., Carvalho, G. G. P., & Mourão, G. B. (2018). Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries*. *Scientific Reports*, *8*, 8851.

Amos, W., Driscoll, E., & Hoffman, J. I. (2011). Candidate genes versus genome-wide associations: Which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society B: Biological Sciences*, *278*(1709), 1183–1188.

Ancans, J., Tobin, D. J., Hoogduijn, M. J., Smit, N. P., Wakamatsu, K., & Thody, A. J. (2001). Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. *Experimental Cell Research*, *268*(1), 26–35.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–265.

Bellono, N. W., Escobar, I. E., & Oancea, E. (2016). A melanosomal two-pore sodium channel regulates pigmentation. *Scientific Reports*, *6*, 26570.

Bellono, N. W., & Oancea, E. V. (2014). Ion transport in pigmentation. *Archives of Biochemistry and Biophysics*, *563*, 35–41.

Berihulay, H., Islam, R., Jiang, L., & Ma, Y. (2019). Genome-wide linkage disequilibrium and the extent of effective population sizes in six Chinese goat populations using a 50K single nucleotide polymorphism panel. *Animals*, *9*(6), E350.

Blackmore, A. C. (2017). Public trust doctrine, research and responsible wildlife management in South Africa. *Bothalia*, *47*(1), a2217.

Bultema, J. J., Boyle, J. A., Malenke, P. B., Martin, F. E., Dell'Angelica, E. C., Cheney, R. E., & Di Pietro, S. M. (2014). Myosin Vc interacts with Rab32 and Rab38 proteins and works in the biogenesis and secretion of melanosomes. *Journal of Biological Chemistry*, *289*(48), 33513–33528.

Burchill, S. A., Ito, S., & Thody, A. J. (1993). Effects of melanocyte-stimulating hormone on tyrosinase expression and melanin synthesis in hair follicular melanocytes of the mouse. *Journal of Endocrinology*, *137*(2), 189–

195.

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Computational Biology*, *8*(12), e1002822.

Caldwell, K. S., Russell, J., Langridge, P., & Powell, W. (2006). Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*, *172*(1), 557–567.

Chao, Y.-K., Schludi, V., Chen, C.-C., Butz, E., Phuong Nguyen, O. N., Müller, M., Krüger, J., Kammerbauer, C., Ben-Johny, M., Vollmar, A. M., Berking, C., Biel, M., Wahl-Schott, C. A., & Grimm, C. (2017). TPC2 polymorphisms associated with a hair pigmentation phenotype in humans result in gain of channel function by independent mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(41), E8595–E8602.

Charon, K. M., & Lipka, K. R. (2015). The effect of a coat colour-associated genes polymorphism on animal health - A review. *Annals of Animal Science*, *15*(1), 3–17.

Cheli, Y., Luciani, F., Khaled, M., Beuret, L., Bille, K., Gounon, P., Ortonne, J.-P., Bertolotto, C., & Ballotti, R. (2009). αMSH and cyclic AMP elevating agents control melanosome pH through a protein kinase A-independent mechanism. *Journal of Biological Chemistry*, *284*(28), 18699–18706.

Chen, B., Cole, J. W., & Grond-Ginsbach, C. (2017). Departure from Hardy Weinberg equilibrium and genotyping error. *Frontiers in Genetics*, *8*, 167.

Chintala, S., Li, W., Lamoreux, M. L., Ito, S., Wakamatsu, K., Sviderskaya, E. V., Bennett, D. C., Park, Y.-M., Gahl, W. A., Huizing, M., Spritz, R. A., Ben, S., Novak, E. K., Tan, J., & Swank, R. T. (2005). *Slc7a11* gene controls production of pheomelanin pigment and proliferation of cultured cells. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(31), 10964–10969.

Cieslak, M., Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Colours of domestication. *Biological Reviews*, *86*(4), 885–899.

Cloete, F. (2017). Did the bubble in the game breeding industry burst? *Wildlife Stud Services: Wildlife Breeders Journal*, 12–13.

Cook, D., Brooks, S., Bellone, R., & Bailey, E. (2008). Missense mutation in exon 2 of SLC36A1 responsible for champagne dilution in horses. *PLoS Genetics*, *4*(9), e1000195.

Coudrier, E. (2007). Myosins in melanocytes: To move or not to move? *Pigment Cell Research*, *20*(3), 153–160.

D'Mello, S. A. N., Finlay, G. J., Baguley, B. C., & Askarian-Amiri, M. E. (2016). Signaling pathways in melanogenesis. *International Journal of Molecular Sciences*, *17*(7), 1144.

Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with

missing genotype data. *Human Heredity*, *66*(2), 87–98.

El-Amraoui, A., Schonn, J.-S., Küssel-Andermann, P., Blanchard, S., Desnos, C., Henry, J.-P., Wolfrum, U., Darchen, F., & Petit, C. (2002). MyRIP, a novel Rab effector, enables myosin VIIA recruitment to retinal melanosomes. *EMBO Reports*, *3*(5), 463–470.

Espigolan, R., Baldi, F., Boligon, A. A., Souza, F. R. P., Gordo, D. G. M., Tonussi, R. L., Cardoso, D. F., Oliveira, H. N., Tonhati, H., Sargolzaei, M., Schenkel, F. S., Calvalheiro, R., Ferro, J. A., & Albuquerque, L. G. (2013). Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics*, *14*, 305.

Fan, Y., Wang, P., Fu, W., Dong, T., Qi, C., Liu, L., Guo, G., Li, C., Cui, X., Zhang, S., Zhang, Q., Zhang, Y., & Sun, D. (2014). Genome-wide association study for pigmentation traits in Chinese Holstein population. *Animal Genetics*, *45*(5), 740–744.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435.

Graf, J., Hodgson, R., & Van Daal, A. (2005). Single nucleotide polymorphisms in the *MATP* gene are associated with normal human pigmentation variation. *Human Mutation*, *25*(3), 278–284.

Grobler, P., Van Wyk, A. M., Dalton, D. L., Van Vuuren, B. J., & Kotzé, A. (2018). Assessing introgressive hybridization between blue wildebeest (*Connochaetes taurinus*) and black wildebeest (*Connochaetes gnou*) from South Africa. *Conservation Genetics*, *19*(4), 981–993.

Gunnarsson, U., Hellström, A. R., Tixier-Boichard, M., Minvielle, F., Bed'hom, B., Ito, S., Jensen, P., Rattink, A., Vereijken, A., & Andersson, L. (2007). Mutations in *SLC45A2* cause plumage color variation in chicken and Japanese quail. *Genetics*, *175*(2), 867–877.

Guo, H., Yang, K., Deng, F., Xing, Y., Li, Y., Lian, X., & Yang, T. (2012). Wnt3a inhibits proliferation but promotes melanogenesis of melan-a cells. *International Journal of Molecular Medicine*, *30*(3), 636–642.

He, L., Vasiliou, K., & Nebert, D. W. (2009). Analysis and update of the human solute carrier (*SLC*) gene superfamily. *Human Genomics*, *3*(2), 195–206.

Hoekstra, H. E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, *97*(3), 222–234.

Hubbard, J. K., Uy, J. A. C., Hauber, M. E., Hoekstra, H. E., & Safran, R. J. (2010). Vertebrate pigmentation: From underlying genes to adaptive function. *Trends in Genetics*, *26*(5), 231–239.

Ito, S., Suzuki, N., Takebayashi, S., Commo, S., & Wakamatsu, K. (2013). Neutral pH and copper ions promote eumelanogenesis after the dopachrome stage. *Pigment Cell and Melanoma Research*, *26*(6), 817–825.

Kondo, T., & Hearing, V. J. (2011). Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Review of Dermatology*, *6*(1), 97–108.

Koseniuk, A., Ropka-Molik, K., Rubiś, D., & Smołucha, G. (2018). Genetic background of coat colour in sheep. *Archives Animal Breeding*, *61*(2), 173–178.

Lamason, R. L., Mohideen, M.-A. P. K., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E., Sinha, S., Moore, J. L., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., McKeigue, P.M., O'Donell, D., Kittles, R., Parra, E. J., Mangini, N. J., Grunwald, D. J., Shriver, M.D., Canfield, V. A., & Cheng, K. C. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, *310*(5755), 1782–1786.

Lao, O., De Gruijter, J. M., Van Duijn, K., Navarro, A., & Kayser, M. (2007). Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Annals of Human Genetics*, *71*(3), 354–369.

Lewis, C. M., & Knight, J. (2012). Introduction to genetic association studies. *Cold Spring Harbor Protocols*, *7*(3), 297–306.

Li, M. H., Tiirikka, T., & Kantanen, J. (2014). A genome-wide scan study identifies a single nucleotide substitution in *ASIP* associated with white versus non-white coat-colour variation in sheep (*Ovis aries*). *Heredity*, *112*(2), 122–131.

Linderholm, A., & Larson, G. (2013). The role of humans in facilitating and sustaining coat colour variation in domestic animals. *Seminars in Cell and Developmental Biology*, *24*(6-7), 583–593.

Maiwashe, N. (2015). Inbreeding *versus* line-breeding in wildlife. *Wildlife Stud Services: Wildlife Breeders Journal*, 22–23.

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. (2019). Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols*, *14*(3), 703–721.

Moorad, J. A., & Linksvayer, T. A. (2008). Levels of selection on threshold characters. *Genetics*, *179*(2), 899–905.

Moorad, J. A., & Promislow, D. E. L. (2011). Evolutionary demography and quantitative genetics: Age-specific survival as a threshold trait. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1702), 144–151.

Mustafa, H., Ahmad, N., Heather, H. J., Eui-Soo, K., Khan, W. A., Ajmal, A., Javed, K., Pasha, T. N., Ali, A., Kim, J. J., & Sonstegard, T. S. (2018). Whole genome study of linkage disequilibrium in Sahiwal cattle. *South African Journal of Animal Science*, *48*(2), 353–360.

Nagarajan, M., Kumar, N., Nishanth, G., Haribaskar, R., Paranthaman, K., Gupta, J., Mishra, M., Vaidhegi, R., Kumar,

S., Ranjan, A. K., & Kumar, S. (2009). Microsatellite markers of water buffalo, *Bubalus bubalis* - Development, characterisation and linkage disequilibrium studies. *BMC Genetics*, *10*, 68.

Namipashaki, A., Razaghi-Moghadam, Z., & Ansari-Pour, N. (2015). The essentiality of reporting Hardy-Weinberg equilibrium calculations in population-based genetic association studies. *Cell Journal*, *17*(2), 187–192.

Nazari-Ghadikolaei, A., Mehrabani-Yeganeh, H., Miarei-Aashtiani, S. R., Staiger, E. A., Rashidi, A., & Huson, H. J. (2018). Genome-wide association studies identify candidate genes for coat color and mohair traits in the Iranian Markhoz Goat. *Frontiers in Genetics*, *9*, 105.

Ohbayashi, N., & Fukuda, M. (2012). Role of Rab family GTPases and their effectors in melanosomal logistics. *Journal of Biochemistry*, *151*(4), 343–351.

Olivier, P. A. (2015). *Colour variation of African Bovidae: Consequences for conservation and the wildlife ranching industry* (Unpublished MSc thesis). University of the Free State, Bloemfontein, South Africa.

Peakall, R., & Smouse, P. E. (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288–295.

Peakall, R., & Smouse, P. E. (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – An update. *Bioinformatics*, *28*(19), 2537–2539.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*(3), 248–249.

Reissmann, M., & Ludwig, A. (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Seminars in Cell & Developmental Biology*, *24*(6), 576–586.

Ren, J., Mao, H., Zhang, Z., Xiao, S., Ding, N., & Huang, L. (2011). A 6-bp deletion in the *TYRP1* gene causes the brown colouration phenotype in Chinese indigenous pigs. *Heredity*, *106*(5), 862–868.

Revell, L. J. (2014). Ancestral character estimation under the threshold model from quantitative genetics. *Evolution*, *68*(3), 743–759.

Rousset, F. (2008). Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106.

Russo, I.-R. M., Hoban, S., Bloomer, P., Kotzé, A., Segelbacher, G., Rushworth, I., Birss, C., & Bruford, M. W. (2019). 'Intentional genetic manipulation' as a conservation threat. *Conservation Genetics Resources*, *11*(2), 237–247.

Smith, D. R., Spaulding, D. T., Glenn, H. M., & Fuller, B. B. (2004). The relationship between Na(+)/H(+) exchanger expression and tyrosinase activity in human melanocytes. *Experimental Cell Research*, *298*(2), 521–534.

Solé, X., Guinó, E., Valls, J., Iniesta, R., & Moreno, V. (2006). SNPStats: A web tool for the analysis of association

studies. *Bioinformatics*, *22*(15), 1928–1929.

Spielmann, M., & Mundlos, S. (2016). Looking beyond the genes: The role of non-coding variants in human disease. *Human Molecular Genetics*, *25*(R2), R157–R165.

Strauss, J. S. (2015). *The long term financial sustainability of the high value game breeding sector in South Africa* (Unpublished MBA thesis). University of the Free State, Bloemfontein, South Africa.

Sturm, R. A., Teasdale, R. D., & Box, N. F. (2001). Human pigmentation genes: Identification, structure and consequences of polymorphic variation. *Gene*, *277*(1–2), 49–62.

Tambling, C., Viljoen, P., & Kotze, A. (2016). A conservation assessment of *Connochaetes taurinus*. In M. F. Child, L. Roxburgh, E. Do Lihn San, D. Raimondo, & H. T. Davies-Mostert (Eds.), *The Red List of Mammals of South Africa, Swaziland and Lesotho*. South African National Biodiversity Institute and Endangered Wildlife Trust, South Africa.

Taylor, W. A., Lindsey, P. A., & Davies-Mostert, H. (2016). *An assessment of the economic, social and conservation value of the wildlife ranching industry and its potential to support the green economy in South Africa*. The Endangered Wildlife Trust, South Africa.

Van Rooyen, L. (2012, July 2). Colour variant game naturally profitable. *Farmer's Weekly*. Retrieved from https://www. farmersweekly.co.za [accessed 5 October 2017].

Wakamatsu, K., Nagao, A., Watanabe, M., Nakao, K., & Ito, S. (2017). Pheomelanogenesis is promoted at a weakly acidic pH. *Pigment Cell and Melanoma Research*, *30*(3), 372–377.

Watabe, H., Valencia, J. C., Yasumoto, K.-I., Kushimoto, T., Ando, H., Muller, J., Vieira, W. D., Mizoguchi, M., Appella, E., & Hearing, V. J. (2004). Regulation of tyrosinase processing and trafficking by organellar pH and by proteasome activity. *Journal of Biological Chemistry*, *279*(9), 7971–7981.

