Genome-wide scan of single nucleotide polymorphisms for parentage analyses in South African indigenous beef breeds

by

Yandisiwe Patience Sanarana

Submitted in fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR (ANIMAL SCIENCE)

In the

Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

October 2022

SUPERVISORY COMMITTEE

Prof E. van Marle-Köster: University of Pretoria
Department of Animal and Wildlife Sciences
Pretoria
Hatfield
0002
South Africa

Prof A. Maiwashe: Agricultural Research Council
Animal Production Institute
Pretoria
Irene
0062
South Africa

Prof D.P. Berry: Teagasc
Animal and Grassland Research, and Innovation Center
Fermoy
County Cork
P61 C996
Ireland

DECLARATION

I, Yandisiwe Patience Sanarana, declare that this thesis/dissertation, which I hereby submit for the degree PhD Animal Science at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: ..Ysanarana.....

DATE: ..07/06/2022.....

ACKNOWLEDGEMENTS

First, I want to express my heartfelt gratitude to the study supervisory committee for providing constructive pieces of advice, support and guidance throughout this research. I want to thank Professor Azwihangwisi Maiwashe and Professor Este van Marle-Köster for enabling me to grow in research and offering opportunities to study further from MSc degree to the now completed PhD. I can't begin to explain how grateful I am to Professor Donagh Berry for all his support and guidance on statistical analysis, and polishing of all the material presented in this thesis. I will always be appreciative to Professor Cuthbert Banga for the opportunities he provided and words of encouragement that kept me sailing through.

From the beginning to the end of this project, a number of institutions contributed through funding, resources, academic experienced expertise, and technical skills. Among the institutions I would like to thank for their contributions are, the Agricultural Research Council, Beef Genomics Project (BGP), Breeder's Societies of the Bonsmara and Drakensberger, National Research Foundation (NRF), South African Studbook, Technology Innovation Agency (TIA), and the University of Pretoria. The TEAGASC Animal and Grassland Research, and Innovation Center in Ireland deserves special recognition for enabling me to undertake research training and learn from its supreme.

"As iron sharpens iron", so are these friends. I want to thank Dikonketso Matjuda and Tumelo Kekana for their precise words that encouraged me to never give up. I also want to express my appreciation to my colleagues, Dr. Simon Lashmar and Ms. Hannelize Swart for their willing to assist with data. Fellow colleagues and friends at Agricultural Research Council and University of Pretoria, I cannot mention you all by names, thank you for your moral support.

Family is everything, thanks to my dad and siblings without your love, patience and support I would not have progressed to this far. Special thanks to my granny, Norih Lolonga who raised the little girl in me to be who I am. I will never forget my aunts who all have supported me following my mother's death (Neliswa Lolonga "MamQocwa" Sanarana) just in the second year of the study. To my daughter, "office and lab mate" Imitha Khazimla Linomtha, thank you for your smiles and warm embraces. I would want to express my gratitude to 3C family in Centurion for helping me maintain my spiritual well-being throughout the years.

To all the farmers, *"Be sure you know the condition of your flocks, give careful attention to your herds" - Proverbs 27:23.*

EXECUTIVE SUMMARY

A number of South Africa (SA) beef cattle including the developed Bonsmara and an indigenous Drakensberger have recently been genotyped using single nucleotide polymorphism (SNP) high-density panels for the establishment of genomic selection programs. This has provided an opportunity to use the SNP data for parentage analyzes in both breeds. The genomic information from the GeneSeek® Genomic Profiler (GGP) Bovine SNP150K BeadChip was used to achieve the aim of the current study. The aim was achieved and organized into three experimental chapters (i.e., Chapter 3, 4 and 5). Before these three chapters, there is Chapter 1 that is an introductory chapter and Chapter 2 that is reviewing a literature related to the objectives on the study. The first experimental Chapter 3 utilized the confirmed sire-offspring pairs to test the efficiency of the ISAG SNP panel and has been published in the journal of Tropical Animal Health and Production. The mismatches of SNP genotypes detected between the confirmed parent-offspring pairs were subsequently used to identify the carriers of hemizygous deletions and regions exhibiting these regions in the genome of these cattle breeds in Chapter 4. The development and testing of the performance of low-cost genotype panels for both breeds are presented in Chapter 5. The thesis is concluded in Chapter 6 with a general discussion, and recommendations for future studies. Except for the published Chapter 3, which has an abstract, a conclusion interwoven into a discussion, and a unique referencing style adhering to the journal's standards, Chapter 4 and 5 are written in a similar format and will be submitted in peer reviewed for publication.

ABSTRACT

Genotyping panels using single nucleotide polymorphism (SNP) markers have superseded the use of microsatellites for parentage testing and other applications. High-density SNP arrays enable accurate parentage assignment but they are not always practical for routine application. The major constraints include high cost coupled with the time spent in analyzing the results. The International Society for Animal Genetics (ISAG) recommends a low-density SNP panel consisting of 200 genomic markers for parentage testing globally. However, SNPs specific to Sanga cattle breeds were not included in constructing the panel, and its utility for use has not been verified in Sanga cattle. This thesis, therefore, firstly tested the usefulness of these ISAG markers for parentage testing in Sanga cattle with verified parentage. The results demonstrated a poor performance of the ISAG panel in both Bonsmara (BON) and Drakensberger (DRB) for parentage verification, with false negatives ranging from 23.4% (BON) to 33% (DRB). This implied that relying on the ISAG panel alone may cause incorrect exclusions and, at times, be unable to determine parentages when closely related candidate parents are considered. The second objective of the thesis was to quantify the accuracy of parentage recording but also the detection of individuals exhibiting hemizygous deletions using 91 185 autosomal SNPs. On average, 8.5% to 10.1% of parentage errors were detected in the recorded BON and DRB pedigree, respectively indicating recording errors on farms. The discovery of the parent-progeny pairs based on SNP genotypes was possible for 69 relationships that had not actually been recorded. Eleven suggestive regions of hemizygosity were detected on 10 chromosomes (3, 6, 7, 9, 10, 11, 17, 24, 25 and 28) in animals mostly used for breeding which included 6 (BON) and 9 (DRB) bulls and as well 10 (BON) and 8 (DRB) cows. Finally, low-density genotype SNP panels for parentage testing consisting of 200 markers were developed. SNPs were chosen to be informative both within and across breeds. The methods to select informative SNPs considered high minor allele frequency (MAF), good clustering quality, and high call rates. SNPs were pruned to reduce the linkage disequilibrium among markers with a minimum distance of one Mb apart. All the panels selected were tested per breed. On average, the genotype panel with SNPs selected across the breeds had a lower MAF of 0.40 compared to 0.48 (DRB) and 0.49 (BON) selected within breeds. SNP markers selected within breed were more accurate at parentage testing with no false negative whereas 4.2% false negatives were observed in the BON using the multi-breed panel. The methods and results presented in this thesis can be used in the construction of parentage SNP panels to provide parentage verification which will contribute to quality control in breeding systems and thus accelerate genetic improvement.

THESIS OUTPUTS

Scientific publications:

Peer-review Journals

Sanarana, **Y.P**., Maiwashe, A., Berry, D.P., Banga, C. & van Marle-Köster, E., 2021. Evaluation of the International Society for Animal Genetics bovine single nucleotide polymorphism parentage panel in South African Bonsmara and Drakensberger cattle. Tropical Animal Health and Production, 53,1, 32. doi.org/10.1007/s11250-020-02481-6.

Congresses:

National

Sanarana, Y.P., Maiwashe, A. & van Marle-Köster, E., 2017. Performance of different parentage SNP panels in South African beef cattle breeds. Proceedings of the 50th South African Society for Animal Science in Port Elizabeth, South Africa, 18 th – 21st September, 2017.

Sanarana, Y.P., Maiwashe, A., Banga, C.B., Berry, D. & van Marle-Köster, E., 2019. Evaluation of low-density SNP panel for parentage verification in South African beef cattle. Proceedings of the 6th Agricultural Research Council National Postgraduate Development Program Conference in Pretoria, South Africa, 7-9th October, 2019.

Sanarana, Y.P., Maiwashe, A., Banga, C.B., Berry, D. & van Marle-Köster, E., 2021. Identification of autosomal hemizygous regions based on Mendelian inconsistencies of parent offspring SNP genotypes in the Bonsmara and Drakensberger populations. Proceedings of the 52nd South African Society for Animal Science in Pretoria, South Africa, 10-12th August, 2021.

International

Sanarana, Y.P., Maiwashe, A., Banga, C.B., Berry, D. & van Marle-Köster, E., 2020. Validation of the ISAG bovine SNP panel for parentage verification in South African Bonsmara and Drakensberger cattle. Proceedings of the Plant and Animal Genomes XXVIII Conference in San Diego, USA, 11-15 January 2020.

TABLE OF CONTENTS

Pages

LIST OF TABLES

LIST OF FIGURES

LIST OF ADDENDA

LIST OF ABBREVIATIONS

AEC	Animal Ethics Committee
AI	Artificial Insemination
ARC	Agricultural Research Council
ARS-USMARC	Agricultural Research Service of the United States Meat Animal Research Center
BGP	Beef Genomics Project
BLUP	Best Linear Unbiased Prediction
CNVs	Copy number of variations
DALRRD	Department of Agriculture, Land Reform and Rural Development
DNA	Deoxyribonucleic Acid
EBV	Estimated breeding value
FAO-UN	Food and Agriculture Organization
GC	GenCall
GGP	GeneSeek Genomic Profiler
GRM	Genomic relationship matrix
GT	GenTrain
GEBV	Genomic estimated breeding value
GenoEx-PSE	Genotype Exchange - Parentage SNP Exchange
Hdel	Hemizygous deletion
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HWE	Hardy Weinberg Equilibrium
ICAR	International Committee for Animal Recording
ISAG	International Society for Animal Genetics
LD	Linkage disequilibrium
LR	Likelihood ratio
MAF	Minor allele frequency
NRF	National Research Foundation
NGS	Next-generation sequencing
PIC	Polymorphic information content
P_E	Probability of parentage exclusion
P_I	Probability of identity
Q-Q	Quantile-quantile
QTL	Quantitative trait loci

RMRD	Red Meat Research Development
ROH	Runs of homozygosity
SA	South Africa
SNP	Single nucleotide polymorphism
TIA	Technology Innovation Agency
UP	University of Pretoria
USDA	United States Department of Agriculture

CHAPTER 1

1.1. BACKGROUND AND MOTIVATION

Beef has been a primary source of animal protein, and its demand continues to increase due to population growth, urbanization, economic development, and shifting consumer preferences (Smith *et al.*, 2018). In spite of production and disease challenges exacerbated by adverse effects of climate change, it is predicted that demand for beef would rise by about 23% in the next decade (FAO., 2021). Although the South African (SA) beef industry only accounts for about 1.4% of the total world beef production, it is the largest (21.4%) beef producer in Africa (USDA Foreign Agricultural Service., 2020). Locally, beef boasts the second fastest growing markets for livestock after poultry, contributing about R320 million (12%) to the total gross value of agricultural products (DALRRD, 2020). This is attributable to a rise in the price of beef compounded by accelerated growth in exports.

A number of developing countries in Africa are confronted with several socio-economic and livestock production challenges and constraints (FAO., 2021). For this reason, beef industries should seize opportunities to contribute to food supply and sustainable economic growth by applying innovative breeding strategies to increase production efficiency. To date, innovative breeding strategies to improve beef production have not been readily available in many African countries to the same extent as in first world countries underpinned by well-established conventional genetic evaluation systems (Mrode *et al.*, 2019). Phenotyping and pedigree recording, which are essential practices for effective breeding and selection, are an ever-present challenge in livestock production systems in many African countries (Marshall *et al.*, 2019; Visser *et al.*, 2020).

South Africa is among the few countries in Africa with a long history of official national beef recording and improvement programs (Bergh, 2010). Sustainable animal recording systems, which provide the means to collect and analyze pedigree and performance data for genetic evaluations, have led to significant genetic progress in beef production (Van Marle-Köster *et al.*, 2013). For the past three decades, genetic evaluations have provided breeders with estimated breeding values (EBVs) as a tool for the selection of the best candidates for breeding (Van Marle-Köster & Visser, 2018a). The reliability of these EBVs depends largely on the accuracy of pedigree data (as well as the quality of phenotypic data). Due to the extensive nature of beef production, accurate parentage recording can be challenging. Errors in parentage causes bias in genetic parameter estimates as well as the genetic evaluations thereby reducing the rate of achievable genetic progress (Israel &Weller, 2000; Baron *et al.*, 2002; Harder *et al.*, 2005). Hence, it is important to ensure the accuracy of pedigree information to achieve high rates of genetic improvement.

Parentage testing using molecular markers has been used for several decades to improve the accuracy of recorded pedigree data. Blood group typing was the first method used in cattle (Stormont, 1967). This method was effective for parent-offspring exclusion, with approximately 96-98 % accuracy (Rendel, 1958), but it was not easy to perform parentage testing due to the intricacy of the procedure and handling of blood samples. Over time, informative DNA-based methods such as multi or bi-allelic markers were discovered (Vignal *et al.*, 2002). In the early 1990s, the multi-allelic microsatellite markers for parentage verification with a reference panel consisting of 12 polymorphic markers were standardized by the International Society for Animal Genetics (ISAG) (Usha *et al.*, 1995). Limitations of the panel have been reported, and thus more markers were required to increase its utility (Kelly *et al.*, 2011). Other studies have highlighted a growing inability of microsatellites to exclude parentages of closely related individuals (Baruch & Weller, 2008; Brenig & Schütz, 2016).

The development of next-generation sequencing (NGS) technologies at the turn of the twentieth century led to the discovery of high-density bi-allelic single nucleotide polymorphisms (SNPs) markers (Matukumalli *et al.*, 2009). The availability of SNP markers facilitated the selection of the best animals for breeding at a young age through genomic selection which could also be used for pedigree validation (Hayes, 2011). Single nucleotide polymorphisms have replaced microsatellite markers for parentage profiling in most countries in the developed world, while the transition has been slow in the developing world, especially in Africa. Major limitations have been associated with costs and delays in establishing genomic selection programs (Van Marle-Köster & Visser, 2018b; Mrode *et al.*, 2019). Research to verify the integrity of livestock pedigree records in Africa has been restricted to microsatellite markers (Kios *et al.*, 2012).

The implementation of the National Beef Genomic Program (BGP) in 2015 presented an opportunity to exploit SNP technology for genomic selection in SA beef breeds (Walsh & Spazzoli, 2018). The BGP database hosts beef cattle genotypes from the GeneSeek Genomic Profiler (GGP) Bovine 150K for SA indigenous breeds of Sanga type such as the Afrikaner, Drakensberger, Nguni, as well as the locally developed Bonsmara. The Bonsmara is numerically the largest indigenous beef cattle breed (SA Stud Book, 2016), making up about 35.8% of the beef cattle population, compared to the Nguni (11.3%), Drakensberger (3.9%) and Afrikaner (2.2%) (Van Marle-Köster & Visser, 2018b). Likewise, the Bonsmara is a well-established breed in the developed sector with substantial genotypic information in the BGP, followed by the Drakensberger, Nguni, and the Afrikaner (Van Marle-Köster *et al.*, 2021).

The structure of the genome-wide SNP data available for the Bonsmara and Drakensberger has provided a valuable resource for pedigree verification and parentage assignment in beef populations as well as for effective use of genomic predictions in breeding programs. Accurate pedigree and having selection candidates more closely related to the reference population increases the reliability of genomic breeding values, but can also contribute to an increase in selection accuracy (Junqueira *et al.*, 2017). The SNP

allows the detection of hemizygous deletions, known as the phenomena of the copy number of variations (CNVs), which appear as false homozygous genotypes after SNP-chip genotyping (Himmelbauer *et al.*, 2019). Detection of false indications of homozygote calling is of interest since they can lead to incorrect parent-offspring exclusion and in turn affect parentage outcomes when they are not well defined in the ancestry of a population.

The genomic tools for parentage testing are advantageous in facilitating optimized breeding programs if they are technically accurate, inexpensive, or require minimal effort for downstream analysis (Berry *et al.*, 2019). Affordable low-density bovine SNP panels have been developed and applied to parentage verification in several countries (Heaton *et al.*, 2002; ISAG, 2013; McClure *et al.*, 2015) to improve the accuracy of breeding values for selection. Following these developments, the International Committee for Animal Recording (ICAR) implemented a service referred to as GenoEx-PSE to expedite the transition to SNP-based parentage verification. This provides a platform for the global exchange of SNP genotypes for parentage analyses of cattle in the Interbull pedigree database (Dürr *et al.*, 2011; Interbull, 2016). Nonetheless, the performance of the low-density ISAG SNP panel recommended for parentage verification by the ISAG-ICAR committee has not been tested in Sanga breeds. SNPs included in the development of the panel were selected non-randomly from highly informative SNPs predominantly from *Bos taurus* and *Bos indicus* populations thereby excluding many of the world's cattle breeds (Strucken *et al.*, 2014). This selection process may impact the informativeness of these SNPs for such under-represented breeds.

1.2. The aim of the research

The aim of this thesis was to perform a genome-wide scan of single nucleotide polymorphisms for parentage analysis in South African Bonsmara and Drakensberger beef cattle breeds.

To achieve the aim of this study, the following objectives were pursued:

Objectives

- To evaluate the effectiveness of the bovine ISAG SNP parentage panel in the SA Bonsmara and Drakensberger beef cattle.
- To validate pedigree records and perform parentage assignment of the two beef cattle breeds using whole-genome data.
- To detect inherited regions with possible hemizygosity in these two cattle populations.
- To develop a low-density SNP genotype panel suitable for SA Bonsmara and Drakensberger cattle breeds.

References

Baruch, E. & Weller, J.I., 2008. Estimation of the number of SNP genetic markers required for parentage verification. Animal Genetics, 39, 474–479.

Baron, E.E., Martinez, M.L., Verneque, R.S. & Coutinho, L.L., 2002. Parentage testing and effect of misidentification on the estimation of breeding value in Gir cattle. Genetics and Molecular Biology, 25, 4, 389-394.

Bergh, L., 2010. The national beef recording and improvement scheme. In: Beef Breeding in South Africa. 2nd ed. Ed. Scholtz, M.M., Agricultural Research Council, Pretoria, South Africa. 55-70.

Berry, D.P., McHugh, N., Wall, E., McDermott, K. & O'Brien, A.C., 2019. Low-density genotype panel for both parentage verification and discovery in a multi-breed sheep population. Irish Journal of Agricultural Food Research, 58, 1-12.

Brenig, B. & Schűtz, E., 2016. Recent development of allele frequencies and exclusion probabilities of microsatellites used for parentage control in the German Holstein Friesian cattle population. BMC Genetics, 17, 1-9.

Department of Agriculture, Land Reforms and Rural Development (DALRRD)., 2020. Trends in Agricultural Sector. Available online:

https://www.dalrrd.gov.za/Portals/0/Statistics%20and%20Economic%20Analysis/Statistical%20Information/Trends%20in%20the%20Agricultural%20Sector%202020.pdf

Dürr, J.W., Forabosco, F., Jakobsen, J. & Zumbach, B., 2011. Interbull Survey on Parentage Verification. Interbull bulletin no. 44. Stavanger, Norway, August 26 – 29.

Food and Agriculture Organization of the United Nations (FAO)., 2021. Africa – Regional Overview of Food Security and Nutrition 2021: Statistics and trends. Available online: https://www.fao.org/documents/card/en/c/cb7496en

Harder, B., Bennewitz, J., Reinsch, N., Mayer, M. & Kalm, E., 2005. Effect of missing sire information on genetic evaluation. Archives Tierzuctch, 48, 219–232.

Hayes, B.J., 2011. Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. Journal of Dairy Science, 94, 2114-2117

Heaton, M.P., Harhay, G.P., Bennett, G.L., Stone, R.T., Grosse, W.M., Casas, E., Keele, J.W., Smith, T.P.L., Chitko-McKown, G.C. & Laegreid, W.W., 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. Mammalian Genome, 13, 272-281.

Himmelbauer, J., Mészáros, G. & Sölkner, J., 2019. Detection of autosomal hemizygous regions in the Fleckvieh population-based on SNP-chip data and parent-offspring pairs. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 67, 6, 1447–1452.

Interbull Meeting., 2016. Puerto Varas, Chile, from October 24th-28th. Available online: https://interbull.org/ib/2016_interbull_meeting#:~:text=The%202016%20Interbull%20Annual%20Meeting,the%2040th%20ICAR%20biennial%20session.

International Society for Animal Genetics (ISAG)., 2013. ISAG cattle core and additional SNP panel. https://www.isag.us/committees.asp

Israel, C. & Weller, J.I., 2000. Effect of misidentification on genetic gain and estimation of breeding values in dairy cattle populations. Journal of Dairy Science, 83, 181-187.

Kelly, A.C., Mateus-Pinilla, N.E., Douglas, M., Douglas, M., Shelton, P. & Novakofski, J., 2011. Microsatellites behaving badly: empirical evaluation of genotyping errors and subsequent impacts on population studies. Genetics and Molecular Research 10, 4, 2534-2553.

Kios, D., van Marle-Köster, E. & Visser, C., 2012. Application of DNA markers in parentage verification of Boran cattle in Kenya. Tropical Animal Health Production, 44, 471-476.

Junqueira, V.S., Cardoso, F.F., Oliveira, M.M., Sollero, B.P., Silva, F.F. & Lopes, P.S., 2017. Use of molecular markers to improve relationship information in the genetic evaluation of beef cattle tick resistance under pedigree-based models. Journal of Animal Breeding and Genetics, 133, 1-13.

Marshall, K., Gibson, J.P., Mwai, O., Mwacharo, J.M., Haile, A., Getachew, T., Mrode, R. & Kemp, S.J., 2019. Livestock genomics for developing countries—African examples in practice. Frontiers in Genetics, 10, 297.

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., 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.

McClure, M.C., McCarthy, J., Flynn, P., Weld, R., Keanne, M., O'Connell, K., Mullen, M.P., Waters, S. & Kearney, J.F., 2015. SNP selection for nationwide parentage verification in beef and dairy cattle. Proceeding of the International Committee of Animal Recording (ICAR), Technical series, 19, 175-181. Krakow, Poland, June 10-12.

Mrode, R., Ojango, J., Okeyo, A. M. & Mwacharo, J. M., 2019. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. Frontiers in genetics, 9, 694.

Rendel, J., 1958. Studies of cattle blood groups. II. Parentage tests. Acta Agriculturae Scandinavica, 8, 131–161.

Smith, S.B., Gotoh, T. & Greenwood, P.L., 2018. Current situation and future prospects for global beef production: overview of special issue. Asian-Australasian Journal of Animal Sciences, 31, 7, 927.

Stormont, C., 1967. Contribution of blood typing to dairy science progress. Journal of Dairy Science, 50, 253-260.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H. & Gondro, C., 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, 45, 572-575.

South African (SA) Stud Book, 2016. SA Stud Book annual report. Available online: http://www.sastudbook.co.za/images/photos/Annual_Report_2016_a.pdf.

United States Department of Agriculture (USDA) Foreign Agricultural Service., 2020. South Africa Unable to Stifle Latest Outbreak of Foot and Mouth Disease. Available online: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=South%20Africa%20Unable%20to%20Stifle%20Latest%20Outbreak%20of%20Foot%20and%20Mouth%20Disease_Pretoria_South%20Africa%20-%20Republic%20of_01-07-2020.

Usha, A. P., Simpson, S. P. & Williams, J. L., 1995. Probability of random sire exclusion using microsatellite markers for parentage verification. Animal Genetics, 26, 155–161.

Van Marle-Köster, E., Visser, C. & Berry, D.P., 2013. A review of genomic selection - Implications for the South African beef and dairy cattle industries. South African Journal of Animal Science, 1, 43, 1-17.

Van Marle-Köster, E. & Visser, C., 2018a. Genetic improvement in South African livestock: can genomics bridge the gap between the developed and developing sectors? Frontier in Genetics, 9, 331.

Van Marle-Köster, E. & Visser, C., 2018b. Genomics for the advancement of livestock production: A South African perspective. South African Journal of Animal Science, 48, 5, 808-817.

Van Marle-Köster, E., Visser, C., Sealy, J. & Frantz, L., 2021. Capitalizing on the potential of South African indigenous beef cattle breeds: A Review. Sustainability, 13, 8, 1-10.

Vignal, A., Milan, D., Sancristobal M. & Eggen A., 2002. SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34, 275-305.

Visser, C., Van Marle-Köster, E., Myburgh, H.C. & De Freitas, A., 2020. Phenomics for sustainable production in the South African dairy and beef cattle industry. Animal Frontiers, 10, 2, 12-18.

Walsh, K. & Spazzoli, R., 2018. Assessing the economic impact of the South African beef genomics program. Available online: http://wagyu.org.za/wp-content/uploads/2018/08/Impact-of-the-Beef-Genomics-Program-FinalReport_060418-Round-1.pdf.

CHAPTER 2

2. Literature review

2.1. Introduction

Genetic evaluations entails using information on selection candidates and their relatives to predict breeding values (EBVs) for decision-making in breeding the next generation (Henderson, 1973). In this process, incorrect parentage records can substantially impact the accuracy of evaluation (Harder *et al.*, 2005). Good pedigree completeness and performance records are essential for effective selection, control of inbreeding, and detection of genetic defects (Santana *et al.*, 2012). The main effect of pedigree misidentifications is a reduction in the rate of genetic gain, predominantly due to erroneous prediction of breeding values, biased heritability estimates or variance components, and the accumulation of inbreeding among relatives (Van Vleck, 1970; Harder *et al.*, 2005). Furthermore, this could also result in a reduction in the ability to detect linkage for genetic disorders and quantitative traits (Belay, 2013). Unfortunately, the extent of pedigree errors are sometimes understated and largely unreported.

Genomic information has many applications facilitating more precise mating and management decisions. This includes parentage assignment, traceability, identification of hemizygous deletions, estimation of co-ancestry, and quantifying breed composition. This chapter examines the role of recent developments in molecular marker technology, with a particular focus on the importance of parentage accuracy in beef cattle populations as a cost-effective breeding strategy for optimal selection.

2.2. Brief overview of genetic improvement in SA beef cattle

Many decades of genetic improvement and selection of cattle adapted to the diverse SA environmental conditions resulted in a rich diversity of approximately 30 registered breeds with well-established breed societies. The rich biodiversity includes imported exotic breeds, indigenous cattle and locally developed composite breeds. Genetic improvement in South Africa has been driven mainly through a National Beef Cattle Performance Testing Scheme (Bergh, 2010). Although performance tests are widely embraced, beef breeds differ in their participation in recording scheme programs, with the number of recorded traits also varying among breeds (SA Studbook, 2016). Consequently, participation in performance recording by different breed societies varies from 32% to 100% (Scholtz *et al.*, 2010).

The type of cattle breeds available comprise the taurine, composites, and Sanga breeds, also known as the *Bos taurus Africanus*, indigenous to SA (Schoeman, 1989). The Sanga cattle breeds are physically characterized by the cervico-thoracic humps that has resulted from historical crossbreeding between the *taurine* and *indicine* subspecies in eastern Africa (Felius *et al.*, 2014). In a commercial production system driven by economic principles, indigenous SA breeds such as the Afrikaner, Drakensberger,

Nguni, and Bonsmara compete with exotic breeds such as the Brahman, Hereford, Simmental, and Angus to improve the production of high-quality beef in tropical regions (Strydom, 2008a).

The current contribution of SA indigenous breeds to the local beef industry is well documented. Several studies have reported their importance as an animal genetic resource that possess adaptive traits associated with unique immune systems for resistance to various diseases and withstand extreme environmental conditions. Studies have indicated resistance to ticks and tick-borne diseases (Marufu *et al.*, 2011; Mapholi *et al.*, 2016), high fertility (Rust & Groeneveld, 2002; Du Plessis *et al.*, 2006; Grobler *et al.*, 2014), and both easy and frequent calving (Schoeman, 1996; Beffa *et al.*, 2009). Moreover, the SA indigenous breeds have been confirmed to have good meat quality for steers raised under extensive or intensive finishing systems (Strydom *et al.*, 2008b). Molecular studies based on SNP (Makina *et al.*, 2014), sequence (Zwane *et al.*, 2019), and microsatellite (Van der Westhuizen *et al.*, 2019) data have been undertaken to understand the underlying genetic variation of these breeds and confirmed small proportions of admixture amongst each other. The current study has focused on the Bonsmara and the Drakensberger breeds. These breeds were considered because they have accumulated modest number of genotypic data, which makes them suitable for this study. A brief overview of the history of the two breeds and their contribution in the National Beef Improvement Scheme is provided below.

The Bonsmara was bred and strictly selected for economically important production traits at Mara and Messina Research Stations in SA between 1937 and 1963 (Bonsma, 1980). The breed was developed nominally on 5/8: 3/16: 3/16 (62%:19%:19%) ratio of the Afrikaner, Hereford, and Shorthorn breeds, respectively (Bonsma, 1980). In 1964, the Bonsmara Breeders Society was formed by twelve breeders, and since then, the breed has expanded globally with over a million performance-tested animals (SA Stud Book, 2016). It is obligatory for all Bonsmara breeders to participate in the beef cattle improvement scheme, with the recording of phenotypes compulsory for many breeders (Van-Marle-Köster *et al.*, 2013). It was the first local breed to receive genomic evaluations through the Beef Genomics Program, resulting from the compulsory recording of several traits, including those with low to moderate heritability (Van Marle-Köster & Visser, 2018a).

The Drakensberger breed was developed out of the black, indigenous cattle of SA. As early as 1497, Vasco da Gama obtained a "fat, black ox" resembling the current breed known from the native people in the Cape (https://drakensbergers.co.za/). The breed became known as the Drakensberger due to its concentration on “sourveld” in the Drakensberg mountain region (Pentz, 2009). The Drakensberger Cattle Breeder’s Society was founded in 1948 and became a full member of SA Studbook in 1972 (Bisschoff & Lotriet, 2013). To date, the breed has spread throughout the country, with more than 12 000 registered animals (SA Studbook, 2016), and genomic estimated breeding values (GEBVs) are routinely available for the breed. The number of animal genotyped is growing, which may allow the application of GEBV in the future for SA beef breeds (Van Marle-Köster *et al.*, 2021).

2.3. Key challenges faced by the beef industry in South Africa

In recent years, beef farming has faced various challenges (e.g. social, economic, climate change, political and management) that resulted in a reduction in the size of the national beef herd which poses a threat to national food security (DALRRD, 2021). The sharp decline in cattle numbers is primarily due to the widespread drought caused by an extreme El Niño that took place between 2015 and 2018 (Agri SA, 2019), which led to the liquidation of herds as a result of a shortage of natural grazing and costs of supplementary feeding. Most farmers did not have any drought mitigation measures in place, so instead, resorted to involuntary culling and retained few heifers for breeding purposes (Mare *et al.*, 2018). The Red Meat Producers Organization (RPO) estimated a decline of 8.4% in production and a reduction of 15% in the national herd (Agri SA, 2019). Although no exact statistics are available, significant implications of involuntary culling on genetic progress are widely documented (Oishi *et al.*, 2013; Dunne *et al.*, 2020). Stock theft and predation, increasing annually with an average of 2.9% (estimated to be worth more than R800 million), threatens to plunge many beef farmers into bankruptcy and put food security at risk (Farmer's Weekly, 2019).

The challenges facing the beef industry in SA demands that producers use genomic tools to avoid inadvertently eliminating superior animals that are key for enhancing production efficiency. However, designing breeding programs that fully utilize the possibilities of DNA technology is difficult due to its extensive nature, compared to the dairy industry, which tends to be more intensive. Breeding strategies are combined from the seedstock to commercial programs with variety of breeding objectives based on various traits (Pollak, 2005). According to Van Eenennaam & Drake (2012), the usefulness and value of DNA information may differ depending on producers' production system and marketing circumstances.

The SA beef industry is divided into a relatively small proportion of approximately 1.5% stud breeders who produce genetic material (seedstock) for the commercial cow-calf operations, who supplies weaners for the feedlots that are responsible for finishing them (Van Marle-Köster & Visser, 2018b). The ownership of cattle can change several times in an animal's life, which may affect the records' accuracy (Figure 2.1). The number of stud animals is estimated at 450 000 or 4% of the beef cattle population (Figure 2.1). The seedstock farmers record pedigree and performance data which are subsequently used for the estimation breeding values (EBVs) for selection to achieve genetic improvement.

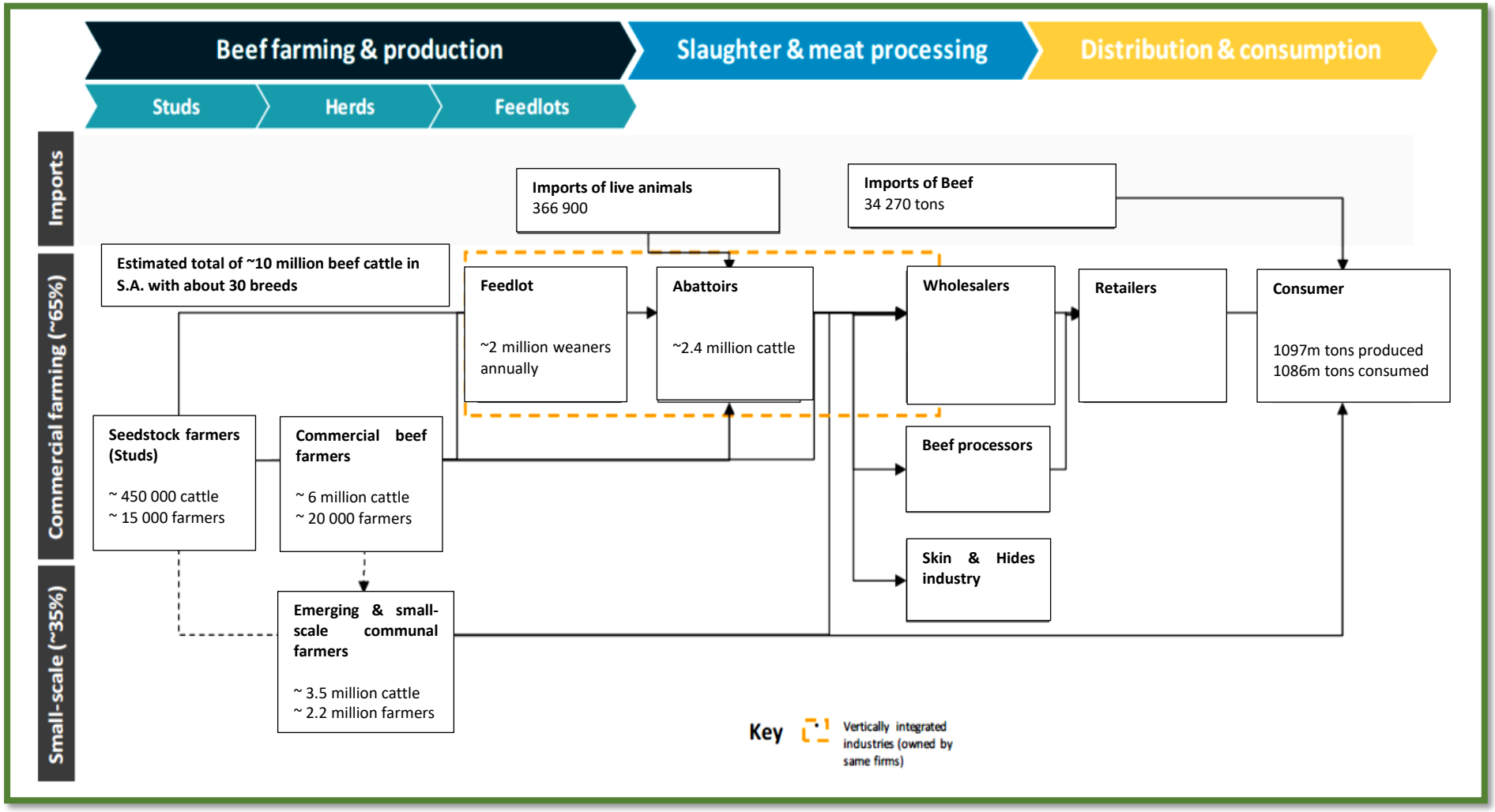


Figure 2.1. South African beef industry value chain (Sources: Walsh & Spazzoli, 2018; DALRRD, 2020; 2021)

1

The commercial sector is driven by selecting traits influencing market value, such as weight and meat quality, which are mostly associated with high profit, and other important traits that influence production efficiency. Most of the calves raised by commercial farmers are sold to feedlots followed by abattoirs (Figure 2.1), which makes it difficult to make genetic improvements for some traits. The fragmentation of the beef industry implies that the breeding objectives of stud breeders and those of the commercial producers and feedlot buyers are not essentially well aligned. The links between the seedstock producers and emerging to small-scale farmers are weak and challenges exist to develop these links (Walsh & Spazzoli, 2018).

Natural service is the most commonly used breeding method, which results in weak genetic linkage between herds and incorrect pedigrees that further compromise the accuracy of genetic predictions. This is mainly due to the extensive management practice in beef production, which leads to ineffective monitoring of mating events and poor adoption of artificial insemination (AI). The small proportion of seedstock farmers and the fragmentation in the value chain have implications for the accuracy of genomic prediction equations for various traits of economic importance (Pollak, 2005). Development of cost-saving genomic breeding strategies depends on the reliability of GEBVs, which in turn depends on the accuracy of the recorded pedigree and genetic linkage among herds along with reliable phenotypic information (Wientjes *et al.*, 2013; Wittenburg *et al.*, 2016).

