IDENTIFICATION OF GENETIC VARIATION IN THE KERATIN-ASSOCIATED PROTEIN AND KRT33A GENES OF THE SWAKARA SHEEP OF NAMIBIA

A THESIS SUBMITTED IN FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN AGRICULTURE (ANIMAL SCIENCE)

OF

UNIVERSITY OF NAMIBIA

BY

NELLIA FELICITY NYONI

201212522

APRIL 2020

SUPERVISOR: DR. THEOPOLINE OMAGANO AMUSHENDJE

(UNIVERSITY OF NAMIBIA)

CO-SUPERVISOR: DR. MARIA NDAKULA TAUTIKO SHIPANDENI

(UNIVERSITY OF NAMIBIA)

ABSTRACT

Keratins are structural intermediate filamentous proteins that constitute about 90% of the total wool fiber in sheep. The protein gene family is divided into two groups which are the Keratin Intermediate-Filament proteins (KRTs) and the Keratin-Associated proteins (KAPs). The KRTs form the skeletal structure of the wool fiber (micro-fibrils) and are embedded in a matrix of KAPs through disulphide cross-linkages. Genetic variation in the keratin genes may be associated with pelt quality traits. Pelt quality is determined by the unique hair formation, which is determined by these protein genes. High demand of exceptional pelts by the fashion industry, has probed this study in determining genetic variation in Keratin genes (KAP1.1, KAP1.3, KAP3.2) and KRT33A). The study also examined allele frequency distribution of the keratin genes in Swakara sheep in Namibia. Blood samples were collected from 168 Swakara sheep randomly selected from four farms; Neudamm, Gellap-Ost, Kalahari and Tsumis. Genomic Deoxyribonucleic acid (DNA) was isolated using the Inqaba biotech-kit protocol. Extracted genomic DNA was confirmed using a Nano drop and amplified using the Polymerase Chain Reaction (PCR). Genetic variation was assessed using PCR-Agarose gel electrophoresis in the KAP1.1 gene. As for the KAP1.3, KAP3.2 and KRT33A, genetic variation was assessed using Polymerase Chain Reaction-Single Strand Conformational Polymorphism (PCR-SSCP). Sequenced keratin genes under investigation and those retrieved from NCBI were grouped together and used to construct phylogenetic trees on Molecular Evolutionary Genetics Analysis (MEGA) version 6.0. At the KAP1.1 locus, the study revealed three alleles; A, B and C with genotype frequencies of 0.13, 0.57 and 0.30, respectively. Mostly the B allele frequency was highly distributed in comparison to the A and C alleles. At the KAP1.1 locus, the study reports statistical significant difference in allele frequency distribution amongst the four farms (P=0.004). The KAP3.2 locus had one allele identified across all four farms and denoted as A. The KAP1.3 locus was statistically significant at P=0.160 with genotype frequencies of 0.5 (AA), 0.35 (AB) and 0.15 (CC). The KRT33A locus was not statistically significant (P=0.402), and had 0.25 (A) and

0.75 (B) allele frequency. The four genes under study showed no significant deviation from Hardy Weinberg Equilibrium. The findings of this study which reports genetic variation in the KAP1.1, KAP1.3 and KRT33A has the potential to help with the identification of genetic markers linked to superior pelts. The use of genetic markers in the selection of Swakara sheep breeding program would potentially increase the accuracy of selection, and compliment visual appraisal technique that is currently being used in the Swakara sheep industry.

Key words: Polymorphism, Keratin genes, Allele frequency, Swakara, Pelt quality.

CONFERENCE PROCEEDINGS

1) Nyoni N.F., Itenge T.O., and Shipandeni M.N.T. (2019). Genetic variation in the KAP1.1 gene and its potential to supplement Indigenous Knowledge Systems within the Swakara sheep in Namibia. In N.P. Iwuanyanwu and N.O. Vhonani - (Ed.), Proceedings of the 5th Annual Conference of the African Association for the Study of Indigenous Knowledge Systems, (pp. 43 - 50). Thohoyandou, Limpopo, South Africa: ISBN: 978-0-86808-767-2.