Williams, D. S., & Lopes, V. S. (2011). The many different cellular functions of MYO7A in the retina. *Biochemical Society Transactions*, *39*(5), 1207–1210.

Yamada, T., Hasegawa, S., Inoue, Y., Date, Y., Yamamoto, N., Mizutani, H., Nakata, S., Matsunaga, K., & Akamatsu, H. (2013). Wnt/β-Catenin and kit signaling sequentially regulate melanocyte stem cell differentiation in UVB-induced epidermal pigmentation. *Journal of Investigative Dermatology*, *133*(12), 2753–2762.

Yamaguchi, H., Aiba, A., Nakamura, K., Nakao, K., Sakagami, H., Goto, K., Kondo, H., & Katsuki, M. (1996). Dopamine D2 receptor plays a critical role in cell proliferation and proopiomelanocortin expression in the pituitary. *Genes to Cells*, *1*(2), 253–268.

Yamaguchi, Y., Brenner, M., & Hearing, V. J. (2007). The regulation of skin pigmentation. *Journal of Biological Chemistry*, *282*(38), 27557–27561.

Yamaguchi, Y., Passeron, T., Hoashi, T., Watabe, H., Rouzaud, F., Yasumoto, K.-I., Hara, T., Tohyama, C., Katayama, I., Miki, T., & Hearing, V. J. (2008). Dickkopf 1 (DKK1) regulates skin pigmentation and thickness by affecting Wnt/B-catenin signaling in keratinocytes. *The FASEB Journal*, *22*(4), 1009–1020.

Zhang, F., & Lupski, J. R. (2015). Non-coding genetic variants in human disease. *Human Molecular Genetics*, *24*(R1), R102-R110.

CHAPTER 4

Study Conclusions

4.1. Overview

Over the past two decades, the South African wildlife ranching industry has developed into a multi-billion ZAR enterprise, mainly due to extensive growth in the breeding and trade subsector (Janovsky, 2016). In recent years, interesting trends have been observed in the ranching of game animals: Game ranching has shifted from breeding large numbers of common antelope species, such as impala (*Aepyceros melampus*) , eland (*Tragelaphus oryx*) and kudu (*Tragelaphus strepsiceros*), towards breeding fewer high-value species (Taylor *et al.*, 2016; Pitman *et al.*, 2016). These high-value species include a growing number of colour variant antelope, such as black impala, white springbuck (*Antidorcas marsupialis*) and golden wildebeest (*Connochaetes taurinus taurinus*; Strauss, 2015). Colour variant antelope are naturally occurring phenomena, which have become more prevalent on wildlife ranches due to deliberate breeding (Taylor *et al.*, 2016). Although colour variant breeding has attracted the attention of both local and international investors, it has also received criticism from key stakeholders and role players in the industry. Some have questioned the conservation consequences of selecting for colour variations that are infrequently seen in the wild (Taylor *et al.*, 2016; Russo *et al.*, 2019). Consequently, research regarding the genetic underpinnings of colour variant game have become extremely important to determine the significance of these concerns and to aid the development of policies and guidelines for the breeding of colour variant game (Olivier, 2015; Russo *et al.*, 2019).

A study by Taylor *et al.* (2016) identified the golden wildebeest, a colour variant of the blue wildebeest (*C. t. taurinus*), as one of the most common colour variant animals that game ranchers breed with in South Africa. Therefore, elucidation of the genetic mechanisms underlying this golden coat colour phenotype is of great interest. Molecular markers have been widely used as genomic tools to unravel the genetic basis of important phenotypic traits. However, despite the commercial value of blue wildebeest in hunting and live sales, the number of molecular markers available for blue wildebeest has been very limited. Fortunately, the advent of next-generation sequencing (NGS) has facilitated the development of several methods that are capable of simultaneously discovering and genotyping large numbers of molecular markers, such as single nucleotide polymorphisms (SNPs), in a single step, even in species for which no or limited genetic information is available

(Davey *et al.*, 2011). The main aim of this study was, therefore, to use a genotyping-by-sequencing (GBS) approach to discover and genotype a large number of SNPs in blue wildebeest, and its golden colour variant, and to conduct a genetic association study with these new markers, to identify SNPs and genes associated with coat colour in the blue wildebeest.

4.2. Summary And Synthesis Of Results

4.2.1. The Development of Genome-Wide SNP Markers in Blue Wildebeest using the DArTseq Platform

Recently, the DArTseq platform, a variant of the GBS method, has become a widely used approach for high-throughput SNP marker discovery and genotyping in non-model animals due to the high quality of the generated markers and the low cost of genotyping per animal (Chen *et al.*, 2018; Nguyen *et al.*, 2018). The second chapter of the current study describes the high-throughput discovery of genome-wide SNP markers in blue wildebeest using the DArTseq platform. This represents the first of such studies in an antelope species commercially farmed in South Africa. This study, therefore, also provided a unique opportunity to determine if the DArTseq platform can offer a solution to the lack of genomic resources currently available in the wildlife ranching industry.

The DArTseq platform identified 20 563 SNP markers, each located in a 69 bp marker sequence, in a sample population of 94 blue wildebeest. This sample population included 35 blue wildebeest with the golden coat colour phenotype. The developed SNP markers showed a high average reproducibility (>99%), and the percentage of missing data (~9.21%) was also considerably less than that generally observed in other reduced representation sequencing (RRS) studies (Campbell *et al.*, 2017; Souza *et al.*, 2017). Furthermore, the typical data characteristics of next-generation sequencing technologies were used to determine SNP quality and the risk of false positives in the dataset. Both the read position of the generated SNPs and the ratio of transitions *versus* transversions (Ts:Tv) across the dataset indicated that the generated SNPs are more likely to be true SNPs than the result of sequencing errors. In addition, the generated SNPs were also assessed based on their information content and their coverage of the genome. Based on allelic frequencies and polymorphic information content (PIC), the DArTseq platform was able to generate a large number of informative SNPs for blue wildebeest. The *in silico* mapping of the marker sequences against the *Bos taurus* genome revealed that the generated SNPs are distributed throughout the whole genome, but coverage of the different bovine chromosomes were uneven. However, unlike other GBS studies that show higher coverage of gene-rich bovine chromosomes, no significant correlation was observed between

DArTseq SNP density and chromosome gene density in the current study (De Donato *et al*., 2013). Nevertheless, the high Ts:Tv observed in the dataset suggests that the DArTseq platform did indeed target gene-rich regions of the blue wildebeest genome. Therefore, the difference in the correlation results between cattle and blue wildebeest could be the result of genome reshuffling that occurred in these species during their differentiation (He *et al*., 2018). Furthermore, the *in silico* mapping of the DArTseq marker sequences also provided the opportunity to identify orthologous genes shared by cattle and blue wildebeest. Moreover, functional analysis of the DArTseq marker sequences revealed a wide variety of different gene ontology (GO) terms, common to even-toed ungulates, associated with the marker sequences.

The generated DArTseq SNPs will be very useful in the ranching of blue wildebeest since it can be used for a wide range of applications. Game ranchers often request proof of relatedness when purchasing game animals from a genetic line considered to be superior (e.*g.* longer horn length, high fertility). In addition, game breeders who have accurate knowledge of the ancestry of their animals are better equipped to avoid production losses associated with inbreeding and unproductive matings (Marx, 2015). Furthermore, individual identification of herd animals is often required in cases where animals lose their identification tags (DeYoung & Honeycutt, 2005). Assessing population structure, parentage and relatedness determination and individual identification are, therefore, some of the key applications these SNPs could be used for (Von Thaden *et al*., 2017). These SNPs will also be useful for identifying populations that could serve as source stocks in order to increase the genetic diversity of captive populations. Since the DArTseq SNP marker sequences revealed a wide variety of different GO terms, these SNPs could also be of great value in studies investigating the genetic basis of phenotypic traits in blue wildebeest. Phenotypic traits of interest to blue wildebeest ranchers, in addition to coat colour, include fertility (*e.g.* age of first calf and inter-calving period for females and scrotal circumference for males), horn traits (*e.g.* horn length, horn circumference), live body weights and carcass traits (Josling, 2017). However, regular and accurate recordkeeping of phenotypic data will be essential to investigate the genetic basis of these traits. Lastly, the application of the DArTseq platform in blue wildebeest, and other game species, could also be a valuable tool for the private wildlife industry regarding forensic cases or insurance claims (Ogden, 2011; Coleman, 2018). The poaching of blue wildebeest for bushmeat is a localised threat in some areas such as the North West Province of South Africa (Tambling *et al*., 2016). Moreover, due to the rapid growth in the trade of high-value game animals, numerous insurance companies have started to provide risk cover for the death of game animals as a result of illness, disease, poaching, post release stress, fire and lightning (Santam, n.d.). Forensic investigations and insurance claims procedures usually involve animal identification from a carcass, through DNA profiling. High

molecular weight DNA is generally required for accurate microsatellite genotyping and by studies utilising conventional RRS approaches such as restriction-site associated DNA sequencing (RADseq) and double digest restriction associated DNA (ddRAD) sequencing (He *et al*. 2011; Souza *et al.*, 2017). In comparison, a study by Schultz *et al.* (2018), which involved the DArTseq genotyping of koala faecal DNA, demonstrated that SNPs could be successfully generated from genomic DNA concentrations as low as 6 ng/µl. Therefore, DArTseq represents a feasible approach for the genotyping of old or degraded DNA. This advantage of the DArTseq approach could make it a suitable approach for individual animal identification from carcass remains.

The DArTseq SNPs identified in blue wildebeest thus represents a new powerful set of tools that can address both management and research questions in blue wildebeest ranching. However, the large number of SNPs generated by the DArTseq platform is not suitable for routine genotyping in blue wildebeest. Therefore, in order to use these SNPs efficiently in breeding and population monitoring, it will be advantageous to mine highly informative SNPs from the DArTseq data that can be included in a reduced SNP assay (Henriques e*t al*., 2018). Fortunately, one of the major advantages of SNP markers is the wide range of genotyping platforms available to address different needs related to marker density, number of samples and costs per sample (Thomson, 2014). Table 4.1 lists a number of SNP genotyping platforms that could be used, depending on the application, to develop a reduced SNP panel for blue wildebeest.

Table 4.1. Genotyping assays suitable for developing a reduced single nucleotide polymorphism (SNP) panel for blue wildebeest (Huang *et al*., 2015; Kleinman-Ruiz *et al*., 2017).

Genotyping Assay	Detection Method	Platform	Throughput	Error Rate	Cost
GoldenGate® (Illumina, https://www.illumina.com/)	Hybridisation	Fluorescence	>1000 SNPs/sample/reaction	Low	Very High
MassARRAY ® system (Agena Bioscience, http://agenabio.com/)	Primer extension	Mass spectrometer	10-400 SNPs/sample/reaction	Low	Moderate
SNPstream® genotyping system (Beckman Coulter, https://www.beckmancoulter.com/)	Primer extension	Fluorescence	48 SNPs/sample/reaction	Low	Moderate
SNaPshot® multiplex system (Thermo Fisher Scientific, https://www.thermofisher.com/)	Primer extension	Fluorescence	10 SNPs/sample/reaction	Low	Moderate

Abbreviations: Single nucleotide polymorphisms (SNPs)

4.2.2. Elucidation of Coat Colour Genetics in Blue Wildebeest

The DArTseq SNPs identified in blue wildebeest enabled the study of the genetic basis of the economically important golden coat colour phenotype. A genome-wide association study (GWAS) was performed with 14 624 DArTseq SNPs in order to identify SNPs or genes significantly associated with coat colour in blue wildebeest. A total of 377 SNPs were found to be significantly associated with coat colour. The marker sequences of the statistically significant SNPs were then functionally annotated based on GO terms. The majority of the SNP marker sequences (90.19%) could not be functionally annotated. These SNPs could either be located in non-coding regions or in genes that have not been full annotated. In addition, since the DArTseq marker sequences were only 69 bp in length, more distant but significant similarities in the sequence databases could have been overlooked. Nonetheless, the GO terms associated with the 37 marker sequences that could be functionally annotated was able to give an overview of the molecular functions, biological processes and cellular components involved in blue wildebeest pigmentation. From this preliminary analysis, membrane transport proteins seem to play an important role in the determination of coat colour. Several previous studies have also identified genes coding for ion transport proteins as key regulators of melanin synthesis (*e.g.* Smith *et al*., 2004; Bellono *et al*., 2016; Chao e*t al*., 2017). Moreover, by matching the marker sequences of the statistically significant SNPs to the orthologous gene identified in cattle (chapter 2), two genes coding for solute carrier proteins [*solute carrier family 6 member 3* (*SLC6A3*) and *solute carrier family 28 member 2* (*SLC28A2*)] were identified as putative candidate genes for coat colour determination. These solute carrier proteins can regulate melanin synthesis through the regulation of melanosomal pH since the activity of tyrosinase, the rate-limiting enzyme in the melanogenesis pathway, depends on the pH level. Tyrosinase enzyme activity is generally stimulated by an increase in melanosomal pH, while the activity of tyrosinase is gradually lost with decreasing pH (Ancans *et al.*, 2001; Cheli *et al.*, 2009). In addition, studies have also found that a slightly acidic pH chemically shifts mixed melanogenesis to more pheomelanic states by suppressing the late stages of eumelanogenesis, that occur after the stages catalysed by tyrosinase (Ito *et al.*, 2013; Wakamatsu *et al.*, 2017).

The other orthologous cattle genes to which a statistically significant SNP mapped were also further investigated and an additional five candidate genes for the determination of coat colour were identified based on literature importance. These putative candidate genes are *myosin VC* (*MYO5C*), *myosin VIIA* (*MYO7A*), *dopamine receptor D2* (*DRD2*), *frizzled class receptor 4* (*FZD4*) and *tyrosinase* (*TYR*). Of the SNPs located in the putative candidate genes, 100076096|F|0--31:A>G (*MYO7A)*, 100025667|F|0--39:T>C (*SLC6A3*) and 100048574|F|0--35:G>A

(*SLC28A2*) had the largest genetic size effects and are, therefore, expected to play the largest role in the determination of coat colour in blue wildebeest. Since multiple genes were found to influence pigmentation in the current study, coat colour in blue wildebeest is suspected to be a quantitative threshold trait mainly determined by these genes with large genetic effect sizes, while the genes with smaller effect sizes contribute to a lesser extent. This specific mode of inheritance is supported by the limited gene-gene interaction observed between the putative candidate genes, and the additive mode of inheritance that best fitted most of the SNPs located in the putative candidate genes.