Nonetheless, challenges and implications addressed herein are not unique to the SA beef industry but have been documented in Australia (Pollak, 2005), the United States (Garrick, 2011) as well as having been recently reported in other African countries (Ibeagha-Awemu *et al.*, 2019; Mrode *et al.*, 2019). Results obtained from the beef genomics programs in Australia and the US suggest that the adoption of genomic technologies significantly improves the productivity and sustainability of the beef industry (Swan *et al.*, 2011; Van Eenennaam *et al.*, 2014). The application of genomic technology to local SA cattle breeds, which are more hardy than exotic cattle breeds, could assist the local economy while lowering the environmental impact of production efficiency (Van Marle-Koster *et al.*, 2021).

2.4. Implications of incorrect parentage in genetic improvement programs

For decades, performance and pedigree information have been used for genetic evaluations to predict EBVs, which are used to select the best candidates for breeding. Since the EBVs provide an estimate of transmitting ability of the parent, accurate parentage remains an important factor for accurate selection. A major concern is the reliability of pedigree information sourced on farm under extensive production systems where natural and multi-sire mating practices are used, making parentage records prone to errors. Classically, the relationship matrix only provides the expected additive genetic relationships among animals and does not account for the difference in actual genomic relationship among individuals resulting from Mendelian sampling (Junqueira *et al.*, 2017).

There are two types of pedigree issues that can influence genetic evaluations. These include incorrect information (i.e., wrong parentage) and missing information (i.e., unknown parentage) (Harder *et al*., 2005). Several factors may contribute to incorrect parentage in extensive beef production systems. Many cows calve in a short space of time making it very difficult to keep track of dam-offspring records resulting in inaccuracies in the dam's genetic merit estimates due to incorrect assignment of dams to their offspring (Van Eenennaam & Drake, 2012). The pedigree error rate for dam-offspring pairs is known to affect multiple generations and has an additive effect on the rate of genetic gain (Harder *et al.*, 2005). Sanders *et al.* (2006) reported wrong sire information to have greater repercussion on response to selection and variance of breeding values compared to missing sire information. Inaccurate prediction of the genetic merit of candidates can lead to the under- or over-estimation of breeding values (Banos *et al.*, 2001). Wrong sire information also has adverse effects on the genetic gain for lowly and moderately heritable traits due to the impact of pedigree information on the accuracy of EBVs estimated using BLUP (Sanders *et al.*, 2006; Rodríguez-Ramilo *et al.*, 2015).

Additive genetic variance and heritability are the most important genetic parameters used in genetic evaluations. Misidentification of parents affects the family means as well as affecting inbreeding rates (Christensen *et al.*, 1982; Hayes *et al.*, 2009) which reduces the additive genetic variance in that population. This alters the heritability (Visscher *et al.*, 2008), and, consequently, negatively affects the prospects for long-term genetic response to selection (Maiwashe *et al.*, 2008). In the absence of any selection pressure, inbreeding rates are lower when the heritability estimates are higher (Visscher *et al.*, 2002).

A reliable pedigree is essential for traits with low and moderate heritability, such as disease resistance, residual feed intake, and reproduction. These traits are highly affected by the genetic effects (i.e. additive or non-additive) and environmental factors (i.e. permanent or temporary) (Pszczola *et al.*, 2012; Junqueira *et al.*, 2017). As for growth traits with high heritability in beef cattle, pedigree errors may have minimal impact on phenotypes, which are representative measures of genetic merit (Peters *et al.*, 2014). Several studies listed in Table 2.1 have investigated the impact of pedigree errors on variance components and genetic parameter estimates for various traits of economic importance.

Table 2.1. Estimates of genetic parameters, heritability and breeding values for traits of economic importance under different pedigree error levels from diverse world cattle populations

Country	Traits	h^2	Pedigree error	[a]RGG / [b]EBVs (%)	Study Reference
Germany	Milk yield	0.50	15%	8.70[a]	Gelderman *et al.*, 1986
	Fat yield	0.20		16.9[a]	
United States, Canada, New Zealand, and Netherlands	Milk, fat and protein yields	0.25	11%	11.0-18.0[a]	Banos *et al.*, 2001
Brazil	Tick resistance	0.15	26.0%	80.0[b]	Junqueira *et al.*, 2017
Mexico	Fertility	0.10	10%	7.00[a]	García-Ruiz *et al.*, 2019
Korea	Carcass weight	0.31	5%	96.0[b]	Nwogwugwu *et al.*, 2020
	Eye muscle area	0.42		96.0[b]	
	Backfat thickness	0.46		97.0[b]	
	marbling score	0.52		97.0[b]	
	Carcass weight	0.18	30%	77.0[b]	
	Eye muscle area	0.18		83.0[b]	
	Backfat thickness	0.32		81.0[b]	
	marbling score	0.29		81.0[b]	
	Carcass weight	0.03	80%	30.0[b]	
	Eye muscle area	0.05		52.0[b]	
	Backfat thickness	0.11		24.0[b]	
	marbling score	0.00		26.0[b]	

[a]RGG - reduction in genetic gain, [b]EBVs – percentage of the accuracies of breeding values predicted from erroneous pedigrees, h^2 - heritability

Relatedness in a small population directly affects the probability of assigning unambiguous parentage (Sherman *et al.*, 2004). However, smaller numbers of elite genotypes selected in small seedstock herds have been reported to pose a danger of increasing undesirable recessive genes due to the strong selection within a population, resulting in inbreeding depression over time (Falconer & Mackay, 1996). The consequential implications of inbreeding depression on beef cattle's productive, reproductive and adaptive traits have been widely documented (Carolino & Gama, 2008; Santana *et al.*, 2012; Pereira *et al.*, 2016). It has been estimated that a 6 to 13% inbreeding coefficient would reduce the selection response by 2 to 3% (Sonesson & Meuwissen 2000; Banos *et al.*, 2001).

2.5. DNA based markers for parentage verification

Historically, many molecular methods were developed for pedigree validation in cattle breeding (Parlato & Van Vleck, 2012). Since the first use of blood group typing in the early 1940s (Ferguson, 1941), parentage analyses have been constantly advancing as new technologies are developed. In the last two decades, there has been an increasing trend to develop time- and cost-saving genetic tools in parentage-related research. DNA-based markers such as microsatellites and, more recently, SNPs have been applied for routine parentage testing in cattle, as summarized in Table 2.2.

Table 2.2. Selected number of parentage studies based on different DNA markers in a number of cattle breeds

DNA marker	Breeds	Number of markers	Reference
Blood typing	Dutch cattle breeds	-	Bouw, 1958
Microsatellite	Boran	11	Kios *et al.*, 2012
SNPs	Angus	32	Heaton *et al.*, 2002
SNPs	(http://www.isag.us/Docs/Cattle-SNP-ISAG-core-additionalpanel-2013.xlsx)	200	ISAG 2013
SNPs	Korean Hanwoo and Wagyu	257 & 245	Strucken *et al.*, 2014
SNPs	Limousin, Charolais, American Angus, Simmental, Hereford, Belgian blue, Salers, Parthenaise, Blonde D'Aquitaine, Aubrac, Pie Rouge Des Plaines, Montbeliarde, Red Angus	500	McClure *et al.*, 2015; McClure *et al.*, 2018
SNPs	Brazilian Red Sindhi	70	Panetto *et al.*, 2017
SNPs	Ankole, Nganda, Small East African Zebu, N'Dama, Nelore, Begait Barka, Danakil Harar, Fogera, Boran, Iringa Red and Singida White	200	Strucken *et al.*, 2017
SNPs	Simmental	50	Zhang *et al.*, 2018
SNPs	African crossbreds, African *Bos taurus*, European *Bos taurus*, *Bos indicus* and African indigenous	200	Gebrehiwot *et al.*, 2021

The DNA-based microsatellite markers were accurate, but they were limited to verifying parentage and less effective for parentage discovery (Buchanan *et al.*, 2016). In addition, microsatellite markers have been reported to be uninformative in highly inbred populations (Brenig & Schütz, 2016). Unlike the previously used methods, a major benefit of SNP-based parentage verification is that the information has multiple other uses over and above simply parentage testing and discovery (Berry, 2019). Highly accurate paternity validation or identification is virtually a no-cost by-product when putative parent-offspring genotypes already exist (Weller *et al.*, 2017). Essentially, if animals are genotyped using a large SNP panel for utilization in genomic evaluations, the SNP profile for parentage is automatically collected, eliminating the need for duplicate DNA applications. Furthermore, the SNP data have recently been recognized to detect potential genome regions that possibly contain a copy number variation (CNV) and define the individuals that might be carriers based on a special method of applying the Mendelian laws (Himmelbauer *et al.*, 2019).

Since the first low-density bovine SNP parentage panel consisting of 121 markers was developed for the *Bos taurus* through a collaboration involving the Agricultural Research Service of the United States Meat Animal Research Center (ARS-USMARC) (Heaton *et al.*, 2002), a number of low-density population-specific panels were compiled and tested. Following these studies, the ISAG advisory committee attempted to agree on a core panel of 100 SNPs derived from the USMARC panel for global parentage testing (CMMPT, 2012). The ISAG 200 SNP markers (full panel) were derived from 4 000 animal genotypes of 23 cattle breeds (ISAG, 2013). These markers are incorporated in almost all commercially available genotyping arrays of the Affymetrix, GeneSeek and Illumina (Affymetrix, 2011; Illumina 2010; 2011). Previous studies that evaluated the core panel on distantly related breeds confirmed less variability, less information content, and a reduced ability to determine parentage accurately (McClure *et al.*, 2015; Buchanan *et al.*, 2016). To correct the issue of ascertainment bias, thus increasing the power of parentage exclusion, the ISAG panel included 100 SNPs selected from the indicine and synthetic breeds (ISAG, 2013). This suggested that SNP parentage panels developed using population parameters from other breed types may have limitations for breeds not represented in the design of the panel.

2.6. Impact of genomic markers on application for genetic evaluations

The use of SNP data in genomic evaluations has increased the accuracy of EBVs and enabled the estimation of the heritability of a trait even in the absence of pedigree information by using actual or realized relationship among relatives (Visscher *et al.*, 2006; Garrick, 2011). Furthermore, as compared to traditional pedigree-based BLUP selection, genome-based selection has the potential to reduce inbreeding per generation (Daetwyler *et al.*, 2007). Mendelian sampling within families can be tracked using genomic-based BLUP approaches, which is simply not possible with pedigree-based BLUP (Bastiaansen *et al.*, 2012). Instead of a numerator relationship matrix, a genomic relationship matrix

(GRM) is used to fix many unidentified pedigree errors. Genomic selection considers marker associations that account for expected genuine genetic linkages (Moore *et al.*, 2019).

The discovery of parentage and pedigree reconstruction using SNP markers has become possible and has increased the accuracy of genetic parameter estimates for genetic evaluations (Veerkamp *et al.*, 2011; García-Ruiz *et al.*, 2019). Based on the frequency of opposing homozygous SNP genotypes between the parent and offspring pairs, Mendelian errors have been utilized to uncover inconsistencies in pedigrees (Ferdosi *et al.*, 2014). Consequently, if the genotype of a parent at a locus is known, the possible genotypes of the offspring at that specific locus can be easily predicted, as shown in Figure 2.2. For instance, the offspring's genotype must be AB when the genotypes of the parents are AA and BB. The expected number of errors between a genotyped parent-offspring pair and the variance of the predicted number is very low when opposing homozygotes only result from genotyping errors (1%).

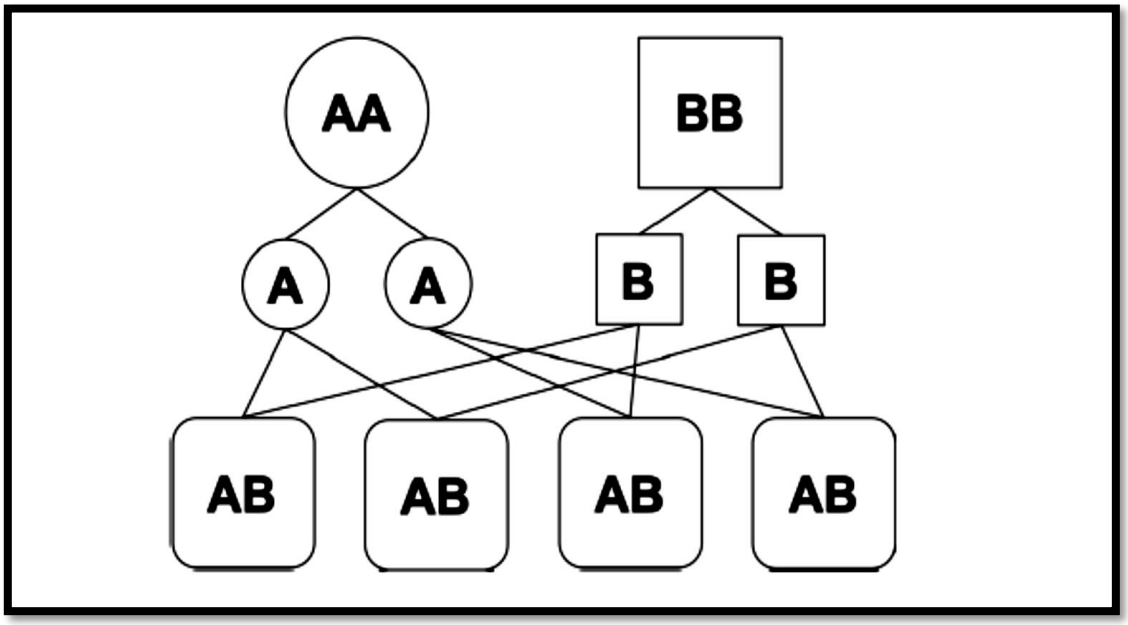


Figure 2.2. An illustration of inferring progeny genotype from parental genotypes (Himmelbauer, 2019)

Junqueira *et al.* (2017) used genomic data to improve relationship information in pedigree-based models for tick resistance evaluations of Hereford and Braford beef cattle. When the accuracy of the relationship information increased from the original pedigree with errors, it influenced the estimation of variance components (Figure 2.3a) and produced higher heritability estimates. Incorporating phantom parent groups into the model reduced the bias in the genetic trend to be insignificant, resulting in increased additive genetic variance and heritability estimates (Theron *et al.*, 2002; Misztal *et al.*, 2013; Silva *et al.*, 2019). The positive trend in these parameters implies that inaccurate or incomplete pedigrees may lead to downward biased estimates of the proportion of the phenotypic variance attributable to additive genetic variables as relationship information improves (Neuner *et al.*, 2009).

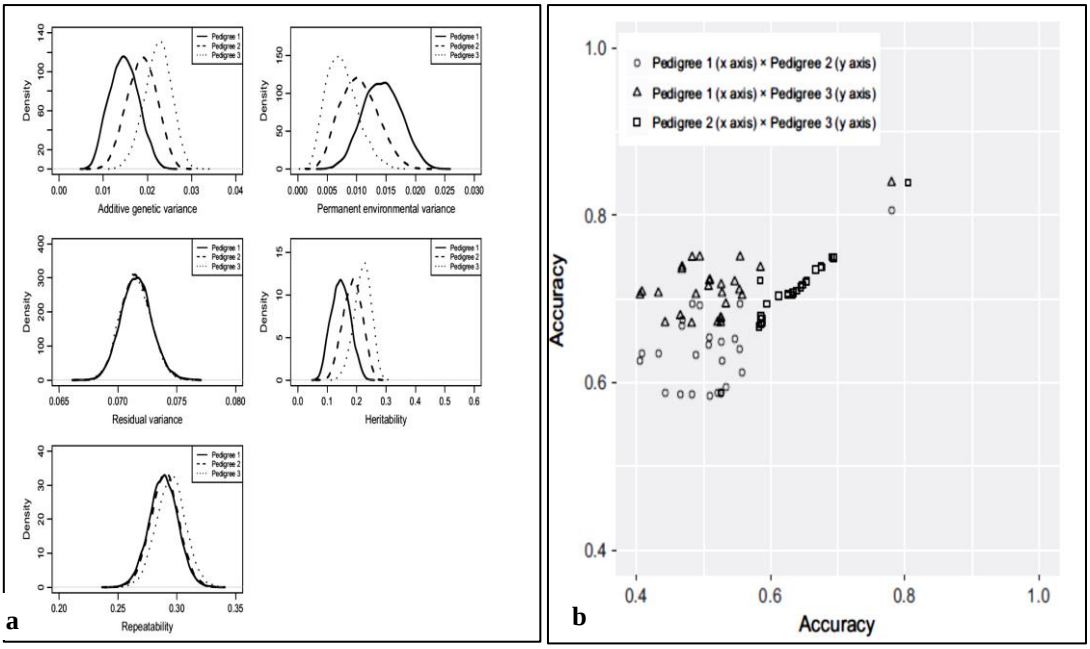


Figure 2.3. Posterior density plots (a) and scatter plots (b) depicting the impact of relationship information on prediction of genetic components and breeding values, respectively (Junqueira *et al.*, 2017).

Likewise, more accurate EBVs were achieved from the corrected and reconstructed pedigrees compared to the original pedigree with errors (Figure 2.3b). It is evident that the accuracy of pedigree has the strongest impact on standard errors of prediction and the increase in its depth has an important role in the precision of genetic parameters. This is supported by previous studies that observed pedigree errors and reported a reduction in the variance of EBVs, which also influenced the genetic gain (Harder *et al.*, 2005; Sanders *et al.*, 2006).

2.7. Detection of the hemizygous deletions in the genome

Hemizygous deletion is when an individual has one allele rather than two copies of an allele at a particular locus (Amos *et al.*, 2003). Thus far, the effects of hemizygous deletions have not been very well explored in livestock genetics. This is because only elite animals used for breeding are targeted to be genotyped (Bradford *et al.*, 2019). Since SNP genotyped animals have pedigree lineage, it has become prudent that detection methods for hemizygous deletions using this information could be beneficial for animal genetics research and application. Hemizygous deletions cannot be detected using called SNP genotype. This is because the region with the deletion is mistakenly displayed as a homozygous region in the genome (Amos *et al.*, 2003). This is a result of the GenTrain algorithm in the GenomeStudio mistyping hemizygous regions as homozygous (Nandolo *et al.*, 2018). Such artefacts may lead to Mendelian inconsistencies and, thus, inaccurate genetic evaluations. A typical scenario is when the sire carries a single copy deletion (A-), where "-" stands for the deleted allele. Then, if mating occurs between the sire with deletion (A-) and the dam with an opposing homozygote (BB) at the same

locus, this may likely result in a progeny with either heterozygote genotype (AB) or a deletion (B-) at that locus. After SNP genotyping, a single-copy deletion (A-) from the sire will be reported as a homozygous AA, while a single-copy deletion from a progeny (B-) will be reported as homozygous BB. According to Mendelian laws, a sire with homozygote AA genotype is expected to produce progeny with genotypes AA or AB at the same locus. Hence, in the presented scenario, hemizygous deletion can mimic Mendelian mismatches while there is none, resulting in inaccurate parentage conclusions.

Information that SNPs are hemizygous for a deletion and appear as homozygous has been verified based on the violation of Mendelian transmission rules using genotypes of parents and offspring. Conrad *et al.* (2006) observed hemizygous deletions using parent-offspring trios in humans and validated the results with comparative genome hybridization shown in Figure 2.4. The possible trio genotype configurations offering Mendelian mismatches were defined as different types that are incompatible and compatible. Thus, potential sites of deletions were defined as regions that at least have more than two Mendelian mismatches and are consistent with a deletion matching to "Type I" in the figure below.

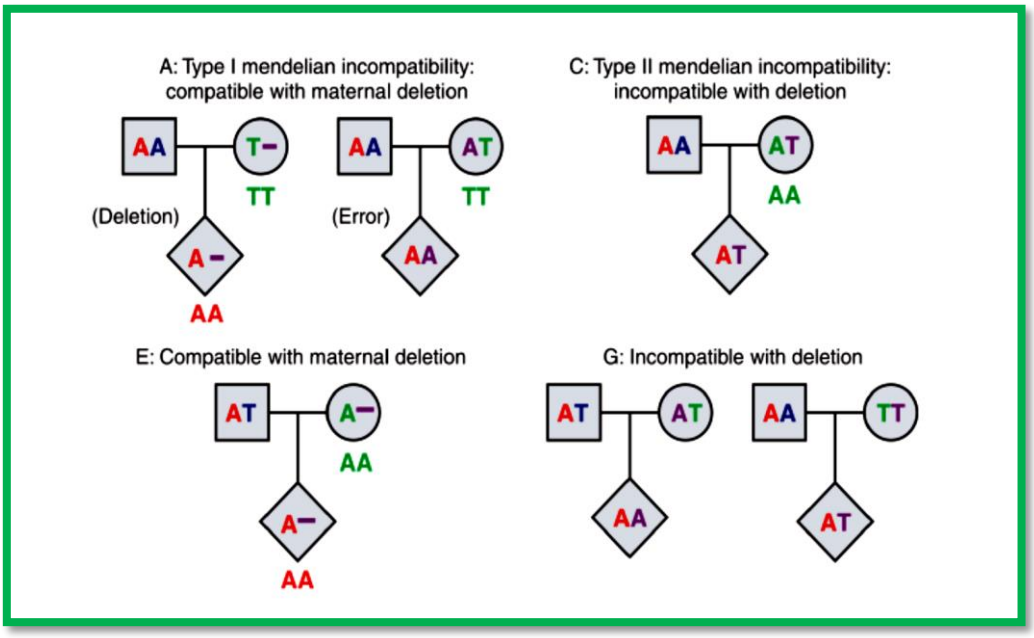


Figure 2.4. An illustration of the detection of hemizygous deletions (Conrad *et al.*, 2006; Himmelbauer, 2019). The symbols inside the diamond and circle represent the actual genotype, whilst those outside represent the genotype determined by SNP genotyping if it differs from the actual genotype. The "-" denotes a deletion, while the other colors represent various alleles.

Some positive and negative effects of deletions in livestock have been documented. Durkin *et al.* (2012), for example, found that deletions can cause color sidedness in calves, whereas Xu et al. (2014a) found that deletions can explain additional genetic variance underlying milk production parameters. In a separate study, Xu *et al.* (2014b) found that a deletion was linked to gastrointestinal nematode resistance in Angus cattle. Other investigations have revealed that large deletions in dairy cattle may exhibit lethal

effects (Charlier *et al.*, 2012; Schütz *et al.*, 2016). In the Nordic Red cattle herd, a 600kb deletion was found to have a beneficial influence on milk production and a detrimental impact on fertility (Kadri *et al.*, 2014), as well as causing stillbirth (Mesbah-Uddin *et al.*, 2018).

2.8. Statistical analysis for parentage testing

2.8.1. Genomic data quality control

Genomic quality control (QC) is defined as the process of identifying and removing DNA samples and markers that can introduce errors into the study (Anderson *et al.*, 2010). Quality of the genotype determines the accuracy of downstream analyses. Genotype quality is important for parentage analyses because it may influence the assignment of the progeny to the parent (ISAG, 2013), which is necessary to estimate the genome-wide impacts of alleles inherited from ancestors in order to anticipate an individual's merit (Ferdosi *et al.*, 2014). The QC for microsatellite-based parentage analysis compares samples defined as the standard or reference profiles of markers with known accuracy and unknown samples, assuming that the latter will perform similarly to the control samples (Schütz & Brenig, 2015). The control samples function as a reference for the pattern of manual inspection of alleles, insertion of true alleles, stutter bands, and deletion of nonspecific regions, making the parentage profiling prone to errors. Requirements for the careful QC for parentage using genomic data (Table 2.3) are documented with important sequential steps after genotype calls from the GenomeStudio (Cooper *et al.*, 2013; McClure *et al.*, 2015; 2018).

Table 2.3. Summary of quality control measures used on cattle parentage studies and for SNP panel development

QC parameters	Criteria	Reference
GenCall (GC) & GenTrain (GT)	≥0.60	Strucken *et al.*, 2015& 2017 Gebrehiwot *et al.*, 2021
SNP call rate	≥ 0.95	Strucken *et al.*, 2016 Buchanan *et al.*, 2017 Panetto *et al.*, 2017
Animal call rate	≥ 0.90	ISAG, 2013 García-Ruiz *et al.*, 2019 Gebrehiwot *et al.*, 2021
Hardy Weinberg Equilibrium (HWE)	$<10^{-6}$ $<10^{-7}$	Strucken *et al.*, 2017 Junqueira *et al.*, 2017
Minor allele frequency (MAF)	≥ 0.01 ≥ 0.25 ≥ 0.30	McClure *et al.*, 2018 McClure *et al.*, 2015 Panetto *et al.*, 2017

Call rate, minor allele frequency (MAF) and the extent of Hardy-Weinberg Equilibrium (HWE) deviations are the most commonly used measures of genotype integrity (Chan *et al.*, 2008). Within the context of obtaining the highest possible genotyping accuracy, others have considered either the SNP GenCall or GenTrain scores (McClure *et al.*, 2015; Berry *et al.*, 2016; Strucken *et al.*, 2017), which reflect the genotype call from the GenomeStudio. The GenTrain (GT) score considers the quality and shape of the genotype clusters and relative distance to each other (More *et al.*, 2019) to devise a statistical score. The GT score is used to mimic evaluations of an expert's visual and cognitive systems. SNPs with GT score >0.55 have been considered for the downstream analyses (Judge *et al.*, 2016; Zhao *et al.*, 2015). The GT score calling algorithm uses the cluster position and shapes for each SNP to estimate GenCall (GC) scores. The GC is used in Bayesian models to filter out poor quality calls, SNPs, or samples (Oliphant *et al.*, 2002). A GC score lower than 0.15 is considered the default threshold in Genome Studio and it signifies failed genotypes. Berry *et al.* (2016) proposed more stringent quality control on the GC score (0.5) to increase the reliability of called genotypes. Some studies have retained genotypes with GC score > 0.50 (Strucken *et al.*, 2017; Dotsev *et al.*, 2018) to increase the accuracy.

Call rate is defined as the proportion of called SNP to the total number of SNP per individual (Purfield *et al.*, 2016). It is the first step and a useful screening tool in data quality control for genomic studies and evaluations (Wiggans *et al.*, 2012). This also involves identifying individuals with discordant sex information, removing duplicates, and identifying individuals with divergent ancestry (Anderson *et al.*, 2010). Low call rates result from poor DNA quality and concentration (Purfield *et al.*, 2016). Individuals with a poor minimum call rate are removed in the analyses, which requires resampling and re-genotyping of the samples, resulting in extra costs (McClure *et al.*, 2018).

The impact of different call rate thresholds on genomic information have been reviewed and described by Cooper *et al.* (2013), and a range of animal call rate has been applied across cattle studies from 0.80 to 0.85 (Purfield *et al.*, 2016); 0.90 (Hayes *et al.*, 2009; ISAG, 2013). Although sometimes choosing low animal call rates maximizes the number of genotyped animals, this could result in errors in pedigree validation as the number of SNP considered in pairs may differ between the individuals in comparison. Purfield *et al.* (2016) detected minimal genotyping errors across genotyped parent-offspring relationships when using high call rate genotypes were considered and regarded the parental-offspring relationships as genomically true compared to low call rate genotypes.

Call rate per marker is another critical step performed by estimating the fraction of called genotypes per SNPs over the total number of samples. This involves identifying and removing non-informative SNPs with a significantly excessive missing genotype of 5% or more. Markers exhibiting significant deviation from HWE are generally removed. These markers indicate genotyping or genotype calling errors and selection. Significant thresholds for declaring SNPs in HWE for parentage studies vary between the *p-value* of 10^{-6} (Panetto *et al.*, 2017) and 10^{-7} (Junqueira *et al.*, 2017).

The final step for QC per marker is eliminating all markers with a very low (MAF). In general, the proficiency of the SNP markers panel depends on the quality and informativeness of each SNP, which highly depends upon MAF. Since only homozygous genotypes in both parents and offspring are considered, SNPs with low MAF have little utility for parentage (Van Doormaal *et al.*, 2016). In contrast, SNPs with high MAF (0.5) generate higher exclusion probabilities and a reduced number of markers to be used (Belay, 2013). The SNP panels selected for high MAF perform best in assigning parentage using opposing homozygote criteria (Strucken *et al.*, 2017). Several parentage studies have selected SNPs with MAF > 0.25 (McClure *et al.*, 2015; 2018; Panetto *et al.*, 2017).

2.8.2. Linkage disequilibrium

Linkage disequilibrium (LD) is the relationship between genotypes at a pair of polymorphic sites (Pritchard & Przeworski, 2001). It occurs when genotypes at two loci are not independent of each other and thus result in a higher frequency of particular haplotypes at the loci than would have been anticipated by chance (Hayes *et al*., 2003). Statistically, the estimated haplotype frequencies are used instead of observed frequencies. The LD between densely spaced, polymorphic genetic markers contains information on historical population size, a relevant parameter in livestock breeding programs. The level of LD influences the accuracy of estimated genetic effects collected by markers in a population (Hayes *et al.*, 2009). Genetic forces of evolution (e.g. genetic drift, inbreeding, migration, mutation, recombination events and selection) and non-genetic factors (e.g. marker ascertainment bias) are known to affect the genome-wide LD level (Biegelmeyer *et al.*, 2016). In some cases, the non-random association of loci may be due to crossbreeding and backcrossing (Hayes *et al.*, 2009) associated with recombination disrupting the physical linkage between chromosomal segments and produces new combination of additional alleles in the next generation (Ferdosi *et al.*, 2014).

The pairwise measure and D estimate are represented by |D'| and r^2. Both have measurements that range from 0 to 1. A |D'| < 1 denotes historical recombination between loci, whereas |D'| = 1 suggests that recombination between the two loci is not possible. Furthermore, in small samples and in the presence of a rare allele, |D'| has a tendency to be inflated (Bohmanova *et al.*, 2010). For SNP markers, r^2 is the most commonly used measure of LD. It is a robust measure that is less sensitive to allele frequency and small sample size, r^2 is the most often used measure of LD for SNP markers (Bohmanova *et al.*, 2010; Wientjes *et al.*, 2013). It represents the correlation between the two loci, such as the presence or absence of a specific allele at the first and second loci (VanLiere & Rosenberg, 2008). The r^2 has a frequency-dependent range and has a strong relationship with minor allele frequency; hence, it is crucial to match loci by allele frequencies before measuring the LD. An $r^2 = 0$ denotes that loci are in perfect linkage equilibrium, while $r^2 = 1$ denotes that loci are in complete linkage disequilibrium with only two haplotypes present due to population bottlenecks or genetic drift (Biegelmeyer *et al.*, 2016).

LD between causal variants and genetic markers is critical for undertaking efficient genome-wide association and genomic selection research (Meuwissen *et al.*, 2001). This is because both approaches rely on the non-random correlation between markers and functional mutations impacting the variables of interest which is handled by LD. The efficacy of a panel is determined by informativeness of markers, which is primarily determined by MAF, as well as the independency of the other loci that make up the panel (Belay, 2013). In parentage analysis, data from several loci is pooled to provide a fair level of exclusion power. The alleles at several loci might not be able to assort separately because the loci could be in LD. Due to the reduced genetic variation needed to determine parentage, this phenomena lowers the estimated likelihood of exclusion (Jones & Ardren, 2003). The physical distance of 20 Mbp between the selected markers is most likely sufficient to avoid a detrimental influence on the power of parentage exclusion and likelihood of identification. In cattle, the physical distance threshold (20Mbp) has been used to create SNP-based parentage panels (Werner *et al.*, 2004; Hara *et al.*, 2010).

2.8.3. Parentage verification and assignment

Parentage is established using either exclusion or probability methods based on allele frequencies (*p* and *q*) (Jones *et al.*, 2010). Exclusions expected for both techniques are dictated by the allele frequencies in the population, genotyping call rates, and count in error rates (Grashei *et al.*, 2018). The exclusion approach only use genotypes with Mendelian inconsistencies, whereas likelihood-based method assume that loci are inherited independently. Thus, the method of exclusion require more loci than likelihood-based method. It is acknowledged that the likelihood-based method is mostly used, whereas the probability of exclusion method is rarely used in cases of related individuals because of the small pool of included alleles (Huisman, 2017; Grashei *et al.*, 2018).

The majority of accessible parentage analysis tools were designed with the premise of tiny datasets when they were first developed for microsatellite data (Jones & Ardren, 2003). Several multifunctional software programs were developed in the post-genomic era, and some of them were modified to identify parent-offspring links within populations where no prior information of family structure existed. The basic approaches and some software used for parentage analysis are summarized in Table 2.4.

Table 2.4. Software available for parentage analysis, including main features and available functions

Packages/tools	Method[a]	Reference
Apparent	GR/GS	Melo & Hale (2019)
Cervus	ML	Kalinowski *et al.* (2007)
Colony	GR/GS/ML	Jones & Wang (2010)
hsphase (pogc)	GR/GS	Ferdosi *et al.* (2014)
Kinship	ML	Goodnight & Queller (1999)
MasterBayes (MCMCped)	BA/ML	Hadfield *et al*. (2006)
Parente	GR/GS	Cercueil *et al.* (2002)
ParentOffspring	GR/GS	Abdel-Haleem *et al.* (2013)
Parfex	GR/GS/ML	Masashi & Shigeho (2012)
Plink	ML	Purcell *et al.* (2007)
SeekParentF90	BA	Aguilar (2014)
SireMatch	ML	Pollak (2006)
Solomon	BA	Christie *et al.* (2013)

[a]BA = Bayesian approach; GR = Genetic relatedness, GS = Genetic similarity, ML = Maximum likelihood

(a) Exclusion

In the exclusion-based method, the hypothesis of parent-offspring relations is rejected when the offspring's candidate parents' genotypes violate Mendel's law of inheritance (Jones & Ardren, 2003; Grashei *et al.*, 2018). The exclusion method has been used to validate, identify novel parentages, and reconstruct pedigrees using microsatellite markers (Dodds *et al.*, 2005) and SNP data (Hayes, 2011; Strucken *et al.*, 2014; Garca-Ruiz *et al.*, 2019). Exclusion is determined by counting the number of opposing homozygotes between parent-offspring pairs divided by the total number of homozygous SNP used for comparison in the panel (Strucken *et al.*, 2014; Purfield *et al.*, 2016). When the putative sire's genotype is BB and the calf's genotype is AA, it signifies the calf did not get either allele from the putative sire at that locus (assuming no hemizygous deletion), and thus the sire is ruled out as the offspring's sire. If then the genotype of the putative sire is BB and the genotype of the calf is AB, there is no exclusion due to the fact that the offspring may have inherited the B allele from the other candidate parent.

When small panels are used for parentage exclusion, one conflict is allowed to account for the possibilities of genotyping errors, mutations, or unknown null alleles. In cases where more than one

candidate parent is linked to a calf, the candidate parent with the fewer mismatches with the calf is assigned the parent, while the rest are assigned as relatives (siblings or grandparents) depending on the exclusion criteria. At times, if there is not a single or multiple animal(s) meet the criteria, then the parentage of the calf is declared unknown (Van Eenennaam *et al.*, 2007). A major limitation with the exclusion-based method is when there is no exclusion for several candidate parents, and therefore consideration should be given to other methods such as a likelihood or increasing number of markers (Huisman, 2017).

There are three general formulae suggested by Jamieson & Taylor (1997) to test the probability of exclusion for markers in a panel. These are based on exclusion of genotypes between an offspring and either of the parents or both parents. The first one is when the genotypes for both parents (PE_1) and an offspring are compared but one parent is excluded. The second one is when the genotype of an offspring and both parents are known but both parents are excluded (PE_2). For PE_3, the genotypes are available only for one parent and the offspring are available for comparison. These formulae are derived from allele frequencies (p_i) of the i^{th} allele, where n is the number of alleles per SNP and calculated as follows:

$$PE1 = 1 - 2\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} p_i^3 + 2\sum_{i=1}^{n} p_i^4 - 3\sum_{i=1}^{n} p_i^5 - 2\left(\sum_{i=1}^{n} p_i^2\right)^2 + 3\sum_{i=1}^{n} p_i^2 \sum_{i=1}^{n} p_i^3$$

$$PE2 = 1 + 4\sum_{i=1}^{n} p_i^4 - 4\sum_{i=1}^{n} p_i^5 - 3\sum_{i=1}^{n} p_i^6 - 8\left(\sum_{i=1}^{n} p_i^2\right)^2 + 8\left(\sum_{i=1}^{n} p_i^2\right)\left(\sum_{i=0}^{n} p_i^3\right) + 2\left(\sum_{i=1}^{n} p_i^3\right)^2$$

$$PE3 = 1 - 4\sum_{i=1}^{n} p_i^2 + 2\left(\sum_{i=1}^{n} p_i^2\right)^2 + 4\sum_{i=1}^{n} p_i^3 - 3\sum_{i=1}^{n} p_i^4$$

The combined (k) probabilities for the power of parentage exclusion over all markers (j) to achieve the sufficient power of exclusion is computed as $P_{Ej} = 1 - (1 - P_1)\,(1 - P_2)\,(1 - P_3)\ldots(1 - P_k)$.