PRESENTATIONS MADE AT CONFERENCES

1. N.F. Nyoni, T.O. Itenge and M.N.T. Shipandeni. Polymorphism of the KAP1.1 Gene in the Swakara sheep of Namibia. Presentation at the 6th Annual Science Research Conference: Scientific Solutions for Economic Development 14-15 November 2018. University of Namibia, Main Campus, Windhoek, Namibia.
2. N.F. Nyoni, T.O. Itenge and M.N.T. Shipandeni. Polymorphism of the KAP1.1 Gene in the Swakara sheep of Namibia. Presentation at the 2nd International Conference on Agriculture and Natural Resources Research and Innovation 15-16 October 2018. University of Namibia, Neudamm Campus, Windhoek, Namibia.
3. N.F. Nyoni, T.O. Itenge, E. Kandiwa and M.N.T. Shipandeni. Identification of genetic variation in the KAP1.1 gene of the Swakara sheep in Namibia. Presentation at the 5th Annual Science Research Conference: Science Today: Global Bridge to the Future 15-16 November, 2017. University of Namibia, Main Campus, Windhoek, Namibia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS OR ACRONYMS

% percentage

ºC degrees Celsius

µL microlitre(s)

µM micromolar

kg kilogram

M molar

mM millimole

mL milliliter

ng nano-gram

V volt

AFLP - amplified fragment length polymorphism

APS- Ammonium persulphate

DNA - Deoxyribonucleic acid

EDTA- Ethylene Diamine Tetracetic Acid

GDP - Gross domestic product

HGT- High glycine-tyrosine

HS- High sulphur

KAP - Keratin Associated Proteins

KRT - Keratin Intermediate Filament Proteins

NCBI- National Center for Biotechnology Information

OD- Optical density

PCR – Polymerase Chain Reaction

PCR-SSCP - Polymerase Chain Reaction-Single Stranded Conformational Polymorphism

RFLP - Restriction Fragment Length Polymorphism

SNP - Single Nucleotide Polymorphism

TAE - Tris-Acetate

TEMED - Tetramethylethylendiamine

TBE – Tris-Borate

UHS- Ultra high sulphur

VAT – Visual appraisal technique

W/v- Weight by volume

ACKNOWLEDGEMENTS

First and foremost, I thank the Lord Almighty for His strength and unconditional love that kept me till the end of my studies. I would like to thank my thesis main supervisor Dr. Theopoline Omagano Amushendje and co-supervisor Dr. Maria Ndakula Tautiko Shipandeni under the Department of Animal science, Faculty of Agriculture and Natural Resources at the University of Namibia. Their doors were always open whenever I ran into trouble or had a question about my thesis write up. Dr. Theopoline Omagano Amushendje and Dr. Maria Ndakula Tautiko Shipandeni consistently allowed this thesis paper to be my own work but steered me in the right direction at all times and for that I am truly grateful. I am thankful for the financial support by the National Commission on Research Science and Technology (NCRST) in Namibia, without their support this research would not have been possible.

I would also like to acknowledge Dr. Emmanuel Lutaaya, Head of Department in Animal Science, as a second reader of this thesis, and I am gratefully indebted to his very valuable comments on this thesis. To the entire Department of Animal Science staff thank you for your assistance and support throughout my studies.

My profound gratitude to the Ministry of Agriculture Water and Forestry for availing the three farms for data collection; Gellap-Ost, Tsumis and Kalahari. My sincere thanks to the following Doctors and an Animal control technician for collecting blood samples on my behalf; Dr. Erick Kandiwa, Dr. Annemie Louwrens, Dr. Thomas Tolmay and Mr. Johan EngelBrecht. My genuine thanks to the Neudamm Small stock management at the University of Namibia for availing Swakara sheep for data collection as well.

I am indebted to Mr. Filemon Nghitilanganye Shindume and Dr. Gerhard Haitembu for being great mentors during the course of my studies, guiding me through my laboratory analysis at the Ministry of Agriculture Water and Forestry and for that I am forever grateful. To my beloved friends Ms. Morga Hamunyela, Mr. Evra Nepaya, Mr. Fillemon Williams, Ms. Abigail Kamwi and Ms. Indileni Mwadulange am most thankful for helping me with my laboratory work. Long hours spent in the laboratory were surely not in vain and I learnt a lot from you all. And for that be blessed forever.

My genuine thankfulness to the transport officer at Neudamm Campus Mr. Willem Goussard for providing transport during the course of my studies. To my academic mentor, Dr. Stephen Barrion am grateful for the encouragement and continuous inquiring about how far I was with my research, am thankful for the wise words you shared with me.

Finally, I must express my very profound gratitude to my parents Mr. Eledge Nyoni and Mrs. Lilian Nyoni for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without you. Thank you Daddy and Mama. My siblings George Harris Nyoni and Lucy Nyoni, thank you for always believing in me and without you two I would be lost. Be blessed forever.

DEDICATION

This thesis is dedicated to the Swakara farming communities in Namibia.

DECLARATION

I, Nyoni Nellia F., hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

No part of this thesis/dissertation may be reproduced, stored in any retrieval system, or transmitted in any form, or by means (e.g. electronic, mechanical, photocopying, recording or otherwise) without the prior permission of the author, or The University of Namibia on my behalf.