Lastly, since animal welfare has become a highly debated topic in the breeding of colour variant game animals, each of the putative candidate genes were investigated to determine if any direct pleiotropic effects related to health have been reported for these genes. Of the putative candidate genes, *TYR* is the only gene in which pleiotropic mutations have been reported. Mutations in this gene have resulted in retinal functional abnormalities, nystagmus and behavioural changes (Reissmann & Ludwig, 2013). However, such negative effects have not been reported in golden wildebeest and, thus, the SNP in the blue wildebeest *TYR* gene is not expected to be deleterious.

4.3. Study Limitations And Future Work

The current study made use of a GWAS approach to identify DArTseq generated SNPs that are associated with coat colour in blue wildebeest. Due to the presence of linkage disequilibrium (LD), a SNP significantly associated with a trait of interest should not be assumed to be the causal variant (*i.e.* a variant responsible for the association signal at a locus; Hormozdiari *et al*., 2015). Based on the LD analysis performed as part of chapter 3, it is possible that numerous SNPs found to be statistically associated with colour in blue wildebeest are merely tag SNPs in high linkage disequilibrium with the true causal variant (Hirschhorn & Daly, 2005; Bush & Moore, 2012). A total of eight SNPs (located in seven different genes) were identified as putative candidate loci that could contribute to coat colour determination in blue wildebeest. To ensure that these identified SNPs are indeed involved in blue wildebeest pigmentation, the sires and dams of the genotyped animals (as presented in Appendix A) could also be genotyped in order to exploit the information provided by family data and confirm the association (Dandine-Roulland & Perdry, 2015).

Reduced-representation approaches, such as DArTseq, are currently used in most genomic studies in non-model organisms, as they are relatively inexpensive and can provide a wealth of sequence information. However, although these methods are useful for the discovery of SNP markers, the generated sequences only represent a small portion of the genome. Therefore, these approaches are likely to fail in the detection of many loci that contribute to phenotypic traits (Hoban *et al.*, 2016). Consequently, the DArTseq approach applied in blue wildebeest could have been unsuccessful in identifying many other loci that contribute to coat colour in blue wildebeest. Indeed, the *melanocortin 1 receptor* (*MC1R*), that has been proven to play a crucial role in coat colour variation in many mammalian species, was not one of the orthologous genes identified in blue wildebeest in the present study. Typically, mutations in *MC1R* that result in the receptor being constitutively active are dominant and are associated with black/brown coat or plumage colour. Conversely, mutations that result in loss-of-function are recessive and are associated with yellow/red phenotypes (Kerje *et al.*, 2003). Loss-of-function mutations in *MC1R*, which result in a golden/yellow coat colour phenotype has been reported in many domesticated and wild animals such as chestnut horses (*Equus ferus caballus*; Marklund *et al.*, 1997), Labrador and Golden Retriever dogs (*Canis lupus familiaris*; Everts *et al*., 2002), Norwegian Forest cat (*Felis catus*; Peterschmitt *et al.*, 2009) and Kermode bear (*Ursus americanus kermodei*; Ritland *et al.*, 2001). Therefore, it could be possible that the golden coat colour of the wildebeest is the result of a mutation in the *MC1R* gene and that, due to the reduced representation sequencing approach employed by DArTseq, the locus was not typed. Moreover, the mode of inheritance exhibited by the golden coat colour phenotype in blue wildebeest fits the recessive *MC1R* mutation that results in a golden/yellow phenotype. Future studies on the effect of the *MC1R* gene on coat colour in blue wildebeest is thus needed to establish whether a mutation in the *MC1R* gene could be the causal variant of the golden coat colour phenotype in blue wildebeest, or if *MC1R* possibly interacts with any of the candidate genes identified in the current study. Since pigmentation genetics is highly conserved across mammals, a candidate gene study utilising the primers originally developed for *MC1R* in cattle, may be useful to investigate the role of *MC1R* in coat colour variation in blue wildebeest (Russo *et al*., 2007).

Furthermore, the cohort of blue wildebeest used for marker discovery and the subsequent GWAS was relatively small and could, therefore, also be considered as a limiting factor. Due to the small study population size, the genetic size effects of the SNPs in the identified pigmentation candidate genes could have been overestimated, and it could also be the reason for the large confidence intervals observed for some of the candidate SNPs. In addition, a quantitative threshold model for inheritance has not been previously reported for coat colour in mammals, and thus the results of the current study represents a novel finding. To confirm and strengthen the

credibility of the study findings, further large-scale replicated investigations of the significant SNPs in other independent populations are thus necessary. Moreover, due to the short flanking sequences of the DArTseq SNPs, it was not possible to accurately determine whether any of the putative candidate SNPs for coat colour determination in blue wildebeest results in an amino acid change. Future studies can assess the functional significance of the SNPs located in the putative candidate genes by determining if any of these SNPs result in amino acid changes or truncated proteins. Nevertheless, the study provided valuable biological insights on the genetic basis of coat colour in blue wildebeest and provides a foundation for further genomic research.

One of the greatest challenges regarding research on non-model species is the lack of a reference genome. The assembly of an accurate reference genome provides a valuable resource for research in non-model species and significantly aids in understanding the underlying genetic mechanisms of phenotypic traits (Hoban *et al.*, 2016). Therefore, a reference genome for blue wildebeest will greatly improve the inference of the genetic basis of important phenotypic traits. In 2016, Miller *et al.* reported that funding had been secured for the sequencing of the blue wildebeest (*C. t. taurinus*) genome. Sequencing of the genome has been completed with assembly of the sequences commencing in 2020 (C. Harper, 2019, personal communication, 25 September). Nevertheless, a recent study by Chen e*t al*. (2019) reports the newly sequenced genomes of 44 ruminant species, which includes the genome of a blue wildebeest. The study was conducted in order to resolve topological controversies in ruminant phylogeny, and to investigate the genetic basis of distinct ruminant traits, such as cranial appendages, a multi-chambered stomach and specialised dentition. It should however be noted that, although the subspecies of blue wildebeest is not reported, the sequenced blue wildebeest is from northern Tanzania and is thus believed to belong to the subspecies *C. t. mearnsi.* Sequencing the genome of *C. t. taurinus* would, thus, still be beneficial for comparison studies between the blue wildebeest subspecies and for research efforts specifically aimed at the Southern African blue wildebeest.

4.4. Final Remarks

This SNP study is the first large scale polymorphism discovery performed in an African antelope species farmed commercially in South Africa. The DArTseq platform generated high-quality, genome-wide SNP markers with desirable polymorphism parameters. The DArTseq platform could, thus, offer a possible solution to the lack of genomic resources currently available in the wildlife ranching industry. In blue wildebeest specifically, the

generated DArTseq markers could be a useful resource for the genetic management of captive stock. Furthermore, these markers could also be used to facilitate research in blue wildebeest regarding traits that are currently being selected for in the industry (*e.g.* horn length, fertility, carcass traits). To demonstrate the utility of these newly generated markers, the DArTseq SNPs were used to investigate the genetic basis of the economically important golden coat colour phenotype in the first genetic association study to be conducted in an African antelope species. Eight SNPs, located in seven different candidate genes (*TYR*, *SLC6A3*, *SLC28A2 MYO5C*, *MYO7A*, *DRD2* and *FZD4*) were identified as putative candidate SNPs. The results of the current study, thus, suggests that coat colour in blue wildebeest is a polygenic trait with an additive mode of inheritance (quantitative threshold trait) and not a single gene trait as originally thought. The study of coat colour in non-model species, such as blue wildebeest, will continue to reveal novel variants and genetic mechanisms that influence pigmentation, which is crucial to ultimately elucidating the network of genes and biological pathways that determine the coat colour phenotype.

REFERENCES

Ancans, J., Tobin, D. J., Hoogduijn, M. J., Smit, N. P., Wakamatsu, K., & Thody, A. J. (2001). Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells. *Experimental Cell Research*, *268*(1), 26–35.

Bellono, N. W., Escobar, I. E., & Oancea, E. (2016). A melanosomal two-pore sodium channel regulates pigmentation. *Scientific Reports*, *6*, 26570.

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Computational Biology*, *8*(12), e1002822.

Campbell, E. O., Davis, C. S., Dupuis, J. R., Muirhead, K., & Sperling, F. A. H. (2017). Cross-platform compatibility of *de novo*-aligned SNPs in a nonmodel butterfly genus. *Molecular Ecology Resources*, *17*(6), e84–e93.

Chao, Y.-K. Schludi, V., Chen, C.-C., Butz, E., Phuong Nguyen, O. N., Müller, M., Krüger, J., Kammerbauer, C., Ben-Johny, M., Vollmar, A. M., Berking, C., Biel, M., Wahl-Schott, C. A., & Grimm, C. (2017). TPC2 polymorphisms associated with a hair pigmentation phenotype in humans result in gain of channel function by independent mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(41), E8595–E8602.

Cheli, Y., Luciani, F., Khaled, M., Beuret, L., Bille, K., Gounon, P., Ortonne, J.-P., Bertolotto, C., & Ballotti, R. (2009). αMSH and cyclic AMP elevating agents control melanosome pH through a protein kinase A-independent mechanism. *Journal of Biological Chemistry*, *284*(28), 18699–18706.

Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., Wang, J., Nie, W., Su, W., Liu, G., Li, Q., Fu, W., Pan, X., Liu, C., Yang, J., Zhang, C., Yin, Y., Wang, Y., Zhao, Y., Zhang, C., Wang, Z., Qin, Y., Liu, W., Wang, B., Ren, Y., Zhang, R., Zeng, Y., Da Fonseca, R. R., Wei, B., Li, R., Wan, W., Zhao, R., Zhu, W., Wang, Y., Duan, S., Gao, Y., Zhang, Y. E., Chen, C., Hvilsom, C., Epps, C. W., Chemnick, L. G., Dong, Y., Mirarab, S., Siegismund, H. R., Ryder, O. A., Gilbert, M. T. P., Lewin, H. A., Zhang, G., Heller, R., & Wang, W. (2019). Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, *364*, eaav6202.

Chen, T., Tantasawat, P. A., Wang, W., Gao, X., & Zhang, L. (2018). Population structure of Chinese southwest wheat germplasms resistant to stripe rust and powdery mildew using the DArT-seq technique. *Ciência Rural*, *48*(4), e20160066.

Dandine-Roulland, C., & Perdry, H. (2015). Where is the causal variant? On the advantage of the family design over the case–control design in genetic association studies. *European Journal of Human Genetics*, *23*(10), 1357–1363.

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*(7), 499–510.

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. (2013). Genotyping-by-sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE*, *8*(5), e62137.

DeYoung, R. W., & Honeycutt, R. L. (2005). The molecular toolbox : Genetic techniques in wildlife ecology and management. *Journal of Wildlife Management*, *69*(4), 1362–1384.

Everts, R. E., Rothuizen, J., & Oost, B. A. (2002). Identification of a premature stop codon in the melanocyte-stimulating hormone receptor gene (*MC1R*) in Labrador and Golden retrievers with yellow coat colour. *Animal Genetics*, *31*(3), 194–199.

He, G., Huang, K., Guo, S., Ji, W., Qi, X., Ren, Y., Jin, X., & Li, B. (2011). Evaluating the reliability of microsatellite genotyping from low-quality DNA templates with a polynomial distribution model. *Chinese Science Bulletin*, *56*(24), 2523–2350.

He, J., Guo, Y., Xu, J., Li, H., Fuller, A., Tait, R. G., Jr., Wu, X.-L., & Bauck, S. (2018). Comparing SNP panels and statistical methods for estimating genomic breed composition of individual animals in ten cattle breeds. BMC Genetics, 19(1), 56.

Henriques, D., Parejo, M., Vignal, A., Wragg, D., Wallberg, A., Webster, M. T., & Pinto, M. A. (2018). Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*). *Evolutionary Applications*, *11*(8), 1270–1282.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, *6*(2), 95–108.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379–397.

Hormozdiari, F., Kichaev, G., Yang, W.-Y., Pasaniuc, B., & Eskin, E. (2015). Identification of causal genes for complex traits. *Bioinformatics*, *31*(12), i206–i213.

Huang, C.-W., Lin, Y.-T., Ding, S.-T., Lo, L.-L., Wang, P.-H., Lin, E.-C., Liu, F.-W., & Lu, Y.-W. (2015). Efficient SNP discovery by combining microarray and lab-on-a-chip data for animal breeding and selection. *Microarrays*, *4*(4), 570–595.

Ito, S., Suzuki, N., Takebayashi, S., Commo, S., & Wakamatsu, K. (2013). Neutral pH and copper ions promote eumelanogenesis after the dopachrome stage. *Pigment Cell and Melanoma Research*, *26*(6), 817–825.

Janovsky, E. (2016). Wildlife industry expected to continue to grow. *Wildlife Stud Services: Wildlife Breeders Journal*, 12–14.

Josling, C. (2017). Breeding goals to meet the game industry's market demands. *Wildlife Stud Services: Wildlife Breeders Journal*, 15–17.

Kerje, S., Lind, J., Schütz, K., Jensen, P., & Andersson, L. (2003). Melanocortin 1-receptor (*MC1R*) mutations are associated with plumage colour in chicken. *Animal Genetics*, *34*(4), 241–248.

Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villaneuva, B., Fernández, J., & Godoy, J. A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, *18*(556), 1–12.

Marklund, L., Johansson Moller, M., Sandberg, K., & Andersson, L. (1997). A missense mutation in the gene for melanocyte-stimulating hormone receptor (*MC1R*) is associated with the chestnut coat color in horses. *Mammalian Genome*, *7*(12), 895–899.

Marx, M. (2015). Basics of DNA genotyping and genomic profiling in game farming. *Wildlife Stud Services: Wildlife Breeders Journal*, 41–46.

Miller, S. M., Clarke, A. B., Bloomer, P., Guthrie, A. J., & Harper, C. K. (2016). Evaluation of microsatellites for common ungulates in the South African wildlife industry. *Conservation Genetics Resources*, *8*(3), 329–341.

Nguyen, N. H., Premachandra, H. K. A., Kilian, A., & Knibb, W. (2018). Genomic prediction using DArT-Seq technology for yellowtail kingfish *Seriola lalandi*. *BMC Genomics*, *19*(107), 1–9.

Ogden, R. (2011). Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources*, *11*(S1), 109–116.

Olivier, P. A. (2015). *Colour variation of African Bovidae: Consequences for conservation and the wildlife ranching industry* (Unpublished MSc thesis). University of the Free State, Bloemfontein, South Africa.

Peterschmitt, M., Grain, F., Arnaud, B., Deléage, G., & Lambert, V. (2009). Mutation in the melanocortin 1 receptor is associated with amber colour in the Norwegian Forest Cat. *Animal Genetics*, *40*(4), 547–552.