(b) Likelihood

Likelihood analysis uses the Likelihood Ratio (LR) statistic to help decide parentage. It is the natural logarithm of the multi-locus likelihood of the candidate parent being the true parent (H_1) divided by the likelihood of the candidate parent not being the true parent (H_2) (Marshall *et al.*, 1998), which is called the LOD score (Meagher 1986). It is computed as follows (Jones *et al.*, 2010):

$$L\,(H_1, H_2|g_m, g_f, g_o) = \frac{T\,(g_o|g_m, g_f)P(g_m)P(g_f)}{T(g_o|g_m)P(g_m)P(g_f)} = \frac{T(g_o|g_m,g_f)}{T(g_o|g_m)}$$

The genotype of a dam, alleged sire and offspring are given by g_m, g_f and g_o, respectively. The Mendelian probabilities of the offspring genotype provided by the dam and alleged sire's genotypes or just the dam's genotype transition probabilities are represented by $T(g_o|g_m,g_f)$ and $T(g_o|g_m)$ (Marshall *et al.*, 1998).

The likelihood approach involves the frequency of both homozygous and heterozygous allele(s) in the offspring that may have come from the candidate parent. Although its technique looks to be a reasonable choice, determining parentage when close relatives are considered also probable parents can be problematic. As a result of this, Delta statistics (Δ) defined as the difference in LOD scores between the most likely parent and the next most likely parent are used (Kalinowski *et al.*, 2007; Grashei *et al.*, 2018). Uncommon alleles are given more weight by the LOD statistic. Different loci are frequently believed to be independent, resulting in a total LOD multiplied across all loci. When the LOD score is equal to 0, the candidate parent is as likely to be the progeny's true parent as any other randomly selected animal. A positive LOD score suggests the candidate parent more likely to be the true parent of the offspring than a randomly chosen individual. A negative LOD score occurs when the putative parent and offspring share a mismatch at one or more loci. Biological parents are less likely to be candidates with low LOD scores (Marshall *et al.*, 1998).

2.8.4. Detection of hemizygous deletion

Since regions of hemizygous deletions (A- or B-) are not correctly identified by SNP genotyping but instead are displayed as homozygous genotypes (AA or BB), the comparison of parent-progeny pairs SNP genotypes based on the Mendelian law leads to mismatches (Winchester *et al.*, 2009; Daetwyler *et al.*, 2014) and thus, affect parentage analysis. A substantial number of adjacent SNPs displaying discordance between the parent-offspring pairs confirms a large deletion (Himmelbauer, 2019). In general, carriers are likely to pass a true deletion on to 50% or more of their progeny. As a result, across the generations available in the population, familial patterns of Mendelian mismatches can be investigated for the detection of hemizygous deletions. This therefore also highlights the importance of retaining SNPs free from Mendelian errors during quality control to improve the accuracy of parentage.

2.9. Conclusion

Parentage verification and discovery provide a cost-effective strategy for ensuring accurate selection. Based on several studies reviewed, it is clear that using SNP genotypes to correct parentage improves the precision of breeding values and, consequently, more accurate selection. The availability of the ISAG-ICAR low-cost genotype panel has provided a good base for parentage testing, but some limitations require optimization and identification of additional markers to enhance its utility in diverse genetic backgrounds. The detection of deletions in the genome is a relatively new field of research, especially in livestock genetics. However, it is becoming of interest for accurate parentage analysis and their effects on breeding. Developing a low-density genotype panel for SA Sanga breeds is, therefore, desirable to perform such analyses towards achieving more accurate genetic evaluations that can provide efficiencies in beef production.

References

Abdel-Haleem, H., Ji, P., Boerma, H.R. & Li, Z., 2013. An R package for SNP marker-based parent-offspring tests. Plant Methods, 9, 1, 1-5.

Affymetrix Inc., 2011. Axiom Genome-Wide BOS1 Array Plate. Available at: http://media.affymetrix.com/support/technical/datasheets/axiom_gw_bos1_arrayplate_datasheet.pdf (Accessed 15 October 2021).

Amos, C.I., Shete, S., Chen, J. & Yu, R.K., 2003. Positional identification of microdeletions with genetic markers. Human Heredity, 56, 107–118.

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P. & Zondervan, K. T., 2010. Data quality control in genetic case-control association studies. Nature Protocols 5, 1564 – 1573.

Agri SA., 2019. "We are in financial drought." Report to the Multi-stakeholder Task Team on the Drought. Pretoria, Agri SA. Available online:

http://www.nstf.org.za/wp-content/uploads/2016/06/Agri-SA-Drought-Report_CS4.pdf

Aguilar, I., 2014. SeekParentF90. Available online:

http://nce.ads.uga.edu/wiki/doku.php?Id=readme.seekparentf90

Banos, G., Wiggans, G.R. & Powell, R.L., 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. Journal of Dairy Science, 84, 2523-2529.

Bastiaansen, J. W. M., Coster, A., Calus, M. P. L., van Arendonk, J. A. M. & Bovenhuis, H., 2012. Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. Genetic Selection Evolution, 44, 1, 1-13.

Beffa, L.M., Van Wyk, J.B. & Erasmus, G.J., 2009. Long-term selection experiment with Afrikaner cattle 1. Environmental factors affecting calf growth traits. South African Journal of Animal Science, 39(2).

Belay, Jr, T.K., 2013. Development of a SNP markers panel for parentage testing and individual identification in the Norwegian white sheep (NWS) population. MSc Thesis. Norwegian University of Life Sciences. Department of Animal and Aquaculture Sciences. Norway.

Bergh, L., 2010. The national beef recording and improvement scheme. In: Beef Breeding in South Africa. 2nd ed. Ed. Scholtz, M.M., Agricultural Research Council, Pretoria, South Africa. 55-70.

Berry, D.P., O'Brien, A., Wall, E., McDermott, K., Randles, S., Flynn, P., Park, S., Grose, J., Weld, R. & McHugh, N., 2016. Inter-and intra-reproducibility of genotypes from sheep technical replicates on Illumina and Affymetrix platforms. Genetics Selection Evolution, 48, 1, 1-5.

Berry, D.P., 2019. Genomic information in livestock has multiple uses in precision breeding and management. Livestock, 24, 1, 30-33.

Biegelmeyer, P., Gulias-Gomes, C.C., Caetano, A.R., Steibel, J.P. &. Cardoso, F.F., 2016. Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. BMC Genetics, 17, 1, 1-12.

Bisschoff, C. & Lotriet, R., 2013, July. The Drakensberger as competitive breed of cattle in the South African beef industry. In 19th International Farm Management Congress, 39-49.

Bonsma, J.C., 1980. Cross-breeding, breed creation and the genesis of the Bonsmara. Livestock production: A global approach. Cape Town: Tafelberg Publishers Ltd, 126-36.

Bohmanova, J., Sargolzaei, M. & Schenke, F.S., 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics, 11, 1, 1-11.

Bouw, J., 1958. Blood group studies in Dutch cattle breeds, Thesis, Utrecht. Veenman en Zn, Wageningen, 84.

Bradford, H.L., Masuda, Y., Cole, J.B., Misztal, I. & VanRaden, P.M., 2019. Modeling pedigree accuracy and uncertain parentage in single-step genomic evaluations of simulated and US Holstein datasets. Journal of Dairy Science, 102, 3, 2308–2318.

Brenig, B. & Schűtz, E., 2016. Recent development of allele frequencies and exclusion probabilities of microsatellites used for parentage control in the German Holstein Friesian cattle population. BMC Genetics, 17, 1, 1-9.

Buchanan, J.W., Woronuk, G.N., Marquess, F.L., Lang, K., James, S.T., Deobald, H., Welly, B.T. & Van Eenennaam, A.L., 2016. Analysis of validated and population specific SNP parentage panels in pedigreed and commercial beef cattle populations. Canadian Journal of Animal Science, 97, 231-240.

Carolino, N. & Gama, L.T., 2008. Inbreeding depression on beef cattle traits: Estimates, linearity of effects and heterogeneity among sire-families. Genetic Selection Evolution, 40, 5, 1-17.

Cattle molecular markers and parentage testing workshop (CMMPT)., 2012. Proceedings of the 33rd Conference of the International Society of Animal Genetics, Cairns, Australia, July15-20.

Cercueil, A., Bellemain, E. & Manel, S., 2002. PARENTE: Computer program for parentage analysis. Journal of Heredity, 93, 458-459.

Chan, E.K.F., Hawken, R. & Reverter, A., 2008. The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. Animal Genetics, 40, 149-156.

Charlier, C., Agerholm, J.S., Coppieters, W., Karlskov-Mortensen, P., Li, W., De Jong, G., Fasquelle, C., Karim, L., Cirera, S., Cambisano, N. & Ahariz, N., 2012. A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. PLoSOne, 7, e43085.

Christensen, L.G., Madsen, P. & Petersen, J., 1982. The influence of incorrect sire-identification on the estimates of genetic parameters and breeding values. Proceedings of the 2nd World Congress Genetics Applied in Livestock Production, Madrid, Spain, 7, 200–208.

Christie, M.R., Tennessen, J.A., Blouin, M.S., 2013. Bayesian parentage analysis with systematic accountability of genotyping error, missing data, and false matching. Bioinformatics, 29, 725–732.

Conrad, D. F, Andrews, T., Carter, N. P., Hurles, M. E. & Pritchard, J. K., 2006. A high-resolution survey of deletion polymorphism in the human genome. In: Nature Genetics 38, 1, 75–81.

Cooper, T. A., Wiggans, G. R. & VanRaden. P. M., 2013. Short communication: Relationship of call rate and accuracy of single nucleotide polymorphism genotypes in dairy cattle. Journal of Dairy Science, 96, 3336–3339.

Daetwyler, H.D., Villanueva, B., Bijma, P. & Woolliams, J.A., 2007. Inbreeding in genome-wide selection. Journal of Animal Breeding and Genetics, 124 369– 376.

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R. F., Liao, X., Djari, A., Rodriguez, S. C., Grohs, C. & Esquerre, D., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nature genetics, 46, 8, 858-865.

Department of Agriculture, Land Reforms and Rural Development (DALRRD)., 2020. Trends in Agricultural Sector. Available online: https://www.dalrrd.gov.za/Portals/0/Statistics%20and%20Economic%20Analysis/Statistical%20Information/Trends%20in%20the%20Agricultural%20Sector%202020.pdf

Department of Agriculture, Land Reforms and Rural Development (DALRRD)., 2021. Statistical and economic publications and reports. Available online: https://www.dalrrd.gov.za/Home/Crop-Estimates/Statistical-Information/Livestock

Dodds, K. G., Tate, M. L. & Sise, J. A., 2005. Genetic evaluation using parentage information from genetic markers. Journal of Animal Science, 83, 2271–2279.

Dotsev, A.V., Deniskova, T.E., Okhlopkov, I.M., Mészáros, G., Sölkner, J., Reyer, H., Wimmers, K., Brem, G. & Zinovieva N.A., 2018. Genome-wide SNP analysis unveils genetic structure and

phylogeographic history of snow sheep (Ovis nivicola) populations inhabiting the Verkhoyansk Mountains and Momsky Ridge (northeastern Siberia). Ecology Evolution, 8, 8000-8010.

Drakensberger breed homepage. https://drakensbergers.co.za/English/ (Accessed 20 January 2022)

Dunne, F.L., Berry, D.P., Kelleher, M.M., Evans, R.D., Walsh, S.W. & Amer, P.R., 2020. An index framework founded on the future profit potential of female beef cattle to aid the identification of candidates for culling. Journal of Animal Science, 98, 11, 1-14.

Du Plessis, I., Hoffman, L.C. & Calitz, F.J., 2006. Influence of reproduction traits and pre-weaning growth rate on herd efficiency of different beef breed types in an arid sub-tropical environment. South African Journal of Animal Science, 36, 89-98.

Durkin, K., Coppieters, W., Drögemüller, C., Ahariz, N., Cambisano, N., Druet, T., Fasquelle, C., Haile, A., Horin, P., Huang, L., Kamatani, Y., Karim, L., Lathrop, M., Moser, S., Oldenbroek, K., Rieder, S., Sartelet, A., Sölkner, J., Stalhammar, H., Zelenika, D., Zhang, Z., Leeb, T., Georges, M. & Charlier, C., 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature, 482, 7383, 81-84.

Falconer, D.S. & Mackay, T.F.C., 1996. Introduction to Quantitative Genetics. Essex, UK: Longman Group.

Farmer's Weekly., 2019. Stealing the future of farming - feature - stock theft. Available online: https://hdl.handle.net/10520/EJC-15a6822a8f

Felius, M., Beerling, M.L., Buchanan, D.S., Theunissen, B., Koolmees, P.A. & Lenstra, J.A., 2014. On the history of cattle genetic resources. Diversity 6, 705-750.

Ferdosi, M.H., Kinghorn, B.P., van der Werf, J.H.J., Lee, S.H. & Gondro, C., 2014. Hsphase: an R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. BMC Bioinformatics, 15, 1, 1-11.

Ferguson, L. C., 1941. Heritable antigens in the erythrocytes of cattle. The Journal of Immunology, 40, 2, 213-242.

García-Ruiz, A., Wiggans, G. R. & Ruiz-López, F. J., 2019. Pedigree verification and parentage assignment using genomic information in the Mexican Holstein population. Journal of Dairy Science, 102, 1806-1810.

Garrick, D. J., 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genetics Selection Evolution, 43, 1, 1-11.

Gebrehiwot, N. Z., Strucken, E. M., Marshall, K., Aliloo, H. & Gibson, J. P., 2021. SNP panels for the estimation of dairy breed proportion and parentage assignment in African crossbred dairy cattle. Genetics Selection Evolution, 53, 1, 1-13.

Goodnight, K.F. & Queller, D.C., 1999. Computer software for performing likelihood tests of pedigree relationship using genetic markers. Molecular Ecology, 8, 1231–1234.

Grashei, K.E., Ødegård, J. & Meuwissen, T.H.E., 2018. Using genomic relationship likelihood for parentage assignment. Genetics Selection Evolution, 50, 1, 1-11.

Grobler, S.M., Scholtz, M.M., Greyling, J.P.C. & Neser, F.W.C., 2014. Reproduction performance of beef cattle mated naturally following synchronization in the Central Bushveld bioregion of South Africa. South African Journal of Animal Science, 44, 5, 70-74.

Hadfield, J.D., Richardson, D.S., Burke, T., 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. Molecular Ecology, 15, 3715–3730.

Hara, K., Watabe, H., Sasazaki, S., Mukai, F. & Mannen, H., 2010. Development of SNP markers for individual identification and parentage test in a Japanese black cattle population. Journal of Animal Science, 81, 2, 152-157.

Harder, B., Bennewitz, J., Reinsch, N., Mayer, M. & Kalm, E., 2005. Effect of missing sire information on genetic evaluation. Archives Tierzuctch, 48, 3, 219–232.

Hayes, B. J., Visscher, P. M., McPartlan, H. C. & Goddard, M. E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research, 13, 4, 635–643.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science, 92, 2, 433–443.

Hayes, B.J., 2011. Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. Journal of Dairy Science, 94, 4, 2114-2117

Heaton, M.P., Harhay, G.P., Bennett, G.L., Stone, R.T., Grosse, W.M., Casas, E., Keele, J.W., Smith, T.P.L., Chitko-McKown, G.C. & Laegreid, W.W., 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. Mammalian Genome, 13, 272-281.

Henderson, C. R., 1973. Sire evaluation and genetic trends. Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush. Champaign, IL., American Society of Animal Science and American Dairy Science Association, 10-41.

Himmelbauer, J., Mészáros, G. & Sölkner, J., 2019. Detection of autosomal hemizygous regions in the Fleckvieh population-based on SNP-chip data and parent-offspring pairs. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 67, 6, 1447–1452.

Himmelbauer, J., 2019. Detection of autosomal hemizygous regions in the Fleckvieh population based on SNP-chip data and parent-offspring pairs. MSc Thesis. University of Natural Resources and Applied Life Sciences. Department of Sustainable Agricultural Systems. Division of Livestock Sciences. Subdivision Animal Breeding and Genetics. Vienna.

Huisman, J., 2017. Pedigree reconstruction for SNP data: parentage assignment, sibship clustering and beyond. Molecular Ecology Resources, 17, 1009–1024.

Ibeagha-Awemu, E.M., Peters, S.O., Bemji, M.N., Adeleke, M.A. & Do, D.N., 2019. Leveraging available resources and stakeholder involvement for improved productivity of African livestock in the era of genomic breeding. Frontiers in Genetics, 10, 357.

Illumina Inc., 2010. BovineHD Genotyping Beadchip. Available at: http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf (Accessed 15 October 2021)

Illumina Inc., 2011. Golden Gate Bovine 3K Genotyping Bead-Chip. Avaialable at: http://www. illumina.com/Documents/products/ datasheets/datasheet_bovine3K.pdf (Accessed 15 October 2021).

International Society for Animal Genetics (ISAG)., 2013. ISAG cattle core and additional SNP panel. https://www.isag.us/committees.asp

Jamieson, A. & Taylor, St CS., 1997. Comparisons of three probability formulae for parentage exclusion. Animal genetics, 28, 6, 397-400.

Jones, A.G. & Ardren, W.R., 2003. Methods of parentage analysis in natural populations. Molecular Ecology 12, 2511−2523.

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

Jones, O.R. & Wang, J., 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources, 10, 551–555.

Judge, M. M., Kearney, J. F., McClure, M. C., Sleator, R. D. & Berry, D. P., 2016. Evaluation of developed low-density genotype panels for imputation to higher density in independent dairy and beef cattle populations. Journal of Animal Science, 94, 949–962.

Junqueira, V.S., Cardoso, F.F., Oliveira, M.M., Sollero, B.P., Silva, F.F. & Lopes, P.S., 2017. Use of molecular markers to improve relationship information in the genetic evaluation of beef cattle tick resistance under pedigree-based models. Journal of Animal Breeding and Genetics, 133, 1-13.

Kadri, N.K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., Sander Nielsen, U., Panitz, F., Pedersen Aamand, G., Schulman, N., Georges, M., Vilkki, J., Lund, M. S. &

Druet, T., 2014. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. Plos Genetics, 10, 1, e1004049.

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

Kios, D., van Marle-Köster, E. & Visser, C., 2012. Application of DNA markers in parentage verification of Boran cattle in Kenya. Tropical Animal Health Production, 44, 471-476.

Maiwashe, A., Nephawe K. A. & Theron, H. E., 2008. Estimates of genetic parameters and effect in inbreeding on milk yield and composition in South African Jersey cows. South African Journal for Animal Science, 38, 119–125.

Makina, S.O., Muchadeyi, F.C., Van Marle-Köster, E., MacNeil, M.D. & Maiwashe, A., 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Frontiers in Genetics, 5, 1-7.

Mapholi, N.O., Maiwashe, A., Matika, O., Riggio, V., Bishop, S.C., MacNeil, M.D., Banga, C., Taylor, J.F. & Dzama, K., 2016. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick-Borne Diseases, 7, 487-497.

Mare, F., Bahta, Y.T. & Van Niekerk, W., 2018. The impact of drought on commercial livestock farmers in South Africa, Development in Practice, 28, 7, 884-898.

Marshall, T.C., Slate, J., Kruuk, L.E.B. & Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology, 7, 639−655.

Marufu, M.C., Chimonyo, M., Mapiye, C. & Dzama, K., 2011. Tick loads in cattle raised on sweet and sour rangelands in the low-input farming areas of South Africa. Tropical Animal Health Production, 43, 307-313.

Masashi, S. & Shigeho, K., 2012. PARFEX v1.0: an EXCEL™ based software package for parentage allocation. Conservation Genetics Resources, 4, 275–278.

McClure, M.C., McCarthy, J., Flynn, P., Weld, R., Keanne, M., O'Connell, K., Mullen, M.P., Waters, S. & Kearney, J.F., 2015. SNP selection for nationwide parentage verification in beef and dairy cattle. Proceeding of the International Committee of Animal Recording (ICAR), Technical series, 19, 175-181. Krakow, Poland, June 10-12.

McClure, M.C., McCarthy, J., Flynn, P., McClure, J. C., Dair, E., O'Connell, D. K. & Kearney, J. F., 2018. SNP data quality control in a national beef and dairy Cattle system and highly accurate SNP based parentage verification and identification. Frontiers in Genetics, 9, 84.

Meagher, T.R. & Thompson, E., 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theoretical Population Biology, 29, 1, 87-106.

Melo, A.T.O. & Hale, I., 2019. 'apparent': a simple and flexible R package for accurate SNP-based parentage analysis in the absence of guiding information. BMC Bioinformatics 20, 1, 1-10.

Mesbah-Uddin, M., Guldbrandtsen, B., Iso-Touru, T., Vilkki, J., De Koning, D.J., Boichard, D., Lund, M.S. & Sahana, G., 2018. Genome-wide mapping of large deletions and their population-genetic properties in dairy cattle. DNA Research, 25, 1, 49-59.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157, 1819–1829.

Misztal, I., Vitezica, Z.G., Legarra, A., Aguilar, I. & Swan, A.A., 2013. Unknown-parent groups in single-step genomic evaluation. Journal of Animal Breeding and Genetics, 130, 4, 252-258.

Moore, K. L., Vilela, C., Kaseja, K., Mrode, R. & Coffey, M., 2019. Forensic use of the genomic relationship matrix to validate and discover livestock pedigrees. Journal of Animal Science, 97, 35–42.

More, M., Gutiérrez, G., Rothschild, M., Bertolini, F. & Ponce de León, F.A., 2019. Evaluation of SNP genotyping in Alpacas using the Bovine HD genotyping Beadchip. Frontier in Genetics, 10, 361.

Mrode, R., Ojango, J., Okeyo, A. M. & Mwacharo, J. M., 2019. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. Frontiers in genetics, 9, 694.

Nandolo, W., Utsunomiya, Y. T., Mészáros, G., Wurzinger, M., Khayadzadeh, N., Torrecilha, R.B.P., Mulindwa, H. A., Gondwa, T.N., Waldmann, P., Ferencakovic, M., Garcia, J., Rosen, B.D., Bickhart, D., Van Tassell, C.P., Curik. I. & Sölkner, J., 2018. Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. Genetics Selection Evolution, 50, 1, 43.

Neuner, S., Edel, C., Emmerling, R., Thaller, G. & Götz, K,U., 2009. Precision of genetic parameters and breeding values estimated in marker assisted BLUP genetic evaluation. Genetic Selection Evolution, 41, 1, 1-10.

Nwogwugwu, C. P., Kim, Y. K., Choi, H. J., Lee, J. H. & Lee, S. H., 2020. Assessment of genomic prediction accuracy using different selection and evaluation approaches in a simulated Korean beef cattle population. Australasian Journal of Animal Sciences, 33, 12, 1912– 1921.

Oishi, K., Kato, Y., Ogino, A. & Hirooka, H., 2013. Economic and environmental impacts of changes in culling parity of cows and diet composition in Japanese beef cow–calf production systems. Agricultural Systems, 115, 95-103.

Oliphant, A., Barker, D.L., Stuelpnagel, J.R. & Chee, M.S., 2002. BeadArray™ technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques, 32(sup), S56-S61.

Panetto, J. C. C., Machado, M. A., Silva, M. V. G. B., Barbosa, R. S., Santos, G. G., Leite, R. M. H. & Peixoto, M.G.C.D., 2017. Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. Livestock Science, 204, 33–38.

Parlato E. & Van Vleck L.D., 2012. Effect of parentage misidentification on estimates of genetic parameters for milk yield in the Mediterranean Italian buffalo population. Journal of Dairy Science, 95, 4059–4064.

Pentz, E., 2009. Marketing perceptions of the Drakensberger breed of cattle. MBA dissertation. North-West University, Potchefstroom, South Africa.

Pereira, R.J., Santana, M.L. Jr., Ayres, D.R., Bignardi, A.B., Menezes, G.R., Silva, L.O.C., Machado, C.H.C., Josahkian, L.A. & Albuquerque, L.G., 2016. Inbreeding depression in Zebu cattle traits. Journal of Animal Breeding and Genetics, 133, 523–533.

Peters S., Kizilkaya K., Garrick D., Fernando R., Pollak E., Enns R., De Donato M., Ajayi O. & Imumorin I., 2014. Use of robust multivariate linear mixed models for estimation of genetic parameters for carcass traits in beef cattle. Journal of Animal Breeding and Genetics, 131, 504–512.

Pollak, E.J., 2005. Application and impact of new genetic technologies on beef cattle breeding: a “real world” perspective. Australian Journal of Experimental Agriculture, 45, 739-748.

Pollak, J., 2006. SireMatch version 2.0.2. Computer Program, Cornell University, Ithaca, NY

Pritchard, J.K. & Przeworski, M., 2001. Linkage disequilibrium in humans: models and data. American Journal of Human Genetics, 69, 1–14.

Pszczola, M., Strabel, T., Mulder, H.A. & Calus, M.P.L., 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. Journal of Dairy Science, 95, 389–400.

Purcell, S., Neale, B. & Todd-Brown, K., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81, 559-575.

Purfield, D.C., McClure, M. & Berry, D.P., 2016. Justification for setting the individual animal genotype call rate threshold at eighty-five percent. Journal of Animal Science, 94, 4558-4569.

Rodríguez-Ramilo, S.T., García-Cortés, L.A. & de Cara, M.Á.R., 2015. Artificial selection with traditional or genomic relationships: consequences in coancestry and genetic diversity. Frontiers in Genetics, 6, 127.

Rust, T. & Groeneveld, E., 2002. Variance component estimation of female fertility traits in two indigenous and two European beef cattle breeds of South Africa. South African Journal of Animal Science, 32, 1, 23-29.

Sanders K., Bennewitz J. & Kalm E., 2006. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. Journal of Dairy Science, 89, 315-321.

Santana, M.L., Eler, J.P., Ferraz, J.B.S. & Mattos, E.C., 2012. Genetic relationship between growth and reproductive traits in Nellore cattle. Animal, 6, 565–570.

Schoeman, S.J., 1989. Recent research into the production potential of indigenous cattle with special reference to the Sanga. South African Journal Animal Science, 19, 55-61.

Schoeman, S.J., 1996. Characterization of beef cattle breeds by virtue of their performances in the National Beef Cattle Performance and Progeny Testing Scheme. South African Journal of Animal Science, 26, 1, 15-19.

Scholtz, M.M., Bosman, D.J., Erasmus, G.J. & Maiwashe, A., 2010. Selection as the base of improvement in beef cattle. Beef Breeding in South Africa, (Ed. 2), 2-9.

Schütz, E. & Brenig, B., 2015. Analytical and statistical consideration on the use of the ISAG-ICAR-SNP bovine panel for parentage control, using the Illumina BeadChip technology: example on the German Holstein population. Genetics Selection Evolution, 47, 1, 1-7.

Schütz, E., Wehrhahn, C., Wanjek, M., Bortfeld, R., Wemheuer, W.E., Beck, J. & Brenig, B., 2016, The Holstein Friesian lethal haplotype 5 (HH5) results from a complete deletion of TBF1M and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of APOB. PLoSOne, 11, e0154602

Sherman, G.B., Kachman, S.D., Hungerford, L.L., Rupp, G.P., Fox, C.P., Brown, M.D., Feuz, B.M. & Holm, T.R., 2004. Impact of candidate sire number and sire relatedness on DNA polymorphism-based measures of exclusion probability and probability of unambiguous parentage. Animal Genetics, 3, 220–226.

Silva, D.A., Costa, C.N., Silva, A.A., Silva, F.F., Lopes, P.S., Santos, G.G., Thompson, G. & Carvalheira, J., 2019. Unknown parent and contemporary groups for genetic evaluation of Brazilian Holstein using autoregressive test-day models. Livestock Science 220, 1–7.

Sonesson, A.K., Woolliams, J.A. & Meuwissen, T.H., 2012. Genomic selection requires genomic control of inbreeding. Genetics Selection Evolution, 44, 27.

Sonesson, A.K. & Meuwissen, T.H.E., 2000. Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genetics Selection Evolution, 32, 3, 231-248.

South African (SA) Stud Book, 2016. SA Stud Book annual report. Available online: http://www.sastudbook.co.za/images/photos/Annual_Report_2016_a.pdf.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H. & Gondro, C., 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, 45, 572-575.

Strucken, E.M., Lee, S.H., Lee, H.K., Song, K.D., Gibson, J.P. & Gondro, C., 2016. How many markers are enough? Factors influencing parentage testing in different livestock populations. Journal of Animal Breeding and Genetics, 133, 1, 13-23.

Strucken, E.M., Al-Mamun, H.A., Esquivelzeta-Rabell, C., Gondro, C., Mwai, O.A. & Gibson, J.P., 2017. Genetic tests for estimating dairy breed proportion and parentage assignment in East African crossbred cattle. Genetic Selection Evolution, 49, 67.

Strydom, P.E., 2008a. Do indigenous Southern African cattle breeds have the right genetics for commercial production of quality meat? Meat Science, 80, 86-93.

Strydom, P. E., Frylinck, L., van der Westhuizen, J. & Burrow, H. M., 2008b. Growth performance, feed efficiency and carcass and meat quality of tropically adapted breed types from different farming systems in South Africa. Australian Journal of Experimental Agriculture, 48, 599-607.

Swan, A. A., Johnston, D. J., Brown, D. J., Tier, B. & Graser, H. U., 2011. Integration of genomic information into beef cattle and sheep genetic evaluations in Australia. Animal Production Science, 52, 126-132.

Theron, H.E., Kanfer, F.H.J. & Rautenbach, L., 2002. The effect of phantom parent groups on genetic trend estimation. South African Journal of Animal Science, 32, 130– 135.

Van der Westhuizen, L., MacNeil, M.D., Scholtz, M.M., Neser, F.W.C., Makgahlela, M.L. & Van Wky, J.B., 2019. Genetic variability and relationships in nine South African cattle breeds using microsatellite markers. Tropical Animal Health Production, 52, 177–184.

Van Doormaal, B.J., Kistemaker, G.J. & Johnston, J., 2016. Parentage Analysis Services for Dairy Cattle in Canada. Interbull Bulletin no. 50. Puerto Varas, Chile, October 24-28.

Van Eenennaam, A.L., Weaber, R.L., Drake, D.J., Penedo, M.C.T., Quaas, R.L., Garrick, D.J. & Pollak, E.J., 2007. DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting. Journal of Animal Science, 85, 3159–69.

Van Eenennaam, A. L. & Drake. D. J., 2012. Where in the beef-cattle supply chain might DNA tests generate value? Animal Production Science, 52, 185–196.

Van Eenennaam, A.L., Weigel, K.A., Young, A.E., Cleveland, M.A. & Dekkers, J.C.M., 2014. Annual Review of Animal Biosciences, 2, 1, 105-139.

Van Marle-Köster, E., Visser, C. & Berry, D.P., 2013. A review of genomic selection - Implications for the South African beef and dairy cattle industries. South African Journal of Animal Science, 1, 43.

Van Marle-Köster, E. & Visser, C., 2018a. Genomics for the advancement of livestock production: A South African perspective. South African Journal of Animal Science, 48, 808-817.

Van Marle-Köster, E. & Visser, C., 2018b. Genetic improvement in South African livestock: can genomics bridge the gap between the developed and developing sectors? Frontier in Genetics, 9, 331.

Van Marle-Köster, E., Visser, C., Sealy, J. & Frantz, L., 2021. Capitalizing on the Potential of South African Indigenous Beef Cattle Breeds: A Review. Sustainability, 13, 4388.

VanLiere, J. M. & Rosenberg, N. A., 2008. Mathematical properties of the r^2 measure of linkage disequilibrium. Theoretical population biology, 74, 1, 130–137.

Van Vleck, L.D., 1970. Misidentification in estimating the paternal-sib correlation. Journal of Dairy Science, 53, 1469–1474.

Veerkamp, R. F., Mulder, H. A., Thompson, R. & Calus, M. P. L., 2011 Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. Journal of Dairy Science, 94, 4189–4197.

Visscher, P.M., Woolliams, J.A., Smith, D. & Williams, J.L., 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. Journal of Dairy Science, 85, 2368-2375.

Visscher, P.M., Medland, S.E., Ferreira, M.A.R., Morley, K.I., Zhu, G., Cornes, B.K., Montgomery, G.W. & Martin, N.G., 2006. Assumption-free estimation of heritability from genome-wide indentity-by-descent sharing between full siblings. PLoS Genetics, 2, 3, e41.

Visscher, P.M., Hill, W.G. & Wray, N.R., 2008. Heritability in the genomics era–concepts and misconceptions. Nature Reviews Genetics, 9 (4), 255-266.

Walsh, K. & Spazzoli, R., 2018. Assessing the economic impact of the South African beef genomics program. Available online: http://wagyu.org.za/wp-content/uploads/2018/08/Impact-of-the-Beef-Genomics-Program-FinalReport_060418-Round-1.pdf.

Werner, F.A.O., Durstewitz, G., Habermann, F.A., Thaller, G., Kramer, W., Kollers, S., Buitkamp, J., Georges, M., Brem, G., Mosner, J. & Fries, R., 2004. Detection and characterization of SNPs

useful for identity control and parentage testing in major European dairy breeds. Animal Genetics, 35, 44–49.

Weller, J.I., Ezra, E. & Ron, M., 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. Journal of Dairy Science, 100, 8633–8644.

Wientjes, Y. C. J., Veerkamp, R.F. & Calus, M.P.L., 2013. The Effect of Linkage Disequilibrium and Family Relationships on the Reliability of Genomic Prediction, Genetics, 193, 1, 621–631.

Wiggans, G. R., Cooper, T. A., VanRaden, P. M., Olson, K. M. & Tooker, M. E., 2012. Use of the Illumina Bovine 3K BeadChip in dairy genomic evaluation. Journal of Dairy Science, 95, 1552–1558.

Winchester, L., Yau, C. & Ragoussis, J., 2009. Comparing CNV detection methods for SNP arrays. Briefings in functional genomics and proteomics, 8(5), 353-366.

Wittenburg, D., Teuscher, F., Klosa, J. & Reinsch, N., 2016. Covariance between genotypic effects and its use for genomic inference in half-sib families. G3: Genes, Genomes, Genetics, 6, 9, 2761-2772.

Xu, L., Hon, Y., Bickhart, B. M., Hou, Y., Song, J., Van Tassell, C. P., Sonstegard, T. S. & Liu, G. E., 2014a. A genome-wide survey reveals a deletion polymorphism associated with resistance to gastrointestinal nematodes in Angus cattle. Functional and integrative genomics, 14, 683.

Xu, L., Cole, J. B., Bickhart, D. M., Hou, Y., Song, J., VanRaden, P. M., Sonstegard, T. S., Van Tassell, C. P. & Liu, G. E., 2014b. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics, 16, 683.

Zhang, T., Guo, L., Shi, M., Xu, L., Chen, Y., Zhang, L., Gao, H., Li, J. & Gao, X., 2018. Selection and effectiveness of informative SNPs for paternity in Chinese Simmental cattle based on a high-density SNP array. Gene, 5, 673, 211-216.

Zhao, F., McParland, S., Kearney, F., Du, L. & Berry, D.P. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genetics Selection Evolution, 47, 49.

Zwane, A.A., Schnabel, R.D., Hoff, J., Choudhury, A., Makgahlela, M.L., Maiwashe, A., van Marle Köster, E. & Taylor, J.F., 2019. Genome-wide SNP discovery in indigenous cattle breeds of South Africa. Frontiers in Genetics, 10, 273.

CHAPTER 3

Evaluation of the International Society for Animal Genetics bovine single nucleotide polymorphism parentage panel in South African Bonsmara and Drakensberger cattle

Yandisiwe P. Sanarana[1,2#] · Azwihangwisi Maiwashe[2] · Donagh P. Berry[1,3] · Cuthbert Banga[2] · Este van Marle-Köster[1]

[1]University of Pretoria, Department of Animal and Wildlife Science, Pretoria, Hatfield, 0002, South Africa

[2]Agricultural Research Council-Animal Production, Pretoria, Irene, 0062, South Africa

[3]Teagasc, Animal and Grassland Research, and Innovation Center, Fermoy, Cork, P61 C996, Ireland

Published in Tropical Animal Health and Production: 23 November 2020 doi.org/10.1007/s11250-020-02481-6

CHAPTER 3

Evaluation of the International Society for Animal Genetics bovine single nucleotide polymorphism parentage panel in South African Bonsmara and Drakensberger cattle

Yandisiwe P. Sanarana[1,2#] · Azwihangwisi Maiwashe[2] · Donagh P. Berry[1,3] · Cuthbert Banga[2] · Este van Marle-Köster[1]

[1]University of Pretoria, Department of Animal and Wildlife Science, Pretoria, Hatfield, 0002, South Africa

[2]Agricultural Research Council-Animal Production, Pretoria, Irene, 0062, South Africa

[3]Teagasc, Animal and Grassland Research, and Innovation Center, Fermoy, Cork, P61 C996, Ireland

Abstract

A panel of 200 single nucleotide polymorphisms (SNPs) have been recommended by the International Society for Animal Genetics (ISAG) for use in parentage verification of cattle. While the SNPs included on the ISAG panel are segregating in European *Bos taurus* and *Bos indicus* breeds, their applicability in South African (SA) Sanga cattle has never been evaluated. This study, therefore, assessed the usefulness of the ISAG panel in SA Bonsmara (BON) and Drakensberger (DRB) cattle. Genotypes of 185 ISAG SNPs from 64 BON and 97 DRB sire-offspring pairs were available, all of which were validated with 119 375 SNPs. Of the 185 ISAG SNPs, 14 and 18 in the BON and DRB, respectively (9 in common to both breeds) were either monomorphic, exhibited at least one discordance between validated sire-offspring pairs, or had poor call rate or clustering issue. The mean minor allele frequency of the 185 ISAG SNPs was 0.331 in the BON and 0.359 in the DRB. The combined probability of parentage exclusion (P_E) was the same (99.46%) for both breeds, while the probability of identity varied from 1.61×10^{-48} (BON) to 1.11×10^{-54} (DRB). Fifteen (23.4%) and 32 (33%) of the already validated sire-offspring pairs for the BON and DRB, respectively, were determined by the ISAG panel to be false-negatives based on a threshold of having at least two discordant SNPs. In comparison to sire discovery using the 119 375 SNPs, sire discovery using only the ISAG panel identified correctly 44 (out of 64 identified using the 119 375 SNPs) unique sire-offspring BON pairs and 62 (out of 97 identified using the 119 375 SNPs) unique sire-offspring DRB when all sires were masked. Five BON and three DRB offspring had >1 sire nominated. This study demonstrated that the use of the ISAG panel may result in incorrect exclusions and multiple candidate sires for a given animal. Selection of more informative SNPs is, therefore, necessary in the pursuit of a low-cost and effective SNP panel for indigenous cattle breeds in SA.