I, Nyoni Nellia F., grant The University of Namibia the right to reproduce this thesis in whole or in part, in any manner or format, which The University of Namibia may deem fit, for any person or institution requiring it for study and research; provided that the University of Namibia shall waive this right if the whole thesis has been or is being published in a manner satisfactory to the University.

__________________ ___________ _____________

Nyoni Nellia Felicity Signature: Date:

Certified by:

_________________ ___________ _____________

Dr. Theopoline Omagano Amushendje Signature: Date:

Main supervisor

Certified by:

_________________ ___________ _____________

Dr. Maria Ndakula Tautiko Shipandeni Signature: Date:

Co-supervisor

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Keratin proteins, the major component of wool fibre consists of three main parts the cuticle, cortex, and medulla which are responsible for most of their structural properties (Kunchi *et al*., 2019; Hediger *et al*., 1991; Onions, 1962). The cortex includes 90% of the wool fibre and has filamentous micro-fibrils (Marshall *et al*., 1991; Powell and Rogers, 1986) through further inter- and intra- molecular disulphide bond formation (Rogers, 2004) but the precise mechanism of linking is still poorly understood (Gong *et al*., 2019). Although the KAPs may have little or no visible effect on Keratin Intermediate Filament structure as their effect on KIF assembly into larger arrays (Plowman, 2003) are considered vital. Gong *et al*. (2016) stated that these KAPs play a vital role in defining the physic-mechanical properties of the wool fibre. Further, divided into two groups known as the Keratin intermediate filament proteins (KIF) and the Keratin associated proteins (KAP) (Yardibi *et al*., 2015), the KIF proteins form filaments through a network of disulphide bonds (Gong *et al*., 2010) that lie within a matrix of the KAP.

There exists hetero-polymers of the type I and type II keratins which are encoded by the KRT1.n and KRT2.n gene families, respectively (Powell and Rogers, 1994) as they form the intermediate filaments of the 'hard' keratins of hair as well as 'soft' keratins found in the epidermis (Rogers, 2004) and mapped uniquely on two separate chromosomes, three and eleven (Hediger *et al*., 1991). Rogers (2004) described that the two keratin families are notably different based on their isoelectric point such that; type I are acidic whereas type II are either basic or neutral. Furthermore, Steinert and Roop (1988) illustrated that equimolar amounts of the two types are required in the formation of the intermediate filaments.

Keratin associated proteins play a critical role in determining the physico-mechanical properties of hair and wool fibres as they form a semi-rigid matrix in which the KIFs are rooted (Li *et al*., 2018; Hua *et al*., 2011; Gong *et al*., 2010; Rogers *et al*., 2007; Powell, 1996; Powell *et al*., 1995). The KAPs are distributed into three groups based on their amino acid compositions; high sulphur proteins (KAP 1.n, KAP 2.n, KAP 3.n, KAP 11, KAP 13 to KAP 16 and KAP 23 to KAP 27 families), ultra high sulphur proteins (KAP 4.n, KAP 5.n, KAP 9, KAP 10, KAP 12 and KAP 17 families) and the high glycine-tyrosine proteins (KAP 6.n, KAP 7.n and KAP 8.n, KAP 18 to KAP 22 families), (Gong *et al*., 2011b; Gong *et al*., 2011; Rogers *et al*., 2005; Plowman, 2003; Parry and Steinert, 1992; Marshall *et al*., 1991). Their genes lack introns and they usually occur in gene clusters (Rogers *et al*., 1994).

Sheep rearing is a common practice in arid and semi-arid regions which contributes significantly to the livelihood and nutritional security of resource poor communities. Sheep often act as a hedge against the losses in agriculture and adversities in the absence of adequate health and social security (Amareswari *et al*., 2018). Namibia, is rated as having the driest climate in sub-Saharan Africa and is predominantly a livestock farming country. Livestock production accounts for approximately 70% of the agricultural output (Ministry of Agriculture Water and Forestry, 2013).

The Karakul sheep (from which the Swakara sheep originated) is believed to be one of the oldest domesticated sheep breed in the world, and is classified as a fat-tailed sheep (Optimal Agricultural Business Systems, 2016; Malesa, 2015; Bravenboer, 2007). Swakara sheep is mainly bred to produce lamb pelts/skins of various colors for the fashion industry (Lundie, 2011; Kevorkian *et al*., 2011). Swakara sheep may also be kept for meat, wool and milk production (Optimal Agricultural Business Systems, 2016; Nsoso and Madimabe, 2003) hence, the advantage of diversification within the breed. The Swakara sheep was imported into Namibia in 1907 (then called Karakul sheep), when the first consignment of two rams, seven ewes and three lambs bred in Bukhara, Uzbekistan were imported from Germany (Visser *et al*., 1998).