Pitman, R. T., Fattebert, J., Williams, S. T., Williams, K. S., Hill, R. A., Hunter, L. T. B., Slotow, R., & Balme, G. A. (2016). The conservation costs of game ranching. *Conservation Letters*, *10*(4), 402–412.

Reissmann, M., & Ludwig, A. (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Seminars in Cell & Developmental Biology*, *24*(6), 576–586.

Ritland, K., Newton, C., & Marshall, H. D. (2001). Inheritance and population structure of the white-phased

"Kermode" black bear. *Current Biology*, *11*(18), 1468–1472.

Russo, I.-R. M., Hoban, S., Bloomer, P., Kotzé, A., Segelbacher, G., Rushworth, I., Birss, C., & Bruford, M. W. (2019). 'Intentional genetic manipulation' as a conservation threat. *Conservation Genetics Resources*, *11*(2), 237–247.

Russo, V., Fontanesi, L., Scotti, E., Tazzoli, M., Dall'Olio, S., & Davoli, R. (2007). Analysis of melanocortin 1 receptor (*MC1R*) gene polymorphisms in some cattle breeds: Their usefulness and application for breed traceability and authentication of Parmigiano Reggiano cheese. *Italian Journal of Animal Science*, *6*(3), 257–272.

Santam (n.d.). *Game risk* [Brochure]. Retrieved from https://www.santam.co.za/media/2680891/36098-santam-game-risk.pdf [accessed 19 August 2019].

Smith, D. R., Spaulding, D. T., Glenn, H. M., & Fuller, B. B. (2004). The relationship between Na(+)/H(+) exchanger expression and tyrosinase activity in human melanocytes. *Experimental Cell Research*, *298*(2), 521–534.

Souza, C. A., Murphy, N., Villacorta-Rath, C., Woodings, L. N., Ilyushkina, I., Hernandez, C. E., Green, B. S., Bell, J. J., & Strugnell, J. M. (2017). Efficiency of ddRAD target enriched sequencing across spiny rock lobster species (Palinuridae: *Jasus*). *Scientific Reports*, *7*(6781), 1–14.

Strauss, J. S. (2015). *The long term financial sustainability of the high value game breeding sector in South Africa* (Unpublished MBA thesis). University of the Free State, Bloemfontein, South Africa.

Tambling, C., Viljoen, P., & Kotze, A. (2016). A conservation assessment of *Connochaetes taurinus*. In M. F. Child, L. Roxburgh, E. Do Lihn San, D. Raimondo, & H. T. Davies-Mostert (Eds.), *The Red List of Mammals of South Africa, Swaziland and Lesotho*. South African National Biodiversity Institute and Endangered Wildlife Trust, South Africa.

Taylor, W. A., Lindsey, P. A., & Davies-Mostert, H. (2016). *An assessment of the economic, social and conservation value of the wildlife ranching industry and its potential to support the green economy in South Africa*. The Endangered Wildlife Trust, South Africa.

Thomson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breeding and Biotechnology*, *2*(3), 195–212.

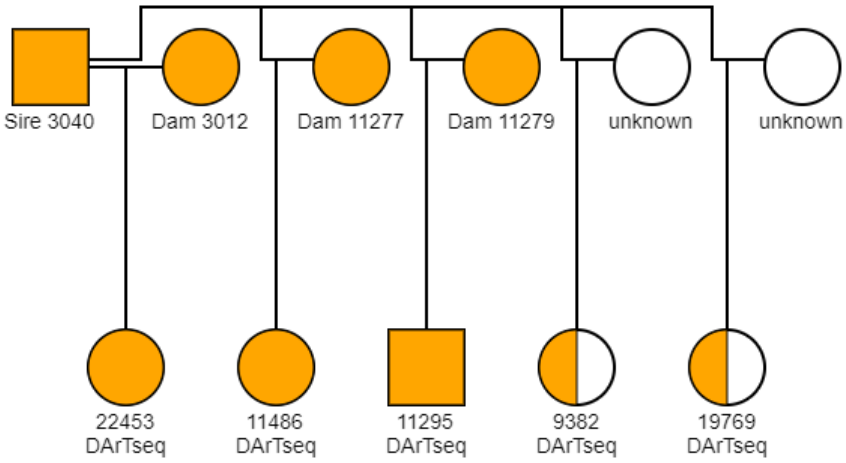
Von Thaden, A., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlidis, A. A., Tiesmeyer, A., Nowak, C., & Muñoz-Fuentes, V. (2017). Assessing SNP genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports*, *7*(1), 10768.

Wakamatsu, K., Nagao, A., Watanabe, M., Nakao, K., & Ito, S. (2017). Pheomelanogenesis is promoted at a weakly acidic pH. *Pigment Cell and Melanoma Research*, *30*(3), 372–377.

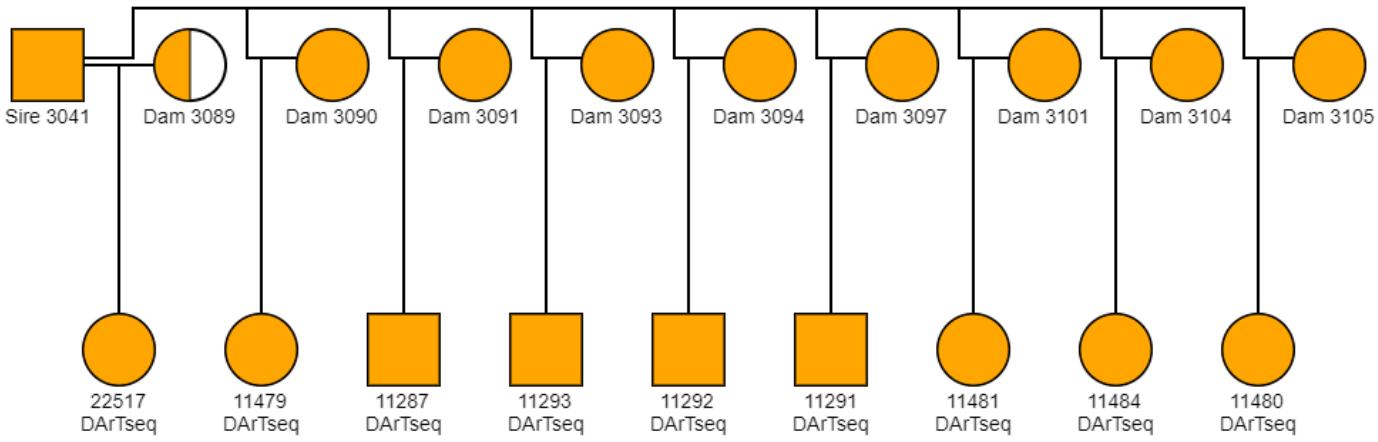
APPENDIX A

Table A1. Family pedigrees of the blue and golden wildebeest selected for DArTseq genotyping. The blue and golden wildebeest could be grouped into one of 15 families based on shared sires. Males are represented by squares and females are represented by circles. Symbols representing a golden wildebeest are shaded (orange). Presumed carriers of the causal variant for golden coat colour are represented with a half-shaded symbol. If available, the unique identifier of the animal is also given. Individuals that have been included in the DArTseq genotyping are represented by "DArTseq".