Keywords: Sanga cattle, parentage verification, GenTrain score, false-negative

#Corresponding author: Sanaranay@gmail.com

Introduction

Correct pedigrees form the basis for accurate genetic evaluation using the Best Linear Unbiased Prediction (BLUP) statistical method (Henderson, 1973). Incorrect parentage, however, occurs in practice and has been reported in cattle (Weller et al. 2004; Kios et al. 2012; McClure et al. 2018). Parentage errors are known to negatively affect genetic improvement as well as influence the variance of estimated breeding values (EBVs), thus reducing response to selection (Sanders et al. 2006). Banos et al. (2001) indicated that an 11% paternal error rate in genetic evaluations reduced genetic gain by 11 to 15%. Other studies have predicted that a 10% parentage error, with a heritability of 0.25, would contribute to a 3% (Visscher et al. 2002) to 4.3% (Israel and Weller 2000) reduction in genetic gain. For the same pedigree error and a heritability of 10%, a 7% reduction in genetic gain was observed (García-Ruiz et al. 2019). The application of molecular information to correct pedigree records has improved the precision of parentage assignment (Van Eenennaam., 2016; McClure et al. 2018). This has contributed to more accurate EBVs for traits of economic importance especially those of low heritability such as fertility (Berry et al. 2014) and tick resistance (Junqueira et al. 2017).

Multi-allelic microsatellite markers, also known as short tandem repeats markers, were traditionally the preferred genetic markers for parentage verification, to trace meat through the entire food chain and other applications in a broad range of livestock species (Teneva et al. 2018). Despite high polymorphic content, these markers have limitations including the presence of null alleles, allele drop-outs, and imperfect repeats caused by mutations; hence microsatellite markers require a dedicated platform and skilled interpretation to ensure consistent accuracy (Buchanan et al. 2016). Furthermore, microsatellite genotypes from different laboratories are not always comparable due to the inconsistencies in allele scoring and size determination, with genotyping error rates estimated to 5% per locus (Weller et al. 2004; Fernández et al. 2013).

The discovery of bi-allelic single nucleotide polymorphism (SNP) markers has revolutionized animal breeding with several associated genomic applications for livestock improvement (Georges et al. 2018). Subsequently, SNP-based parentage has become the common practice for parentage validation. The superior performance of SNPs over microsatellite markers in parentage validation has been demonstrated in several studies (Fisher et al. 2009; Fernández et al. 2013; Buchanan et al. 2016). In an effort to expedite the shift from microsatellite markers to SNP-based parentage, the International Society for Animal Genetics (ISAG), together with the International Committee for Animal Recording (ICAR), developed a cattle consensus panel of SNP markers for global application, mainly selected for high minor allelic frequency (ISAG, 2013). The panel consists of a core panel of 100 SNP markers, combined with an additional 100 markers (ISAG200). The core 100 SNP panel, mostly derived from European breeds, have been demonstrated to be insufficient for parentage verification in several breeds,

whereas the ISAG200 panel including SNPs from the *indicine* cattle breeds has been demonstrated to have better resolution power (Strucken et al. 2014; Lyons et al. 2015; Van Doormaal et al. 2016).

In the establishment of SNP-based marker systems, a critical factor to consider is whether the chosen SNPs are informative in multiple cattle populations. This emerges from a non-random selection method that preferentially selects alleles segregating at high frequencies in European *Bos taurus* and *Bos indicus* breeds known as ascertainment bias. The efficiency of the SNP panel varies by breed, depending on the respective minor allele frequency (MAF) (Nielsen and Signorovitch, 2003). The restricted number of breeds used in the establishment of the ISAG panel may consequently influence the applicability of the panel in under-represented breeds. Despite concerns on marker ascertainment bias for the interpretation of parentage results, the quality of the genotype remains the basis for further analysis and accuracy. The applicability of the ISAG SNP panel for parentage testing has not been evaluated in South African Sanga cattle breeds or locally developed composites. Therefore, the aim of this study was to evaluate the effectiveness of the bovine ISAG SNP parentage panel in both SA Bonsmara and Drakensberger beef cattle.

Materials and methods

Genotype data and quality control

Ethical approval was received for the use of external data by the Animal Ethics Committee (AEC) of the University of Pretoria (EC066-16 AEC). Genotypes from 1 567 Bonsmara and 1 022 Drakensberger beef cattle were available from the SA Beef Genomics (BGP) and Red Meat Research Development (RMRD) projects. The South African Stud Book and Animal Improvement Association provided the pedigree data for the genotyped animals. Data edits and calculations were all performed using *R* software (R Development Core Team, 2018). All animals were genotyped with the GeneSeek® Genomic Profiler (GGP) Bovine 150K BeadChip with 138 888 SNP markers in common for both the BON and DRB. A total of 119 375 autosomal SNPs with genomic positions assigned based on UMD3.1 bovine genome build were kept for further analysis. Only 185 SNP markers recommended by ISAG for parentage verification existed in the common dataset. This was satisfactory as 90% of the SNPs were present for the basis of verifying an offspring with one parent (ISAG., 2013). A total of 80 individual samples with more than 10% missing SNP genotypes were not considered further.

Evaluation of the ISAG SNP markers

Mean allele frequency per breed was estimated for each ISAG SNP in the BON and DRB separately. The allele and genotype frequencies were used to estimate different genetic parameters. The call rate, minor allele frequency, Hardy-Weinberg Equilibrium (HWE) *p-value*, the level of heterozygosity (expected and observed) and the polymorphic information content (PIC) per SNP were estimated within

each breed separately. The probability of parentage exclusion (P_E) per SNP for the scenario of one known parent (i.e. sire) was calculated according to Jamieson and Taylor (1997) as:

$$P_E = 1 - 4\sum_{i=1}^{n} p_i^2 + 2\left(\sum_{i=1}^{n} p_i^2\right)^2 + 4\sum_{i=1}^{n} p_i^3 - 3\sum_{i=1}^{n} p_i^4$$

where p_i is the MAF for marker i and n the number of alleles per SNP ($n = 2$ since SNP are biallelic markers). The combined probability of parentage exclusion (P_E) over all the SNPs was calculated as P_{Ej} = 1-(1-P_1) (1-P_2) (1- P_3)….(1-P_k) where j is the probability of exclusion per marker and k is the number of loci. The probability of identity (P_I) that two randomly selected individuals in a population would possess identical SNP genotypes was calculated per SNP. The combined P_I, which is the multiple product of each SNP marker probability, was estimated according to Waits et al. (2001) as follows:

$$P_I = \prod_{i=1}^{N} (p_i^2 + 2p_i^2 q_i^2 + q_i^2)$$

where N is the number of SNPs with p^2, $2pq$, and q^2 as the relative genotype frequencies of AA, AB, and BB respectively, p and q being the A and B allele frequencies.

Pedigree verification and sire-offspring genotype mismatches

Based on the available pedigree information, 50 putative sire-offspring pairs in the BON and 93 in the DRB were available, with the number of offspring per sire ranging from 1 to 21. The (in)validation of these sire-offspring pairs was based on the count of opposing homozygous genotypes between the sire and offspring. This was determined per animal (across all SNPs), using all 119 375 autosomal SNPs. An opposing homozygote was defined as when a sire had a homozygous genotype (AA) and the offspring was also homozygous but for the other allele (BB), and vice versa (Hayes, 2011). Sire parentage exclusion was considered whenever the rate of the mismatches exceeded 1%. A total of 124 male animals (71 BON and 53 DRB) were available in the dataset. In a separate analysis, these were considered as candidate sires where all the known sires of the animals were actually masked and, subsequently, predicted using initially the 119 375 SNPs to identify the true sire-offspring pairs and then using the ISAG panel. This new dataset, comprising of validated 64 BON and 97 DRB sire-offspring pairs, was used to evaluate the ISAG SNP panel for sire discovery. False-negatives were estimated as the proportion of wrongly excluded sire-offspring relationships divided by the total number of true sire-offspring pairs.

Results

Of the 185 ISAG SNPs, 175 had a call rate ≥ of 0.95 (Fig. 1), with just five and eight SNPs in the BON and DRB, respectively, exhibiting a call rate < 0.95.

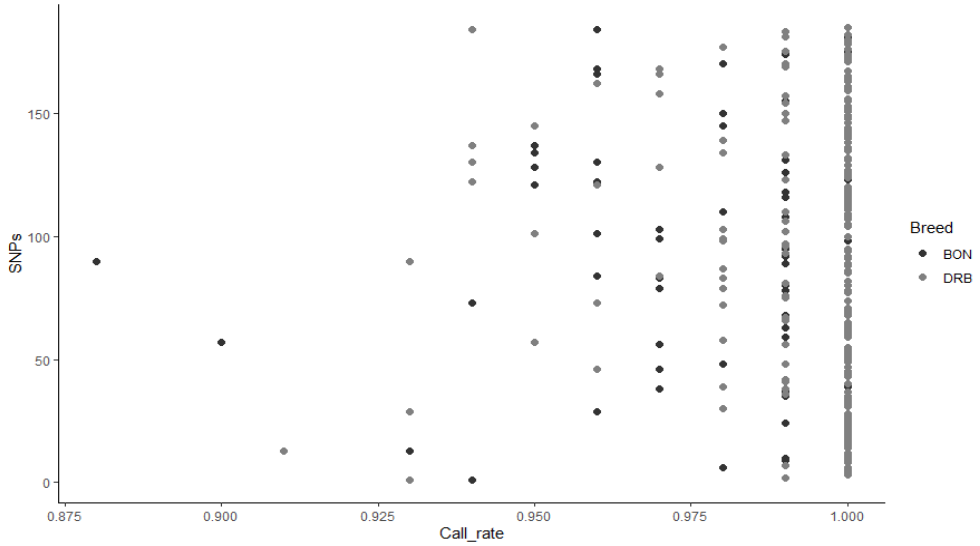


Figure 3.1. A scatter-plot depicting the call rate of the 185 International Society for Animal Genetics single nucleotide polymorphisms (SNPs) in Bonsmara (BON) and Drakensberger (DRB) cattle breeds

There were two SNPs, ARS-BFGL-NGS-11383 (rs110577061) and ARS-USMARC-Parent-DQ786762 (rs29010772), with a low GenTrain (GT) score of < 0.55 (Fig. 2), one in the BON and both in the DRB. The GT score of the 185 SNPs ranged between 0.49 and 0.90 with a mean of 0.78.

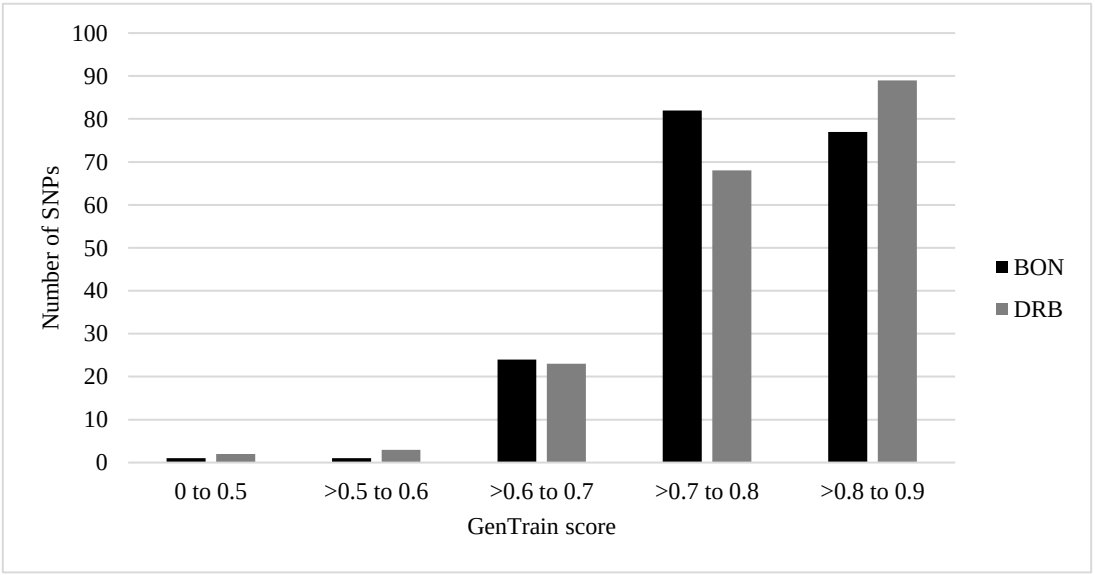


Figure 3.2. GenTrain score distributions for the 185 International Society for Animal Genetics single nucleotide polymorphisms (SNPs) for the Bonsmara (BON) and Drakensberger (DRB) cattle breeds

The majority of the 185 SNPs had a MAF between 0.1 and 0.5 (Fig. 3) with just three monomorphic SNPs per breed. Of the 185 ISAG markers, 135 and 115 SNPs had a MAF ≥ 0.3 in the DRB and BON, respectively.

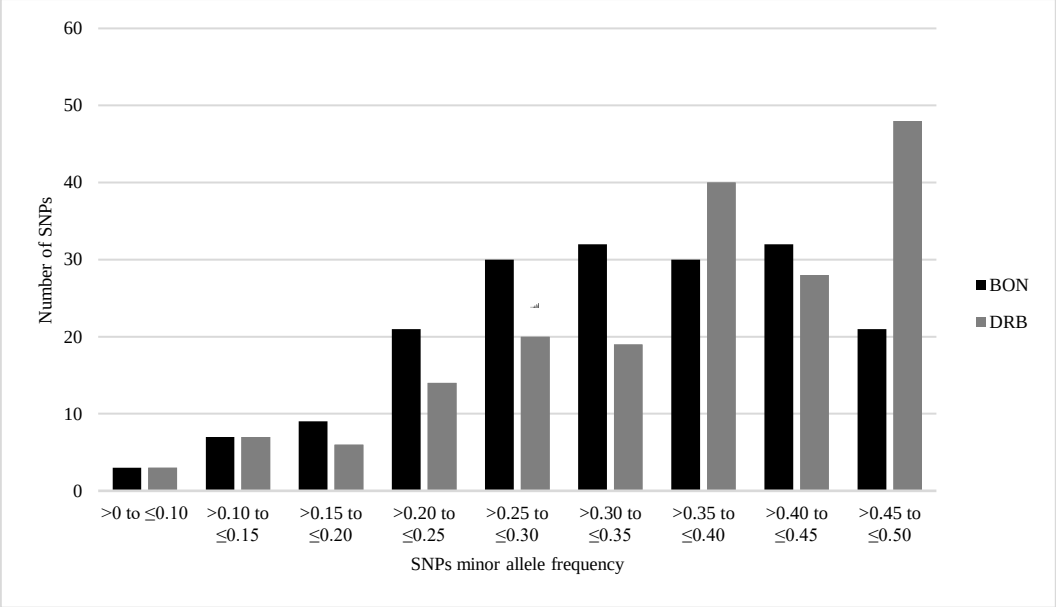
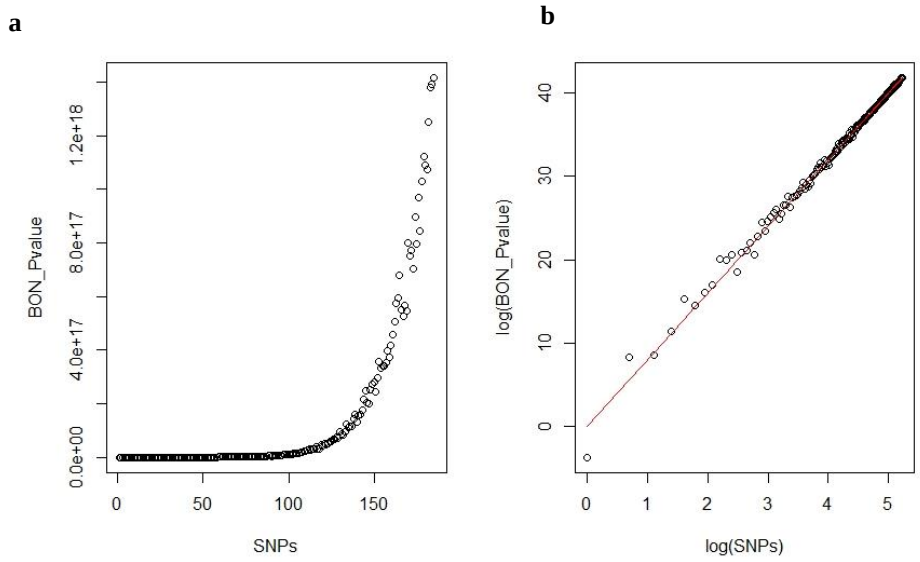


Figure 3.3. Minor allele frequency distributions for the 185 International Society for Animal Genetics single nucleotide polymorphisms (SNPs) for the Bonsmara (BON) and Drakensberger (DRB) cattle breeds

The distribution of the HWE *p-values* for the ISAG SNPs is represented by quantile-quantile (Q-Q) plots in Fig. 4. Eight and twelve SNPs in the BON and DRB, respectively, deviated (P<0.001) from a uniform distribution.



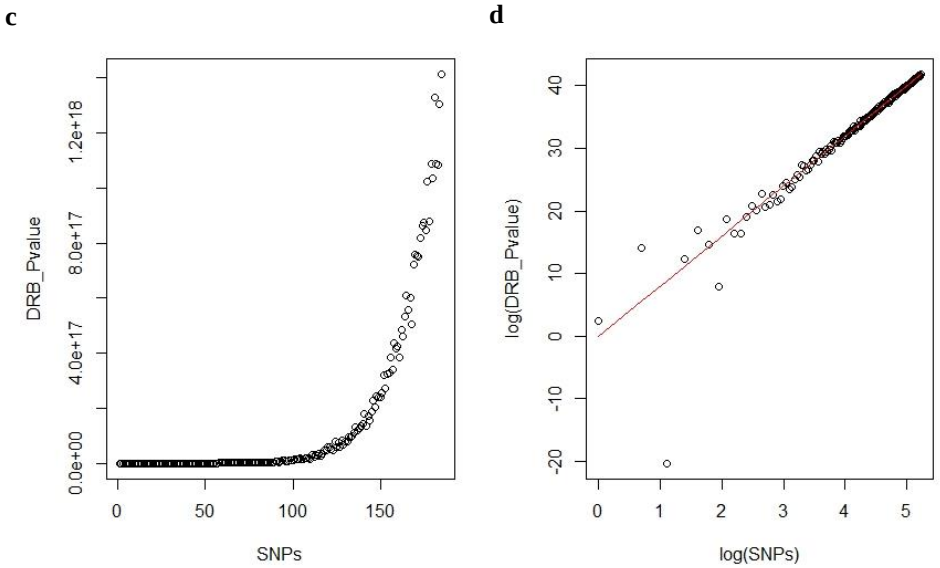


Figure 3.4. Q-Q plots of HWE *p-values* for 185 International Society for Animal Genetics single nucleotide polymorphism (SNP) panel. **a,b** Bonsmara (BON) and **c,d** Drakensberger (DRB) cattle breeds

The average MAF, H_E and PIC in the two breeds are summarised in Table 1. Each parameter was similar in each breed, except for a slightly higher MAF and H_E in the DRB. The P_E was 99.46% in both breeds and, based on the genotype frequencies, the estimated P_I values ranged from 1.61×10^{-48} to 1.11×10^{-54} in the BON and DRB, respectively.

Table 3.1. Descriptive statistics for the SA Bonsmara and Drakensberger breeds using the ISAG markers

Populations	MAF	H_O	H_E	PIC	P_E	P_I
BON	0.331	0.212	0.423	0.330	0.9946	1.61E-48
DRB	0.359	0.214	0.437	0.338	0.9946	1.11E-54

MAF, minor allele frequency; H_E, expected heterozygosity; H_O, observed heterozygosity, PIC, polymorphic information content; P_E, the combined probability of parentage exclusion for the scenario of one known parent; P_I, probability of identity

The number of discordant SNPs between sires and their validated offspring is in Fig.5. These discordances occurred in 13 SNPs with two SNPs in common for both breeds. The frequency of discordances per SNP ranged from one to five in the BON and up to 14 in the DRB.

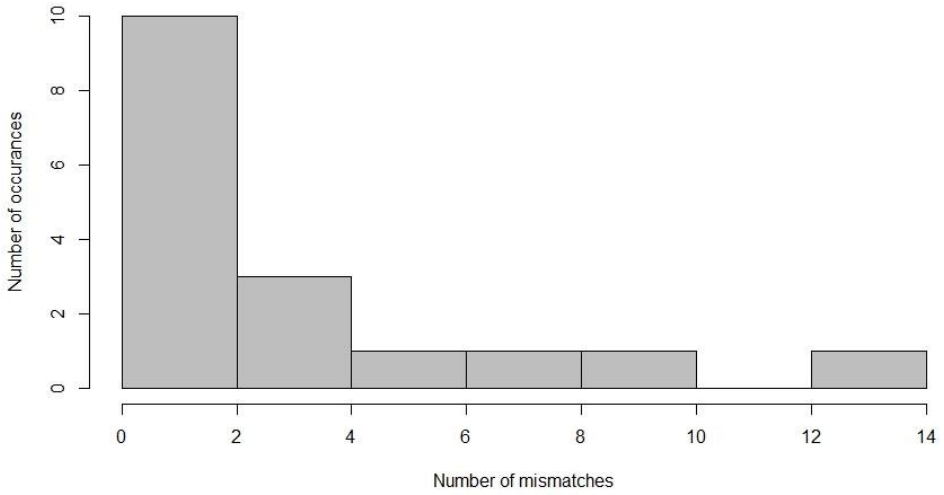


Figure 3.5. Frequency distribution of the number of mismatches between animals and the sires

Of the 50 BON and 93 DRB sire-offspring pairs recorded in the pedigree and verified with 119 375 SNPs, 5 (10.0%) and 16 (17.2%) parentages, respectively, were inconsistent with that reported in the pedigree. In addition, 19 and 20 additional sire-offspring pairs in the BON and DRB, respectively, were discovered when all the sires were masked and predicted using the 119 375 SNPs. Of the total 64 and 97 sire-offspring pairs in the BON and DRB, respectively, detected with the 119 375 SNP panel, the ISAG panel indicated 15 (23.4%) BON and 32 (33%) DRB sire-offspring pairs as false-negatives. Accurate sire discovery with only one candidate sire was possible for 44 BON and 62 DRB animals using the ISAG 185 SNP genotypes. The ISAG panel assigned multiple sires for a further 5 BON and 3 DRB animals all of which also contained the true sire as determined by the 119 375 SNP panel; a single unique sire was identified for these 8 animals when using the 119 375 SNPs.

Discussion

The value of accurate parentage verification for effective selection and breeding has been demonstrated in several studies (Banos et al. 2001; Visscher et al. 2002; Weller et al. 2004). In developing countries such as SA, routine genotyping is not feasible for all cattle breeds and the adoption of genomic evaluations remains limited. Lower density SNP panels hold the potential for exploiting genomic information at a lower cost. A measure of the applicability or informativeness of a SNP is the extent to which the SNP is segregating in the population within which it will be used. The International Society of Animal Genetics (ISAG) took the initiative to propose a 200-SNP panel that could be used internationally for parentage verification. The present study is the first study to evaluate the effectiveness of these SNPs in SA Bonsmara and Drakensberger beef cattle. In the development of the ISAG panel, SNPs specific to African *Bos taurus* (Sanga) cattle breeds were not included and, thus, this evaluation is crucial prior to recommending the use of this panel in these breeds.

While not all Sanga breeds were included in the present study, a representative sample of a composite and an indigenous cattle of the BON and DRB originated from 22 and 10 herds, respectively, were included to evaluate the efficiency of the ISAG panel. Due to variation in the number of SNP genotypes created during the genotyping process, genotypes on only 185 of the 200 ISAG markers panel were available. The call rate, minor allele frequency, and extent of Hardy-Weinberg equilibrium deviations have been the most commonly used measures to define the integrity of the genotype (Chan et al. 2008). In the pursuit of improving the genotype quality for parentage tests, others (McClure et al. 2015; Berry et al. 2019) have also considered the SNP GenTrain score and the number of discordant SNPs between validated parent-offspring.

The value of a SNP for parentage verification highly depends on the MAF within a population of animals that are being compared. Since only homozygous genotypes in both the parent and offspring are informative, a SNP with low MAF has a limited value for parentage verification (Van Doormaal et al. 2016). A total of 62.2% and 73.0% of the ISAG markers in BON and DRB cattle, respectively, were segregating strongly (i.e., MAF > 0.3) with a mean minor allele frequency of the ISAG SNP panel of 0.331 and 0.359 in the BON and DRB, respectively. This was supported with heterozygosity and polymorphic information content values that were close to 0.5. Of the six monomorphic SNPs identified in the present study, two i.e. ARS-USMARC-Parent-DQ786764 (rs109943112) and ARS-USMARC-Parent-EF034087 (rs110665639) were already reported by McClure *et al.* (2015) to be monomorphic in an Irish multi-breed cattle population. The GenTrain score calculated from the GenomeStudio for each SNP provides a measure that takes into account the quality and shape of the genotype clusters and the relative distance from one another. Both the ARS-BFGL-NGS-11383 (rs110577061) and ARS-USMARC-Parent-DQ786762 (rs29010772) SNPs exhibited poor genotype clustering with a GenTrain score of 0.50 and 0.52, respectively, as well as suffering from low call rates (i.e. <0.93). A total of 13 SNPs had at least one discordant genotype between the validated sire-offspring pairs. SNPs exhibiting some parent-offspring discordances, as well as those with low call rates or poor GenTrain score < 0.55 are often discarded prior to the downstream genomic analyses (Zhao et al. 2015; Judge et al. 2016; Berry et al. 2019). Some SNPs also deviated from HWE indicating either the occurrence of genotype errors or the loss of heterozygotes (Chan et al. 2008). There was no difference in the power of probability of parentage exclusion (P_E) (99.46%) between the two breeds. The P_E value observed in this study corresponds with that reported for Hanwoo and Wagyu cattle in Australia (0.99) using 195 and 199 ISAG SNP markers, respectively (Strucken et al. 2014). The probability of identity was, however, lower in the DRB (1.11×10^{-54}) compared to the BON (1.61×10^{-48}) and suggests that the ISAG SNP panel differed in applicability between the two SA breeds. The values observed in the 185 ISAG SNPs were low compared to those reported in purebred American Angus (2.0×10^{-13}) and multi-breed composite populations (1.9×10^{-10}) with only 32 SNPs (Heaton et al. 2002), in Black Japanese cattle (2.7×10^{-34}) based on 87 SNPs (Hara et al. 2010) and 10^{-44} reported in purebred Angus population in

Argentina with 116 SNPs (Fernández et al. 2013). Given the difference in the number of SNPs tested, a lower probability of identity value would be expected in the present study.

DNA-based parentage verification is used to improve the quality of pedigrees and, consequently, of genetic evaluations. Parent-offspring mismatches between the genotypes of the animal and putative sires provide means to detect errors in pedigree records. Pedigree errors can occur for several reasons including incorrect recording on-farm, multi-sire breeding, mistakes in genotyping laboratories, or even from the genotype format (Weller et al. 2004; McClure et al. 2018). The parentage errors of 10% and 17.2% reported in the present study based on 119 375 SNP genotypes for the BON and DRB, respectively, does indicate issues with recording at the farms.

It is, therefore, imperative that the markers used to assign parentage should be consistently accurate, in order to minimize pedigree errors. The 13 SNPs with some discordances between validated sire-offspring pairs within the ISAG SNP markers suggests that they may in fact not be informative, and could actually be counter-informative within a parentage panel. The false-negative parentages, and failure to exclude parentages that were already validated with the 119 375 SNP genotypes, demonstrated that the use of the ISAG panel in Sanga breeds may lead to parentage errors. This may also mean that, at times, it might not be possible to conclude parentage when closely related candidate sires are linked. These results show the need for the selection of additional informative SNP markers for parentage testing in Sanga breeds.

Acknowledgments

The authors would like to acknowledge funding given by the Technology Innovation Agency (TIA) for the Beef Genomics Project (BGP). We wish to thank the South African Bonsmara and Drakensberger Cattle Breeders Societies for providing access to their genotype and pedigree databases available at the Agricultural Research Council (ARC) and South African Studbook. Authors express their gratitude to projects funded by the Red Meat Research Development (RMRD) under the University of Pretoria (UP) that donated the genotypes. This research project was also financially supported by the ARC, National Research Foundation (NRF), and UP.

Conflict of interest: The authors declare no conflict of interest.

References

Banos, G., Wiggans, G.R. and Powell, R.L., 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. Journal of Dairy Science, 84, 2523-2529.

Berry, D.P., Wall, E. and Pryce, J., 2014. Genetics and genomics of reproductive performance in dairy and beef cattle. Animal, 8(S1), 105-121.

Berry, D.P., McHugh, N., Wall, E., McDermott, K. and O'Brien, A.C., 2019. Low-density genotype panel for both parentage verification and discovery in a multi-breed sheep population. Ireland Journal of Agricultural and Food Research, 58, 1-12.

Buchanan, J.W., Woronuk, G.N., Marquess, F.L., Lang, K., James, S.T., Deobald, H., Welly, B.T. and Van Eenennaam, A.L., 2016. Analysis of validated and population specific SNP parentage panels in pedigreed and commercial beef cattle populations. Canadian Journal of Animal Science, 97, 231-240.

Chan, E.K.F., Hawken, R. and Reverter, A., 2008. The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. Animal Genetics, 40, 149-156.

Fernández, M.E., Goszczynski, D.E., Lirón, J.P., Villegas-Castagnasso, E.E., Carino, M.H., Ripoli, M.V., Rogberg-Muñoz, A., Posik, M.D., Peral-García, P. and Giovambattista, G., 2013. Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Journal of Genetics and Molecular Biology, 36, 185-191.

Fisher, P.J., Malthus, B., Walker, M.C., Corbett, G. and Spelman, R.J., 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. Journal of Dairy Science, 92, 369-374.

García-Ruiz, A., Wiggans, G.R. and Ruiz-López, F.J., 2019. Pedigree verification and parentage assignment using genomic information in the Mexican Holstein population. Journal of Dairy Science, 102, 1806-1810.

Georges, M., Charlier, C. and Hayes, B., 2018. Harnessing genomic information for livestock improvement. Nature Reviews Genetics, 20, 135-156.

Hara, K., Watabe, H., Sasazaki, S., Mukai, F. and Mannen, H., 2010. Development of SNP markers for individual identification and parentage test in a Japanese black cattle population. Journal of Animal Science, 81, 152-157.

Hayes, B.J., 2011. Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. Journal of Dairy Science, 94, 2114-2117

Heaton, M.P., Harhay, G.P., Bennett, G.L., Stone, R.T., Grosse, W.M., Casas, E., Keele, J.W., Smith, T.P.L., Chitko-McKown, G.C. and Laegreid, W.W., 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. Mammalian Genome, 13, 272-281

Henderson, C.R., 1973. Sire evaluation and genetic trends. Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush. Champaign, IL., American Society of Animal Science and American Dairy Science Association, 10-41.

International Society for Animal Genetics (ISAG)., 2013. ISAG cattle core and additional SNP panel.www.isag.us/ Docs/Cattle-SNP-ISAG-core-additional-panel-2013.xlsx

Israel, C. and Weller, J.I., 2000. Effect of misidentification on genetic gain and estimation of breeding values in dairy cattle populations. Journal of Dairy Science, 83, 181-187.

Jamieson, A. and Taylor C.S., 1997. Comparison of three probability formulae for parentage exclusion. Animal Genetics, 28, 397-400.

Judge, M.M., Kearney, J.F., McClure, M.C., Sleator, R.D. and Berry, D.P., 2016. Evaluation of developed low-density genotype panels for imputation to higher density in independent dairy and beef cattle populations. Journal of Animal Science, 94 (3), 949-962.

Junqueira, V.S., Cardoso, F.F., Oliveira, M.M., Sollero, B.P., Silva, F.F. and Lopes, P.S., 2017. Use of molecular markers to improve relationship information in the genetic evaluation of beef cattle tick resistance under pedigree-based models. Journal of Animal Breeding and Genetics, 133, 1-13.

Kios, D., van Marle-Köster, E. and Visser, C., 2012. Application of DNA markers in parentage verification of Boran cattle in Kenya. Tropical Animal Health Production, 44, 471-476

Lyons, R.E., Buttsworth, S., Waine, D. and Kelly, M., 2015. SNP-based parentage in an Australian cattle industry context: Does one size fit all? Proc. Association for Advancement of Animal Breeding and Genetics, 21, 249-252.

McClure, M.C.., McCarthy, J., Flynn, P., Weld, R., Keanne, M., O'Connell, K., Mullen, M.P., Waters, S. and Kearney, J.F., 2015. SNP selection for nationwide parentage verification in beef and dairy cattle. International Committee of Animal Recording (ICAR) Technical series. No 19. www.icar.org/wp-content/uploads/2015/11/ICARTechnical-Series-19-Krakow-2015-Proceedings.pdf. (Accessed 20 January 2020)

McClure, M.C., McCarthy, J., Flynn, P., McClure, J. C., Dair, E., O'Connell, D. K. and Kearney, J. F., 2018. SNP data quality control in a national beef and dairy Cattle system and highly accurate SNP based parentage verification and identification. Frontiers in Genetics, 9, 84.

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

R Core Team., 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Sanders K., Bennewitz J. and Kalm E., 2006. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. Journal of Dairy Science, 89, 315-321.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H. and Gondro, C., 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, 45, 572-575.

Teneva, A., Todorovska, E., Petrović, M.P., Kusza, S., Perriassamy, K., Petrović, V.C., Andrić, D.O. and Gadjev, D., 2018. Short tandem repeats (STR) in cattle genomics and breeding. Biotechnology in Animal Husbandry, 34, 127-147.

Van Doormaal, B.J., Kistemaker, G.J. and Johnston, J., 2016. Parentage Analysis Services for Dairy Cattle in Canada. Interbull Bulletin no. 50. Puerto Varas, Chile, October 24-28. www.icar.org/Documents/DNA-WG/Docs/2017-02-20%20Meeting%20Package.pdf. (Accessed 12 October 2019)

Van Eenennaam, A.L., 2016. Use of genetic marker information in beef cattle selection. Applied Reproductive Strategies in Beef Cattle-Des Moines, Iowa, 7-8 September.

Visscher, P.M., Woolliams, J.A., Smith, D. and Williams, J.L., 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. Journal of Dairy Science, 85, 2368-2375.

Waits, LP., Luikart, G. and Taberlet, P., 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology, 10 (1), 249-256.

Weller, J.I., Feldmesser, E., Golik, M., Tager-Cohen, I., Domochovsky, R., Alus, O. and Ron M., 2004. Factors affecting incorrect paternity assignment in the Israeli Holstein population. Journal of Dairy Science, 87 (8), 2627-2640.

Zhao, F., McParland, S., Kearney, F., Du, L. and Berry, D.P. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genetics Selection Evolution, 47, 49.

CHAPTER 4

Parentage validation and pedigree reconstruction of the Bonsmara and Drakensberger cattle using genomic information

Abstract

Genomic data on individual animals is now becoming routinely available in South Africa. The objective of the present study was to validate recorded ancestry of Bonsmara (BON) and Drakensberger (DRB) animals but also discover unknown genomic relationships using the single nucleotide polymorphism (SNP) markers. Genomic relationships were validated based on the count of autosomal Mendelian conflicts between the parent-progeny pairs, which were sequentially used to detect possible regions with hemizygous deletion (hDEL) in the genome. A high frequency of Mendelian conflicts was observed in chromosomes 1 and 6 in the BON and DRB, respectively. On average, 8.5% and 10.1% of parentage errors were observed in the pedigrees of the BON and DRB, respectively. Errors detected in dam-progeny pairs varied from 7.0% (DRB) to 7.6% (BON), whereas for the sire-progeny pairs, error rates were relatively higher and ranged from 10.0% (BON) to 17.2% (DRB). SNP-based parent-progeny discovery was possible for 69 relationships which were not previously recorded in the parentage information. This included 19 and 20 sire-offspring pairs; 12 and 18 dam-offspring pairs as well as 4 and 9 half-siblings of the BON and DRB. The suspected regions to contain deletions based on the approach of using the SNP-chip data and Mendelian mismatches between the parent-progeny pairs were detected from 11 regions of 10 chromosomes (3, 6, 7, 9, 10, 11, 17, 24, 25 and 28). These suggestive regions were traced from the most used breeding animals, which were 7 (BON) and 5 (DRB) bulls, 2 (BON) and 7 (DRB) cows. This study demonstrated how SNP markers can be utilized to confirm and determine parentage; they can also be used to identify locations in carriers' genomes that are linked to hDEL. It is advised to use such methods to genotype database accumulations to enhance genetic and genomic analyses. Although the SNP density of 91 185 SNPs used in the present study provided a high precision in performing these analyses, a much lower density panel is desirable in practice to perform parentages. On the other hand, a large familial whole genome SNP-chip based data is necessary to identify the regions of homozygosity. Due to the differences of the required properties SNPs, these analyses may be undertaken independently.