In the course of time, the sheep made a vital contribution to the development of small stock industry in Namibia, as seen in the infrastructural development on farms in the Southern and Western areas of the country (Visser *et al*., 1998). The Swakara sheep farming communities have embraced the breed as one of their own.

Swakara sheep pelt production is quite sensitive such that the euthanasia methods of lamb slaughter are complied with and adhered to in Namibia as stated in the Code of Practice (CoP) for the Care and Handling of Karakul Sheep in Namibia (Karakul Annual Report, 2010). Swakara lamb pelts are harvested within 48 hours post parturition if not, the curl structure of the pelt deteriorates with increase in age (Martins and Peters, 1992) and harvesting of pelts is administered by well-trained personnel. The Swakara (then Karakul) board of Namibia carries out regular routine inspections on farmers registered for pelt production exports to ensure compliance to the CoP when harvesting pelts (Kruger *et al*., 2013). Furthermore, the Namibia Swakara sheep slaughter procedures concurs with the international standards and Namibia attained the Original Assured label from the international Fur trade Federation for all Swakara pelts sold. Hence, the humane treatment of sheep also promotes sales on the international market as well as success of Swakara farming in Namibia.

The black Swakara pelt was once termed the “black diamond” of the Namib Desert because it resembled the original black color of the sheep (Itenge and Shipandeni, 2015; Bravenboer, 2007). Pelt production in 1925 was 23 000 pelts with a gross income of approximately R 34 500 and increased to 1 975 683 pelts produced in 1960 with a gross income of R 8 613 978 (Kirsten, 1966). Schoeman (1998) illustrated that around 1970, more than 95% of the total 4 400 000 million sheep in Namibia were Karakul. In 1976, a total income of N$50.2 million was obtained from approximately three million pelts which were exported to the European markets. In a turn of events, around late 1970’s and early 1980’s the worldwide anti-fur campaigns became highly vocal which led to the collapse of the fur industry in many countries, including Namibia (Schoeman, 1998). Moreover, the pelt industry went through a rough phase when it was contended by mink, which represented 70% of all fur sales worldwide in the 1980s’ (Anonymous, 1998). This forced farmers to diversify with a resultant increase in mutton sheep breeds in Namibia and other Karakul farming nations in the world. In the early 90’s the Swakara pelt market recovered slowly (Visser *et al*., 1998).

Thereafter, the Karakul sheep industry in Namibia has focused on strategic breeding programmes in order to produce unique quality pelt (short hair, exceptional pattern and better hair texture) as shown in Figure 1.1. The government of Namibia renamed the Karakul sheep as Swakara sheep in 2012 (Optimal Agricultural Business Systems, 2016). Swakara became the new trade name for all pelts produced in three Southern African countries; Namibia, South Africa and Botswana which are auctioned in Copenhagen, Denmark bi-annually in April and September.

Figure 1.4: Upgrading of the karakul sheep from 1907 to the current Swakara sheep with unique pelt. Source: Itenge and Shipandeni, 2015.

To our knowledge no genetic markers that are associated with pelt quality traits in Swakara sheep have been identified so far. Itenge (2002) stated that genetic markers are not affected by environmental noise and this would allow sheep breeders to select animals with improved wool characteristics at an early age and cull the non-desirable lambs.

Numerous studies showed that there has been an increase in the number of KAP genes defined in humans and sheep species as well as progressive accounts of variation in these genes (Hua *et al*., 2012). Rogers *et al*. (2005) illustrated that keratin associated proteins are encoded by many genes which are attributed to be highly polymorphic. Between two to nine alleles have been identified in KAP genes among *Ovies aries* species (Zhou *et al*., 2012; Gong *et al*., 2010). Gene expression can be affected by the structure and function of the encoded proteins (Elmaci *et al*, 2013), and variation in KAP genes has significantly influenced wool traits (Yardibi *et al*., 2015) in most ovine species. Genetic variation has become an objective tool traditionally used for improving animal species (Melus *et al*., 2009). This study aimed at identifying genetic variation within the KAP1.1, KAP1.3, KAP3.2 and KRT33A loci, which could lead to the development of potential genetic markers associated with pelt quality traits.

1.2 Statement of the problem

The Swakara sheep industry currently practices the Visual appraisal technique (VAT) to select animals with superior pelt quality traits and the breeding herd. However, VAT does not give accurate genetic potential, as it is quite subjective. Some of the phenotypic characteristics looked at when grading/sorting pelts are hair length, pelt weight and curl type by experts in the Swakara sheep industry.

Moreover, scientific literature is deficient as a source of information, as it primarily contains information on the contribution of keratin genes to wool structure and development in other sheep breeds in depth, however less is known about the association of keratin genes with superior pelt quality in Swakara sheep. Hence, prompting this study to further investigate at molecular level by identifying genetic variation in keratin genes of the Swakara sheep in Namibia that could potentially be developed into gene markers. Hence, genotyping of the progeny could be done earlier to determine the genetic potential at an early stage and complement the VAT used.