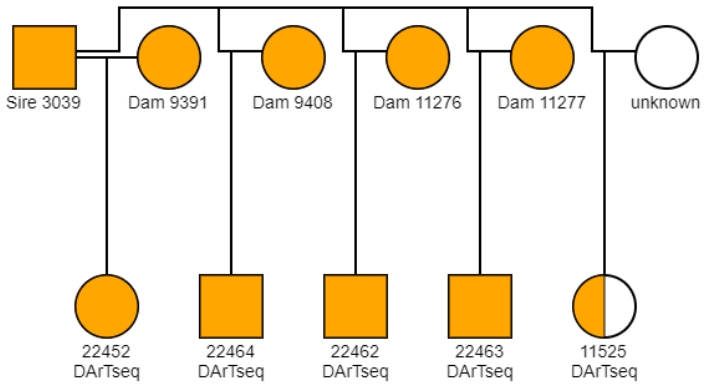
Family 1	Sire 3040



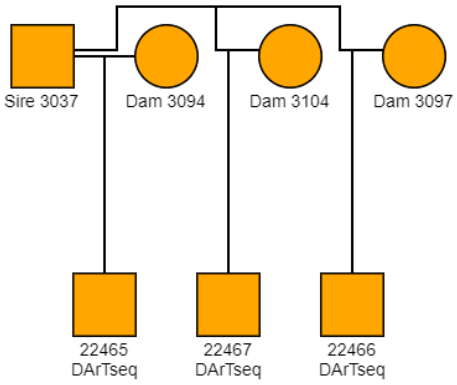
Family 2	Sire 3041



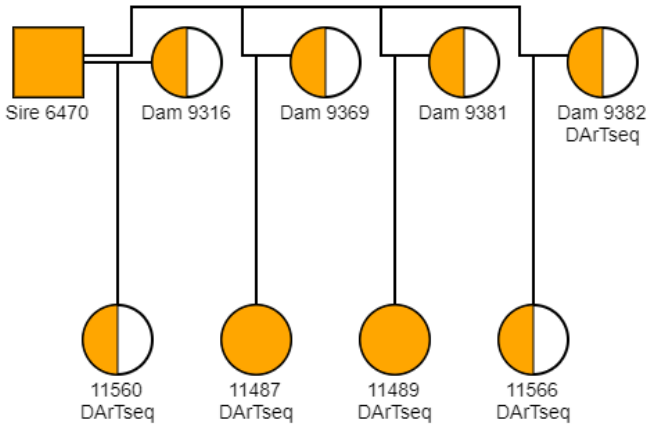
Family 3	Sire 3039



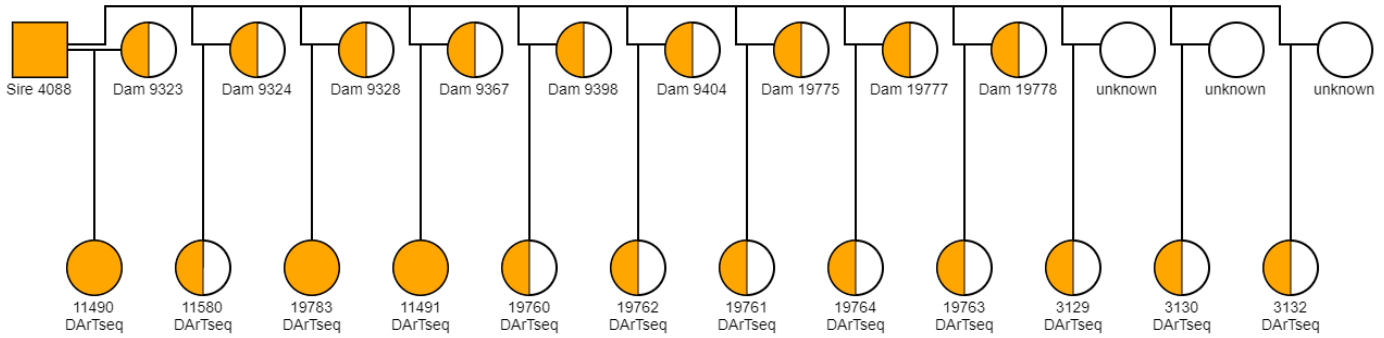
Family 4	Sire 3037



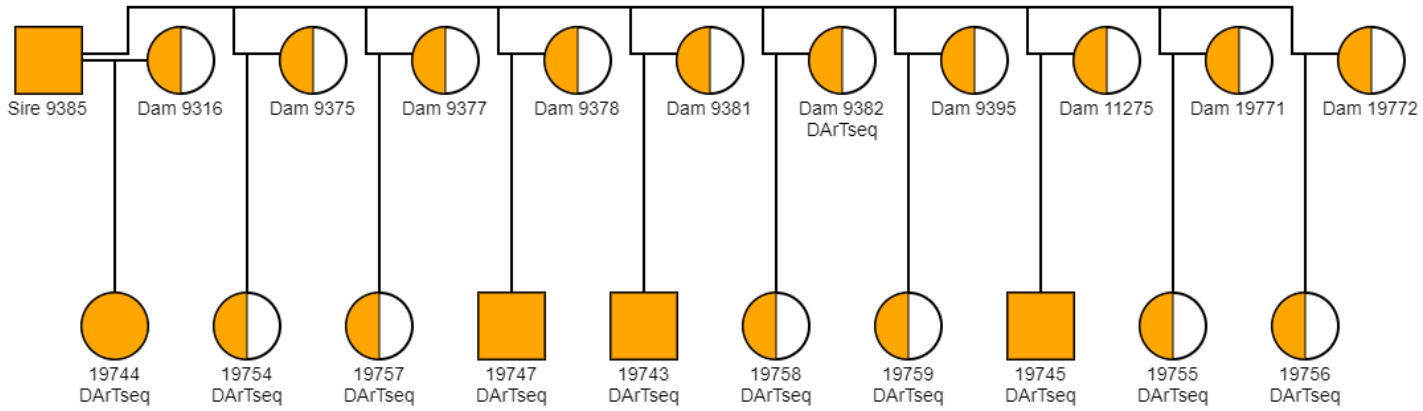
Family 5	Sire 6470



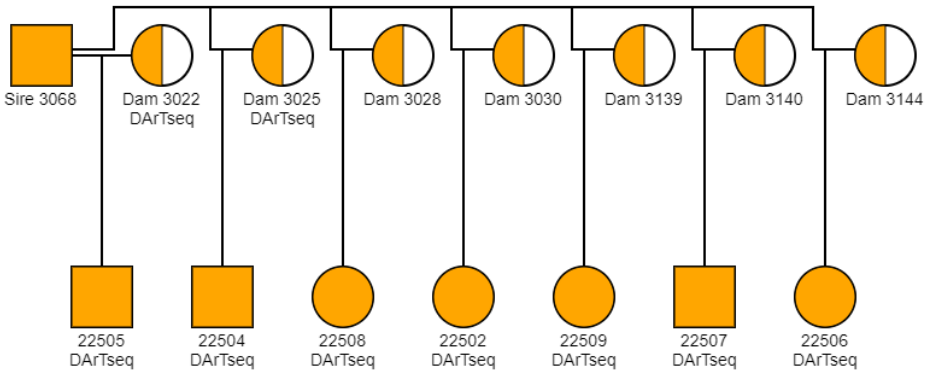
Family 6	Sire 4088



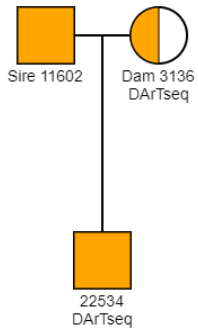
Family 7	Sire 9385



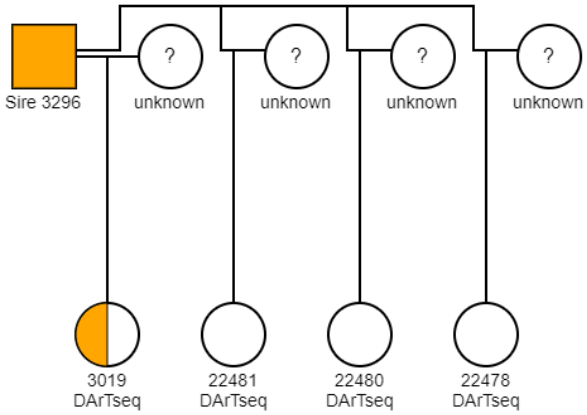
Family 8	Sire 3068



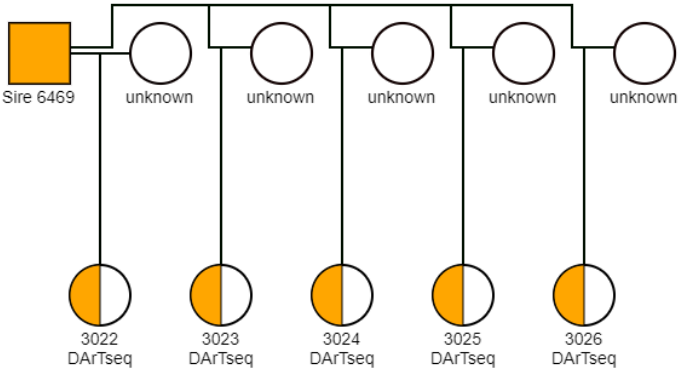
Family 9	Sire 11602



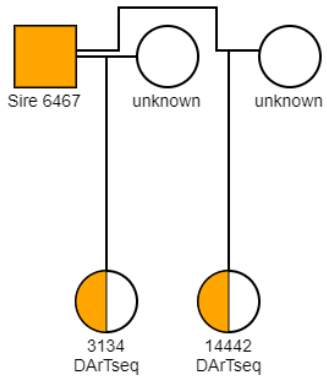
Family 10	Sire 3296



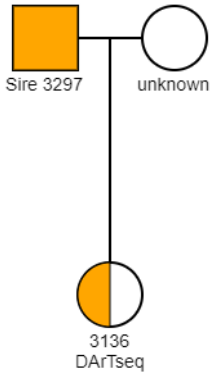
Family 11	Sire 6469



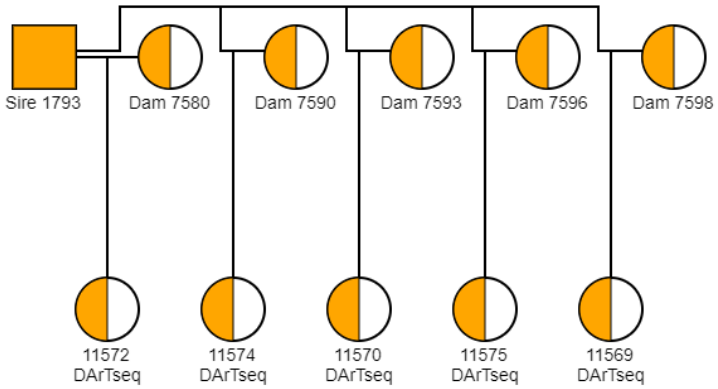
Family 12	Sire 6467

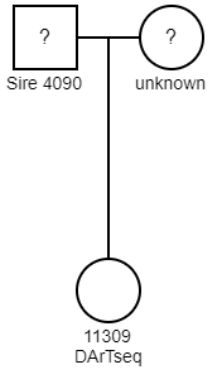


Family 13	Sire 3297



Family 14	Sire 1793



Family 15	Sire 4090
	
Blue wildebeest heifers, included in the DArTseq genotyping, with the wild-type coat colour. Both the sire and dam for these animals are unknown.	
11306; 11307; 11308; 11309; 11310; 11311; 11312; 11313; 11314; 11315; 11316; 19753; 22478; 22479; 22480; 22481; 22482; 22483; 22486; 22487; 22488; 22494; 22497; 22498	

APPENDIX B

Table A2. Single nucleotide polymorphisms (SNPs) that were significantly associated ($p \leq 0.001$) with coat colour in blue wildebeest and the *Bos taurus* orthologous gene to which the SNP sequence mapped.

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100025667|F|0--39:T>C	1.65E-12	*SLC6A3*	4.16E-07
100038227|F|0--48:G>C	4.30E-10	No match	Not applicable
100056649|F|0--14:C>G	4.57E-09	No match	Not applicable
100060588|F|0--26:C>T	8.15E-09	No match	Not applicable
100076096|F|0--31:A>G	2.94E-08	*MYO7A*	3.20E-02
100044257|F|0--50:A>G	3.70E-08	Uncharacterised protein	1.45E-06
100060085|F|0--39:C>T	4.81E-07	Uncharacterised protein	2.80E-09
100043298|F|0--9:C>T	5.47E-07	*ST14*	2.63E-03
100050912|F|0--6:C>T	9.12E-07	No match	Not applicable
100074542|F|0--25:G>C	1.39E-06	No match	Not applicable
100064058|F|0--16:G>A	1.48E-06	Uncharacterised protein	2.30E-10
100073674|F|0--15:C>T	1.58E-06	Uncharacterised protein	8.03E-10
100043573|F|0--25:G>A	1.63E-06	*TNS2*	9.16E-03
100017173|F|0--24:C>T	1.74E-06	*CEP170*	6.17E-05
100073779|F|0--31:C>T	1.82E-06	Uncharacterised protein	8.03E-10
100067322|F|0--24:G>C	1.90E-06	No match	Not applicable
100023127|F|0--33:C>T	2.64E-06	No match	Not applicable
100024357|F|0--6:G>A	2.71E-06	No match	Not applicable
100030072|F|0--58:T>C	2.74E-06	*TAOK1*	2.63E-03
100025991|F|0--26:G>A	3.04E-06	Uncharacterised protein	9.78E-09
100018821|F|0--42:G>A	3.21E-06	No match	Not applicable
100015994|F|0--41:C>T	3.27E-06	No match	Not applicable
100033709|F|0--39:A>C	4.17E-06	No match	Not applicable
100074108|F|0--27:A>G	4.89E-06	No match	Not applicable
100068125|F|0--40:T>C	4.90E-06	Uncharacterised protein	9.16E-03
100078116|F|0--6:C>T	5.02E-06	*ZBTB8OS*	4.16E-07
100029827|F|0--63:G>A	8.22E-06	No match	Not applicable
100068768|F|0--16:T>C	8.39E-06	No match	Not applicable
100043697|F|0--25:C>T	8.47E-06	*SEC22B*	2.15E-04
100012609|F|0--68:G>T	8.49E-06	No match	Not applicable
100051962|F|0--40:C>T	8.87E-06	No match	Not applicable
100041787|F|0--5:A>G	9.07E-06	No match	Not applicable
100074742|F|0--26:G>A	1.10E-05	Uncharacterised protein	5.07E-08
100061843|F|0--30:C>G	1.11E-05	No match	Not applicable

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100058789|F|0--19:G>C	1.12E-05	*CCDC14*	1.45E-06
100059698|F|0--34:C>T	1.18E-05	Uncharacterised protein	6.59E-11
100055854|F|0--15:G>T	1.24E-05	No match	Not applicable
100024101|F|0--33:C>T	1.26E-05	No match	Not applicable
100055025|F|0--36:G>A	1.26E-05	No match	Not applicable
100044449|F|0--57:G>A	1.26E-05	No match	Not applicable
100068094|F|0--19:G>C	1.30E-05	*CCDC14*	1.45E-06
100034477|F|0--41:T>C	1.35E-05	*PARK2-201*	4.16E-07
100037541|F|0--10:C>T	1.57E-05	No match	Not applicable
100070348|F|0--17:C>T	1.60E-05	*ATP8A1*	6.17E-05
100067257|F|0--41:G>A	1.76E-05	*ANGPT2*	1.19E-07
100066914|F|0--20:G>A	1.81E-05	No match	Not applicable
100056494|F|0--27:G>A	1.87E-05	No match	Not applicable
100055558|F|0--32:G>A	1.92E-05	*ZBTB8OS*	7.52E-04
100055076|F|0--24:C>T	1.93E-05	No match	Not applicable
100017837|F|0--43:G>A	1.93E-05	No match	Not applicable
100055844|F|0--16:A>C	1.99E-05	No match	Not applicable
100070021|F|0--56:C>A	2.03E-05	*GLA*	4.16E-07
100051129|F|0--22:T>G	2.03E-05	No match	Not applicable
100049382|F|0--58:C>T	2.12E-05	*PTGER3*	1.66E-18
100033796|F|0--51:G>A	2.26E-05	No match	Not applicable
100055147|F|0--46:C>T	2.27E-05	No match	Not applicable
100067445|F|0--32:G>A	2.28E-05	No match	Not applicable
100057950|F|0--13:A>G	2.35E-05	Uncharacterised protein	1.19E-07
100011071|F|0--55:C>T	2.39E-05	No match	Not applicable
100069935|F|0--68:T>C	2.50E-05	No match	Not applicable
100055342|F|0--23:C>A	2.60E-05	No match	Not applicable
100034289|F|0--20:A>G	2.67E-05	No match	Not applicable
100014059|F|0--53:C>G	2.92E-05	No match	Not applicable
100025596|F|0--30:G>A	3.07E-05	*GLA*	1.80E-09
100023672|F|0--36:A>G	3.35E-05	*SDH*	2.46E-16
100048278|F|0--46:G>A	3.37E-05	No match	Not applicable
100048483|F|0--6:G>A	3.65E-05	No match	Not applicable
100057738|F|0--27:G>A	3.75E-05	*ZBTB8OS*	2.63E-03
100038960|F|0--27:G>A	3.75E-05	*ZBTB8OS*	2.62E-03
100034316|F|0--32:C>T	4.05E-05	No match	Not applicable
100046152|F|0--11:G>A	4.15E-05	No match	Not applicable
100074964|F|0--61:G>A	4.29E-05	No match	Not applicable

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100061428|F|0--46:A>G	4.35E-05	No match	Not applicable
100068557|F|0--18:G>A	4.40E-05	*PARK2*	5.41E-12
100049889|F|0--58:G>A	4.54E-05	*SULT2B1*	3.20E-02
100045609|F|0--33:C>T	4.60E-05	No match	Not applicable
100019683|F|0--8:A>G	4.97E-05	No match	Not applicable
100032051|F|0--45:T>A	5.41E-05	*ANKRD26*	9.16E-03
100077754|F|0--13:G>T	5.49E-05	Uncharacterised protein	2.15E-04
100023893|F|0--14:T>C	5.51E-05	No match	Not applicable
100060872|F|0--34:C>T	5.57E-05	*HCAR1*	2.30E-29
100056827|F|0--11:T>C	5.67E-05	No match	Not applicable
100055930|F|0--30:G>C	5.72E-05	No match	Not applicable
100026664|F|0--59:T>C	6.23E-05	No match	Not applicable
100067122|F|0--38:A>G	6.29E-05	No match	Not applicable
100060024|F|0--63:A>G	6.30E-05	*NECTIN2*	8.03E-10
100045405|F|0--67:C>A	6.52E-05	Uncharacterised protein	2.02E-17
100054380|F|0--18:T>C	6.57E-05	Uncharacterised protein	9.16E-03
100041117|F|0--26:G>A	6.61E-05	*PARK2-201*	1.45E-06
100012176|F|0--60:C>T	6.61E-05	*PAOX*	4.44E-12
100062911|F|0--5:T>G	6.63E-05	No match	Not applicable
100042513|F|0--50:G>A	6.73E-05	*CCDC14*	5.07E-06
100077887|F|0--18:T>C	6.81E-05	Uncharacterised protein	3.41E-07
100068723|F|0--34:T>A	6.97E-05	Uncharacterised protein	3.26E-03
100018403|F|0--50:C>T	7.24E-05	No match	Not applicable
100057057|F|0--49:C>T	7.24E-05	*ASCC2*	7.52E-04
100034603|F|0--41:T>G	7.54E-05	No match	Not applicable
100068678|F|0--45:G>A	7.56E-05	*RNF144B*	5.06E-25
100021667|F|0--6:C>T	7.57E-05	No match	Not applicable
100034418|F|0--10:T>C	7.61E-05	No match	Not applicable
100054910|F|0--15:G>A	7.67E-05	No match	Not applicable
100057048|F|0--29:G>C	7.89E-05	*PARK2-201*	2.63E-03
100071958|F|0--9:A>G	7.91E-05	*ASPM*	3.20E-02
100057659|F|0--21:G>A	7.97E-05	*CCDC14*	7.52E-04
100042751|F|0--34:G>A	7.98E-05	No match	Not applicable
100037109|F|0--28:A>G	8.11E-05	No match	Not applicable
100047775|F|0--46:C>A	8.16E-05	No match	Not applicable
100039200|F|0--34:C>T	8.53E-05	*CCNT1*	9.16E-03
100038465|F|0--44:C>T	8.67E-05	No match	Not applicable
100050487|F|0--35:A>G	8.73E-05	Uncharacterised protein	5.07E-06