Keywords: Genomic evaluations, genealogy, hemizygosity, pedigree discovery, half-siblings

Introduction

Genetic improvement in livestock begins with establishing selection goals. Genetic gain is realized by recording pedigree and performance data and applying genetic evaluation models to select the next generation's parents (Dürr *et al.*, 2011). Accurate genetic evaluations rely on precise estimates of genetic parameters to predict breeding values, which in turn are directly influenced by the precision of relationships represented in the pedigree (Hayes & Goddard, 2008); accurate genetic evaluations also require accurate performance and pedigree information (as well as accurate recoding of the contributing systematic environmental effects). A number of factors impact the integrity of pedigree records under extensive livestock management systems where natural mating and multiple-sire breeding practices are used, making it relatively easy for parentage errors to occur (Van Eenennaam & Drake, 2012; Buchanan *et al.*, 2016). Such errors are troublesome when evaluations are computed based on genetic relationships recorded on-farm without genomic information to validate pedigree records.

There are two forms of parentage errors influencing genetic parameters and genetic evaluations and, thus, the prediction of breeding values: incorrect parentage information (i.e., wrong parentage) and missing parentage information (i.e., unknown parentage) (Harder *et al.*, 2005). There is relatively little information reported on the latter form of errors. Previous studies have investigated the amount and the consequences of wrong paternity information (Baron *et al.*, 2002; Weller *et al.*, 2004; Sanders *et al.*, 2006), but relatively few studies have reported on wrong dam information (Purfield *et al.*, 2016; Junqueira *et al.*, 2017). Generally, only sire information is perceived as unknown in practice, which is not always the case. In commercial herds, the fraction of incorrect maternity and missing parentage information can be substantial (Harder *et al.*, 2005; Sanders *et al.*, 2006).

Single nucleotide polymorphisms (SNPs) genotyping provides an unprecedented ability to correct and reconstruct pedigrees (Hayes & Goddard, 2008; Wiggans *et al.*, 2009; Habier *et al.*, 2010). Recent studies have reported improved estimates of breeding values, genetic gain, and selection accuracy when genomic data was used to correct parentage records (García-Ruiz *et al.*, 2019; Nwogwugwu *et al.*, 2020). However, the high cost of genotyping using the high-density panels results in not all breeding animals being genotyped, suggesting that not all pedigree records can be fully verified. Thus, pedigree reconstruction captures relationships not detailed in the pedigree by calculating the extent of relationships among the genotyped animals based on the count of Mendelian conflicts and allowing the detection of full- or half-siblings (Ferdosi *et al.*, 2014; Huisman, 2017). Considering more relationships within the population has been found to influence the estimation of genetic evaluations, as females are not frequently genotyped, but the maternal grandsires (VanRaden *et al.*, 2013; Moore *et al.*, 2019). This makes genomic information a better fit for the management strategy that uses multiple-sire mating groups, particularly beef production.

Several studies have documented the effective use of opposing homozygotes to detect mismatches between SNP genotypes of parent-offspring pairs in cattle (Hayes, 2011; Strucken *et al.*, 2014; Junquiera *et al.*, 2017; García-Ruiz *et al.*, 2019; Sanarana *et al.*, 2021). Prior to its use in cattle, opposing homozygotes were used to identify hemizygous deletion (hDEL) in the human genome (Conrad *et al.*, 2006; Kohler & Cutler, 2007) and, more recently, in cattle (Nandolo *et al.*, 2018; Himmelbauer *et al.*, 2019). Hemizygous deletion is when an individual possesses only a single copy of an allele at a given locus instead of two copies (Amos *et al.*, 2003). The hDEL loci mimic homozygous genotypes in the called genotype output file resulting from the mistyping in GenTrain algorithm (Nandolo *et al.*, 2018). Thus, these cause Mendelian mismatches where there are actually none (Winchester *et al.*, 2009; Daetwyler *et al.*, 2014; Rafter *et al.*, 2018) and potentially lead to doubtful or inaccurate parentage outcomes. These deletions are also known to hold important information on factors affecting production traits, genetic defects and diseases (Kadri *et al.*, 2014; McDaneld *et al.*, 2014).

Over the past two decades, microsatellite markers have been used in South Africa to improve pedigree records (Van Marle-Köster & Visser, 2018). The recent implementation of the South African Beef Genomics Program (BGP) in 2015, followed by the availability of genotypes in 2016, provided an opportunity to demonstrate the effective use of dense single nucleotide polymorphisms for correcting genetic relationships recorded in pedigrees along with the detection of hemizygous regions. Several beef cattle breeds have been genotyped with the GeneSeek Genomic Profiler (GGP) Bovine 150K to facilitate genomic evaluations in South Africa. The available genomic data contains useful data to validate the accuracy of pedigrees, pedigree reconstruction, and diagnosis of hemizygous deletions. Therefore, the objectives of this study were to (1) quantify the accuracy of pedigree information, (2) illustrate the use of Mendelian inconsistencies to detect hemizygous deletions and (3) identify individuals that might be the carriers in the Bonsmara and Drakensberger cattle populations using the genomic data.

Materials and Methods

Ethical approval for the external data use was granted by the Animal Ethics Committee (AEC) of the University of Pretoria (EC066-16 AEC). The Bonsmara and Drakensberger cattle Breeders' Societies provided consent to use genotypes generated within the Beef Genomics Project (BGP). Pedigree records of the breeds were obtained from the South African Studbook and Animal Improvement Association.

Origin of pedigree and genotypes

The data used in this study originated from approximately 22 BON and 10 DRB herds with pedigree records that included 26 281 and 6 178 animals and 1 632 and 1 203 genotypes, respectively. All genotypes were from the Illumina 150K Bovine BeadChip panel comprised of 138 888 single nucleotide polymorphisms (SNPs). The pedigrees were investigated to identify the number of genotyped sire-offspring and dam-offspring pairs. Based on the recorded pedigree information,

genotypes were available from three (DRB) to five (BON) generations, born between 1977 through 2017. The number of progeny per sire ranged from 1 to 10 and 1 to 26 for the BON and DRB, respectively. The number of progeny per dam ranged from 1 to 2 and 1 to 4 for the BON and DRB, respectively. Approximately 88% (BON) and 69% (DRB) of the animals with genotypes had unknown parents. The number of putative sire-offspring pairs for the BON and DRB was 50 and 93, while 79 and 205 were dam-offspring pairs. Of these sire-offspring and dam-offspring pairs, only 16 and 39 genotyped trios were available for the BON and DRB.

Genotype data quality control

Data editing and quality control were performed using custom R software scripts (R Development Core Team, 2021). The data were quality controlled by eliminating SNPs from both the sex chromosome and mitochondrial DNA along with SNPs with no recorded position. All SNPs with positions based on the UMD3.1 genome build remained for the downstream analyses. Additional edits applied to the data included the removal of genotypes with median GenCall (GC), GenTrain (GT) scores lower than 0.60 and 0.55, respectively, and a sample-wise call rate less than 90%. All SNPs that departed from Hardy–Weinberg equilibrium ($p < 0.001$), had a missing genotype rate of more than 5% and where the minor allele frequency (MAF) was <0.05 across the populations were discarded. After all the edits, 2 585 animals (BON = 1 563 and DRB = 1 022) with 92 835 autosomal SNPs remained for subsequent analyses.

Step1: Identification of Mendelian inconsistencies

The accuracy of the relationships recorded in the pedigree was verified based on Mendelian conflicts by counting the number of opposing homozygous SNP genotypes between the putative parent and progeny (Wiggans *et al.*, 2009) using the following method:

$$\%Mendelian\ conflicts = \frac{\#Conflict_{HomSNPs}}{\#Tested_{HomSNPs}} * 100$$

where $\#Conflict_{HomSNPS}$ is the number of opposite homozygous SNP genotypes between the parent and an offspring, and $\#Tested_{HomSNPs}$ is the number of SNP tested where both the animal and potential parent have homozygous genotypes. The exclusion method did not consider heterozygous (AB) or uncalled (--) genotypes in either pair. The realized distribution of opposing homozygotes defined the threshold for declaring correct or incorrect parentage. Parentage was excluded when the rate of conflicts exceeded 1% genotype error tolerance (Calus *et al.*, 2011). Results were classified as confirmed, conflicted and unmatched. After checking the Mendelian conflicts, the genotyped parents and offspring pairs that retained the same sire or dam were confirmed. The conflicts referred to animals where the reported parent was incorrect but no alternative genotyped sire or dam could be detected. The unmatched

included all the animals without a genotyped parent or progeny. When comparing candidate parents and offspring, the recorded dates of birth of all animals were used, so the candidate parents had to be at least 10 months older than the possible offspring.

Step 2: Pedigree validation and reconstruction

Since most of the genotyped individuals their parents were either unknown, not recorded on pedigree records or did not match with any of the potential parents, an approach of reconstructing the pedigree to find possible new relationships was performed. For this step, 1 563 BON and 1 022 DRB genotypes were subjected to the *Hsphase* package (Ferdosi *et al*., 2014) in R software, which was used to construct the kinship genomic relationship matrix. A shared phantom parent was allocated to each detected family of genotyped individuals which had no sires recorded in the original pedigree. As a result, new genetic half-sibling groups were created. The pedigree reconstruction, errors and new assignments were visualized on a heatmap per breed. New relationships were summarized and reported for animals with parents, grandparents, trios, and half-siblings.

Step 3: Detection of hemizygous regions based on Mendelian conflicts of parent-offspring

Hemizygous regions were detected following the method described by Himmelbauer *et al.* (2019) from the verified parent-progeny pairs. The data set of parent-progeny pairs with 119 375 SNPs was used to detect hemizygous deletions. This data was not subjected to quality control (QC) to avoid SNPs with deletion that may be removed during QC. The number of SNPs with a high error rate was calculated using a distribution of Mendelian conflicts on the confirmed parent-progeny pairs. If the mismatches were not attributable to random genotyping errors (1%), the error's inheritance was tracked down through generations, from the parent to its progeny. Thus, the possible regions were presumed where a parent mismatched with more than one of its progeny, grandchildren and great-grandchildren at a certain locus and position.

Results

The distribution of the number of SNP markers with Mendelian conflicts from putative parent-offspring pairs examined with 92 835 SNPs is presented in Figure 4.1. A total of 18 827 SNPs had conflicts between the verified Bonsmara (BON) parent-offspring pairs, while 22 367 were observed in the verified Drakensberger (DRB) parent-offspring pairs. The greatest number of SNPs with Mendelian conflicts were observed on chromosomes 1 (BON) and 6 (DRB). SNPs with the high number of conflicts observed on these chromosomes included SNP ARS-BFGL-NGS-45078 (BON) and SNP BovineHD0600011074 (DRB). These SNP conflicts were detected on verified 17 and 89 parent-progeny pairs in BON and DRB, respectively. In general, more conflicts per chromosome were detected in DRB than in BON.

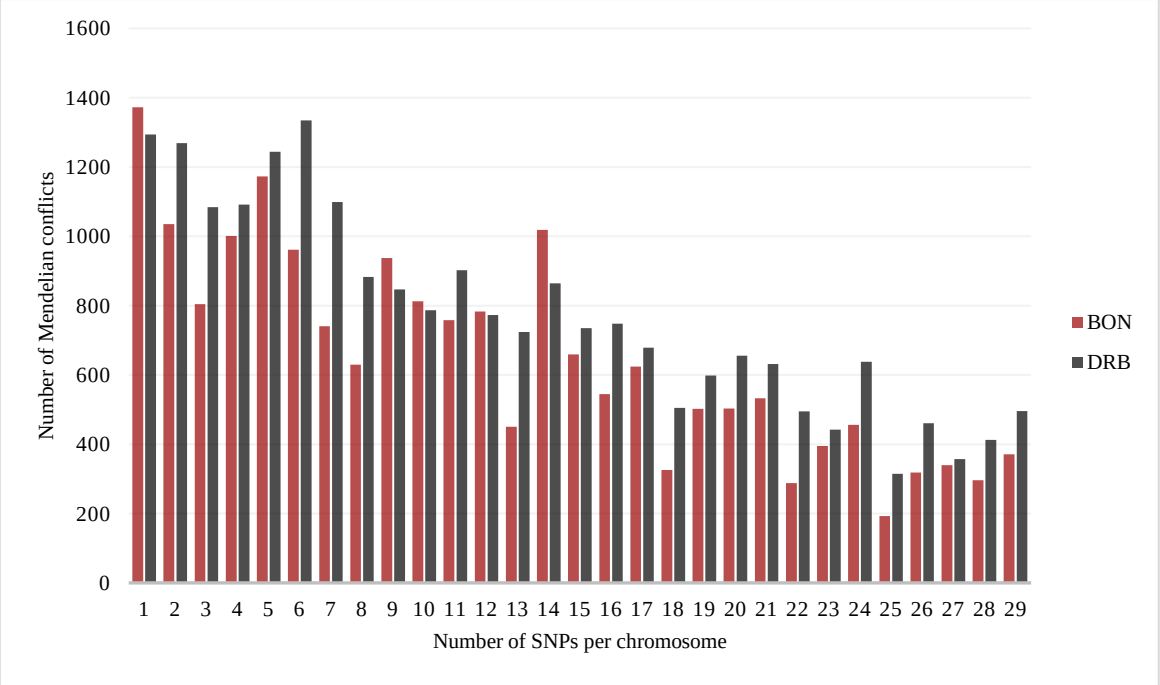


Figure 4.1. Distribution of Mendelian conflicts based on 92 835 SNPs from the parent-progeny pairs of the Bonsmara (BON) and Drakensberger (DRB) cattle breeds

For the confirmed and mismatching parent–progeny pairs, the distribution of the number of SNP markers with Mendelian conflicts were significantly different (Figure 4.2). The average number of conflicting genotypes was 156,73±67,7 for 395 matching pairings, and 4549,33±2241,4 for 42 mismatching pairs. The minimum number of Mendelian conflicts for the assigned parent-progeny pairs varied between 34 and 622, corresponding to 0.03% and 0.68% (less than 1% threshold allowed for genotyping errors) of the total number of SNPs. The SNP conflicts of mismatching pairs ranged from the minimum of 937 to the maximum number of 9 173. This was equivalent to 1.03% and 10.1% (greater than the 1% threshold allowed for genotyping errors) of the total number of SNPs.

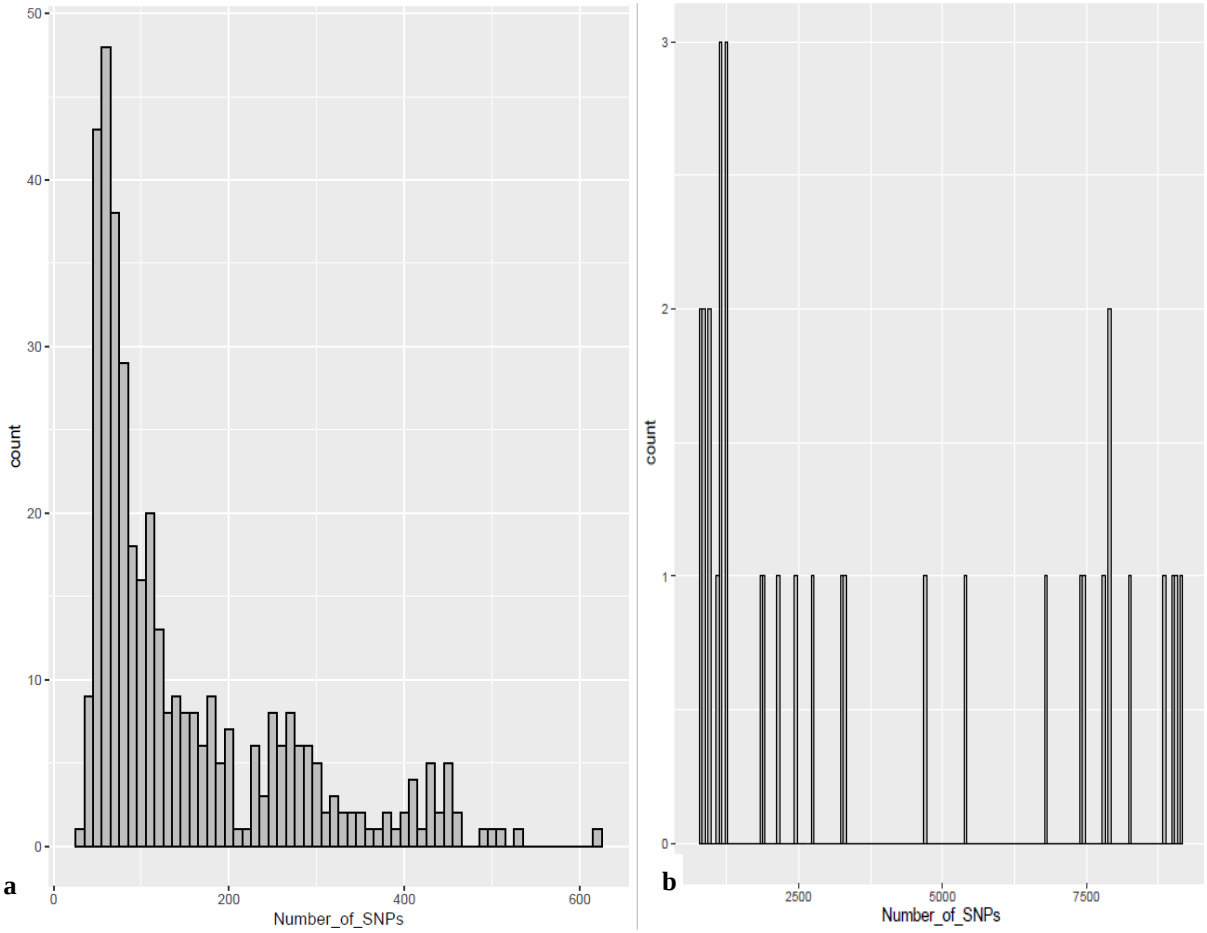


Figure 4.2. Distribution of the number of markers with Mendelian conflicts for parent-progeny pairs from the original pedigree with (a) confirmed and (b) mismatching relationships based on SNP markers.

The parentage errors and new relationships of the BON and DRB cattle were calculated and are presented in Table 4.1. Of the 437 parent-progeny pairs tested, 395 were confirmed, while 42 were conflicts. Parentage error rates among parent-progeny pairs were 8.5% (11/129) and 10.1% (31/308) in the BON and DRB, respectively. Paternal errors were higher (10.0% and 17.2% for BON and DRB, respectively) than maternal errors (7.6% and 7.0% for BON and DRB respectively).

Table 4.1. Number of parent-offspring pairs from the pedigree verified within the available genotypes of the Bonsmara (BON) and Drakensberger (DRB) cattle

Breed	Relationship	*N* pairs[1]	*N* conflicted[2]	*N* confirmed[3]	*N* discovered[4]
BON	Dam-offspring	79	6 (7.60%)	73 (92.4%)	12
	Sire-offspring	50	5 (10.0%)	45 (90.0%)	19
	Trios	16	3 (18.7%)	13 (81.3%)	2
	Half-siblings	0	0 (0%)	2 (100%)	2
	Grandparents	3	0 (0%)	3 (100%)	4
DRB	Dam-offspring	215	15 (7.0%)	200 (90.0%)	20
	Sire-offspring	93	16 (17.2%)	77 (82.8%)	18
	Trios	39	4 (10.3%)	35 (89.7.0%)	1
	Half-siblings	8	2 (25.0%)	6 (75.0%)	3
	Grandparents	6	1 (16.7%)	5 (83.3%)	3

[1]Number of parent-offspring pairs that had been available based on the pedigree file, [2]Number of animals with incorrect parents [3]Number of genotyped parent–progeny pairs maintained the same sire or dam, [4]Number of individuals that had no sire or dam in the pedigree file and a compatible matching sire or dam was identified

Following the reconstruction of pedigrees based on the genomic relationship matrix method, 69 parent-progeny pairs were discovered, leading to new relationships being confirmed as grandparents (4) and trios (3). Of the 69 parent-progeny relationships discovered, 40 (~58%) were sire-offspring pairs, while 29 (42%) were dam-offspring pairs. This reduced the number of animals with unknown parents from 88% to 84% in the BON and 69% to 61% in the DRB. A total of 13 half-sibling family groups are shown in Figure 4.3 were generated, which were four in the BON and nine in the DRB, represented by the dark square blocks in the diagonals of Figure 4.3.

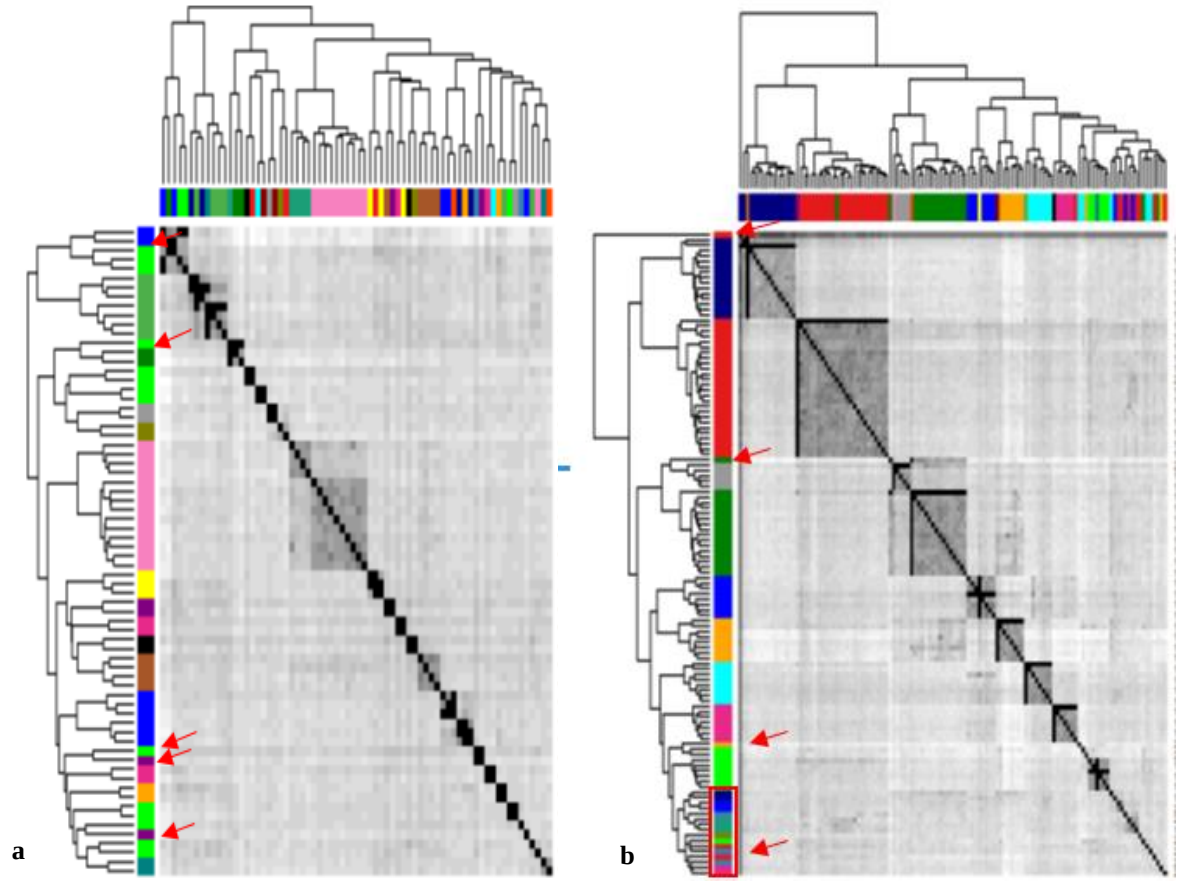


Figure 4.3. Kinship matrix shows the true and incorrect relationships recorded in pedigree records determined based on the count of opposing homozygotes for the (a) Bonsmara (BON) and (b) Drakensberger (DRB). The original pedigree relationships with errors are displayed on the vertical bars of the plots (pointed with arrows). The horizontal bars (top) show the inferred pedigree and correct pedigree, with animals correctly assigned to sires.

On ten different chromosomes (3, 6, 7, 9, 10, 11, 17, 24, 25 and 28), Mendelian conflicts between the validated parent-progeny pairs regularly occurred at eleven regions that were thought to be the origin of hemizygous deletions. Due to the close familial structure of the DRB data, the pattern of mismatches over the generations was noticeable compared to the BON data. These were likely identified on animals mostly used for breeding, including 7 and 5 bulls; 2 and 7 cows of the BON and DRB, respectively. Pedigree charts showing selected lineages that had Mendelian mismatches at SNP BovineHD0600011074 (DRB) and SNP ARS-BFGL-NGS-45078 (BON) are presented in Figure 4.4.

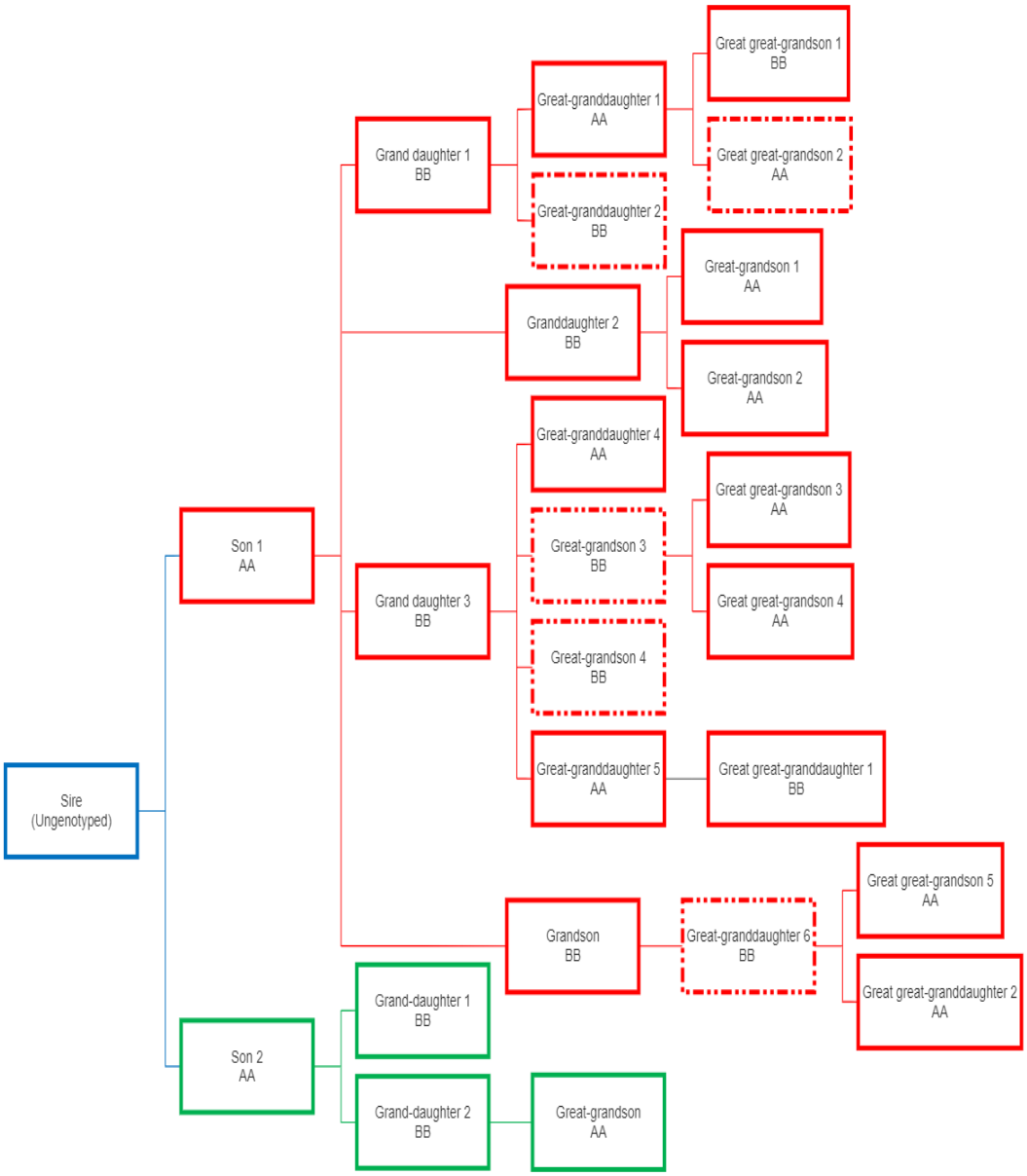


Figure 4.4. (a) Pedigree chart showing the familial structure of Drakensberger (DRB) on chromosome 6, SNP BovineHD0600011074, and position 40638197. Boxes made with continuous lines are for animals that had mismatches with their parent(s), and those with broken lines are for animals without mismatches

Son 1 and son 2 were identified as half-siblings from the father marked in the pedigree chart as the sire. The sire was ungenotyped and is listed in the pedigree as the father of the two sons (1 and 2). The relationship between son 1 and son 2 was confirmed based on the genomic relationship matrix and the comparison of Mendelian mismatch patterns analyses. Son 1 had mismatches with three of its daughters (great-granddaughters to the sire), while son 2 had mismatches with its son (great-grandson to the sire) and two daughters (great-granddaughters to the sire), as shown in Figure 4.4a. Consequently, the same

mismatches were observed on their six out of seven great-grandchildren (great-great-grandchildren to the sire) in the same SNP and position shown in the figure.

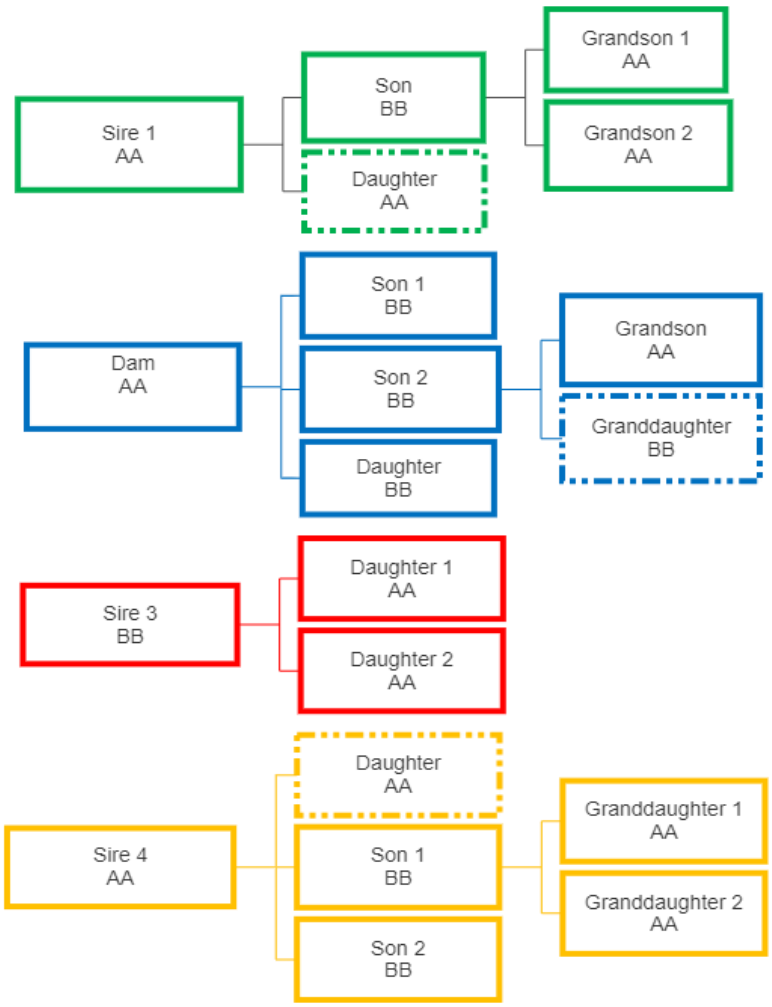


Figure 4.4 (b) Pedigree chart showing the familial structure of Bonsmara (BON) on chromosome 1, SNP ARS-BFGL-NGS-45078, position 112584531. Boxes made with continuous lines are for animals that had mismatches with their parent(s), and those with broken lines are for animals without mismatches

Mendelian mismatches on chromosome 1 of the BON parent-offspring pairs involved three sires (1, 3 and 4) and one dam. Unlike in the DRB, where mismatches consisted of a sire with a long lineage, the pattern of mismatches observed in the BON was traced from the parents to the offspring and limited to grandchildren. None of the parents of the sires and the dam were genotyped. Based on the genomic relationship matrix, sire 1 and the dam were found to be maternal half-siblings. In addition, sire 1 and the dam also had more mismatches on the same chromosome 1, which were identified from three SNPs BovineHD0100040682, BovineHD0100041396 and BovineHD0100043727 located in positions

141961081; 143811841 and 150741812, respectively. The pattern of mismatches between parents and progeny to grandchildren were observed.

Discussion

Correct parentage information is essential in selection programs to distinguish genetically elite from inferior individuals. Parentage mistakes may occur for various reasons, including human (recording) error, technical errors during sampling and recording from the farms, genotype format, or extensive breeding systems where multiple-sire natural mating is practised (Weller *et al.*, 2004; McClure *et al.*, 2018). In some circumstances, pedigree information may be unattainable, incomplete or contain errors (Holroyd *et al.*, 2002; Van Eenennaam *et al.*, 2007; Kios *et al.*, 2012), which may have a large impact on genetic evaluations, thereby impacting genetic progress (Baron *et al.*, 2002; Sanders *et al.*, 2006; Parlato & Van Vleck, 2012). Molecular-based information is, therefore, crucial in this regard, provided that the markers are sufficiently dense to verify the on-farm pedigree records and to assign parents retrospectively (Hayes & Goddard, 2008). Parentage exclusion was determined based on the count of Mendelian inconsistencies between the parent-progeny pairs, which were also used to discover animals exhibiting hemizygous deletions in the genomes. The present study is the first to perform such analyses using SNP markers in South African cattle populations.

The validation of pedigrees based on the identification of Mendelian conflicts was performed using 91 185 SNPs in two SA local beef breeds. In general, the number of opposing homozygotes is expected to be more when an animal is compared with a parent that is not its sire or dam than its true parents (Hayes, 2011). In the current study, the verified relationships had low conflict rates (between 0.03% and 0.68%), far below the 1% conflict tolerance threshold. There was a remarkable gap between the maximum number of SNPs with conflicts for the confirmed parent-progeny pairs (622) and the minimum number of SNPs (937) for the parent-progeny pairs that had conflicts. Other studies have also observed an obvious difference between the maximum number of SNPs with conflicts for the confirmed pairs and the number of SNPs with conflicts for mismatching pairs. For example, Wiggans *et al*. (2009) used 37 811 SNPs in the Holstein population and reported a mean value of 2.3 SNPs (range: 0–89) for the true relationships and 2 411 SNPs (range: 754–3507) for the relationships that had mismatched. Based on 41 045 SNPs, Junqueira *et al*. (2017) observed a mean value of 17.73 (range: 0–75) for the matching relationships and 2 704.13 (range: 595–4993) for the mismatching relationships in Hereford and Braford cattle populations.

Identifying mismatches between parent and progeny SNP genotypes provides a powerful tool for detecting pedigree errors (Wiggans *et al.*, 2009; Calus *et al.*, 2011). A mismatch rate of up to 5% has been considered to have minimal effect on estimated breeding values; however, the effect increases with increasing error rates (Van Vleck, 1970). Based on this information, the average pedigree error rates detected in BON (8.5%) and DRB (10.1%) should affect estimated breeding values. Maternal

conflicts were 7.0% (DRB) and 7.6% (BON), while paternal conflicts were higher at 10.0% (BON) and 17.2% (DRB). Dam errors in commercial herds could be substantial, with negative and additive effects on the rate of genetic gain (Harder *et al.*, 2005). However, the present study's dam-progeny errors were lower (14.3%) than in Brazilian beef populations (Junqueira *et al.*, 2017). Paternity errors reported in this study were similar to 10% reported in the United Kingdom (Visscher *et al.*, 2002), 11% reported in the United States (Banos *et al.*, 2001), and 12% reported in Israel (Weller *et al.*, 2004) but lower than 27.2% reported in Brazilian populations (Junqueira *et al.*, 2017). Most of the identified parent-offspring pairs were related to sire information (58%) compared to the dams (42%), suggesting some recording issues in the farms due to the practice of multiple sire breeding and the switching of calves at birth. Nonetheless, in genetic evaluations there are fewer missing pedigree records than pedigree records with wrong parentage assignment (Sander *et al.*, 2006). Identifying new half-siblings has been reported to increase additive genetic variance and heritability estimates (Visscher *et al.*, 2008; Junqueira *et al.*, 2017).