1.3 Objectives of the study

a) To identify genetic variation in Keratin genes of the Swakara sheep; KRT33A, KAP1.1, KAP1.3 and KAP3.2.

b) To compare the level of polymorphism of the keratin genes from different study sites situated in three regions of Namibia (Khomas, Hardap, Karas).

1.4 Null Hypotheses

a) H_0: There is no genetic variation that exists within the keratin genes of the Swakara sheep.

b) H_0: No level of polymorphism exists in the different study sites (Neudamm, Gellap-Ost, Tsumis and Kalahari).

1.5 Significance of the study

Swakara is the trade name for all Swakara sheep raised in three Southern-African countries; Namibia, South Africa and Botswana. The distinct pelts harvested 48 hours post-parturition are delicate and highly on demand in the fashion industry, which probed the current study to investigate genetic variability in the keratin genes of the Swakara sheep in Namibia. The study provides knowledge on genetic variability and organization of the KAPs genes (KAP 1.1, KAP 1.3 and KAP3.2) and KRT gene (KRT33A) found on chromosome eleven in Swakara sheep. Polymorphism at these loci, would possibly expedite the identification of possible genetic markers associated with superior pelt quality traits. The use of genetic markers in Swakara sheep breeding programs would potentially increase the accuracy of selection and in-turn increase revenue earned from Swakara sheep pelts sold hence success of the small stock industry in Namibia, Botswana and South Africa. Since the Swakara sheep breed is vital for Namibia's production of good quality pelts, it will ensure sustainability of the farming enterprise. This study would provide crucial baseline information that other researchers can use in better utilization and improvement of the Swakara's role in Namibia's Agriculture industry and in the fashion industry.

1.6 Limitation of the study

Due to financial constraint, allelic diversity through Sanger sequencing could not be fully exploited in this study as only a total of 26 samples were sequenced, without duplications. Challenges such as failure of Polyacrylamide gel electrophoresis cassettes to split when conducting Polymerase Chain Reaction-Single Strand Conformational (PCR-SSCP) analysis.

1.7 Delimitation of the study

From an overall sample size of 205 the study assessed genetic variation from 168 Swakara in the KAP1.1 locus using Polymerase chain reaction-Agarose gel electrophoresis. The reduction in number was as a result of poorly amplified Polymerase chain reaction samples therefore they were indefinitely disqualified from the study. A total of 20 samples from each farm randomly selected were subjected to the PCR-SSCP analysis on three loci (KAP1.3, KAP3.2 and KRT33A) on polyacrylamide gel electrophoresis. Due to cost implications only six samples within the KAP1.1 locus were sequenced while 20 from KAP1.3, KAP3.2 and KRT33A were sequenced.

CHAPTER 2

LITERATURE REVIEW

2.1.0 Small ruminants in Namibia

Sheep and goats are small ruminants kept as livestock and they form the backbone of most rural livelihood (Ahlawat *et al.*, 2014; FAO, 2009) in agricultural industries across the globe (Clark *et al.*, 2017). The Namibian population depends on agriculture either directly or indirectly and (Namibia's 5th National Development Plan, 2017) states that the agriculture sector significantly contributes approximately 3.8% to GDP and remains a strategic sector that supports above 70% of the Namibian population. The Namibian Small Stock industry constitutes of sheep (Karakul, Dorper and others) and goats (Angora, Boer, indigenous and others) reared in both commercial and communal areas. According to MAWF 2017 census reports, Namibia had approximately 3 651 143 small stock; with 1 624 834 goats and 2 026 309 sheep distributed across the country, with most small stock found in the Southern regions of the country; Hardap and Karas. Karakul sheep were approximately 119 608 as per the 2017 MAWF census report (Nambahu, 2019).

2.1.1 The conception of Karakul (now Swakara) sheep in Namibia

The Karakul sheep was introduced to Namibia in 1907, from the former Soviet Republic of Uzbekistan in Bukhara, in Central Asia (Bravenboer, 2007; Devandra and McLeroy, 1982) for experimental trials. It was reported that two rams, seven ewes and three lambs were imported to Namibia (Bravenboer, 2007; Visser *et al.*, 1998) via the port of Swakopmund, however, Neubert (1989) reported a total of four rams and 28 ewes. The Bukhara region in Uzbekistan is in temperate latitudes. Though Namibia is characterized by low and irregular rainfall the hardy karakul breed has successfully adapted to this environment (Duffield-Harding, 2005;

Bravenboer, 2007). Nsoso and Madimabe (1999) noted that most areas in Namibia where there is Karakul farming have similar climates to the parts of Botswana where the breed is farmed.