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100055034|F|0--13:C>T	8.94E-05	No match	Not applicable
100064549|F|0--22:G>A	9.19E-05	No match	Not applicable
100067431|F|0--15:C>T	9.22E-05	No match	Not applicable
100052523|F|0--50:C>T	9.27E-05	*PAOX*	2.63E-03
100065921|F|0--11:C>T	9.33E-05	No match	Not applicable
100024463|F|0--13:C>G	9.34E-05	*IKZF2*	2.80E-09
100075867|F|0--25:T>G	9.42E-05	*KIF27*	3.20E-02
100064070|F|0--47:G>A	9.47E-05	Uncharacterised protein	3.41E-08
100048304|F|0--45:C>T	1.04E-04	No match	Not applicable
100061463|F|0--8:A>G	1.06E-04	No match	Not applicable
100037556|F|0--13:G>A	1.07E-04	*PARK2-201*	2.15E-04
100043157|F|0--62:C>G	1.09E-04	No match	Not applicable
100071267|F|0--5:C>T	1.11E-04	No match	Not applicable
100026626|F|0--7:T>C	1.14E-04	Uncharacterised protein	2.46E-16
100067795|F|0--24:T>C	1.18E-04	No match	Not applicable
100058414|F|0--8:T>G	1.20E-04	No match	Not applicable
100067776|F|0--34:T>C	1.23E-04	*CCNT1*	9.16E-03
100056477|F|0--46:G>A	1.26E-04	No match	Not applicable
100061995|F|0--16:A>C	1.31E-04	No match	Not applicable
100012981|F|0--21:G>A	1.31E-04	No match	Not applicable
100056729|F|0--10:T>C	1.31E-04	No match	Not applicable
100067568|F|0--13:A>G	1.38E-04	No match	Not applicable
100015355|F|0--52:A>G	1.38E-04	No match	Not applicable
100018422|F|0--24:C>G	1.39E-04	No match	Not applicable
100067402|F|0--23:C>T	1.42E-04	No match	Not applicable
100056129|F|0--28:T>C	1.43E-04	*AKAP11*	2.63E-03
100065529|F|0--19:A>G	1.47E-04	No match	Not applicable
100065022|F|0--15:G>A	1.49E-04	No match	Not applicable
100067534|F|0--31:C>A	1.49E-04	Uncharacterised protein	1.77E-05
100015839|F|0--55:G>A	1.50E-04	*GRIK4*	3.20E-02
100073946|F|0--6:C>T	1.51E-04	*CCNT1*	2.80E-09
100067709|F|0--39:C>T	1.52E-04	Uncharacterised protein	5.07E-06
100048581|F|0--34:G>A	1.54E-04	Uncharacterised protein	5.06E-25
100028247|F|0--11:A>G	1.54E-04	No match	Not applicable
100024577|F|0--62:T>G	1.57E-04	Uncharacterised protein	4.74E-19
100028638|F|0--44:C>T	1.61E-04	Uncharacterised protein	1.04E-14
100042962|F|0--53:G>A	1.61E-04	No match	Not applicable
100050710|F|0--14:T>C	1.69E-04	*ANGPT2*	8.03E-12

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100060988|F|0--29:C>T	1.70E-04	*VPS28*	3.20E-02
100042040|F|0--14:C>A	1.70E-04	No match	Not applicable
100049312|F|0--23:C>T	1.73E-04	Uncharacterised protein	5.41E-12
100067745|F|0--44:T>C	1.76E-04	No match	Not applicable
100066888|F|0--40:T>C	1.77E-04	*PLXNA4*	3.20E-02
100057679|F|0--38:C>T	1.77E-04	Uncharacterised protein	3.20E-02
100050704|F|0--32:C>T	1.80E-04	No match	Not applicable
100051199|F|0--18:C>T	1.81E-04	No match	Not applicable
100047582|F|0--29:A>G	1.83E-04	No match	Not applicable
100070145|F|0--39:G>T	1.83E-04	*PAOX*	3.20E-02
100043995|F|0--57:C>T	1.85E-04	No match	Not applicable
100071798|F|0--48:A>G	1.85E-04	*SLC6A3*	4.16E-07
100063882|F|0--26:G>A	1.99E-04	No match	Not applicable
100012104|F|0--63:C>T	2.01E-04	No match	Not applicable
100060085|F|0--28:A>G	2.01E-04	*CCNT1*	5.41E-12
100063200|F|0--33:G>A	2.03E-04	No match	Not applicable
100042397|F|0--17:A>C	2.04E-04	Uncharacterised protein	8.57E-16
100078873|F|0--55:A>G	2.07E-04	Uncharacterised protein	2.20E-17
100073138|F|0--26:G>A	2.08E-04	Uncharacterised protein	9.78E-09
100076079|F|0--9:C>T	2.15E-04	Uncharacterised protein	5.78E-18
100061588|F|0--5:C>G	2.24E-04	No match	Not applicable
100066537|F|0--42:G>A	2.28E-04	No match	Not applicable
100037206|F|0--14:C>G	2.30E-04	*PTGER3*	2.80E-09
100037106|F|0--32:G>A	2.33E-04	No match	Not applicable
100016322|F|0--62:C>T	2.34E-04	No match	Not applicable
100056741|F|0--37:G>T	2.36E-04	No match	Not applicable
100056664|F|0--11:G>A	2.41E-04	No match	Not applicable
100021610|F|0--19:A>G	2.46E-04	No match	Not applicable
100049166|F|0--46:G>A	2.58E-04	No match	Not applicable
100066587|F|0--6:A>G	2.62E-04	No match	Not applicable
100038776|F|0--16:G>A	2.63E-04	No match	Not applicable
100044320|F|0--59:C>G	2.63E-04	No match	Not applicable
100064535|F|0--11:G>A	2.67E-04	No match	Not applicable
100023760|F|0--36:C>A	2.67E-04	No match	Not applicable
100035360|F|0--53:A>C	2.69E-04	No match	Not applicable
100071743|F|0--16:G>A	2.69E-04	*ATP2B2*	9.78E-09
100055996|F|0--46:G>A	2.72E-04	No match	Not applicable
100041313|F|0--29:C>T	2.76E-04	*SNAP91*	7.52E-04

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100067222|F|0--19:A>G	2.78E-04	Uncharacterised protein	9.16E-03
100059481|F|0--29:C>T	2.88E-04	*SNAP91*	7.52E-04
100042873|F|0--42:C>T	2.91E-04	*GPR12*	3.20E-02
100031184|F|0--65:A>C	2.98E-04	No match	Not applicable
100055604|F|0--29:A>T	3.10E-04	No match	Not applicable
100070339|F|0--59:G>A	3.12E-04	*GLA*	1.19E-07
100045037|F|0--53:C>T	3.14E-04	No match	Not applicable
100073637|F|0--44:G>A	3.25E-04	Uncharacterised protein	1.89E-11
100070041|F|0--55:G>A	3.28E-04	Uncharacterised protein	2.02E-17
100058816|F|0--26:C>G	3.32E-04	*MOK*	2.63E-03
100066183|F|0--25:G>A	3.36E-04	No match	Not applicable
100074480|F|0--22:C>G	3.39E-04	No match	Not applicable
100047732|F|0--19:C>T	3.40E-04	No match	Not applicable
100024601|F|0--56:G>A	3.41E-04	*SULT2B1*	1.04E-14
100073112|F|0--13:A>G	3.46E-04	No match	Not applicable
100056826|F|0--13:G>A	3.47E-04	*PTGER3*	9.78E-09
100015583|F|0--36:C>T	3.49E-04	No match	Not applicable
100067326|F|0--17:A>G	3.51E-04	No match	Not applicable
100017915|F|0--20:A>G	3.52E-04	No match	Not applicable
100022408|F|0--51:G>A	3.55E-04	*FSIP2*	8.57E-16
100069609|F|0--60:G>A	3.56E-04	No match	Not applicable
100070130|F|0--49:C>T	3.58E-04	*VPS28*	9.18E-03
100073101|F|0--62:C>T	3.60E-04	No match	Not applicable
100057825|F|0--31:C>T	3.60E-04	*CCNT1*	5.07E-06
100071790|F|0--24:T>C	3.60E-04	No match	Not applicable
100036296|F|0--44:A>G	3.63E-04	*STAG2*	3.20E-02
100065612|F|0--28:G>A	3.65E-04	No match	Not applicable
100064823|F|0--11:G>C	3.66E-04	*PHLDA2*	9.78E-09
100024800|F|0--34:A>C	3.70E-04	Uncharacterised protein	2.15E-04
100015807|F|0--8:G>A	3.77E-04	*PRKAG1*	3.20E-02
100028006|F|0--14:C>T	3.81E-04	*KPNA1*	2.63E-03
100073662|F|0--12:C>A	3.84E-04	Uncharacterised protein	1.19E-07
100061928|F|0--12:C>A	3.99E-04	No match	Not applicable
100048935|F|0--8:A>G	4.02E-04	*TLN2*	1.55E-12
100066792|F|0--27:A>G	4.03E-04	No match	Not applicable
100045078|F|0--16:C>T	4.06E-04	No match	Not applicable
100071218|F|0--14:T>C	4.07E-04	No match	Not applicable
100065969|F|0--22:G>C	4.10E-04	No match	Not applicable

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100047777|F|0--17:T>C	4.12E-04	No match	Not applicable
100046894|F|0--60:G>A	4.12E-04	No match	Not applicable
100066262|F|0--24:A>G	4.13E-04	No match	Not applicable
100047104|F|0--37:T>C	4.16E-04	*NECTIN2*	1.27E-13
100025753|F|0--65:A>C	4.17E-04	Uncharacterised protein	1.45E-06
100056064|F|0--24:G>A	4.18E-04	No match	Not applicable
100028731|F|0--42:C>G	4.24E-04	No match	Not applicable
100051796|F|0--49:G>A	4.24E-04	No match	Not applicable
100049227|F|0--37:T>C	4.33E-04	No match	Not applicable
100049705|F|0--62:T>G	4.34E-04	No match	Not applicable
100012669|F|0--60:G>A	4.35E-04	No match	Not applicable
100051420|F|0--16:T>C	4.36E-04	No match	Not applicable
100047158|F|0--54:C>T	4.37E-04	No match	Not applicable
100066210|F|0--39:C>T	4.37E-04	No match	Not applicable
100065982|F|0--37:T>C	4.43E-04	No match	Not applicable
100015619|F|0--33:G>A	4.51E-04	*CDK5RAP2*	3.20E-02
100072346|F|0--52:C>A	4.57E-04	No match	Not applicable
100074461|F|0--14:C>T	4.64E-04	No match	Not applicable
100033976|F|0--5:C>A	4.75E-04	No match	Not applicable
100045189|F|0--39:C>G	4.77E-04	No match	Not applicable
100058657|F|0--23:C>T	4.79E-04	No match	Not applicable
100051055|F|0--13:A>G	4.81E-04	No match	Not applicable
100045386|F|0--38:C>T	4.81E-04	No match	Not applicable
100020447|F|0--41:T>A	4.81E-04	No match	Not applicable
100020114|F|0--27:C>T	4.81E-04	*ANKIB1*	1.45E-25
100066667|F|0--6:A>C	4.85E-04	No match	Not applicable
100068765|F|0--23:C>T	4.89E-04	*IPO5*	3.20E-02
100036514|F|0--7:G>C	4.89E-04	No match	Not applicable
100013053|F|0--14:G>A	4.89E-04	No match	Not applicable
100043044|F|0--45:G>A	4.89E-04	No match	Not applicable
100016197|F|0--28:G>A	4.93E-04	No match	Not applicable
100071018|F|0--11:C>T	4.96E-04	No match	Not applicable
100025870|F|0--26:G>A	5.00E-04	Uncharacterised protein	9.78E-09
100013634|F|0--36:T>C	5.00E-04	No match	Not applicable
100065877|F|0--19:G>A	5.01E-04	No match	Not applicable
100065941|F|0--14:T>G	5.02E-04	No match	Not applicable
100051603|F|0--57:T>A	5.03E-04	No match	Not applicable
100065648|F|0--23:C>G	5.06E-04	No match	Not applicable

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100066370|F|0--5:C>T	5.10E-04	Uncharacterised protein	2.15E-04
100033737|F|0--23:T>G	5.12E-04	No match	Not applicable
100025456|F|0--50:C>T	5.14E-04	*SULT2B1*	5.07E-06
100048484|F|0--19:G>A	5.16E-04	No match	Not applicable
100073520|F|0--29:G>A	5.21E-04	No match	Not applicable
100012047|F|0--38:C>A	5.29E-04	No match	Not applicable
100061731|F|0--66:G>A	5.29E-04	*KLHDC3*	9.16E-03
100040107|F|0--37:C>T	5.30E-04	No match	Not applicable
100045869|F|0--37:G>A	5.36E-04	*ANGPT2*	4.44E-13
100038118|F|0--11:A>C	5.40E-04	No match	Not applicable
100023047|F|0--59:G>C	5.41E-04	No match	Not applicable
100064439|F|0--63:C>T	5.41E-04	*ZBTB8OS*	2.15E-04
100064950|F|0--24:T>C	5.42E-04	No match	Not applicable
100043735|F|0--17:T>C	5.44E-04	*ATP2B2*	6.17E-05
100028101|F|0--43:C>A	5.46E-04	*MYO5C*	2.63E-03
100027693|F|0--46:C>T	5.48E-04	No match	Not applicable
100077192|F|0--14:T>C	5.50E-04	No match	Not applicable
100015940|F|0--27:G>A	5.52E-04	No match	Not applicable
100043347|F|0--13:T>C	5.54E-04	No match	Not applicable
100046907|F|0--56:A>T	5.60E-04	*ACTR1B*	1.04E-14
100070603|F|0--21:A>G	5.68E-04	No match	Not applicable
100066500|F|0--22:T>C	5.73E-04	No match	Not applicable
100066592|F|0--37:A>C	5.76E-04	Uncharacterised protein	1.77E-05
100013407|F|0--18:C>T	5.77E-04	No match	Not applicable
100058010|F|0--16:G>A	5.78E-04	No match	Not applicable
100020917|F|0--34:G>A	5.79E-04	Uncharacterised protein	9.78E-09
100057066|F|0--14:T>G	5.85E-04	No match	Not applicable
100022261|F|0--33:C>A	5.85E-04	*RUNDC3A*	7.52E-04
100025271|F|0--41:G>C	5.85E-04	*PTGER3*	1.89E-11
100071010|F|0--9:G>T	5.87E-04	No match	Not applicable
100068727|F|0--12:C>T	5.97E-04	*NECTIN2*	4.16E-07
100044770|F|0--60:C>T	6.07E-04	No match	Not applicable
100062872|F|0--65:T>C	6.13E-04	*PTGER3*	3.41E-08
100058375|F|0--39:G>T	6.20E-04	No match	Not applicable
100077892|F|0--52:T>C	6.22E-04	Uncharacterised protein	2.80E-09
100054589|F|0--42:C>G	6.22E-04	Uncharacterised protein	7.52E-04
100050453|F|0--23:G>C	6.22E-04	*CD55*	3.41E-08
100066607|F|0--40:T>C	6.23E-04	No match	Not applicable