High-quality genetic relationships are known to increase the accuracy of breeding value predictions of selection candidates. The high number of animals, 61% (DRB) and 84% (BON), that did not have genotyped parents can either be associated with the current limited genomic testing, a lack of biological samples for older (non-existing) animals, or the short period of establishment of the genomic selection program. The percentage of animals without genotyped parents was in the same range (79%) as reported in Mexican Holstein cattle (García-Ruiz *et al.*, 2019).

Several Mendelian conflicts identified in SNP positions from this study were traced from two or more progeny per parent. Assumptions that the high number of mismatches occurred in SNPs, ARS-BFGL-NGS-45078 and BovineHD0600011074 of the previously validated parent-progeny pairs could have resulted from either mutations or genotyping errors are unreliable. According to Himmelbauer *et al.* (2019), the likelihood of errors accruing in the progeny due to genotyping errors occurring by chance is unlikely. Other studies have defined possible deletions as the regions with at least two Mendelian mismatches (Himmelbauer *et al.*, 2019) following human-based simulation studies (e.g. Conrad *et al.*, 2006) and empirical data (e.g. Kohler & Cutler, 2007) that had confirmed this as a satisfactorily measure. Nandolo *et al.* (2018) reported a considerable fraction of runs of homozygosity (ROH) islands in the bovine genome as artefacts that are misidentified due to SNP coverage gaps, which supports the hypothesis that the region around the mismatches may include deletions.

Of the ten chromosomes harbouring regions of deletions reported in the current study, seven (6, 7, 10, 17, 23 and 28) were also reported in Austrian cattle breeds (Nandolo *et al.*, 2018; Himmelbauer *et al.*, 2019) and therefore, supports reasonable grounds to suspect deletions in those chromosomes. Common regions detected in this study from the parent–progeny combinations to grandchildren and great-grandchildren illustrated that these were not just mismatches but deletions transmitted from parents to

next generations. Certainly, these may affect parentage analysis and lead to discordant and doubtful parentages if reduced panels are used without considering hemizygous deletions in populations. The deletion regions identified in the current study require further investigation to confirm the extent of coverage of the deletions and be used to detect their association with the production, reproduction, and health traits of these breeds.

Conclusion

The SNP markers used in this study provided sufficient information for parentage validation, pedigree reconstruction and detection of hemizygous regions. Hemizygous regions need to be taken into account and should be considered with great care when validating parentage analysis for genomic evaluations. Even though the analysis was successfully performed based on 91 185 SNPs due to the datasets originally generated for the genomic beef selection, in practice, a significantly lower density panel would be required to execute these analyses resulting in a much lower investment in genotyping. On the other hand, a large familial whole genome SNP-chip data is required to identify homozygosity regions. The method and results of this study can be used for parentage verification SA beef cattle breeding programs.

References

Amos, C.I., Shete, S., Chen, J. & Yu, R.K., 2003. Positional identification of micro deletions with genetic markers. Human Heredity, 56, 107–118.

Banos, G., Wiggans, G.R. & Powell, R.L., 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. Journal of Dairy Science, 84, 2523–2529.

Baron, E.E., Martinez, M.L., Verneque, R.S. & Coutinho, L.L., 2002. Parentage testing and effect of misidentification on the estimation of breeding value in Gir cattle. Genetics and Molecular Biology, 25, 4, 389-394.

Buchanan, J.W., Woronuk, G.N., Marquess, F.L., Lang, K., James, S.T., Deobald, H., Welly, B.T. & Van Eenennaam, A.L., 2016. Analysis of validated and population specific SNP parentage panels in pedigreed and commercial beef cattle populations. Canadian Journal of Animal Science, 97, 231-240.

Calus, M.P.L., Mulder, H.A. & Bastiaansen, J.W.M., 2011. Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genetic Selection Evolution, 43, 34.

Conrad, D.F., Andrews, T., Carter, N.P., Hurles, M.E. & Pritchard, J.K., 2006. A high-resolution survey of deletion polymorphism in the human genome. Nature Genetics, 38, 1, 75–81.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R. F., Liao, X., Djari, A., Rodriguez, S. C., Grohs, C. & Esquerre, D., 2014. Whole-genome sequencing of 234

bulls facilitates mapping of monogenic and complex traits in cattle. Nature genetics, 46, 8, 858-865.

Dürr, J.W., Forabosco, F., Jakobsen, J. & Zumbach B., 2011. Interbull Survey on Parentage Verification. Interbull bulletin no. 44. Stavanger, Norway, August 26 – 29. Available online: https://journal.interbull.org/index.php/ib/article/view/1822

Ferdosi, M.H., Kinghorn, B.P., van der Werf, J.H.J., Lee, S.H. & Gondro, C., 2014. Hsphase: an R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. BMC Bioinformatics, 15, 172.

García-Ruiz, A., Wiggans, G.R. & Ruiz-López, F.J., 2019. Pedigree verification and parentage assignment using genomic information in the Mexican Holstein population. Journal of Dairy Science, 102, 1806-1810.

Habier, D., Tetens, J., Seefried, F., Lichtner, P. & Thaller, G., 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution, 42, 5.

Harder, B., Bennewitz, J., Reinsch, N., Mayer, M & Kalm, E., 2005. Effect of missing sire information on genetic evaluation. Archieves Animal Breeding, 48, 219–232.

Hayes, B.J. & Goddard, M.E., 2008. Technical note: Prediction of breeding values using marker derived relationship matrices. Journal of Animal Science 86, 2089–2092.

Hayes, B.J., 2011. Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. Journal of Dairy Science, 94, 2114–2117.

Himmelbauer, J., Mészáros, G. & Sölkner, J., 2019. Detection of autosomal hemizygous regions in the Fleckvieh population-based on SNP-chip data and parent-offspring pairs. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 67, 6, 1447–1452.

Holroyd, R.G., Doogan, V.J., De Faveri, J., Fordyce, G., McGowan, M.R., Bertram, J.D., Vankan, D.M., Fitzpatrick, L.A., Jayawardhana, G.A. & Miller, R.G., 2002. Bull selection and use in Northern Australia 4. Calf output and predictors of fertility of bulls in multiple-sire herds. Animal Reproduction Science 71, 67–79.

Huisman, J., 2017. Pedigree reconstruction for SNP data: parentage assignment, sibship clustering and beyond. Molecular Ecology Resources, 17, 1009–1024.

Junqueira, V.S., Cardoso, F.F., Oliveira, M.M., Sollero, B.P., Silva, F.F. & Lopes, P.S., 2017. Use of molecular markers to improve relationship information in the genetic evaluation of beef cattle tick resistance under pedigree-based models. Journal of Animal Breeding and Genetics, 134, 14-26.

Kadri, N.K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., Nielsen, U.S., Panitz, F., Aamand, G.P., Schulman, N. & Georges, M., 2014. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. PLoS Genetics, 10, 1, e1004049.

Kios, D., van Marle-Köster, E. & Visser, C., 2012. Application of DNA markers in parentage verification of Boran cattle in Kenya. Tropical Animal Health Production, 44, 471-476.

Kohler, J.R. & Cutler, D.J., 2007. Simultaneous discovery and testing of deletions for disease association in SNP genotyping studies. The American Journal of Human Genetics, 81, 4, 684–699.

McClure, M.C., McCarthy, J., Flynn, P., McClure, J.C., Dair, E., O'Connell, D.K., Kearney, J.F., 2018. SNP data quality control in a national beef and dairy cattle system and highly accurate SNP based parentage verification and identification. Frontiers in Genetics, 9, 84.

McDaneld, T.G., Kuehn, L.A., Thomas, M.G., Pollack, E.J. & Keele, J.W., 2014. Deletion on chromosome 5 associated with decreased reproduction efficiency in female cattle. Journal of Animal Science, Volume 92, 4, 1378-1384.

Moore, K.L., Vilela, C., Kaseja, K., Mrode, R. & Coffey, M., 2019. Forensic use of the genomic relationship matrix to validate and discover livestock pedigrees. Journal of Animal Science, 97, 35–42.

Nandolo, W., Utsunomiya, Y.T., Mészáros, G., Wurzinger, M., Khayadzadeh, N., Torrecilha, R.B.P., Mulindwa, H.A., Gondwa, T.N., Waldmann, P., Ferencakovic, M., Garcia, J., Rosen, B.D., Bickhart, D., Van Tassell, C.P., Curik. I. & Sölkner, J., 2018. Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. Genetics Selection Evolution, 50, 1, 43.

Nwogwugwu, C.P., Kim, Y.K., Choi, H.J., Lee, J.H. & Lee, S.H., 2020. Assessment of genomic prediction accuracy using different selection and evaluation approaches in a simulated Korean beef cattle population. Australasian Journal of Animal Sciences, 33, 12, 1912– 1921.

Parlato, E. & Van Vleck, L.D., 2012. Effect of parentage misidentification on estimates of genetic parameters for milk yield in the Mediterranean Italian buffalo population. Journal of Dairy Science 95, 4059–4064.

Purfield, D.C., McClure, M. & Berry, D.P., 2016. Justification for setting the individual animal genotype call rate threshold at eighty-five percent. Journal of Animal Science, 94, 4558-4569.

R Core Team., 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Rafter, P., Purfield, D.C., Berry, D.P., Parnell, A.C., Gormley, I.C., Kearney, J.F., Coffey, M.P. & Carthy, T.R., 2018. Characterization of copy number variants in a large multibreed population of beef and dairy cattle using high-density single nucleotide polymorphism genotype data. Journal Animal Science, 96, 4112–4124.

Sanarana, Y.P., Maiwashe, A., Berry, D.P., Banga, C. & Van Marle-Köster, E., 2021. Evaluation of the International Society for Animal Genetics bovine single nucleotide polymorphism parentage panel in South African Bonsmara and Drakensberger cattle. Tropical Animal Health Production, 53, 32.

Sanders, K., Bennewitz, J. & Kalm, E., 2006. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. J. Dairy Sci. 89, 315–321.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H. & Gondro, C., 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, 45, 572–575.

Van Eenennaam, A.L., Weaber, R.L., Drake, D.J., Penedo, M.C. Quaas, R.L., Garrick, D.J. & Pollack, E.J., 2007. DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting. Journal of Animal Science, 85, 3159–3169.

Van Eenennaam, A.L. & Drake, D.J., 2012. Where in the beef-cattle supply chain might DNA tests generate value? Animal Production Science, 52, 185–196.

Van Marle-Köster, E. & Visser, C., 2018. Genetic improvement in South African livestock: can genomics bridge the gap between the developed and developing sectors? Frontier in Genetics, 9, 331.

VanRaden, P.M., Cooper, T.A., Wiggans, G.R., O'Connell, J.R. & Bacheller, L.R., 2013. Confirmation and discovery of maternal grandsires and great-grandsires in dairy cattle. Journal of Dairy Science, 96, 1874–1879.

Van Vleck, L.D., 1970. Misidentification in estimating the paternal sib correlation. Journal of Dairy Science, 53, 1469–1474.

Visscher, P.M., Woolliams, J.A., Smith, D. & Williams, J.L., 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. Journal of Dairy Science, 85, 2368–2375.

Visscher, P.M, Hill, W.G. & Wray, N.R., 2008. Heritability in the genomics era – concepts and misconceptions. Nature Reviews Genetics, 9, 4, 255–266.

Weller, J.I., Feldmesser, E., Golik, M., Tager-Cohen, I., Domochovsky, R., Alus, O. & Ron. M., 2004. Factors affecting incorrect paternity assignment in the Israeli Holstein population. Journal Dairy Science, 87, 2627–2640.

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R. D., Taylor, J.F., Schenkel, F.S. & Van Tassell, C. P., 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science, 92, 3431–3436.

Winchester, L., Yau, C. & Ragoussis, J., 2009. Comparing CNV detection methods for SNP arrays. Briefings in functional genomics and proteomics, 8, 5, 353-366.

CHAPTER 5

Development of a low-density SNP genotype panels for parentage verification in South African Bonsmara and Drakensberger cattle

Abstract

Genotyping panels exploiting single nucleotide polymorphism (SNP) markers have superseded the use of microsatellites for parentage testing and other applications. Although, high-density SNP arrays allow accurate parentage assignment but are not practical for routine application. Major constraint are high cost, time spent in running and analyzing the samples. In this study, low-density SNP panels with high resolution power were developed to perform accurate parentage tests for the Bonsmara (BON) and Drakensberger (DRB) cattle breeds. A genomic data consisted of 38 888 variants from the GeneSeek® Genomic Profiler (GGP) Bovine 150K BeadChip was available for 2 835 animal genotypes representing both breeds. The data included 161 sire-progeny and 313 dam-progeny pairs that were previously validated using 119 375 SNP markers. Informative SNPs were selected based on high minor allele frequencies (MAFs), clustering quality, and high call rates. To reduce linkage between markers, SNPs were spread across and within chromosomes with a minimum distance of one mega base pair (Mb) apart. A total of 200 markers were selected accurately to develop a multi-breed and population specific low-density genotype panels. On average, the genotype panel with SNPs selected across the breeds had less minor allele frequency (MAF) of 0.40 while on breed-specific panels were slightly higher with an average of 0.48 and 0.49 in the DRB and BON, respectively. The panels were validated on parent-offspring pairs based on the method of parentage exclusion of sharing at least one allele between parent–offspring pairs. The results revealed that population-specific SNP marker panels had high accuracy for parentage exclusion without any false-negatives, while the multi-breed panel did not perform well. There were no parental-offspring discrepancies in the relations of the DRB when the multi-breed panel was used, while 4.2% false-negatives were observed in the BON. This study demonstrated that markers selected based on multi-breeds may not be sufficient, and at times may fail to exclude parentages of the inbred populations. The use of population-specific panel is therefore, recommended for accurate parentages.

Keywords: Parentage exclusion, pedigree verification, Mendelian conflicts, ISAG panel

Introduction

Different genetic markers have been used for parentage verification and assignment in cattle. In recent years, technological advancements in high-throughput DNA sequencing has led to the use of highly dense single nucleotide polymorphism (SNP) markers (Matukumalli *et al.*, 2009; Weller *et al.*, 2010). The highly dense commercial SNP arrays facilitate accurate parentage verification and assignment but are not always practical for routine parentage application due to the relatively high costs and the large amount of data, which can become computationally unwieldy. Hence, smaller panels with highly selected informative markers may be preferable to reduce costs and time spent on analysis (Tokarska *et al.*, 2009; McClure *et al.*, 2012). Previous studies have demonstrated the usefulness of low-density SNP panels in cattle parentage tests which, in turn, have contributed to accelerated genetic progress (Senneke *et al.*, 2004; Anderson & Garza, 2006; Hara *et al.*, 2010).

To date, an array of parentage panels of different densities have been developed for many cattle populations globally. Initially, Heaton *et al.* (2002) described a minimum set of 32 SNP markers that has sufficient power for identification and paternity analyses in a variety of popular cattle breeds and crossbred populations in the United States. Fisher *et al.* (2009) suggested that for parentage testing in New Zealand dairy herds, a SNP panel consisting of 40 markers could be sufficient. The International Society for Animal Genetics (ISAG) standardized a SNP panel with 200 SNPs from *Bos taurus and Bos indicus* breeds to ensure consistency of parentage results between laboratories throughout the world (ISAG, 2013). However, some of the studies have found that not all the markers included in the ISAG panel were informative for parentage use in most cattle populations and suggested that more informative SNPs are required. For example, McClure *et al.* (2015) reported about 18 non-informative SNPs out of the 200 ISAG SNP markers in the Irish multi-breed cattle population. Furthermore, Sanarana *et al.* (2021) identified 19 non-informative SNPs in two South African cattle breeds associated with low minor allele frequency (MAF), probe clustering patterns, poor call rate or causing some discordant. These studies suggested that the ISAG panel may cause parentage errors and numerous parents predicted for a particular animal.

The information contained in a SNP panel may vary significantly between breeds and populations hence it is necessary to develop a SNP panel with sufficient resolution power for diverse populations. The cost of obtaining a genotype must be balanced against the accuracy with which parentage can be determined. These are likely influenced by the quantity and informativeness of genomic markers (Strucken *et al.*, 2016). While genotyping a few hundred SNPs is inexpensive, the accuracy of parentage exclusion with such a small number of SNPs is uncertain. Recently, Strucken *et al.* (2017) used Illumina panel data from 735K SNPs to create a smaller panel for parentage in East African crossbred dairy cattle, and reported that correctly selected breed-specific 200 SNPs were able to assign parentages more reliably than multi-breed panels. McClure *et al.* (2015; 2018) analyzed different SNP densities for *Bos*

taurus cattle in Ireland and recommended a 500-SNP panel for parentage validation and 800 SNPs for predicting a parent from a large reference population.

Genomic selection is still at its infancy in South Africa with some beef populations genotyped with the Bovine 150K BeadChip. In the absence of routine genotyping for a high number of SNP markers for use in genome-wide-enabled selection, low-cost genomic technologies that verify parentage can be valuable to improve genetic gain. Thus, the aim of this study was to select informative SNP markers from the 150K SNP panel currently used for genomic selection to develop low-density genotype panels for parentage testing in the local breeds of the BON and DRB cattle.

Materials and Methods

The Animal Ethics Committee (AEC) at the University of Pretoria granted its approval for the study to use external data (EC066-16 AEC).

Genotype data and quality control

Animal genotypes consisting of 138 888 bi-allelic SNPs from the Illumina Bovine SNP 150K BeadChip array (Illumina Inc., San Diego, USA) were available on 1 563 Bonsmara and 1 022 Drakensberger cattle. Data editing and analysis were performed using custom scripts in R software (R Development Core Team, 2021). All animal genotypes had call rate higher than 90%. After removing duplicate SNPs, those on the X, Y, and MT chromosomes, and those with unknown locations, a total of 119 375 autosomal SNPs based on the UMD3.1 bovine genome remained.

Parentage of the available genotyped animals was confirmed in a separate analysis using the high-density SNP panel based on number of mismatches between the SNP genotypes of a parent and its progeny. Thus, the data used in the current study was subdivided into two groups (1) a calibration dataset which included animals without genotyped parents, and (2) a test population which included animals with genotyped parents. The test population consisted of 118 and 273 animals sired by 25 and 9 bulls and from 68 and 163 dams in the Bonsmara and Drakensberger breeds, respectively. A total of 1 936 animals (BON = 1 357 and DRB = 579) without genotyped parents were used in the calibration dataset for SNP editing and the selection of informative SNPs.

The GenCall (GC) and GenTrain (GT) scores were used to edit the genotypes based on the accuracy of the calls. A minimum threshold of 0.60 for the GC and 0.55 for the GT scores was applied to determine the validity of the genotypes for each SNP. A set of 160 sire-offspring pairs with previously verified parentage were used to identify SNPs with opposing homozygote genotypes (Chapter 4). A total of 3 575 autosomal SNPs not adhering to Mendelian inheritance were removed. Following the removal of SNPs with a call rate of less than 99%, the minor allele frequency (MAF) per SNP within each breed was calculated, and all SNPs with MAF less than 0.05 were removed. A test for Hardy-Weinberg

Equilibrium (HWE) was performed within population and SNPs that departed ($p<0.001$) from HWE were discarded. After quality control, 78 286 candidate autosomal SNPs remained for selection of informative SNPs to compile a low-density parentage verification panel. Additional genetic diversity parameters including the expected heterozygosity (H_E), and polymorphic information loci (PIC) were calculated per SNP.

SNP selection for panel development

The number of SNPs chosen per chromosome was directly proportional to the chromosome length, which was measured from the first to the last SNP's genomic position. Selection of SNPs based on the block method was shown to increase the accuracy of the low-density panel (O'Brien *et al.*, 2019). The block approach divides each chromosome into equal-length blocks, with the number of blocks per chromosome equal to the chromosome's predetermined number of SNPs. SNPs with both high MAF and high call rate are known to be informative and produce the highest number of opposing homozygotes between unrelated individuals (Van Doormaal *et al.*, 2016). All SNPs were ranked on MAF and the SNP with the highest MAF per block was selected. To minimize linkage disequilibrium (LD) among selected SNPs, the selected SNPs had to be at least 1 Mb apart. Finally, a total of 200 informative SNPs were compiled for the low-density genotype parentage panels.

Parentage verification

Since parentage for each sire- and dam-offspring pair in the test dataset were previously validated, the verified parentages based on 91 185 SNPs were used as the point of reference for accuracy of the 200 selected markers for parentage exclusion. The SNP panels with markers selected across the genome of the BON and DRB were validated in comparison with the breed-specific selected SNPs and the ISAG panel. False-negative parentage assignment rates were assessed from the total number of relationships.

Results

The distribution of the 200 selected SNPs were located across the 29 autosomes (Figure 5.1). As expected, a large number of markers were observed on chromosome 1, while the fewest SNPs were chosen on shorter chromosomes (i.e. 25, 27 and 29). Details of the selected 200 SNPs are provided in Supplementary Table 7.5.

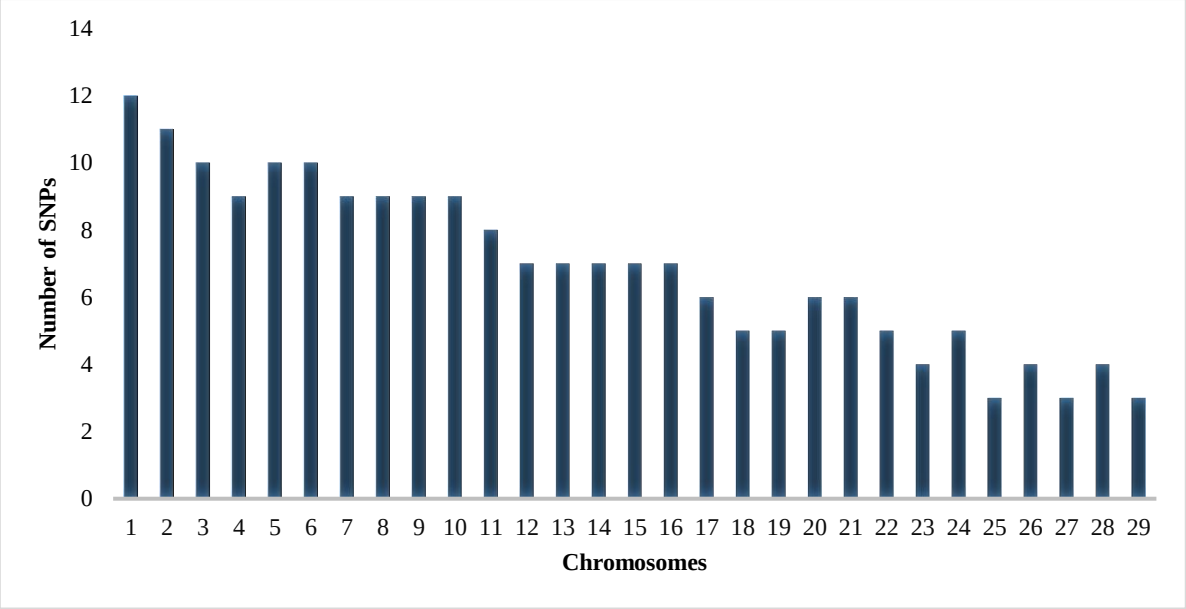


Figure 5.1. The distribution of the 200 selected SNPs included in 29 chromosomes of the low-density panel

The boxplots of the mean MAFs per breed or breed combination for each panel are presented in Figure 5.2. The MAF values per SNP for the multi-breed panel were within the range of 0.30 to 0.50 with an average of 0.40. The mean MAF values were 0.48 and 0.49 in the DRB and BON, respectively in the SNP panels developed per breed. The MAF per SNP ranged from 0.42 to 0.50 in the BON and 0.46 to 0.50 in the DRB. In comparison to the multi-breed and breed-specific panels, the ISAG panel showed low MAF values of 0.33 in the BON and 0.36 in the DRB with an interquartile range between 0.01 and 0.50.

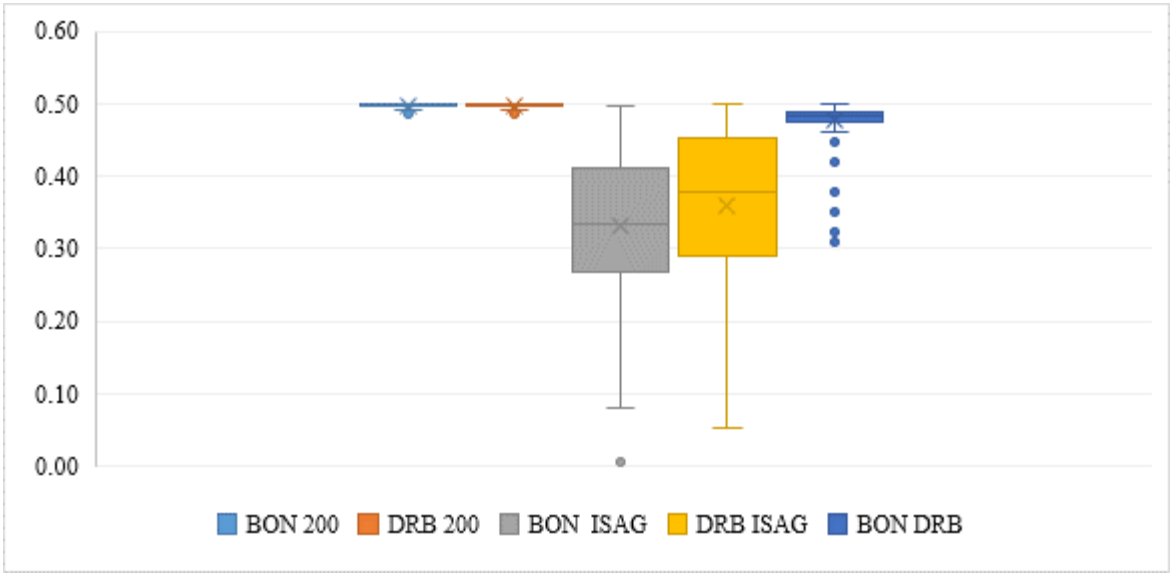


Figure 5.2. Boxplot of the minor allele frequencies (MAF) of SNP markers selected within and across breeds in relation to the MAF of the ISAG panel

The performance of the within and multi-breed SNP panels are presented in Table 5.1. As expected, based on the biased selection of markers across the breeds, a total of 5 discordant SNPs were observed in the BON population while 2 were observed in the DRB population when the multi-breed panel was tested. The multi-breed panel also showed difficulties in excluding five parent–offspring relations (4.2%) out of 118 in the BON. Even though some discordant SNPs were observed in the DRB, there were no false-negatives.

Table 5.1. Performance of SNP panels for parentage testing in the Bonsmara and Drakensberger breeds within and across breeds

Population	Panel	MAF	H_E	PIC	Disc	Fn (%)
Bonsmara	Multi-breed	0.38	0.46	0.35	5	4.2
	BON	0.49	0.50	0.37	0	0
Drakensberger	Multi-breed	0.40	0.44	0.35	2	0
	DRB	0.48	0.48	0.36	0	0

BON - SNP panel for the Bonsmara; DRB - SNP panel for the Drakensberger; MAF - minor allele frequency; H_E - expected heterozygosity; PIC - polymorphic information content; Disc - number of discordant SNPs; Fn - false negatives

Breed-specific panels performed better than multi-breed panel for parentage testing. The BON and DRB genotype panels were free from any false- positives or -negatives when one mismatch was allowed. The average H_E and PIC values were also comparatively higher in the BON and DRB panels than multi-breed.

Discussion

The importance of correct parentage is well established in terms of more precise genetic selection through accurate estimates of genetic parameters, EBVs, and genetic trend (Van Vleck, 1970; Geldermann *et al.*, 1986). Pedigree errors occur in beef production systems and have been reported in several populations (Kios *et al.*, 2012; Panetto *et al.*, 2017; Moore *et al.*, 2019). Loss in genetic gain can be recovered through better pedigree accuracy, enabling fair comparisons among animals (Bradford *et al.*, 2019). Accurate and cost-effective molecular based parentage verification tools are relevant for genetic improvement and optimal selection in breeding programs. While there is currently no international standard on the number SNPs to use for parentage outside of the ISAG set, a robust, technically precise and reasonably priced tool for determining parentage could be advantageous in facilitating optimized breeding programs and accelerate the rates of genetic gain (Berry *et al.*, 2019).

Several studies in the developed world have demonstrated the use of genomic markers to accelerate the rate of genetic gain (Junqueira *et al.*, 2017; García-Ruiz *et al.*, 2019). East and West African countries

have recently embarked on developing genotype parentage panels for Sanga and crossbred cattle populations for routine applications (Strucken *et al.*, 2017; Gebrehiwot *et al.*, 2021). The present study is the first attempt to provide valuable insights into the development of low-density panel for parentage testing from southern African Sanga cattle breeds i.e. Bonsmara and Drakensberger cattle breeds.

A number of studies have selected different number of SNP markers for different populations (McClure *et al.*, 2015; Panetto *et al.*, 2017; Hu *et al.*, 2021). The 200 markers selected to develop the low-density SNP panels in this study were larger than the 71 and 50 parentage panels that were selected for the Brazilian Red Sindhi (Panetto *et al.*, 2017) and Chinese Simmental cattle (Zhang *et al.*, 2018), respectively. Other studies have considered higher or similar number of markers as in the current study e.g. 257 and 245 SNPs in Hanwoo and Wagyu cattle (Strucken *et al*., 2014) and 200 in East and West African cattle populations (Strucken *et al*., 2017; Gebrehiwot *et al*., 2021).

The MAF of a SNP in a population determines how useful it is for parentage analysis (McClure *et al.*, 2018). As a result, SNPs with a low MAF are less relevant for parentage analysis, because only homozygous genotypes in the parent and progeny are considered informative (Van Doormaal *et al.*, 2016). The number of parentage SNPs used may differ depending on the cost and level of willingness to risk for parentage errors. The number of SNPs used in parentage testing based on exclusion method is not always equal to the total number of SNPs in the panel. This is mainly because neither heterozygous SNP genotypes nor uncalled genotypes are taken into account. Thus, the value of MAF observed in the panels of 200 SNPs selected based on high MAF in this study have performed best in assigning parentage using opposing homozygote criteria.

It was interesting to note that the accuracy of parentage testing increased with an increase in MAF values. For example, the breed-specific panels of the BON and DRB had higher MAF values than the multi-breed and the ISAG panels. Furthermore, breed-specific panels verified parentage accurately without any false-negatives detected making them better panels than both the multi-breed and the ISAG panels. The false-negatives based on the use of the multi-breed panel are probably due to either selective genotyping of the samples available in the database or the small sample population size of the BON. At times, small panels do not have enough power to exclude potentially closely related candidates (Strucken *et al.*, 2014).

Conclusion

This study identified 200 SNP markers for parentage testing for BON and DRB cattle populations. These results also suggested that small panels designed for a particular breed are better than a set of markers selected across breeds. Thus, in order to achieve parentage accuracy and assignment across breeds, a larger marker panel may be required. The low-density panels developed will assist towards achieving more accurate genetic evaluation in South African beef cattle, which will likely contribute to

more efficient beef production. It is, however, not known if the panels developed will be efficient for other local breeds not considered in the current study.

References

Anderson, E.C. & Garza, J.C., 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics, 172, 4, 2567–2582.

Berry, D.P., McHugh, N., Wall, E., McDermott, K. & O'Brien, A.C., 2019. Low-density genotype panel for both parentage verification and discovery in a multi-breed sheep population. Ireland Journal of Agricultural and Food Research, 58, 1-12.

Bradford, H.L., Masuda, Y., Cole, J.B., Misztal, I. & VanRaden, P.M., 2019. Modeling pedigree accuracy and uncertain parentage in single-step genomic evaluations of simulated and US Holstein datasets. Journal of Dairy Science, 102, 2308–2318.

Fisher, P.J., Malthus, B., Walker, M.C., Corbett, G. & Spelman, R.J., 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. Journal of Dairy Science, 92, 369-374.

García-Ruiz, A., Wiggans, G.R. & Ruiz-López, F.J., 2019. Pedigree verification and parentage assignment using genomic information in the Mexican Holstein population. Journal of Dairy Science, 102, 1806-1810.

Geldermann, H., Pieper U. & Weber, W., 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. Journal of Animal Science, 63, 1759–1768.

Gebrehiwot, N.Z., Strucken, E.M., Marshall, K., Aliloo, H. & Gibson, J.P., 2021. SNP panels for the estimation of dairy breed proportion and parentage assignment in African crossbred dairy cattle. Genetics Selection Evolution, 53, 21.

Hara, K., Watabe, H., Sasazaki, S., Mukai, F. & Mannen, H., 2010. Development of SNP markers for individual identification and parentage test in a Japanese black cattle population. Journal of Animal Science, 81, 152-157.

Heaton, M.P., Harhay, G.P., Bennett, G.L., Stone, R.T., Grosse, W.M., Casas, E., Keele, J.W., Smith, T.P.L., Chitko-McKown, G.C. & Laegreid, W.W., 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. Mammalian Genome, 13, 272-281.

Hu, L.R., Li, D., Chu, Q., Wang, Y.C., Zhou, L., Yu, Y., Zhang, Y., Zhang, S.L., Usman, T., Xie, Z.Q. & Hou, S.Y., 2021. Selection and implementation of single nucleotide polymorphism markers for parentage analysis in crossbred cattle population. Animal, 15, 1, 100066-100071.

International Society for Animal Genetics (ISAG), cattle core and additional SNP panel., 2013. www.isag.us/Docs/Cattle-SNP-ISAG-core-additional-panel-2013.xlsx

Junqueira, V.S., Cardoso, F.F., Oliveira, M.M., Sollero, B.P., Silva, F.F. & Lopes, P.S., 2017. Use of molecular markers to improve relationship information in the genetic evaluation of beef cattle tick resistance under pedigree-based models. Journal of Animal Breeding and Genetics, 134, 14-26.

Kios, D., van Marle-Köster, E. & Visser, C., 2012. Application of DNA markers in parentage verification of Boran cattle in Kenya. Tropical Animal Health Production, 44, 471-476.

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., 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.

McClure, M., Sonstegard, T., Wiggans, G. & Van Tassell, C.P., 2012. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. Frontier in Genetics, 3, 140, 1-9.

McClure, M.C., McCarthy, J., Flynn, P., Weld, R., Keanne, M., O'Connell, K., Mullen, M.P., Waters, S. & Kearney, J.F., 2015. SNP selection for nationwide parentage verification in beef and dairy cattle. International Committee of Animal Recording (ICAR) Technical series. No 19.

McClure, M.C., McCarthy, J., Flynn, P., McClure, J. C., Dair, E., O'Connell, D. K. & Kearney, J. F., 2018. SNP data quality control in a national beef and dairy Cattle system and highly accurate SNP based parentage verification and identification. Frontiers in Genetics, 9, 84.

Moore, K.L., Vilela, C., Kaseja, K., Mrode, R. & Coffey, M., 2019. Forensic use of the genomic relationship matrix to validate and discover livestock pedigrees. Journal of Animal Science, 97, 35–42.

O'Brien, A.C., Judge, M.M., Fair, S. & Berry, D.P., 2019. High imputation accuracy from informative low-to-medium density single nucleotide polymorphism genotypes is achievable in sheep. Journal of Animal Science, 97, 1550–1567.

Panetto, J.C.C., Machado, M.A., Silva, M.V.G.B., Barbosa, R.S., Santos, G.G., Leite, R.M.H. & Peixoto, M.G.C.D., 2017. Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. Livestock Science, 204, 33–38.

R Core Team., 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Sanarana, Y.P., Maiwashe, A., Berry, D.P., Banga, C. & van Marle-Köster, E., 2021. Evaluation of the International Society of Animal Genetics bovine single nucleotide polymorphism parentage

panel in South African Bonsmara and Drakensberger cattle. Tropical Animal Health and Production, 53, 32.

Senneke, S.L., MacNeil, M.D. & Van Vleck, L.D., 2004. Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. Journal of Animal Science, 82, 2307–2312.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H. & Gondro, C., 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, 45, 572-575.

Strucken, E.M., Lee, S.H., Lee, H.K., Song, K.D., Gibson, J.P. & Gondro, C., 2016. How many markers are enough? Factors influencing parentage testing in different livestock populations. Journal of Animal Breeding and Genetics, 133, 1, 13-23.

Strucken, E.M., Al-Mamun, H.A., Esquivelzeta-Rabell, C., Gondro, C., Mwai, O.A. & Gibson, J.P., 2017. Genetic tests for estimating dairy breed proportion and parentage assignment in East African crossbred cattle. Genetic Selection Evolution, 49, 67.

Tokarska, M., Marshall, T., Kowal-czyk, R., Wojcik, J.M., Pertoldi, C., Kristensen, T. N., Loeschcke, V., Gregersen, V.R. & Bendixen, C., 2009. Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. Heredity 103, 326–332.

Van Doormaal, B.J., Kistemaker, G.J. & Johnston, J., 2016. Parentage Analysis Services for Dairy Cattle in Canada. Interbull Bulletin no. 50. Puerto Varas, Chile, October 24-28.

Van Vleck, L.D., 1970. Misidentification in estimating the paternal sib correlation. Journal of Dairy Science 53, 1469–1474.

Weller, J.I., Glick, G., Ezra, E., Zeron, Y., Seroussi, E. & Ron, M., 2010. Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. Animal genetics, 41, 5, 551-553.

Zhang, T., Guo, L., Shi, M., Xu, L., Chen, Y., Zhang, L., Gao, H., Li, J. & Gao, X., 2018. Selection and effectiveness of informative SNPs for paternity in Chinese Simmental cattle based on a high-density SNP array. Gene, 5, 673, 211-216.