The Swakara sheep has transitioned over the past century from the classic pipe curl type that was bred. In the course of time, breeders have developed and upgraded with commercial lines facilitated by crossing Swakara with the indigenous Afrikaner and Black-head Persian (Bravenboer, 2007). A census carried out in 1913, reported that, approximately 21 000 Swakara sheep were crossbred with these indigenous breeds. One of the government stations, Neudamm played a vital role in the contentious breeding program which resulted in a retard curl development with shallow and water silk (WS) curl types. The Swakara sheep upgrading over the years is thought to have contributed to a change in the genotype when compared to the Karakul from the native country, Uzbekistan. The white color is not inherent among Swakara sheep, it arose from crossing the breed with the white-wooled Persian sheep (Swakara Annual Report, 2011; Bravenboer, 2007). Swakara is the trade name for all lamb pelts produced in Southern Africa and they are popular with the international fur trade.

2.1.2 Swakara sheep breed

Swakara sheep are classified as a fat-tailed breed that has been reared for the past 10 000 years (Heren, 2000). A mature Swakara sheep is characterized by long, hairy fleece that is shorn at-least twice a year. Bravenboer (2007) observed that both polled and horned herds exists among the Swakara sheep. On average adult rams weigh between 60-80 kg whilst ewes weigh 40-55 kg. Swakara sheep have strong herd instincts which protects them from being preyed on, and thrive well in areas that receive approximately 250 mm or less of rainfall (Ministry of Agriculture Water and

Forestry, 2013; Bravenboer, 2007). Swakara sheep is deemed the golden story of Namibia as it engages large parts of the rural communities with both employment and food (www.swakara.net) in the Southern part of the country. In the global livestock sector Swakara sheep significantly contributes to pelt, meat, milk and wool with pelt production being the most predominant practice (Clark, 2017; Bravenboer, 2007). Non-food products from Swakara sheep such as pelts, wool, fibre and manure can be further processed into clothes; blankets; carpets and bio gas (FAO, 2009; Bravenboer, 2007; Heren, 2000; Ensminger, 1991).

2.1.3 Distribution of pelts

The International Fur Federation (IFF) constitutes of 42 member associations and organization from 35 countries, as it represents almost every fur producer and fur consuming country in the world. Swakara board of Namibia is the only member from the African continent (Swakara Annual Report, 2015/2016; Karakul Board Annual Report, 2008/2009) that is part of the IFF. Such that, the only Swakara pelt producing countries in Africa are Namibia, South Africa and Botswana with pelt production of approximately 84%, 14% and 0.29%, respectively (Swakara Annual Report, 2015/2016). Afghanistan, Uzbekistan and other Asian countries produce Karakul skins but when compared to Swakara pelts they are heavier with long hair and of lower quality (Karakul Board Annual Report, 2008/2009). Negative perception towards Karakul pelt production from Afghanistan promulgated the change in name hence, the Swakara Board of Namibia in conjunction with the government of Namibia renamed the karakul sheep to Swakara in 2012 (Farmers Weekly, 2012). The intent was to dissociate and stand alone as a nation that humanely breeds Swakara sheep.

2.1.4 Unique pelt quality

The Namibian Swakara has a unique S-shaped fibre pattern (Duffield-Harding, 2005; Schoeman, 1998). The reverse S-formation produces delicate patterns which when coupled with the silky extra fibre and the retarded curl, reflects light (Duffield-Harding, 2005; Schoeman, 1998) and is regarded as top quality Swakara pelt. Furthermore, Swakara pelt is characterised by short hair, exceptional patterns and better hair texture (Bravenboer, 2007) with a flat, silky and elegant touch (www.swakara.net). Bravenboer (2007) illustrated the importance of traceability in Karakul pelts which is a mechanism that allows all pelts sold internationally to be traced back to the owner and farm where they were produced. Distinct pelt with curls forms a surface tracery unique like a fingerprint to each sheep.

The natural colors found in the Namibian Swakara consist of black, white, brown, grey and spotted (Bravenboer, 2007) are depicted on Figure 2.1, however, over 200 different variations in a tone have been observed (www.swakara.net). A variety of these colors and shades are versatile in fashion and are of economic importance as different combinations of colors (silver-tipped chocolate and brown-grey) may arise due to mutations and alterations within the genome. White is the most preferred color as it is easier to dye into alternative colors for clothing and other pelt uses (Muchadeyi *et al*., 2015; Bravenboer, 2007). In general the price of white pelts is almost double the price of black or spotted pelts. During the 2016 auction, the most priced white pelt cost N$ 1 823.81 and the highest black pelt fetched a price of N$1780.38. The top Namibian producer sold 250 pelts and achieved an average price of N$ 856.04 per pelt (www.swakara.net).