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100015098|F|0--46:G>A	6.23E-04	Uncharacterised protein	8.03E-10
100063506|F|0--39:G>A	6.25E-04	No match	Not applicable
100011713|F|0--27:T>C	6.29E-04	No match	Not applicable
100036979|F|0--35:T>C	6.31E-04	No match	Not applicable
100056343|F|0--36:G>A	6.33E-04	No match	Not applicable
100018608|F|0--61:G>C	6.38E-04	No match	Not applicable
100054991|F|0--32:C>T	6.40E-04	No match	Not applicable
100065947|F|0--32:C>T	6.40E-04	*CMYA5*	3.20E-02
100011470|F|0--20:T>C	6.44E-04	No match	Not applicable
100044214|F|0--34:T>G	6.44E-04	TRIM45	3.20E-02
100051478|F|0--62:T>C	6.45E-04	No match	Not applicable
100042947|F|0--50:T>C	6.49E-04	Uncharacterised protein	1.04E-14
100029558|F|0--8:A>G	6.63E-04	No match	Not applicable
100075787|F|0--66:C>T	6.64E-04	No match	Not applicable
100047612|F|0--38:T>C	6.70E-04	Uncharacterised protein	1.45E-06
100065002|F|0--38:G>A	6.74E-04	No match	Not applicable
100052715|F|0--52:G>A	6.76E-04	No match	Not applicable
100060690|F|0--48:A>G	6.76E-04	No match	Not applicable
100057291|F|0--42:G>A	6.77E-04	*ATP10D*	4.16E-07
100054791|F|0--14:C>A	6.81E-04	No match	Not applicable
100058211|F|0--40:C>T	6.82E-04	*CCNT1*	2.63E-03
100062417|F|0--40:C>T	6.91E-04	Uncharacterised protein	5.07E-06
100034867|F|0--31:A>G	6.92E-04	No match	Not applicable
100022122|F|0--20:T>C	6.94E-04	No match	Not applicable
100056915|F|0--11:C>T	6.94E-04	*PARK2-201*	1.77E-05
100059941|F|0--50:C>T	6.95E-04	No match	Not applicable
100022182|F|0--53:A>G	7.00E-04	*CCDC14*	2.63E-03
100044246|F|0--36:T>G	7.03E-04	No match	Not applicable
100046843|F|0--40:C>T	7.16E-04	*GRIK4*	2.15E-04
100049734|F|0--28:A>G	7.18E-04	No match	Not applicable
100035928|F|0--6:G>A	7.18E-04	No match	Not applicable
100042835|F|0--14:A>G	7.22E-04	No match	Not applicable
100022325|F|0--27:A>G	7.23E-04	No match	Not applicable
100044939|F|0--27:T>A	7.33E-04	No match	Not applicable
100057511|F|0--6:A>G	7.39E-04	No match	Not applicable
100064531|F|0--39:C>G	7.40E-04	No match	Not applicable
100041922|F|0--11:C>G	7.41E-04	No match	Not applicable
100054316|F|0--9:T>C	7.45E-04	No match	Not applicable

Significant SNPs	Association p-value	*Bos taurus* Orthologous Gene	Alignment E-value
100058092|F|0--50:C>G	7.46E-04	Uncharacterised protein	8.03E-10
100068436|F|0--16:G>A	7.56E-04	No match	Not applicable
100061723|F|0--22:G>T	7.66E-04	*MYH9*	7.52E-04
100060979|F|0--59:C>G	7.84E-04	Uncharacterised protein	9.16E-03
100062798|F|0--29:C>T	7.86E-04	*PTGER3*	1.45E-06
100024494|F|0--59:T>C	8.13E-04	*GLA*	1.04E-14
100049296|F|0--11:C>G	8.22E-04	No match	Not applicable
100069754|F|0--8:G>T	8.23E-04	*USP35*	9.16E-03
100051532|F|0--10:G>A	8.24E-04	No match	Not applicable
100066568|F|0--38:C>T	8.25E-04	No match	Not applicable
100051328|F|0--11:A>G	8.35E-04	No match	Not applicable
100035514|F|0--31:G>T	8.37E-04	No match	Not applicable
100059756|F|0--7:C>T	8.42E-04	No match	Not applicable
100076323|F|0--44:C>T	8.54E-04	*PTGER3*	3.41E-08
100059768|F|0--16:G>A	8.61E-04	No match	Not applicable
100073513|F|0--40:G>A	8.88E-04	No match	Not applicable
100015397|F|0--20:T>C	8.91E-04	No match	Not applicable
100069392|F|0--27:C>G	9.00E-04	*CACNA1G*	3.20E-02
100073819|F|0--19:C>T	9.03E-04	No match	Not applicable
100059615|F|0--6:C>T	9.03E-04	*PARK2-201*	3.41E-08
100077917|F|0--46:T>G	9.05E-04	No match	Not applicable
100031741|F|0--63:C>A	9.09E-04	Uncharacterised protein	2.63E-03
100012731|F|0--67:T>G	9.13E-04	No match	Not applicable
100071695|F|0--27:C>T	9.15E-04	*NUP205*	3.20E-02
100055333|F|0--6:T>A	9.22E-04	*CKAP2*	3.20E-02
100067031|F|0--28:A>G	9.29E-04	No match	Not applicable
100073505|F|0--6:A>G	9.31E-04	No match	Not applicable
100029512|F|0--31:A>G	9.39E-04	No match	Not applicable
100013832|F|0--42:T>C	9.48E-04	Uncharacterised protein	2.80E-09
100034667|F|0--7:T>A	9.57E-04	*PCSK4*	3.20E-02
100051706|F|0--9:A>G	9.59E-04	*CDC37*	2.30E-10
100064434|F|0--41:G>A	9.66E-04	No match	Not applicable
100052334|F|0--12:C>A	9.67E-04	No match	Not applicable
100067048|F|0--16:T>C	9.84E-04	No match	Not applicable
100048574|F|0--35:G>A	9.86E-04	*SLC28A2*	3.20E-02
100073886|F|0--11:G>C	9.86E-04	*VPS28*	1.19E-07

[1]Only the abbreviations of the orthologous genes are given.

APPENDIX C

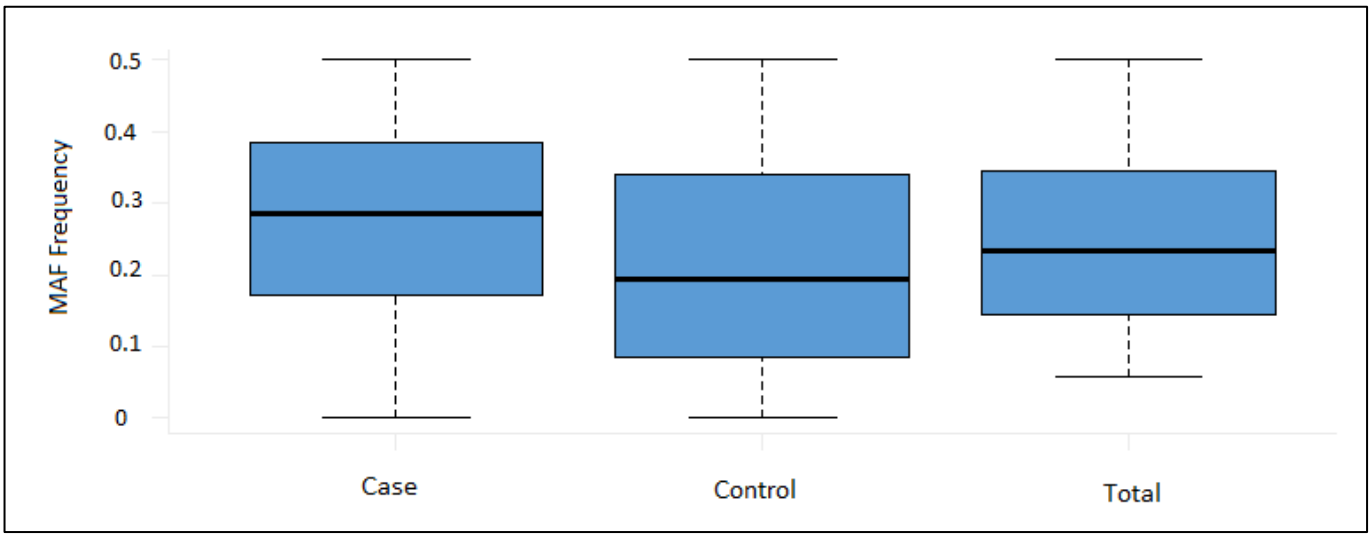


Figure A1. Minor allele frequency (MAF) distribution in the golden wildebeest (case) population, blue wildebeest (control) population and the total cohort.

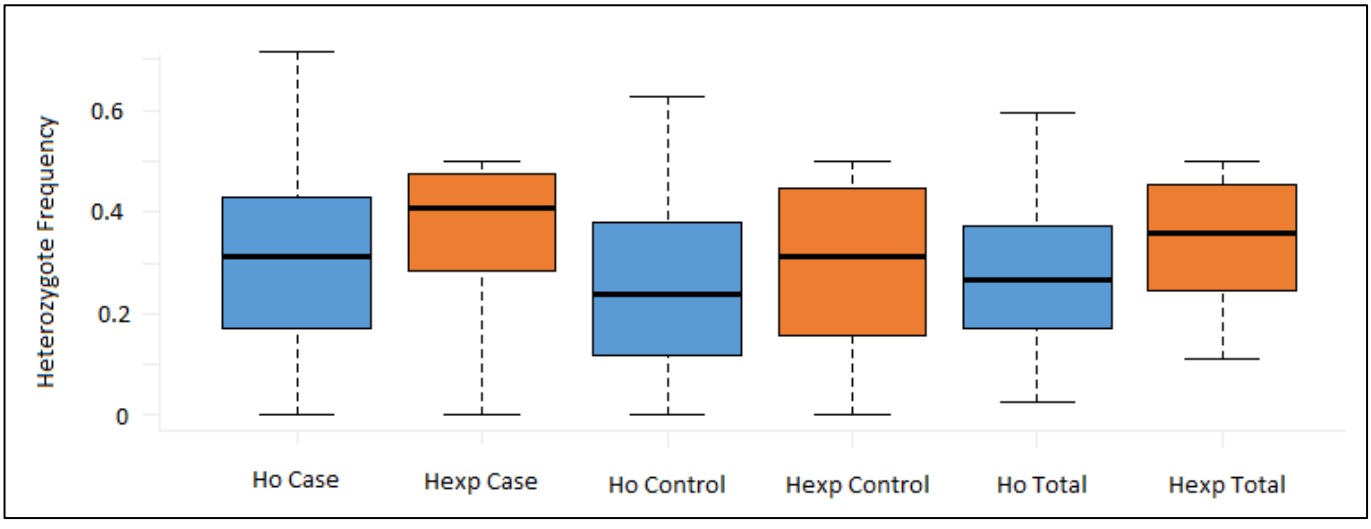


Figure A2. Observed (Ho) and expected (Hexp) heterozygote frequency distribution in the golden wildebeest (case) population, the blue wildebeest (control) population and the total cohort.

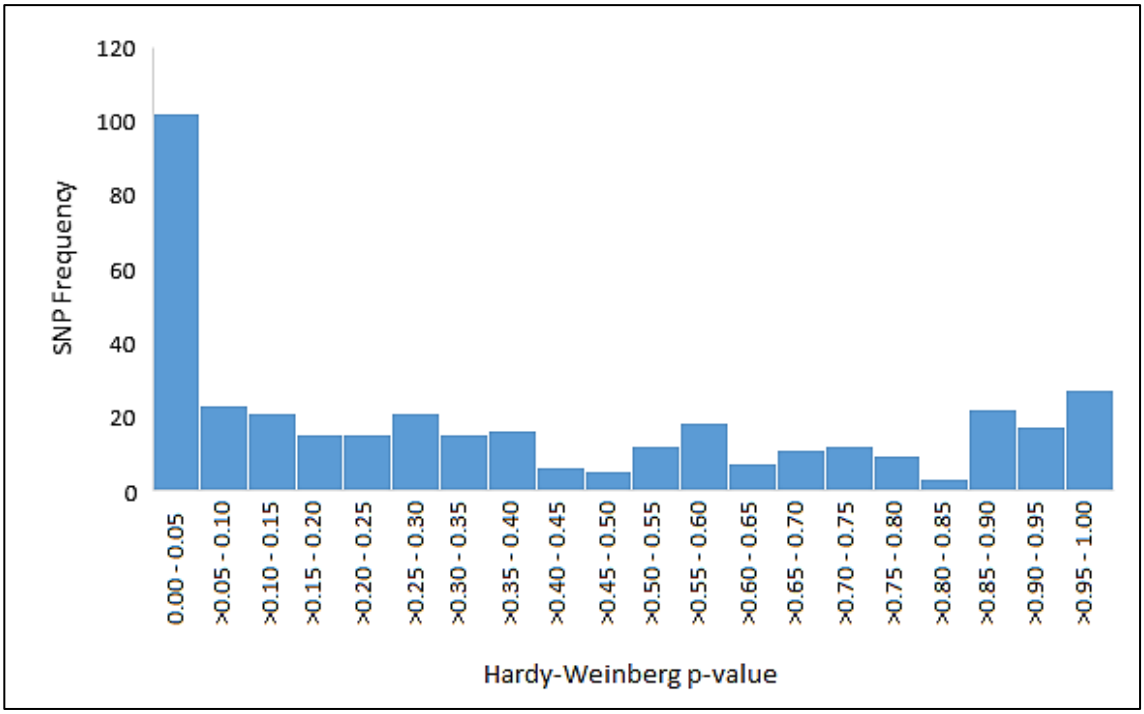


Figure A3. Distribution of Hardy-Weinberg p-values in the golden wildebeest (case) population.

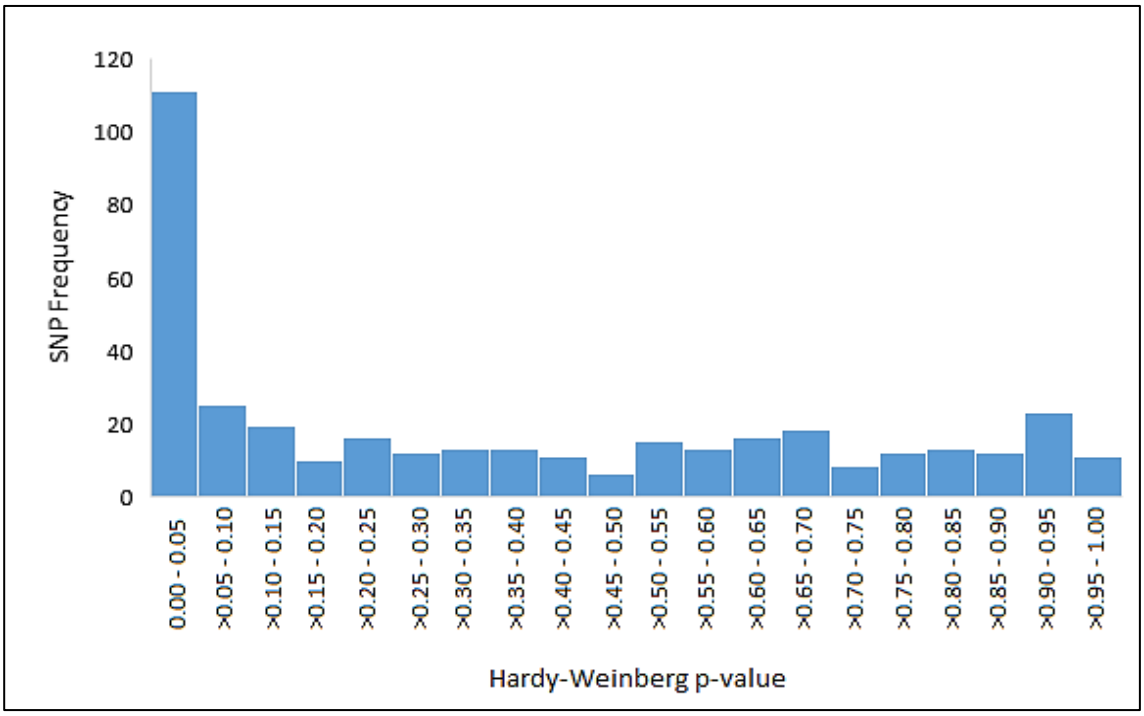


Figure A4. Distribution of Hardy-Weinberg p-values in the blue wildebeest (control) population.