CHAPTER 6

6.1. Critical discussion and recommendations

The use of genomic information has played a major role in parentage verification and assignment in cattle populations and enhanced the rates of genetic progress in developed countries, while in developing countries such a tool is often not available to the same extent (Mrode *et al.*, 2019). Scientific reviews have highlighted major limitations associated with lack of accurate or complete records, poor performance recording along with delays in the establishment of genomic selection programs (Van Marle-Köster &Visser, 2018; Ibeagha-Awemu *et al.*, 2019). However, the latter offers applications to circumvent some of the limitations and thus potentially facilitating more precise breeding and management decisions. The current study has utilized genomic information from the GeneSeek® Genomic Profiler (GGP) Bovine SNP150K Bead Chip to perform parentage analysis in South African Bonsmara (BON) and Drakensberger (DRB) cattle populations. This is the first study to analyze the pedigree records of the SA Bonsmara and Drakensberger cattle populations using SNPs.

The 200 SNPs recommended for parentage verification of cattle globally by ISAG-ICAR are embedded within the Bovine SNP 150K Bead Chip. The ISAG panel was evaluated on Sanga cattle as the first objective of the thesis. The results revealed that although the populations were genotyped with the same Bovine 150K Bead Chip, not all ISAG markers were segregating in these populations. Similar observations were made when the ISAG panel was used in the East Asia Hanwoo and Wagyu populations (Strucken *et al.*, 2014); Australian Brahman and Brahman crosses (Lyons *et al.*, 2015) and East African dairy crossbred and Sanga cattle populations (Strucken *et al.*, 2017).

The effectiveness of parentage testing panels depend on marker informativeness. The analysis in the current study demonstrated that the ISAG SNP markers have some genotyping issues such as the clustering patterns, very low MAF and call rate, with some causing discordances between the parent-offspring pairs. McClure *et al.* (2015) observed similar shortcomings with 18 SNPs in a multi-breed Irish cattle population and recommended the addition of more SNPs for accurate parentage exclusion and assignment. It is important to note that SNPs identified with genotyping issues in the current study were not the same SNPs in the BON and DRB breeds indicating that the performance of ISAG may vary significantly between breeds. The ISAG panel poorly performed with 23.4% and 33.0% false-negatives detected in the BON and DRB, respectively and failed to correctly assign unique sire-offspring pairs validated with 119 375 markers when sires were masked. These results further revealed that not all SNP markers included in the ISAG panel have sufficient resolution power for parentage exclusion hence, at times, the ISAG panel may fail to confirm parentage and result in multiple parent candidates for a single animal. The poor performance of the ISAG panel was not unexpected since the

panel was developed for parentage analysis predominantly in *Bos taurus* and *Bos indicus* populations, and did not include the Sanga breeds (ISAG, 2013; Strucken *et al.*, 2014). Thus, adding informative markers was suggested with the interest of developing a low-cost and effective SNP density panels for local cattle breeds.

The second objective of this thesis validated parentage and identified genomic relationships that were not recorded in pedigrees using the genomic relationship matrix. The analysis allowed the detection of hemizygous deletions (hDEL) in the genome and individual carriers using the Mendelian mismatch method. Considering the nature of the beef production system involving multiple sire mating and less use of artificial insemination, parentage errors observed in the present study were expected. The results revealed, on average, parentage errors of 8.5% and 10.1% in the BON and DRB, respectively. Parentage errors were relatively higher in the sire-progeny pairs (10.0% in BON and 17.2% in DRB) than in dam-offspring pairs (7.0% in BON and 7.6% in DRB). Likewise, of the 69 relationships that were discovered through the genomic relationship matrix, ~57% were related to the sire information while ~53% were related to the dam information. The genomic relationship matrix is a useful tool in validating and discovering pedigrees but it relies on the ability to define certain threshold values, which require a sufficient number of genotyped animals and pedigree validated before it can be used (Moore *et al.*, 2019).

Even though just a few relationships were corrected, and just a few discovered in the present study, the use of SNP-based parentage verification in South African cattle will help mitigate the repercussions of incorrect pedigree information. The effect of wrong sire information is known to have twice the impact of missing sire information for both reliability and genetic gain (Woolliams, 2006). On the other hand, the effect of wrong dam information can be substantial, with negative and additive effects on the rate of genetic gain (Harder *et al.*, 2005). It is anticipated that the percentage of animals with parentage test will increase in the future as the number of animals genotyped for genomic evaluations is increasing. In addition, the Beef Genomics Program (BGP) has created a bank for hair samples, which could increase the possibility of future parentage tests, which in turn will assist the precision of genomic evaluations.

The method of analysis using Mendelian mismatches to define the potential regions of deletions based on SNP data is considered as a reliable procedure because all mismatches identified from the progeny were traced back to the parents especially in the DRB breed, which exhibited close familial structure. This information provided useful details to identify suspect deletions in the region at which the Mendelian mismatches frequently occurred (Himmelbauer *et al.*, 2019). It is recommended that these results should be validated either with a larger genomic sample size, or with the next generation sequencing (NGS) data to capture the genomic coverage gaps of the regions identified. Additionally,

large genomic sample size from different breeds is required to investigate the potential effects of hemizygous deletions on different traits of economic importance in beef production.

The final objective focused on the compilation of low-density genotype panels for parentage verification. Results from this objective indicated that breed-specific selected SNP markers are more efficient for parentage testing without any false-negatives compared with panels developed based on selection of markers across the two breeds. The multi-breed panel had some disconcordants but performed better than the ISAG when it was applied in both breeds. This was not surprising as some of the SNPs on the ISAG parentage panel were not informative in both breeds. The number of selected markers to develop the panel was rather large compared to other studies (Panetto *et al.*, 2017; Zhang *et al.*, 2018) but smaller than those selected by McClure *et al.* (2015) for the Irish multi-breed populations and equivalent to those selected for crossbred in East and West Africa (Strucken *et al.*, 2017; Gebrehiwot *et al.*, 2021).

Based on the results from the current and other studies, a large number of markers is recommended to develop a panel for multi-breed to cover the difference in allele frequencies across breeds. However, the only concern about maximizing the number of SNPs is the cost of genotyping, as the final panel will be used for routine parentage testing. Therefore, the number of SNPs utilized by each laboratory, breed society, or national valuation centre likely depend on cost and level of acceptable risk for a parentage error (McClure *et al.*, 2018). The low-density panel developed in the current study can be useful for other applications such as animal forensics to solve stock theft, and meat authentication for food safety by only applying slight modifications to the SNP panel.

6.2. Future research studies

Using SNP data to confirm conflicts between parent-offspring disputes improves relationship and, as a result, the accuracy of breeding value predictions of candidate animals (Junqueira *et al.*, 2017). Although there is a variety of cattle breeds in South Africa, the current study was limited to Bonsmara and Drakensberger cattle breeds because of the amount of genomic data available in these two breeds compared to the other breeds. In addition, the analyses of the study required a large dataset needed for calibration and test population sets for validation.

Future studies should consider more breeds as more genotypes become available for development of multi-breed parentage genotype panels and the detection of hemizygous deletions. Since many breeds exist in South Africa, a study that will develop a reduced SNP panel for the estimation of breed proportion will be desirable. This will be helpful to identify individuals of a particular breed in a cost-effective routine genotyping (Kumar *et al.*, 2019).

Routine genotyping of cows is encouraged for the testing of dams which will assist to evaluate the effect of using the discovered maternal pedigrees specifically maternal grandsires and great grandsires when the dam is not known in genetic prediction (Bradford *et al.*, 2019). Thus, more research is necessary to understand the use of discovered maternal pedigrees in genetic predictions. The effect of parentage errors on genetic gain was not considered in the current study because of selective genotyping and therefore research is required in this area. These studies could help to improve the genetic evaluation programs and allow breeders to make informed accurate selection decisions.

6.3. Conclusions

Parentage analyses should carefully consider accuracy of the SNP panel to ensure true relationships are determined. The parentage panel recommended by the ISAG-ICAR committee showed some limitations in Sanga breeds which may result in parentage mis-assignments. This indicated that parentage validation from SNP markers requires careful consideration of marker selection to ensure accuracy in pedigree records of the Bonsmara and Drakensberger cattle breeds.

The analysis and interpretation of genetic relationships from the genomic relationship matrix based on opposing homozygote genotypes can be used to validate and discover missing and identify wrong pedigree information, which may play a role in increasing the rates of genetic improvement and selection. The method based on Mendelian mismatches in parent-offspring pairs showed effectiveness in identifying regions of possible deletions, which is a great advantage due to the fact that SNP data are commonly used and available for cattle many populations.

A total of 200 SNP markers were identified for parentage testing based on the performance in the Bonsmara and Drakensberger. However, it is unknown how well these markers will perform in other Sanga cattle populations in South Africa. Although not all the South African cattle breeds were included in the current study, it is important to highlight that the methodology and findings from this thesis can be applied to databases of the current genotypes to improve genetic and genomic evaluations.

References

Bradford, H.L., Masuda, Y., Cole, J.B., Misztal, I. & VanRaden, P.M., 2019. Modelling pedigree accuracy and uncertain parentage in single-step genomic evaluations of simulated and US Holstein datasets. Journal of Dairy Science, 102, 2308–2318.

Gebrehiwot, N.Z., Strucken, E.M., Marshall, K., Aliloo, H. & Gibson, J.P., 2021. SNP panels for the estimation of dairy breed proportion and parentage assignment in African crossbred dairy cattle. Genetics Selection Evolution, 53, 1, 1-18.

Himmelbauer, J., Mészáros, G. & Sölkner, J., 2019. Detection of autosomal hemizygous regions in the Fleckvieh population-based on SNP-chip data and parent-offspring pairs. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 67, 6, 1447–1452.

Harder, B., Bennewitz, J., Reinsch, N., Mayer, M. & Kalm, E., 2005. Effect of missing sire information on genetic evaluation. Archives Tierzuctch, 48, 219–232.

Ibeagha-Awemu, E.M., Peters, S.O., Bemji, M.N., Adeleke, M.A. & Do, D.N., 2019. Leveraging available resources and stakeholder involvement for improved productivity of African livestock in the era of genomic breeding. Frontiers in Genetics, 10, 357.

International Society for Animal Genetics (ISAG), cattle core and additional SNP panel., 2013. www.isag.us/Docs/Cattle-SNP-ISAG-core-additional-panel-2013.xlsx

Junqueira, V.S., Cardoso, F.F., Oliveira, M.M., Sollero, B.P., Silva, F.F. & Lopes, P.S., 2017. Use of molecular markers to improve relationship information in the genetic evaluation of beef cattle tick resistance under pedigree-based models. Journal of Animal Breeding and Genetics, 133, 1-13.

Kumar, H., Panigrahi, M., Chhotaray, S., Parida, S., Chauhan, A., Bhushan, B., Gaur, G.K., Mishra, B.P. & Singh, R.K., 2021. Comparative analysis of five different methods to design a breed-specific SNP panel for cattle. Animal Biotechnology, 32, 1, 130-136.

Lyons, R.E., Buttsworth, S., Waine, D. & Kelly, M., 2015. SNP-based parentage in an Australian cattle industry context: Does one size fit all. Association for Advancement of Animal Breeding and Genetics, 21, 249-252.

McClure, M.C., McCarthy, J., Flynn, P., Weld, R., Keanne, M., O'Connell, K., Mullen, M.P., Waters, S. & Kearney, J.F., 2015. SNP selection for nationwide parentage verification in beef and dairy cattle. Proceeding of the International Committee of Animal Recording (ICAR), Technical series, 19, 175-181. Krakow, Poland, June 10-12.

McClure, M.C., McCarthy, J., Flynn, P., McClure, J. C., Dair, E., O'Connell, D. K. & Kearney, J. F., 2018. SNP data quality control in a national beef and dairy Cattle system and highly accurate SNP based parentage verification and identification. Frontiers in Genetics, 9, 84.

Moore, K.L., Vilela, C., Kaseja, K., Mrode, R. & Coffey, M., 2019. Forensic use of the genomic relationship matrix to validate and discover livestock pedigrees. Journal of Animal Science, 97, 35–42.

Mrode, R., Ojango, J., Okeyo, A. M. & Mwacharo, J. M., 2019. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. Frontiers in Genetics, 9, 694.

Panetto, J.C.C., Machado, M.A., Silva, M.V.G.B., Barbosa, R.S., Santos, G.G., Leite, R.M.H. & Peixoto, M.G.C.D., 2017. Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. Livestock Science, 204, 33–38.

Strucken, E.M., Gudex, B., Ferdosi, M.H., Lee, H.K., Song, K.D., Gibson, J.P., Kelly, M., Piper, E.K., Porto-Neto, L.R., Lee, S.H. & Gondro, C., 2014. Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, 45, 572-575.

Strucken, E.M., Al-Mamun, H.A., Esquivelzeta-Rabell, C., Gondro, C., Mwai, O.A. & Gibson, J.P., 2017. Genetic tests for estimating dairy breed proportion and parentage assignment in East African crossbred cattle. Genetic Selection Evolution, 49, 67.

Van Marle-Köster, E. & Visser, C., 2018. Genetic improvement in South African livestock: can genomics bridge the gap between the developed and developing sectors?. Frontier in Genetics, 9, 331.

Woolliams, J.A., 2006. Technical Note: A Note on the differential impact of wrong and missing sire information on reliability and gain. Journal of Dairy Science, 89, 4901–4902.

Zhang, T., Guo, L., Shi, M., Xu, L., Chen, Y., Zhang, L., Gao, H., Li, J. & Gao, X., 2018. Selection and effectiveness of informative SNPs for paternity in Chinese Simmental cattle based on a high-density SNP array. Gene, 673, 211-216.

ADDENDA

Addendum 1 Electronic supplementary material ESM_1 (Online) of Chapter 3 (https://doi.org/10.1007/s11250-020-02481-6)

SNPNAME	BON	DRB
ARS-BFGL-BAC-19454	0.379	0.471
ARS-BFGL-BAC-27364	0.149	0.269
ARS-BFGL-BAC-35552	0.375	0.383
ARS-BFGL-NGS-10035	0.447	0.413
ARS-BFGL-NGS-101456	0.265	0.392
ARS-BFGL-NGS-102169	0.259	0.460
ARS-BFGL-NGS-106015	0.294	0.219
ARS-BFGL-NGS-111053	0.345	0.483
ARS-BFGL-NGS-111076	0.412	0.350
ARS-BFGL-NGS-111114	0.231	0.281
ARS-BFGL-NGS-112094	0.409	0.479
ARS-BFGL-NGS-112325	0.411	0.500
ARS-BFGL-NGS-11383	0.371	0.357
ARS-BFGL-NGS-114006	0.191	0.288
ARS-BFGL-NGS-115514	0.390	0.409
ARS-BFGL-NGS-117319	0.401	0.441
ARS-BFGL-NGS-118319	0.485	0.439
ARS-BFGL-NGS-118340	0.313	0.314
ARS-BFGL-NGS-119662	0.453	0.495
ARS-BFGL-NGS-14740	0.352	0.395
ARS-BFGL-NGS-15731	0.483	0.463
ARS-BFGL-NGS-24419	0.318	0.368
ARS-BFGL-NGS-26517	0.481	0.414
ARS-BFGL-NGS-27577	0.398	0.444
ARS-BFGL-NGS-31640	0.270	0.388
ARS-BFGL-NGS-31807	0.322	0.292
ARS-BFGL-NGS-36513	0.405	0.393
ARS-BFGL-NGS-38423	0.262	0.318
ARS-BFGL-NGS-38620	0.398	0.301
ARS-BFGL-NGS-39978	0.239	0.309
ARS-BFGL-NGS-42283	0.345	0.205
ARS-BFGL-NGS-55943	0.291	0.393
ARS-BFGL-NGS-57711	0.281	0.258
ARS-BFGL-NGS-58613	0.290	0.398
ARS-BFGL-NGS-67146	0.486	0.360
ARS-BFGL-NGS-70946	0.288	0.129
ARS-BFGL-NGS-72471	0.185	0.370
ARS-BFGL-NGS-76191	0.235	0.361

ARS-BFGL-NGS-76330	0.249	0.352
ARS-BFGL-NGS-93119	0.248	0.221
ARS-BFGL-NGS-96125	0.246	0.235
ARS-BFGL-NGS-99210	0.446	0.394
ARS-USMARC-569	0.481	0.465
ARS-USMARC-Parent-AY761135-rs29003723	0.465	0.280
ARS-USMARC-Parent-AY776154-no-rs	0.401	0.208
ARS-USMARC-Parent-AY841151-rs29003466	0.280	0.197
ARS-USMARC-Parent-AY842472-rs29001941	0.299	0.311
ARS-USMARC-Parent-AY842473-rs29001956	0.424	0.409
ARS-USMARC-Parent-AY842474-rs29003226	0.387	0.495
ARS-USMARC-Parent-AY842475-rs29002127	0.164	0.333
ARS-USMARC-Parent-AY844963-rs17871338	0.329	0.407
ARS-USMARC-Parent-AY849381-rs29003287	0.473	0.434
ARS-USMARC-Parent-AY850194-no-rs	0.390	0.381
ARS-USMARC-Parent-AY851162-no-rs	0.498	0.384
ARS-USMARC-Parent-AY851163-rs17871661	0.470	0.487
ARS-USMARC-Parent-AY853302-no-rs	0.444	0.393
ARS-USMARC-Parent-AY853303-no-rs	0.275	0.203
ARS-USMARC-Parent-AY856094-rs17871190	0.474	0.353
ARS-USMARC-Parent-AY858890-rs29002256	0.433	0.447
ARS-USMARC-Parent-AY860426-no-rs	0.300	0.371
ARS-USMARC-Parent-AY863214-rs17871744	0.389	0.468
ARS-USMARC-Parent-AY914316-rs17871403	0.321	0.462
ARS-USMARC-Parent-AY916666-no-rs	0.439	0.297
ARS-USMARC-Parent-AY919868-rs29002211	0.204	0.385
ARS-USMARC-Parent-AY929334-no-rs	0.381	0.403
ARS-USMARC-Parent-AY937242-rs17872223	0.428	0.368
ARS-USMARC-Parent-AY939849-rs17870274	0.315	0.337
ARS-USMARC-Parent-AY941204-rs17872131	0.353	0.206
ARS-USMARC-Parent-AY943841-rs17871566	0.356	0.474
ARS-USMARC-Parent-DQ381152-rs29002408	0.436	0.485
ARS-USMARC-Parent-DQ381153-rs29012842	0.282	0.212
ARS-USMARC-Parent-DQ404149-no-rs	0.480	0.356
ARS-USMARC-Parent-DQ404150-rs29012530	0.438	0.457
ARS-USMARC-Parent-DQ404151-rs29019282	0.378	0.407
ARS-USMARC-Parent-DQ404152-rs29022245	0.203	0.220
ARS-USMARC-Parent-DQ404153-no-rs	0.110	0.289
ARS-USMARC-Parent-DQ435443-rs29010802	0.317	0.476
ARS-USMARC-Parent-DQ451555-rs29010795	0.300	0.230
ARS-USMARC-Parent-DQ468384-rs29003967	0.208	0.097
ARS-USMARC-Parent-DQ470475-no-rs	0.292	0.401
ARS-USMARC-Parent-DQ489377-rs29026932	0.297	0.373
ARS-USMARC-Parent-DQ647186-rs29014143	0.092	0.196
ARS-USMARC-Parent-DQ647189-rs29012226	0.315	0.189
ARS-USMARC-Parent-DQ647190-rs29013632	0.205	0.401

ARS-USMARC-Parent-DQ650635-rs29012174	0.336	0.206
ARS-USMARC-Parent-DQ650636-rs29024525	0.134	0.359
ARS-USMARC-Parent-DQ786757-rs29019900	0.479	0.362
ARS-USMARC-Parent-DQ786758-rs29024430	0.302	0.307
ARS-USMARC-Parent-DQ786761-rs29012840	0.345	0.373
ARS-USMARC-Parent-DQ786762-rs29010772	0.229	0.497
ARS-USMARC-Parent-DQ786763-rs29020472	0.272	0.323
ARS-USMARC-Parent-DQ786764-no-rs	0.153	0.072
ARS-USMARC-Parent-DQ786765-rs29009858	0.471	0.493
ARS-USMARC-Parent-DQ786766-rs29012070	0.457	0.454
ARS-USMARC-Parent-DQ789028-rs29017713	0.241	0.389
ARS-USMARC-Parent-DQ837643-rs29018818	0.274	0.052
ARS-USMARC-Parent-DQ837644-rs29010468	0.371	0.414
ARS-USMARC-Parent-DQ837645-rs29015870	0.258	0.381
ARS-USMARC-Parent-DQ839235-rs29012691	0.442	0.459
ARS-USMARC-Parent-DQ846688-rs29023691	0.269	0.109
ARS-USMARC-Parent-DQ846690-no-rs	0.297	0.285
ARS-USMARC-Parent-DQ846691-rs29019814	0.370	0.466
ARS-USMARC-Parent-DQ846693-rs29017621	0.428	0.447
ARS-USMARC-Parent-DQ866817-no-rs	0.281	0.491
ARS-USMARC-Parent-DQ866818-rs29011701	0.479	0.461
ARS-USMARC-Parent-DQ888309-rs29013741	0.300	0.298
ARS-USMARC-Parent-DQ888310-rs29012422	0.226	0.471
ARS-USMARC-Parent-DQ888311-rs29017313	0.222	0.403
ARS-USMARC-Parent-DQ888313-no-rs	0.213	0.476
ARS-USMARC-Parent-DQ916057-rs29009979	0.269	0.328
ARS-USMARC-Parent-DQ916058-rs29016146	0.388	0.497
ARS-USMARC-Parent-DQ916059-rs29009907	0.313	0.403
ARS-USMARC-Parent-DQ984825-rs29012457	0.183	0.279
ARS-USMARC-Parent-DQ984826-rs29027559	0.234	0.479
ARS-USMARC-Parent-DQ984827-rs29012019	0.327	0.493
ARS-USMARC-Parent-DQ990832-rs29015065	0.259	0.100
ARS-USMARC-Parent-DQ990833-rs29010147	0.446	0.378
ARS-USMARC-Parent-DQ990834-rs29013727	0.215	0.110
ARS-USMARC-Parent-DQ995976-no-rs	0.449	0.409
ARS-USMARC-Parent-DQ995977-rs29020834	0.269	0.278
ARS-USMARC-Parent-EF026085-rs29021607	0.272	0.452
ARS-USMARC-Parent-EF026086-rs29013660	0.378	0.129
ARS-USMARC-Parent-EF026087-rs29011643	0.428	0.485
ARS-USMARC-Parent-EF028073-rs29014953	0.216	0.259
ARS-USMARC-Parent-EF034080-rs29024749	0.374	0.474
ARS-USMARC-Parent-EF034081-rs29009668	0.431	0.385
ARS-USMARC-Parent-EF034082-rs29013532	0.336	0.493
ARS-USMARC-Parent-EF034083-rs29018286	0.363	0.338
ARS-USMARC-Parent-EF034084-rs29016185	0.146	0.370
ARS-USMARC-Parent-EF034085-rs29025677	0.305	0.316

ARS-USMARC-Parent-EF034086-no-rs	0.409	0.253
ARS-USMARC-Parent-EF034087-no-rs	0.005	0.105
ARS-USMARC-Parent-EF042090-no-rs	0.483	0.369
ARS-USMARC-Parent-EF042091-rs29014974	0.129	0.420
ARS-USMARC-Parent-EF089234-rs29020870	0.176	0.361
ARS-USMARC-Parent-EF093509-rs29015170	0.339	0.231
ARS-USMARC-Parent-EF093511-rs29012316	0.441	0.484
ARS-USMARC-Parent-EF141102-rs29015783	0.419	0.419
ARS-USMARC-Parent-EF150946-rs29023666	0.284	0.261
BTA-100621-no-rs	0.378	0.378
BTA-11701-rs29017459	0.181	0.494
BTA-30857-no-rs	0.224	0.489
BTA-37062-no-rs	0.391	0.447
BTA-73768-no-rs	0.344	0.446
BTA-92021-no-rs	0.369	0.305
BTB-00188171	0.318	0.474
BTB-00394801	0.080	0.120
BTB-00420215	0.177	0.206
BTB-01057979	0.428	0.396
BTB-01285245	0.434	0.458
BTB-01371672	0.440	0.269
BTB-01416427	0.468	0.458
BTB-01478115	0.339	0.462
BTB-01980499	0.392	0.406
Hapmap24215-BTA-163266	0.373	0.397
Hapmap31098-BTA-136127	0.346	0.374
Hapmap34424-BES10_Contig566_926	0.258	0.346
Hapmap36588-SCAFFOLD90561_9460	0.322	0.308
Hapmap39425-BTA-70290	0.456	0.292
Hapmap39461-BTA-109898	0.428	0.466
Hapmap40148-BTA-92999	0.377	0.326
Hapmap40729-BTA-40319	0.299	0.418
Hapmap41591-BTA-59790	0.195	0.475
Hapmap42648-BTA-71195	0.367	0.362
Hapmap43057-BTA-80741	0.498	0.409
Hapmap43142-BTA-107561	0.331	0.495
Hapmap43792-BTA-122725	0.322	0.243
Hapmap43953-BTA-83292	0.283	0.489
Hapmap46550-BTA-103548	0.145	0.366
Hapmap46653-BTA-47447	0.382	0.448
Hapmap47281-BTA-40051	0.344	0.328
Hapmap49452-BTA-112834	0.220	0.158
Hapmap50598-BTA-122724	0.334	0.327
Hapmap51227-BTA-41809	0.146	0.152
Hapmap51527-BTA-97415	0.415	0.278
Hapmap51908-BTA-63031	0.430	0.491

Hapmap52240-rs29013844	0.451	0.448
Hapmap54020-rs29023153	0.310	0.168
Hapmap54313-rs29012632	0.400	0.339
Hapmap54547-rs29012198	0.362	0.495
Hapmap55441-rs29010990	0.212	0.261
Hapmap59876-rs29018046	0.433	0.450
Hapmap60017-rs29023471	0.296	0.473
UA-IFASA-5034	0.375	0.452
UA-IFASA-6532	0.329	0.255
AVERAGE	0.331	0.359

Addendum 2 Number of SNPs genotyped and left after quality control

Breeds	Initial SNPs	SNPs removed based on:						SNPs remained after QC	SNPs lost after merging datasets	Final SNPs
		Unknown position	GC Score (<0.6)	GT Score (<0.55)	Call rate (<95%)	MAF (<0.05)	HWE (p<0.001)			
Bonsmara	138 888	19 513	6 527	0	2 826	10 711	611	98 617	5 782	92 835
Drakensberger	138 888	19 513	8 677	11	1 271	7 439	13	101 964	9 129	92 835

Addendum 3 R Script used in Chapter 4 for quality control

Removal of SNPs without positions, on sex chromosome and mitochondrial DNA

genotype_file<- genotype_file [!(genotype_file $Chr==0),]

genotype_file <- genotype_file [!(genotype_file $Chr=="Y"),]

genotype_file <- genotype_file [!(genotype_file $Chr=="X"),]

genotype_file <- genotype_file [!(genotype_file $Chr=="MT"),]

###SNP MAF calculation

genotype_fileFreq1<-subset(count(genotype_file, c("SNPName", "AlleleAB_Ani")), (AlleleAB_Ani=="AA" | AlleleAB_Ani=="AB" | AlleleAB_Ani=="BB"))

SNPFreq_tot<-aggregate(freq ~ SNPName, data = genotype_fileFreq1, sum)

colnames(SNPFreq_tot)[colnames(SNPFreq_tot)=='freq']<-"Total_occurance"

SNPFreq_tot$Total_occurance<-SNPFreq_tot$Total_occurance*2

SNPFreqAA<-subset(genotype_fileFreq1, AlleleAB_Ani=="AA")

SNPFreqAB<-subset(genotype_fileFreq1, AlleleAB_Ani=="AB")

SNPFreqBB<-subset(genotype_fileFreq1, AlleleAB_Ani=="BB")

SNPFreq_AA<-aggregate(freq ~ SNPName, data = SNPFreqAA, sum)

```
colnames(SNPFreq_AA)[colnames(SNPFreq_AA)=='freq']<-"AA_occurance"
SNPFreq_AB<-aggregate(freq ~ SNPName, data = SNPFreqAB, sum)
colnames(SNPFreq_AB)[colnames(SNPFreq_AB)=='freq']<-"AB_occurance"
SNPFreq_BB<-aggregate(freq ~ SNPName, data = SNPFreqBB, sum)
colnames(SNPFreq_BB)[colnames(SNPFreq_BB)=='freq']<-"BB_occurance"
SNPFreq_tot<-merge(SNPFreq_tot,SNPFreq_AA, by ="SNPName", all=TRUE)
SNPFreq_tot<-merge(SNPFreq_tot,SNPFreq_AB, by ="SNPName", all=TRUE)
SNPFreq_tot<-merge(SNPFreq_tot,SNPFreq_BB, by ="SNPName", all=TRUE)
SNPFreq_tot[is.na(SNPFreq_tot)]<-0
SNPFreq_tot$A_Freq<-round((((SNPFreq_tot$AA_occurance*2)+(SNPFreq_tot$AB_occurance))/SNPFreq_tot$Total_occurance), digits=3)
SNPFreq_tot$B_Freq<-round((((SNPFreq_tot$BB_occurance*2)+(SNPFreq_tot$AB_occurance))/SNPFreq_tot$Total_occurance), digits=3)
SNPFreq_tot$Minor_Allele<-ifelse(SNPFreq_tot$A_Freq<SNPFreq_tot$B_Freq, "A","B")
SNPFreq_tot$Minor_Freq<-ifelse(SNPFreq_tot$A_Freq<SNPFreq_tot$B_Freq,SNPFreq_tot$A_Freq,SNPFreq_tot$B_Freq)
Drb_MAF<-SNPFreq_tot[order(SNPFreq_tot$Minor_Freq),]
Drb_MAF<-subset(SNPFreq_tot, select=c("SNPName", "Minor_Freq"))
write.csv(Drb_MAF, file = "Drb_MAF.csv",row.names=FALSE)

###SNP callrate calculation
SNP_Callrate1<- subset(genotype_file, select=c("SNPName", "AlleleAB_Ani"))
SNP_Callrate1Freq <- count(SNP_Callrate1, c("SNPName", "AlleleAB_Ani"))
SNP_Callrate1Freq_tot <- aggregate(freq ~  SNPName, data = SNP_Callrate1Freq, sum)
colnames(SNP_Callrate1Freq_tot)[colnames(SNP_Callrate1Freq_tot)=="freq"] <- "Total_occurance"
SNP_Callrate1Freq1 <- subset(SNP_Callrate1Freq, (AlleleAB_Ani == "AA" | AlleleAB_Ani == "AB" | AlleleAB_Ani == "BB"))
SNP_Callrate1Freq_call <- aggregate(freq ~  SNPName, data = SNP_Callrate1Freq1, sum)
```

```
colnames(SNP_Callrate1Freq_call)[colnames(SNP_Callrate1Freq_call)=="freq"] <- "calls"
SNP_Callrate1Freq_tot <- merge(SNP_Callrate1Freq_tot, SNP_Callrate1Freq_call, by ="SNPName", all = TRUE)
SNP_Callrate1Freq_tot$call_rate <- round((((SNP_Callrate1Freq_tot$calls)/(SNP_Callrate1Freq_tot$Total_occurance))), digits=3)
SNP_Callrate1Freq_tot<-subset(SNP_Callrate1Freq_tot, select=c("SNPName", "call_rate"))
write.csv(SNP_Callrate1Freq_tot, file = "Drb_callrate.csv",row.names=FALSE)

###Calculation of Hardy Weinberg Equilibrium, Chi SquareDist and P_value
genotype_fileFreq1<-subset(count(genotype_file, c("SNPName", "AlleleAB_Ani")), (AlleleAB_Ani=="AA" | AlleleAB_Ani=="AB" | AlleleAB_Ani=="BB"))
SNPFreq_tot<-aggregate(freq ~ SNPName, data = genotype_fileFreq1, sum)
colnames(SNPFreq_tot)[colnames(SNPFreq_tot)=='freq']<-"Total_occurance"
SNPFreq_tot$Total_occurance<-SNPFreq_tot$Total_occurance*2
SNPFreqAA<-subset(genotype_fileFreq1, AlleleAB_Ani=="AA")
SNPFreqAB<-subset(genotype_fileFreq1, AlleleAB_Ani=="AB")
SNPFreqBB<-subset(genotype_fileFreq1, AlleleAB_Ani=="BB")
SNPFreq_AA<-aggregate(freq ~ SNPName, data = SNPFreqAA, sum)
colnames(SNPFreq_AA)[colnames(SNPFreq_AA)=='freq']<-"AA_occurance"
SNPFreq_AB<-aggregate(freq ~ SNPName, data = SNPFreqAB, sum)
colnames(SNPFreq_AB)[colnames(SNPFreq_AB)=='freq']<-"AB_occurance"
SNPFreq_BB<-aggregate(freq ~ SNPName, data = SNPFreqBB, sum)
colnames(SNPFreq_BB)[colnames(SNPFreq_BB)=='freq']<-"BB_occurance"
SNPFreq_tot<-merge(SNPFreq_tot,SNPFreq_AA, by ="SNPName", all=TRUE)
SNPFreq_tot<-merge(SNPFreq_tot,SNPFreq_AB, by ="SNPName", all=TRUE)
SNPFreq_tot<-merge(SNPFreq_tot,SNPFreq_BB, by ="SNPName", all=TRUE)
```

```
SNPFreq_tot[is.na(SNPFreq_tot)]<-0

SNPFreq_tot$A_Freq<-round((((SNPFreq_tot$AA_occurance*2)+(SNPFreq_tot$AB_occurance))/SNPFreq_tot$Total_occurance), digits=3)

SNPFreq_tot$B_Freq<-round((((SNPFreq_tot$BB_occurance*2)+(SNPFreq_tot$AB_occurance))/SNPFreq_tot$Total_occurance), digits=3)

SNPFreq_tot$Number_Individuals<- (SNPFreq_tot$Total_occurance)/2

SNPFreq_tot$AA_ExpFreq<-(SNPFreq_tot$A_Freq)^2

SNPFreq_tot$AB_ExpFreq<-2*(SNPFreq_tot$A_Freq*SNPFreq_tot$B_Freq)

SNPFreq_tot$BB_ExpFreq<-(SNPFreq_tot$B_Freq)^2

SNPFreq_tot$AA_Exp<- (SNPFreq_tot$AA_ExpFreq*SNPFreq_tot$Number_Individuals)

SNPFreq_tot$AB_Exp<- (SNPFreq_tot$AB_ExpFreq*SNPFreq_tot$Number_Individuals)

SNPFreq_tot$BB_Exp<- (SNPFreq_tot$BB_ExpFreq*SNPFreq_tot$Number_Individuals)

SNPFreq_tot$ChiSq_Value<-((SNPFreq_tot$AA_occurance-SNPFreq_tot$AA_Exp)^2/SNPFreq_tot$AA_Exp)+((SNPFreq_tot$AB_occurance-
SNPFreq_tot$AB_Exp)^2/SNPFreq_tot$AB_Exp)+((SNPFreq_tot$BB_occurance-SNPFreq_tot$BB_Exp)^2/SNPFreq_tot$BB_Exp)

SNPFreq_tot$P_value<-pchisq(SNPFreq_tot$ChiSq_Value, df=1, lower.tail=FALSE)

SNPFreq_tot<- subset(SNPFreq_tot, select=c("SNPName","P_value"))

write.csv(SNPFreq_tot, file = "Pvalue",row.names=FALSE)

###Quality control

BonQC_fileGC<-BonQC_file[BonQC_file$GC_Score>=0.6,]

BonQC_fileGT<-BonQC_fileGC[BonQC_fileGC$GT_Score>=0.55,]

BonQC_fileCR<-BonQC_fileGT[BonQC_fileGT$call_rate>=0.95,]

BonQC_fileMAF<-BonQC_fileCR[BonQC_fileCR$Minor_Freq>=0.05,]

BonQC_fileHWE<-BonQC_fileMAF[BonQC_fileMAF$P_value>0.001,]


write.csv(BonQC_fileCR, file = "BonQC_fileCR.csv",row.names=FALSE)

write.csv(BonQC_fileGC, file = "BonQC_fileGC.csv",row.names=FALSE)
```

```
write.csv(BonQC_fileGT, file = "BonQC_fileGT.csv",row.names=FALSE)
write.csv(BonQC_fileHWE, file = "BonQC_fileHWE.csv",row.names=FALSE)
write.csv(BonQC_fileMAF, file = "BonQC_fileMAF.csv",row.names=FALSE)

DrbQC_fileGC<-DrbQC_file[DrbQC_file$GC_Score>=0.6,]
DrbQC_fileGT<-DrbQC_fileGC[DrbQC_fileGC$GT_Score>=0.55,]
DrbQC_fileCR<-DrbQC_fileGT[DrbQC_fileGT$call_rate>=0.95,]
DrbQC_fileMAF<-DrbQC_fileCR[DrbQC_fileCR$Minor_Freq>=0.05,]
DRbQC_fileHWE<-DrbQC_fileMAF[DrbQC_fileMAF$P_value>0.001,]

###Merge files
QC_fileMerged<- merge(BonQC_fileHWE, DrbQC_fileHWE, by ="SNPName", all = TRUE)
QCSNPCommon_fileMerged<- merge(BonQC_fileHWE, DrbQC_fileHWE, by ="SNPName")
```