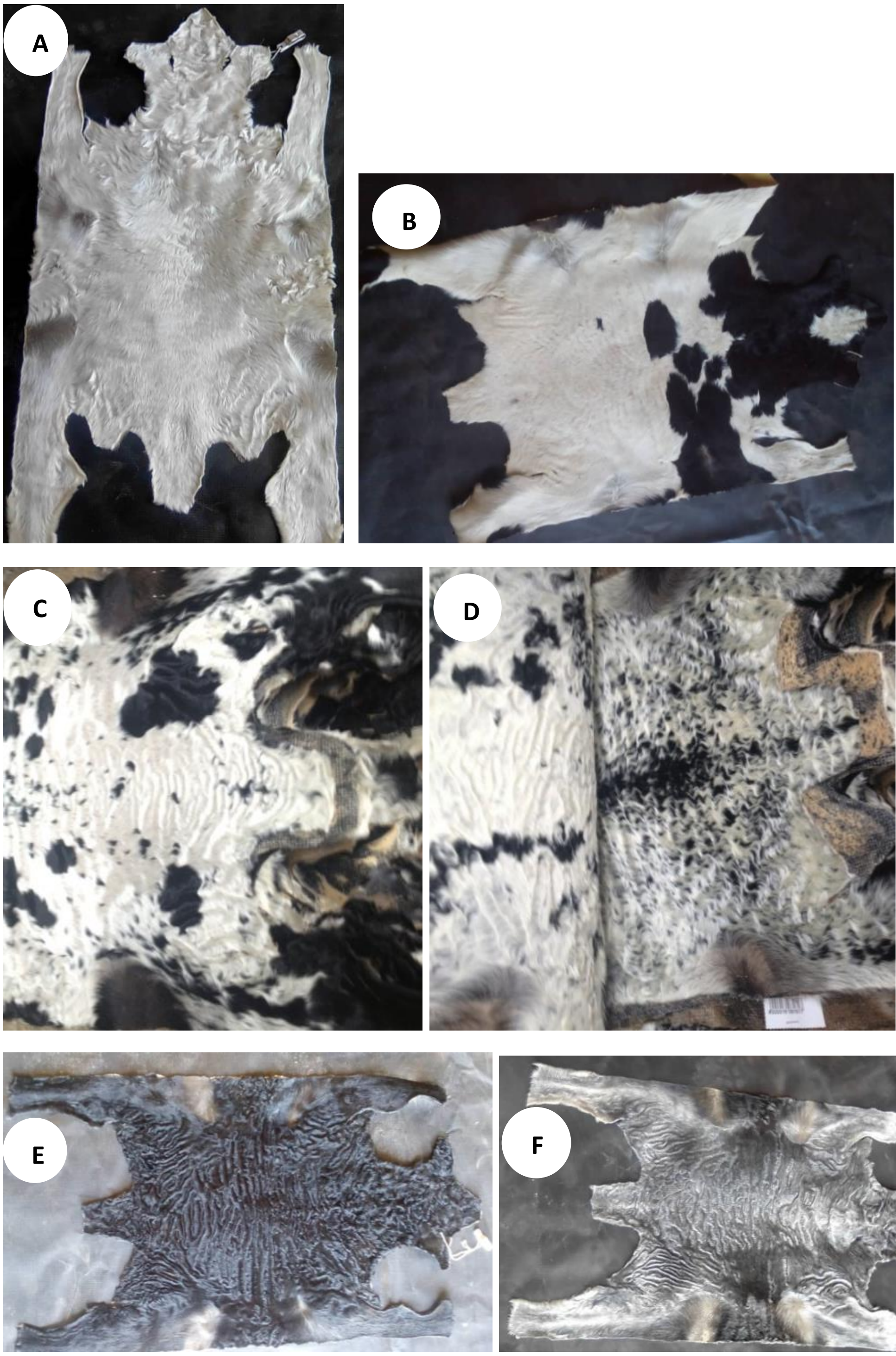


Figure 5.1: Namibian Swakara pelt colors: White pelt (A), Spotted pelt (B, C and D), Black pelt (E) and grey pelt (F).