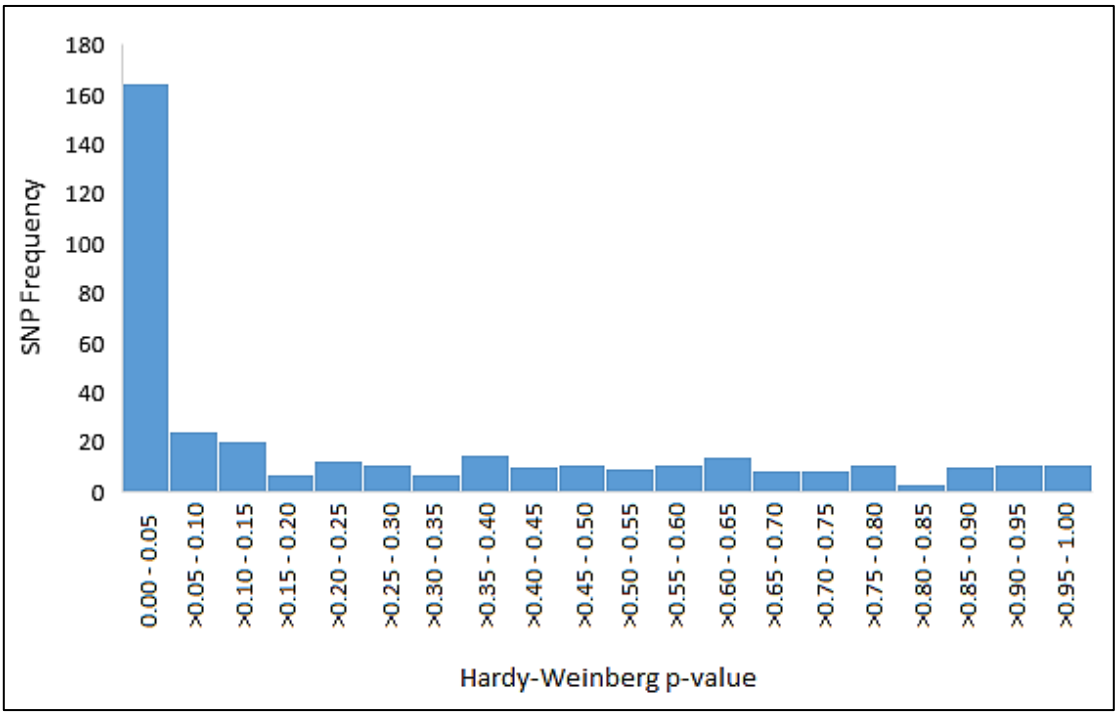


Figure A5. Distribution of Hardy-Weinberg p-values in the total blue wildebeest cohort.

APPENDIX D

Table A3. Association between SNPs, located in putative coat colour and genomic stability candidate genes, and coat colour in blue wildebeest under multiple models of inheritance.

SNP ID	Model	Genotype	Case	Control	Crude Analysis		AIC	BIC
					OR (95% CI)	P-value		
100028101|F|0--43:C>A	Co-dominant	C/C	17 (48.6%)	9 (15.5%)	1	0.0015	116.2	123.8
		A/C	14 (40%)	31 (53.5%)	4.18 (1.50-11.66)			
		A/A	4 (11.4%)	18 (31%)	8.50 (2.20-32.84)			
	Dominant	C/C	17 (48.6%)	9 (15.5%)	1	0.0006	115.5	120.6
		A/C-A/A	18 (51.4%)	49 (84.5%)	5.14 (1.95-13.59)			
	Recessive	C/C-A/C	31 (88.6%)	40 (69%)	1	0.025	122.1	127.2
		A/A	4 (11.4%)	18 (31%)	3.49 (1.07-11.36)			
	Over-dominant	C/C-A/A	21 (60%)	27 (46.5%)	1	0.21	125.6	130.7
		A/C	14 (40%)	31 (53.5%)	1.72 (0.74-4.03)			
	Log-additive	---	---	---	3.12 (1.57-6.17)	0.0004	**114.8**	**119.8**
100076096|F|0--31:A>G	Co-dominant	A/A	16 (47.1%)	52 (100%)	1	<0.0001	80.2	87.6
		G/A	8 (23.5%)	0 (0%)	0.00 (0.00-NA)			
		G/G	10 (29.4%)	0 (0%)	0.00 (0.00-NA)			
	Dominant	A/A	16 (47.1%)	52 (100%)	1	<0.0001	**78.2**	**83.1**
		G/A-G/G	18 (52.9%)	0 (0%)	0.00 (0.00-NA)			
	Recessive	A/A-G/A	24 (70.6%)	52 (100%)	1	<0.0001	98.8	103.7
		G/G	10 (29.4%)	0 (0%)	0.00 (0.00-NA)			
	Over-dominant	A/A-G/G	26 (76.5%)	52 (100%)	1	0.0001	103.3	108.2
		G/A	8 (23.5%)	0 (0%)	0.00 (0.00-NA)			
	Log-additive	---	---	---	0.00 (0.00-NA)	<0.0001	**78.2**	**83.1**
100025667|F|0--39:T>C	Co-dominant	T/T	35 (100%)	16 (27.6%)	1	<0.0001	69.4	77
		T/C	0 (0%)	34 (58.6%)	NA (0.00-NA)			
		C/C	0 (0%)	8 (13.8%)	NA (0.00-NA)			
	Dominant	T/T	35 (100%)	16 (27.6%)	1	<0.0001	**67.4**	**72.5**
		T/C-C/C	0 (0%)	42 (72.4%)	NA (0.00-NA)			
	Recessive	T/T-T/C	35 (100%)	50 (86.2%)	1	0.0047	119.2	124.2
		C/C	0 (0%)	8 (13.8%)	NA (0.00-NA)			
	Over-dominant	T/T-C/C	35 (100%)	24 (41.4%)	1	<0.0001	83.7	88.8
		T/C	0 (0%)	34 (58.6%)	NA (0.00-NA)			
	Log-additive	---	---	---	NA (0.00-NA)	<0.0001	**67.4**	**72.5**

SNP ID	Model	Genotype	Case	Control	Crude Analysis		AIC	BIC
					OR (95% CI)	P-value		
100071798|F|0--48:A>G	Co-dominant	A/A	7 (20.6%)	35 (61.4%)	1	0.0004	110.6	118.2
		A/G	20 (58.8%)	14 (24.6%)	0.14 (0.05-0.40)			
		G/G	7 (20.6%)	8 (14%)	0.23 (0.06-0.84)			
	Dominant	A/A	7 (20.6%)	35 (61.4%)	1	0.0001	**109.3**	**114.3**
		A/G-G/G	27 (79.4%)	22 (38.6%)	0.16 (0.06-0.44)			
	Recessive	A/A-A/G	27 (79.4%)	49 (86%)	1	0.42	123.6	128.6
		G/G	7 (20.6%)	8 (14%)	0.63 (0.21-1.93)			
	Over-dominant	A/A-G/G	14 (41.2%)	43 (75.4%)	1	0.0011	113.6	118.6
		A/G	20 (58.8%)	14 (24.6%)	0.23 (0.09-0.57)			
	Log-additive	---	---	---	0.40 (0.22-0.75)	0.0028	115.3	120.4
100048574|F|0--35:G>A	Co-dominant	G/G	29 (82.9%)	27 (46.5%)	1	0.0005	114	121.6
		G/A	6 (17.1%)	25 (43.1%)	4.48 (1.59-12.58)			
		A/A	0 (0%)	6 (10.3%)	NA (0.00-NA)			
	Dominant	G/G	29 (82.9%)	27 (46.5%)	1	0.0003	114.4	119.4
		G/A-A/A	6 (17.1%)	31 (53.5%)	5.55 (2.00-15.38)			
	Recessive	G/G-G/A	35 (100%)	52 (89.7%)	1	0.015	121.3	126.3
		A/A	0 (0%)	6 (10.3%)	NA (0.00-NA)			
	Over-dominant	G/G-A/A	29 (82.9%)	33 (56.9%)	1	0.008	120.2	125.2
		G/A	6 (17.1%)	25 (43.1%)	3.66 (1.32-10.17)			
	Log-additive	---	---	---	5.06 (1.93-13.26)	0.0001	**112.6**	**117.6**
100028737|F|0--18:G>A	Co-dominant	G/G	10 (28.6%)	34 (58.6%)	1	0.0044	118.3	125.9
		G/A	17 (48.6%)	21 (36.2%)	0.36 (0.14-0.94)			
		A/A	8 (22.9%)	3 (5.2%)	0.11 (0.02-0.50)			
	Dominant	G/G	10 (28.6%)	34 (58.6%)	1	0.0044	119.1	124.1
		G/A-A/A	25 (71.4%)	24 (41.4%)	0.28 (0.11-0.69)			
	Recessive	G/G-G/A	27 (77.1%)	55 (94.8%)	1	0.012	120.8	125.9
		A/A	8 (22.9%)	3 (5.2%)	0.18 (0.05-0.75)			
	Over-dominant	G/G-A/A	18 (51.4%)	37 (63.8%)	1	0.24	125.8	130.9
		G/A	17 (48.6%)	21 (36.2%)	0.60 (0.26-1.41)			
	Log-additive	---	---	---	0.34 (0.17-0.67)	0.001	**116.3**	**121.4**
100070025|F|0--27:C>T	Co-dominant	C/C	18 (54.5%)	44 (80%)	1	0.017	114.3	121.8
		C/T	7 (21.2%)	8 (14.6%)	0.47 (0.15-1.48)			
		T/T	8 (24.2%)	3 (5.5%)	0.15 (0.04-0.64)			
	Dominant	C/C	18 (54.5%)	44 (80%)	1	0.012	114.1	119.1
		C/T-T/T	15 (45.5%)	11 (20%)	0.30 (0.12-0.78)			
	Recessive	C/C-C/T	25 (75.8%)	52 (94.5%)	1	0.011	114	118.9

SNP ID	Model	Genotype	Case	Control	Crude Analysis		AIC	BIC
					OR (95% CI)	P-value		
100075122|F|0 --28:G>A		T/T	8 (24.2%)	3 (5.5%)	0.18 (0.04-0.74)			
	Over-dominant	C/C-T/T	26 (78.8%)	47 (85.5%)	1	0.43	119.8	124.8
		C/T	7 (21.2%)	8 (14.6%)	0.63 (0.21-1.94)			
	Log-additive	---	---	---	0.41 (0.21-0.78)	0.0046	**112.4**	**117.4**
	Co-dominant	G/G	18 (51.4%)	39 (69.6%)	1	0.057	121.5	129.1
		G/A	11 (31.4%)	15 (26.8%)	0.63 (0.24-1.64)			
		A/A	6 (17.1%)	2 (3.6%)	0.15 (0.03-0.84)			
	Dominant	G/G	18 (51.4%)	39 (69.6%)	1	0.082	122.2	127.3
		G/A-A/A	17 (48.6%)	17 (30.4%)	0.46 (0.19-1.11)			
	Recessive	G/G-G/A	29 (82.9%)	54 (96.4%)	1	0.028	120.4	125.4
		A/A	6 (17.1%)	2 (3.6%)	0.18 (0.03-0.94)			
	Over-dominant	G/G-A/A	24 (68.6%)	41 (73.2%)	1	0.63	125	130.1
		G/A	11 (31.4%)	15 (26.8%)	0.80 (0.32-2.02)			
	Log-additive	---	---	---	0.47 (0.24-0.92)	0.024	**120.2**	**125.2**
100030072|F|0--58:C>T	Co-dominant	C/C	12 (40%)	43 (87.8%)	1	<0.0001	90.5	97.6
		T/C	7 (23.3%)	3 (6.1%)	0.12 (0.03-0.53)			
		T/T	11 (36.7%)	3 (6.1%)	0.08 (0.02-0.32)			
	Dominant	C/C	12 (40%)	43 (87.8%)	1	<0.0001	**88.7**	**93.4**
		T/C-T/T	18 (60%)	6 (12.2%)	0.09 (0.03-0.29)			
	Recessive	C/C-T/C	19 (63.3%)	46 (93.9%)	1	0.0006	97.1	101.8
		T/T	11 (36.7%)	3 (6.1%)	0.11 (0.03-0.45)			
	Over-dominant	C/C-T/T	23 (76.7%)	46 (93.9%)	1	0.028	104.1	108.8
		T/C	7 (23.3%)	3 (6.1%)	0.21 (0.05-0.91)			
	Log-additive	---	---	---	0.24 (0.12-0.50)	<0.0001	89.7	94.4
100057057|F|0--49:C>T	Co-dominant	C/C	23 (65.7%)	56 (96.5%)	1	0.0002	112.5	120.1
		C/T	11 (31.4%)	2 (3.5%)	0.07 (0.02-0.36)			
		T/T	1 (2.9%)	0 (0%)	0.00 (0.00-NA)			
	Dominant	C/C	23 (65.7%)	56 (96.5%)	1	0.0001	110.8	115.8
		C/T-T/T	12 (34.3%)	2 (3.5%)	0.07 (0.01-0.33)			
	Recessive	C/C-C/T	34 (97.1%)	58 (100%)	1	0.16	125.2	130.3
		T/T	1 (2.9%)	0 (0%)	0.00 (0.00-NA)			
	Over-dominant	C/C-T/T	24 (68.6%)	56 (96.5%)	1	0.0002	112.9	118
		C/T	11 (31.4%)	2 (3.5%)	0.08 (0.02-0.38)			

SNP ID	Model	Genotype	Case	Control	Crude Analysis		AIC	BIC
					OR (95% CI)	P-value		
	Log-additive	---	---	---	0.07 (0.02-0.35)	<0.0001	**110.5**	**115.6**