Addendum 4 Used Script in R for parentage validation in Chapter 4

```
####Import data parent offspring file (paroff_file)

###Count the number of opposing homozygotes

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AA" &   paroff_file$AlleleAB_parent == "AA" ,1,0)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AA" &   paroff_file$AlleleAB_parent == "BB" &   paroff_file$Mismatch == 0  ,-1,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="BB" &   paroff_file$AlleleAB_parent == "BB" &   paroff_file$Mismatch == 0  ,1,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="BB" &   paroff_file$AlleleAB_parent == "AA" &   paroff_file$Mismatch == 0  ,-1,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NA" &   paroff_file$AlleleAB_parent == "AA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NA" &   paroff_file$AlleleAB_parent == "BB" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AA" &   paroff_file$AlleleAB_parent == "NA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="BB" &   paroff_file$AlleleAB_parent == "NA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NA" &   paroff_file$AlleleAB_parent == "NA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NANA" &   paroff_file$AlleleAB_parent == "NANA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NANA" &   paroff_file$AlleleAB_parent == "AA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)
```

```
paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NANA" &   paroff_file$AlleleAB_parent == "BB" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AA" &   paroff_file$AlleleAB_parent == "NANA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="BB" &   paroff_file$AlleleAB_parent == "NANA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="0" &   paroff_file$AlleleAB_parent == "AA" &   paroff_file$Mismatch == 0  ,-9, paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="0" &   paroff_file$AlleleAB_parent == "BB" &   paroff_file$Mismatch == 0  ,-9, paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AA" &   paroff_file$AlleleAB_parent == "0" &   paroff_file$Mismatch == 0  ,-9, paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="BB" &   paroff_file$AlleleAB_parent == "0" &   paroff_file$Mismatch == 0  ,-9, paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="0" &   paroff_file$AlleleAB_parent == "0" &   paroff_file$Mismatch == 0  ,-9, paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AB" &   paroff_file$AlleleAB_parent == "AB" &   paroff_file$Mismatch == 0  ,9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AB" &   paroff_file$AlleleAB_parent == "NA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AB" &   paroff_file$AlleleAB_parent == "NANA" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AB" &   paroff_file$AlleleAB_parent == "0" &   paroff_file$Mismatch == 0  ,-9, paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NANA" &   paroff_file$AlleleAB_parent == "AB" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="NA" &   paroff_file$AlleleAB_parent == "AB" &   paroff_file$Mismatch == 0  ,-9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="0" &   paroff_file$AlleleAB_parent == "AB" &   paroff_file$Mismatch == 0  ,-9, paroff_file$Mismatch)
```

```
paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AB" &   paroff_file$AlleleAB_parent == "AA" &   paroff_file$Mismatch == 0  ,9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AB" &   paroff_file$AlleleAB_parent == "BB" &   paroff_file$Mismatch == 0  ,9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="AA" &   paroff_file$AlleleAB_parent == "AB" &   paroff_file$Mismatch == 0  ,9,
paroff_file$Mismatch)

paroff_file$Mismatch <-ifelse( paroff_file$AlleleAB_Offspring =="BB" &   paroff_file$AlleleAB_parent == "AB" &   paroff_file$Mismatch == 0  ,9,
paroff_file$Mismatch)

paroff_file[is.na( paroff_file)]<-0

paroff_file_correct<- filter( paroff_file, Mismatch ==1)

paroff_file_wrong<- filter( paroff_file, Mismatch ==-1)

paroff_file_correctAB<- filter( paroff_file, Mismatch ==9)

paroff_file_MissingG<- filter( paroff_file, Mismatch ==-9)

count( paroff_file,c('Mismatch'))

paroff_file$Mismatch[ paroff_file$Mismatch %in% "0"] <- "-9"

count( paroff_file,c('Mismatch'))

paroff_file_correct[is.na( paroff_file_correct)]<-0

 paroff_file_wrong[is.na( paroff_file_wrong)]<-0

 paroff_file_correctAB[is.na( paroff_file_correctAB)]<-0

 paroff_file_MissingG[is.na( paroff_file_MissingG)]<-0

Offspringmal_parent_Correct <-aggregate(cbind(count = (Mismatch)) ~ Offspringmal_ID + parent_ID, data =  paroff_file_correct, FUN = function(x){NROW(x)})
```

```
Offspringmal_parent_Wrong <-aggregate(cbind(count = (Mismatch)) ~ Offspringmal_ID +parent_ID, data =  paroff_file_wrong, FUN = function(x){NROW(x)})
Offspringmal_parent_CorrectAB <-aggregate(cbind(count = (Mismatch)) ~ Offspringmal_ID +parent_ID, data =  paroff_file_correctAB, FUN = function(x){NROW(x)})
Offspringmal_parent_MissingG <-aggregate(cbind(count = (Mismatch)) ~ Offspringmal_ID + parent_ID, data =  paroff_file_MissingG, FUN = function(x){NROW(x)})
Offspringmal_parent_Check<-merge(Offspringmal_parent_Correct,Offspringmal_parent_Wrong,by=c("Offspringmal_ID","parent_ID"),all=TRUE)
setnames(Offspringmal_parent_Check,"count.x","SNP_Correct")
setnames(Offspringmal_parent_Check,"count.y","SNP_Wrong")
Offspringmal_parent_Check<-merge(Offspringmal_parent_Check,Offspringmal_parent_CorrectAB,by=c("Offspringmal_ID","parent_ID"),all=TRUE)
setnames(Offspringmal_parent_Check,"count","SNP_CorrectAB")
Offspringmal_parent_Check<-merge(Offspringmal_parent_Check,Offspringmal_parent_MissingG,by=c("Offspringmal_ID","parent_ID"),all=TRUE)
setnames(Offspringmal_parent_Check,"count","SNP_Missing")
library(data.table)
library(dplyr)
library(plyr)
Offspringmal_parent_Check[is.na(Offspringmal_parent_Check)]<-0
Offspringmal_parent_Check$Proportion<-(Offspringmal_parent_Check$SNP_Wrong/Offspringmal_parent_Check$SNP_Correct)
Offspringmal_parent_Check$Percent<-(Offspringmal_parent_Check$Proportion*100)
Offspringmal_parent_Check$Percent<-round(Offspringmal_parent_Check$Percent, 3)
Offspringmal_parent_Check$Matches<-round(100-Offspringmal_parent_Check$Percent, 3)
```

```
write.csv(Offspringmal_parent_Check, file = "Offspringmal_parent_Check",row.names=FALSE)

write.csv(Offspringmal_parent_Correct, file = "Offspringmal_parent_Correct",row.names=FALSE)

write.csv(Offspringmal_parent_Wrong, file = "Offspringmal_parent_Wrong",row.names=FALSE)

write.csv(Offspringmal_parent_MissingG, file = "Offspringmal_parent_MissingG",row.names=FALSE)

write.csv( paroff_file, file = " paroff_file_Mismatch",row.names=FALSE)

write.csv(Column_mean, file = "Column_Mean",row.names=FALSE)

paroff_file_SNPwrong<-count( paroff_file_wrong,"SNPName")

write.csv( paroff_file_SNPwrong, file = " paroff_file_SNPwrong",row.names=FALSE)

paroff_file_SNPcorrect<-count( paroff_file_correct,"SNPName")

write.csv( paroff_file_SNPcorrect, file = " paroff_file_SNPcorrect",row.names=FALSE)

paroff_file_SNPMissingG<-count( paroff_file_MissingG,"SNPName")

write.csv( paroff_file_SNPMissingG, file = " paroff_file_SNPMissingG",row.names=FALSE)

Wrong_parentages<-Offspringmal_parent_Check[Offspringmal_parent_Check$Matches<99,]

write.csv(Wrong_parentages, file = "Wrong_parentages",row.names=FALSE)

Correct_parentages<-Offspringmal_parent_Check[Offspringmal_parent_Check$Matches>=99,]

write.csv(Correct_parentages, file = "Correct_parentages",row.names=FALSE)
```

Addendum 5 Supplementary table of the 200 SNPs selected for the BON and DRB

BON							DRB				
SNPName	**Chr**	**Position**	**GC_Score**	**GT_Score**	**Callrate**	**MAF**	**SNPName**	**GC_SCore**	**GT_Score**	**Callrate**	**MAF**
ARS-BFGL-NGS-105113	1	7943430	0.78	0.77	1.00	0.48	BovineHD0100001444	0.88	0.86	1.00	0.47
Hapmap48613-BTA-112066	1	17296638	0.75	0.76	1.00	0.48	BovineHD0100005349	0.89	0.86	0.99	0.48
BovineHD0100010461	1	36547637	0.86	0.85	1.00	0.49	BovineHD0100011399	0.76	0.77	1.00	0.48
BovineHD0100012920	1	45230725	0.83	0.81	1.00	0.49	BTB-00132601	0.81	0.80	1.00	0.48
BTB-00033025	1	64667900	0.80	0.77	0.99	0.49	BovineHD0100016412	0.81	0.80	1.00	0.49
BovineHD0100021450	1	74713515	0.86	0.83	1.00	0.49	BovineHD0100021004	0.78	0.78	1.00	0.49
BovineHD0100026611	1	93844902	0.87	0.85	1.00	0.50	BovineHD0100023752	0.72	0.74	1.00	0.47
BovineHD0100027939	1	97850386	0.89	0.85	1.00	0.48	BovineHD0100027429	0.76	0.75	1.00	0.48
BTB-01877866	1	119446658	0.71	0.73	1.00	0.49	BovineHD0100032409	0.76	0.77	1.00	0.48
BTB-01662109	1	130217213	0.72	0.73	0.99	0.48	BovineHD0100037619	0.92	0.89	1.00	0.50
BovineHD0100037698	1	132347489	0.89	0.86	0.99	0.48	BovineHD0100039458	0.87	0.85	1.00	0.49
BovineHD0100042553	1	147185998	0.78	0.78	1.00	0.48	BovineHD0100042713	0.85	0.83	1.00	0.47
BovineHD0200001065	2	3966227	0.70	0.73	0.99	0.49	BovineHD0200002454	0.86	0.84	1.00	0.49
BovineHD0200005509	2	19128970	0.77	0.79	1.00	0.49	BovineHD0200004419	0.81	0.80	1.00	0.48
ARS-USMARC-Parent-AY776154-no-rs	2	26997623	0.76	0.76	1.00	0.48	BovineHD0200007776	0.78	0.78	1.00	0.48
BovineHD0200012044	2	41477309	0.84	0.81	1.00	0.49	BovineHD0200014124	0.88	0.85	1.00	0.48
BovineHD0200017638	2	61389361	0.92	0.89	1.00	0.50	BovineHD0200016608	0.80	0.78	1.00	0.48
Hapmap41723-BTA-101688	2	68386619	0.87	0.84	1.00	0.49	Hapmap50002-BTA-47754	0.88	0.86	1.00	0.48
BovineHD0200025799	2	90714173	0.78	0.77	1.00	0.48	BovineHD0200021155	0.76	0.77	1.00	0.48
BovineHD0200029130	2	101510681	0.92	0.88	1.00	0.49	BovineHD0200027202	0.91	0.88	1.00	0.48
BovineHD0200031349	2	108789552	0.89	0.87	1.00	0.49	BovineHD0200031587	0.88	0.85	1.00	0.47
BovineHD0200034687	2	119777560	0.74	0.75	1.00	0.48	ARS-BFGL-NGS-19012	0.93	0.90	1.00	0.49
BovineHD0200039118	2	134132026	0.92	0.88	1.00	0.50	BovineHD0200040001	0.93	0.92	1.00	0.48
BovineHD0300000803	3	2811753	0.89	0.85	1.00	0.50	BovineHD0300001128	0.75	0.75	1.00	0.48
BovineHD0300007257	3	23212939	0.69	0.72	1.00	0.50	BovineHD0300008621	0.80	0.79	1.00	0.48
ARS-BFGL-NGS-113072	3	30853504	0.66	0.70	0.99	0.49	ARS-BFGL-NGS-109648	0.74	0.75	1.00	0.48

BovineHD0300012473	3	40968839	0.79	0.79	1.00	0.49	BovineHD0300011769	0.84	0.82	0.99	0.48
ARS-USMARC-Parent-DQ435443-rs29010802	3	58040470	0.81	0.79	1.00	0.47	BovineHD0300017284	0.77	0.77	0.99	0.48
BovineHD0300017797	3	59208089	0.75	0.78	0.99	0.49	BTB-00130588	0.90	0.88	1.00	0.48
ARS-BFGL-NGS-57959	3	88579850	0.80	0.76	1.00	0.49	BTA-68587-no-rs	0.65	0.69	1.00	0.48
Hapmap47494-BTA-108688	3	91583640	0.80	0.78	1.00	0.50	Hapmap30329-BTA-163245	0.89	0.86	1.00	0.47
BovineHD0300029039	3	101489726	0.80	0.78	1.00	0.48	ARS-BFGL-NGS-43577	0.66	0.70	0.99	0.48
BovineHD0300033867	3	116623778	0.80	0.79	1.00	0.49	BovineHD0300034518	0.77	0.77	1.00	0.49
BovineHD0400001652	4	5822335	0.90	0.88	1.00	0.50	BovineHD0400001472	0.78	0.78	1.00	0.49
BovineHD0400003686	4	12354572	0.80	0.78	1.00	0.50	BovineHD0400006340	0.84	0.82	0.99	0.48
Hapmap59011-rs29027498	4	26490406	0.83	0.82	1.00	0.50	BovineHD0400008091	0.78	0.78	1.00	0.48
BovineHD0400014174	4	51202778	0.80	0.78	1.00	0.50	BovineHD0400012296	0.91	0.88	0.99	0.48
BTB-00188171	4	57787437	0.90	0.85	1.00	0.44	BovineHD0400016186	0.82	0.81	1.00	0.47
BovineHD0400020147	4	72740198	0.76	0.76	1.00	0.50	ARS-BFGL-NGS-21842	0.85	0.82	1.00	0.48
ARS-BFGL-NGS-19836	4	82671406	0.89	0.85	1.00	0.50	ARS-BFGL-NGS-92104	0.69	0.71	0.99	0.48
Hapmap48060-BTA-71676	4	95645699	0.72	0.74	1.00	0.50	BovineHD0400028397	0.88	0.86	1.00	0.48
BTA-98250-no-rs	4	117913533	0.82	0.79	1.00	0.49	BovineHD0400032240	0.82	0.81	1.00	0.47
BovineHD0500002037	5	7256354	0.76	0.76	1.00	0.50	BovineHD0500001593	0.86	0.84	1.00	0.49
ARS-BFGL-NGS-5790	5	26986116	0.71	0.74	1.00	0.50	BovineHD0500006625	0.80	0.79	1.00	0.48
BovineHD0500011267	5	39350451	0.88	0.84	1.00	0.50	ARS-USMARC-635	0.81	0.80	1.00	0.49
BovineHD0500013064	5	45373404	0.79	0.78	1.00	0.49	Hapmap58641-rs29010249	0.81	0.80	0.99	0.47
Hapmap27803-BTA-66694	5	65095289	0.81	0.79	1.00	0.50	Hapmap34023-BES4_Contig227_995	0.81	0.80	1.00	0.48
BovineHD0500020558	5	72710101	0.89	0.85	1.00	0.50	BovineHD0500020690	0.89	0.86	1.00	0.48
BovineHD0500026293	5	92576706	0.81	0.79	1.00	0.49	BovineHD0500022134	0.83	0.83	1.00	0.48
BovineHD0500027355	5	96342660	0.75	0.75	1.00	0.50	BovineHD0500029644	0.83	0.82	1.00	0.49
BovineHD0500030651	5	106744180	0.75	0.75	1.00	0.50	BovineHD0500030709	0.76	0.77	1.00	0.49
Hapmap41631-BTA-75177	5	114405063	0.78	0.77	1.00	0.50	BovineHD0500032001	0.93	0.90	1.00	0.49
BovineHD0600000384	6	1530938	0.77	0.77	0.99	0.50	BTB-01454593	0.81	0.80	1.00	0.48
BovineHD0600005392	6	19429113	0.88	0.85	1.00	0.50	BTB-00247385	0.87	0.85	1.00	0.47
ARS-BFGL-NGS-103009	6	34889220	0.92	0.88	1.00	0.50	BovineHD0600009168	0.75	0.76	1.00	0.48

BTB-00260450	6	40063618	0.92	0.89	1.00	0.50	BovineHD0600010587	0.74	0.75	0.99	0.48
BovineHD4100004791	6	42908206	0.71	0.73	0.99	0.50	BTB-00251655	0.71	0.73	0.99	0.48
BovineHD0600016358	6	59523327	0.75	0.75	0.99	0.50	BovineHD0600013744	0.86	0.85	1.00	0.47
BovineHD0600019408	6	70145579	0.83	0.82	1.00	0.49	BovineHD0600019712	0.76	0.76	1.00	0.49
ARS-BFGL-NGS-40151	6	92690455	0.70	0.72	1.00	0.49	BovineHD0600026212	0.70	0.73	1.00	0.48
BovineHD0600029135	6	104443533	0.69	0.72	1.00	0.50	BovineHD0600029383	0.82	0.80	1.00	0.48
BovineHD0600031495	6	111365596	0.82	0.79	1.00	0.49	BovineHD0600031543	0.81	0.80	1.00	0.48
BovineHD0700001189	7	4186426	0.74	0.74	1.00	0.49	BovineHD0700003344	0.78	0.78	1.00	0.48
BovineHD0700003400	7	13103114	0.85	0.82	1.00	0.49	BovineHD0700003428	0.85	0.83	1.00	0.48
BovineHD0700006502	7	23729977	0.91	0.86	1.00	0.50	BovineHD0700007504	0.76	0.77	1.00	0.47
BovineHD0700010751	7	37222679	0.77	0.77	1.00	0.49	ARS-BFGL-NGS-110899	0.66	0.69	1.00	0.49
BovineHD0700017949	7	62208719	0.70	0.73	1.00	0.50	BovineHD0700017295	0.88	0.85	1.00	0.48
BovineHD0700018956	7	65132003	0.70	0.73	0.99	0.49	BovineHD0700018930	0.71	0.73	1.00	0.48
BovineHD0700024081	7	82451796	0.75	0.75	1.00	0.49	BovineHD0700024108	0.66	0.70	0.99	0.48
BovineHD0700027735	7	95137796	0.74	0.75	1.00	0.50	BTA-21118-no-rs	0.84	0.82	1.00	0.47
BovineHD0700032856	7	112516157	0.92	0.88	1.00	0.50	BovineHD0700030547	0.93	0.89	1.00	0.48
BovineHD0800000174	8	640491	0.77	0.77	1.00	0.50	BovineHD0800000174	0.78	0.78	1.00	0.49
BTB-01369338	8	19967642	0.88	0.84	1.00	0.50	BovineHD0800003482	0.77	0.77	0.99	0.49
BovineHD0800007360	8	24491223	0.77	0.77	1.00	0.50	BovineHD0800007246	0.84	0.83	1.00	0.48
BTB-01721163	8	43202753	0.68	0.71	0.99	0.50	BovineHD0800012119	0.91	0.88	1.00	0.48
BovineHD0800016797	8	55710550	0.80	0.79	0.99	0.50	BovineHD0800013594	0.86	0.83	1.00	0.47
ARS-BFGL-NGS-111488	8	65569786	0.88	0.85	1.00	0.50	Hapmap38451-BTA-97230	0.85	0.84	1.00	0.49
BTB-00358369	8	75347512	0.87	0.84	0.99	0.50	BovineHD0800022187	0.76	0.77	1.00	0.48
ARS-BFGL-NGS-73943	8	89707967	0.88	0.83	1.00	0.50	BovineHD0800025413	0.76	0.77	1.00	0.49
Hapmap57802-rs29009571	8	104383991	0.76	0.77	1.00	0.50	BTA-82825-no-rs	0.76	0.76	1.00	0.49
BovineHD0900002904	9	11535001	0.78	0.78	1.00	0.50	Hapmap36664-SCAFFOLD50340_7682	0.88	0.85	1.00	0.47
ARS-BFGL-NGS-14740	9	14436985	0.89	0.85	1.00	0.47	BovineHD0900005705	0.78	0.78	1.00	0.48
BovineHD0900009028	9	33019902	0.89	0.85	1.00	0.50	BovineHD0900009546	0.85	0.82	1.00	0.47
BovineHD0900013887	9	50418639	0.67	0.71	0.99	0.50	BovineHD0900010463	0.89	0.86	1.00	0.47

BovineHD0900015752	9	57672149	0.87	0.83	1.00	0.50	BovineHD0900016490	0.90	0.87	1.00	0.49
ARS-BFGL-NGS-60933	9	62907869	0.76	0.76	1.00	0.50	Hapmap58190-rs29022461	0.92	0.89	1.00	0.48
BovineHD0900024168	9	86290150	0.81	0.79	1.00	0.50	ARS-BFGL-NGS-10386	0.85	0.83	1.00	0.48
ARS-BFGL-NGS-100350	9	97760489	0.73	0.75	1.00	0.50	ARS-BFGL-NGS-118060	0.64	0.70	0.99	0.49
BovineHD0900030448	9	103788327	0.86	0.83	1.00	0.49	BovineHD0900028282	0.81	0.79	0.99	0.48
BovineHD1000031115	10	5154634	0.89	0.86	1.00	0.49	BovineHD1000002593	0.83	0.81	1.00	0.48
BovineHD1000005684	10	16975162	0.76	0.76	1.00	0.50	BovineHD1000005577	0.74	0.76	0.99	0.48
BovineHD1000007311	10	22442359	0.64	0.70	1.00	0.50	BovineHD1000006603	0.90	0.88	1.00	0.48
BovineHD1000011295	10	36041911	0.85	0.81	1.00	0.50	BovineHD1000012212	0.85	0.84	1.00	0.48
BovineHD1000016604	10	55925288	0.92	0.88	1.00	0.50	BovineHD1000016688	0.85	0.83	1.00	0.47
BovineHD1000018336	10	63412898	0.85	0.83	1.00	0.50	BTB-01693207	0.82	0.80	1.00	0.49
BovineHD1000021473	10	75518089	0.79	0.78	1.00	0.49	BovineHD1000023172	0.87	0.85	1.00	0.47
BTB-00442438	10	89826995	0.90	0.86	1.00	0.50	BovineHD1000025775	0.61	0.68	1.00	0.48
BovineHD1000028629	10	99000284	0.79	0.77	1.00	0.49	ARS-BFGL-NGS-107482	0.88	0.86	1.00	0.48
BovineHD1100000602	11	1688557	0.78	0.77	0.99	0.50	ARS-BFGL-NGS-57976	0.74	0.75	1.00	0.48
ARS-BFGL-NGS-116521	11	15427710	0.63	0.69	0.99	0.50	BovineHD1100006210	0.90	0.87	1.00	0.48
Hapmap51172-BTA-122698	11	23803086	0.86	0.83	0.99	0.50	BovineHD1100010780	0.77	0.78	1.00	0.49
BovineHD1100013781	11	47123538	0.85	0.82	1.00	0.50	BovineHD1100012620	0.74	0.75	1.00	0.48
ARS-BFGL-NGS-38946	11	67592479	0.71	0.73	1.00	0.50	BovineHD1100016207	0.85	0.84	0.99	0.49
BovineHD1100022791	11	79476215	0.82	0.80	1.00	0.50	BovineHD1100020749	0.90	0.87	1.00	0.48
BovineHD1100026305	11	90771548	0.86	0.82	1.00	0.50	BovineHD1100025363	0.89	0.87	1.00	0.49
BovineHD1100029028	11	99956241	0.91	0.87	1.00	0.49	BovineHD1100030349	0.79	0.78	1.00	0.48
BovineHD1200001588	12	5402254	0.72	0.75	1.00	0.50	BovineHD1200000375	0.77	0.77	1.00	0.48
ARS-USMARC-Parent-DQ786763-rs29020472	12	11824653	0.80	0.79	1.00	0.45	BovineHD1200006329	0.71	0.73	1.00	0.48
ARS-BFGL-NGS-100956	12	37110042	0.79	0.78	1.00	0.49	BovineHD1200009722	0.84	0.82	1.00	0.49
BovineHD1200014050	12	51090648	0.79	0.78	1.00	0.50	BovineHD1200011868	0.86	0.83	1.00	0.49
BovineHD1200015312	12	55257511	0.85	0.84	1.00	0.49	BTB-01499620	0.90	0.88	0.99	0.48
ARS-BFGL-NGS-112325	12	79643103	0.84	0.80	1.00	0.42	Hapmap28306-BTA-142098	0.88	0.85	1.00	0.48
BovineHD1200026173	12	89360668	0.75	0.75	0.99	0.50	BovineHD1200026107	0.75	0.76	1.00	0.48

BovineHD1300001042	13	3881631	0.72	0.73	0.99	0.50	BovineHD1300002479	0.75	0.76	1.00	0.48
BovineHD1300006768	13	23254232	0.79	0.78	1.00	0.50	BovineHD1300004130	0.91	0.88	1.00	0.49
BovineHD1300009872	13	33893791	0.88	0.85	0.99	0.49	ARS-BFGL-NGS-113540	0.78	0.78	1.00	0.47
BovineHD1300012831	13	43950694	0.88	0.84	1.00	0.49	ARS-BFGL-NGS-13061	0.79	0.79	1.00	0.48
BovineHD1300014266	13	48974104	0.73	0.74	0.99	0.49	BovineHD1300015621	0.72	0.74	1.00	0.48
BovineHD1300020070	13	70231274	0.76	0.76	1.00	0.50	ARS-BFGL-NGS-97229	0.85	0.82	0.99	0.48
BovineHD1300021259	13	73763974	0.83	0.80	1.00	0.50	ARS-BFGL-NGS-82178	0.83	0.81	1.00	0.48
Hapmap31564-BTC-007633	14	9606819	0.82	0.80	1.00	0.50	BovineHD1400001446	0.75	0.76	1.00	0.48
BovineHD4100011193	14	12567227	0.88	0.85	0.99	0.50	BovineHD1400003992	0.85	0.83	1.00	0.49
BovineHD1400007433	14	25708285	0.82	0.80	1.00	0.50	BovineHD1400006600	0.85	0.83	1.00	0.48
BovineHD1400010363	14	35982826	0.83	0.80	1.00	0.49	Hapmap49092-BTA-24990	0.90	0.87	1.00	0.46
BovineHD1400016259	14	58570368	0.70	0.72	0.99	0.50	BovineHD1400014328	0.81	0.80	1.00	0.48
BovineHD1400018435	14	65910953	0.63	0.71	1.00	0.50	BovineHD1400016259	0.72	0.74	0.99	0.47
BovineHD1400022075	14	78761666	0.79	0.79	1.00	0.49	BovineHD1400021348	0.78	0.78	1.00	0.48
Hapmap47928-BTA-37003	15	3735247	0.61	0.69	1.00	0.50	BovineHD1500000615	0.90	0.87	1.00	0.48
BovineHD1500004285	15	16834250	0.81	0.80	1.00	0.50	ARS-BFGL-NGS-5589	0.70	0.72	1.00	0.48
Hapmap38594-BTA-36288	15	29648882	0.90	0.87	1.00	0.50	BovineHD1500009754	0.82	0.81	1.00	0.48
BovineHD1500012376	15	44435747	0.78	0.77	1.00	0.50	BovineHD1500011406	0.85	0.82	1.00	0.48
BTA-37062-no-rs	15	51528617	0.80	0.77	1.00	0.45	BovineHD1500016358	0.62	0.68	0.99	0.48
BovineHD1500019376	15	67280833	0.74	0.75	1.00	0.49	BovineHD1500020944	0.84	0.82	1.00	0.49
BovineHD1500022839	15	78573349	0.79	0.78	1.00	0.48	BovineHD1500021562	0.78	0.78	1.00	0.48
BovineHD1600001175	16	4171120	0.67	0.71	1.00	0.49	Hapmap49982-BTA-39672	0.89	0.87	1.00	0.48
ARS-BFGL-BAC-20631	16	16270589	0.90	0.86	1.00	0.50	BovineHD1600004345	0.90	0.87	1.00	0.47
Hapmap54632-rs29021962	16	34519715	0.77	0.76	1.00	0.49	BovineHD1600010279	0.76	0.76	1.00	0.49
Hapmap54267-rs29023167	16	47850219	0.77	0.77	1.00	0.50	BovineHD1600011460	0.87	0.85	1.00	0.48
ARS-BFGL-NGS-116784	16	50737342	0.77	0.83	0.99	0.48	ARS-BFGL-NGS-112035	0.82	0.81	1.00	0.48
BovineHD1600018082	16	64306609	0.92	0.89	1.00	0.50	BovineHD1600018604	0.84	0.83	1.00	0.46
BovineHD1600023636	16	80913615	0.80	0.79	1.00	0.50	BovineHD1600022275	0.74	0.76	0.99	0.48
BovineHD1700000819	17	3463993	0.77	0.77	1.00	0.50	BovineHD1700000498	0.86	0.83	1.00	0.48
BovineHD1700004208	17	14713664	0.79	0.77	1.00	0.50	ARS-BFGL-NGS-28104	0.84	0.82	1.00	0.48

ARS-BFGL-BAC-18831	17	26010339	0.69	0.71	1.00	0.49	BovineHD1700009372	0.85	0.82	1.00	0.48
BovineHD1700013069	17	46771661	0.88	0.84	1.00	0.50	BovineHD1700012915	0.81	0.81	1.00	0.49
ARS-BFGL-NGS-118636	17	60366064	0.89	0.86	1.00	0.49	BovineHD1700016373	0.77	0.77	1.00	0.48
BovineHD1700018547	17	64317906	0.80	0.78	1.00	0.49	BovineHD1700021086	0.82	0.80	1.00	0.49
BovineHD1800002811	18	8154617	0.69	0.73	0.99	0.50	BTB-01040787	0.86	0.84	1.00	0.48
BovineHD1800006398	18	20589044	0.83	0.81	1.00	0.50	BovineHD1800005348	0.80	0.79	1.00	0.48
BovineHD1800007973	18	25881447	0.79	0.79	0.99	0.49	BovineHD1800009360	0.89	0.86	1.00	0.48
BovineHD1800013130	18	43575986	0.76	0.76	1.00	0.50	BovineHD1800011656	0.76	0.76	1.00	0.48
BovineHD1800018548	18	63970871	0.78	0.76	1.00	0.50	BTA-23545-no-rs	0.82	0.80	1.00	0.49
BovineHD1900000459	19	2027297	0.89	0.85	1.00	0.50	BovineHD1900018655	0.77	0.78	0.99	0.48
BovineHD1900007774	19	26247577	0.90	0.86	1.00	0.50	ARS-BFGL-NGS-71297	0.89	0.86	1.00	0.49
ARS-BFGL-NGS-116815	19	37845999	0.79	0.77	1.00	0.50	BovineHD1900010701	0.62	0.68	1.00	0.48
BovineHD1900011640	19	40906205	0.79	0.78	1.00	0.50	Hapmap47855-BTA-119202	0.80	0.80	0.99	0.48
BovineHD1900018179	19	62932571	0.79	0.77	1.00	0.50	BovineHD1900017317	0.76	0.76	1.00	0.48
Hapmap34051-BES7_Contig165_112	20	5044097	0.76	0.77	1.00	0.50	BovineHD2000000578	0.91	0.89	1.00	0.48
BovineHD2000003706	20	11515684	0.73	0.75	1.00	0.49	BTB-00771463	0.77	0.78	0.99	0.48
BovineHD2000006739	20	22425353	0.87	0.83	0.99	0.50	BovineHD2000005925	0.89	0.86	1.00	0.48
BovineHD2000011164	20	39263700	0.94	0.90	1.00	0.50	Hapmap57276-ss46526009	0.82	0.81	1.00	0.48
BovineHD2000016907	20	60246308	0.81	0.79	1.00	0.49	BovineHD2000014001	0.89	0.86	1.00	0.48
BovineHD2000020099	20	69035226	0.79	0.80	1.00	0.50	BovineHD2000020360	0.80	0.79	1.00	0.49
BovineHD2100000779	21	4608753	0.69	0.72	1.00	0.50	ARS-BFGL-NGS-118735	0.67	0.71	1.00	0.48
BovineHD2100003656	21	14011434	0.68	0.72	1.00	0.50	BovineHD2100004960	0.80	0.79	1.00	0.49
BovineHD2100008370	21	29030826	0.82	0.81	1.00	0.50	BovineHD2100010477	0.89	0.86	1.00	0.49
BovineHD2100013468	21	47084001	0.89	0.84	0.99	0.50	BovineHD2100011342	0.87	0.85	1.00	0.49
BovineHD2100016933	21	58797963	0.73	0.74	1.00	0.49	BovineHD2100017359	0.80	0.79	1.00	0.48
BovineHD2100021040	21	71573501	0.86	0.83	1.00	0.50	ARS-BFGL-NGS-21241	0.78	0.78	1.00	0.48
BovineHD2200000108	22	511111	0.81	0.79	1.00	0.50	BovineHD2200002037	0.71	0.74	1.00	0.48
ARS-BFGL-NGS-24419	22	11756783	0.74	0.75	1.00	0.42	BovineHD2200003493	0.84	0.83	1.00	0.47
Hapmap50386-BTA-53775	22	22004775	0.92	0.88	1.00	0.49	BovineHD2200008940	0.81	0.80	0.99	0.48
BovineHD2200014084	22	49192681	0.80	0.78	1.00	0.50	BTB-00848473	0.74	0.76	1.00	0.47

ARS-BFGL-NGS-16048	22	56448122	0.77	0.76	0.99	0.50	ARS-BFGL-NGS-56282	0.67	0.72	1.00	0.49
BovineHD2300002182	23	8855928	0.85	0.82	1.00	0.49	ARS-BFGL-BAC-35859	0.76	0.77	1.00	0.48
BovineHD2300003310	23	13152167	0.77	0.78	1.00	0.50	BovineHD2300003681	0.62	0.68	1.00	0.49
Hapmap53349-ss46526376	23	27305227	0.79	0.83	1.00	0.49	BovineHD2300009593	0.83	0.81	1.00	0.48
BovineHD2300012089	23	41994287	0.74	0.75	1.00	0.50	UA-IFASA-8930	0.71	0.73	1.00	0.48
ARS-USMARC-Parent-DQ995977-rs29020834	24	1854953	0.70	0.72	1.00	0.48	BovineHD2400002785	0.85	0.83	1.00	0.49
BovineHD2400007076	24	26094338	0.66	0.70	0.99	0.50	BovineHD2400004603	0.68	0.70	1.00	0.48
BovineHD2400007553	24	27679319	0.89	0.84	1.00	0.50	BovineHD2400007562	0.70	0.73	1.00	0.48
BovineHD2400012466	24	45519177	0.72	0.75	1.00	0.50	BovineHD2400011641	0.94	0.92	1.00	0.48
ARS-BFGL-NGS-10638	24	59890947	0.77	0.77	1.00	0.50	ARS-BFGL-NGS-3577	0.75	0.76	1.00	0.48
BovineHD2500001344	25	5572319	0.76	0.77	1.00	0.48	ARS-BFGL-BAC-43578	0.87	0.86	1.00	0.49
BovineHD2500006341	25	22687549	0.88	0.85	1.00	0.50	BovineHD2500005001	0.80	0.79	1.00	0.48
BovineHD2500009664	25	35128465	0.77	0.77	1.00	0.50	BovineHD2500011750	0.67	0.71	1.00	0.47
BovineHD2600000493	26	2898621	0.86	0.82	1.00	0.49	BovineHD2600000059	0.70	0.73	1.00	0.49
BovineHD2600006985	26	26335390	0.84	0.80	1.00	0.50	BovineHD2600004773	0.72	0.73	1.00	0.48
BovineHD2600009148	26	33976565	0.88	0.84	1.00	0.50	ARS-BFGL-NGS-20281	0.85	0.84	1.00	0.48
BovineHD2600013668	26	47448679	0.74	0.76	1.00	0.50	BTA-61751-no-rs	0.79	0.78	1.00	0.48
UA-IFASA-1517	27	7561409	0.89	0.85	1.00	0.50	BovineHD2700002707	0.88	0.85	1.00	0.48
BovineHD2700005290	27	18522077	0.90	0.86	1.00	0.50	Hapmap39675-BTA-66683	0.78	0.78	1.00	0.48
BovineHD2700013137	27	45125791	0.77	0.77	1.00	0.50	BovineHD2700010621	0.74	0.75	1.00	0.48
BovineHD2800000088	28	220322	0.89	0.85	1.00	0.49	BovineHD2800001322	0.87	0.85	1.00	0.48
BovineHD2800004190	28	14938753	0.90	0.85	0.99	0.50	BovineHD2800006610	0.81	0.81	1.00	0.48
BovineHD2800008807	28	33148192	0.71	0.73	1.00	0.49	BovineHD2800009995	0.91	0.87	1.00	0.49
UA-IFASA-5633	28	35354394	0.82	0.79	1.00	0.50	BovineHD2800012820	0.87	0.85	1.00	0.48
BovineHD2900014980	29	5806437	0.72	0.73	0.99	0.50	BovineHD2900001260	0.82	0.82	0.99	0.49
BovineHD2900007698	29	26364780	0.87	0.83	1.00	0.50	ARS-BFGL-NGS-24259	0.91	0.89	1.00	0.48
ARS-BFGL-NGS-110853	29	36188416	0.77	0.77	1.00	0.50	BovineHD2900009868	0.68	0.71	0.99	0.48