2.1.5 Code of practice and pelt grading/sorting at Agra Pro-Vision

Kruger *et al*. (2013) reported that Swakara producers in Namibia adhere to the Code of Practice (CoP); for example ill-practice like induced premature birth is strictly prohibited. Essentially the CoP primarily focuses on the welfare of Swakara sheep as it sets out strict humane handling, stunning of lambs to minimize stress and preparation of the pelts prior to being sent to AGRA. Such that natural mating and artificial insemination in Swakara sheep is legal practice (Bravenboer, 2007). Swakara sheep farmers are expected to practice good husbandry as routine checks are carried out by the Swakara board of Namibia (The Karakul Board of Namibia, 2006) and well trained personnel handle all the processes from stunning until the final pelt product. Namibian Swakara pelt received international recognition for being a sustainable natural product hence, became a part of the 'Original Assured^{TM}' label issued by the International Fur Trade Federation (IFTF). The IFTF indicated that only fur from a country which abides to local ethical regulations can be a part of the 'Original Assured^{TM}' label.

AGRA Pro-Vision is the only pelt grading/sorting center for all pelts in Southern African. Mostly the visual appraisal technique is used which is defined as the selection method were producers visually inspect animals and judge which individuals are closest to the ideal for the desired traits (Awgichew *et al.,* 2007). When grading pelts most farmers understand hair quality, pattern excellency, pelt weight and classification of curl development however, hair length remains a mystery to most. Oftentimes, there is misconception of hair length is a subjective topic despite the lessons provided by a team of experts at AGRA Pro-Vision to farmers.

Nel (1966) and Schoeman (1968) illustrated that curl type is the degree of curl development and often differs from smooth (galliac) to more curly (pipe curl) and is negatively associated with pelt price. Tribute to the late A. D. Thompson the pioneer in selection for less developed types, which resulted in the smooth types for which Swakara became highly popular with the fashion industry. Anonymous (2005) illustrated five stages of curl development which are recognized in the Swakara industry as: Galliac broadtail characterized by super fine pattern with short fibre length; broadtail often characterized by watered-silk patterns; flat characterized by generally flat appearance with a slightly raised pattern; semi flat has a raised pattern with flat spaces in between and lastly, curl which has a raised pattern without flat spaces however, with short to medium fibre length. Campbell (2007) demonstrated the four main curl types as the water silk, shallow, developed shallow and pipe curl. However, short descriptions in order to cater for the four main curl types are used to minimize bias during pelt sorting and grading. These include; Galliac (GAL), Water-silk galliac (WS/GAL), Shallow watered-silk (VL/WS), Shallow (VL), Shallow developed (VO), Developed shallow (OV), Developed pipe curl (OV/PK) and pipe curl.

Hair length in Swakara sheep has become shorter over the past 50 years due to improved breeding and is measured using a ruler which is pushed into the hair forcing the hair to rise up against the ruler. Furthermore, hair length is a vital component during grading such that, for every millimeter gained the leather thickens by adding weight and bulkiness when manufacturing garments and directly affects pelt price at the Copenhagen auction in Denmark. Hence, uniformity in pelts graded enhances easy manufacturing of various end products. Fibre formation is influenced by follicle arrangement and there are two main formations, which are drawn and lyre.

Drawn is defined as straight fiber whilst lyre is the S-fiber formation (www.swakara.net). Simplified results from hair length exercise in September 1949 (Table 2.1).

Table 2.1: Hair length characteristic exercise of Swakara (then Karakul) pelts in 1949

Classification type	**Description**	**Millimeter (s)**
D	Moiré	6
S	Galliac	6.3
C	Ribbed/ broken curl	7.3
B	Flat broadtail	7.5
Q	Small curl	7.5
A	Broadtail	8
J/M	Small curl	8.3
O	Large flat curl	8.5
G	Medium curl	9.1
K	Large curl	10.2
N	Extra-large curl	10.6
RF	Overgrown flat	11
RC	Overgrown curl	13

Source: Karakul Annual report, 2010

The current hair length scale (in millimeter) used in the production of the Swakara modern assortment for the market is presented in Table 2.2 (Karakul Annual report, 2010). Campbell (2007) noted that breeding practices have improved from the old pipe-curl pelt types that were popular with the fashion industry in the 1950's and 1960's has reduced to a less developed type like water-silk that is currently on demand. Table 2.3 relates the direct relationship of typical weight per pelt, in grams, for each regular hair length stage (Karakul Annual report, 2010).

Table 2.2: Hair length grading in Swakara sheep

Hair length	Millimeter (s)	Grade description	Extreme line
Premature	0-2	Reject	Shortest
Under-developed	2-3	Low grade	
Extra short	3-4	Regular	
Short	4-5		
Medium	5-6		
Long	6-9		
Overgrown	9-19	Low grade	
Outgrown	Over 19	Reject	Longest

Source: Karakul Annual report, 2010

Table 2.3: Pelt weight relationship to hair length

Hair length	Pelt weight in grams
Extra short	100
Short	120
Medium	130
Long	145

Source: Karakul Annual report, 2010

2.1.6 Keratin protein classification

Early attempts made in identifying and classifying wool proteins was in 1935 (Goddard and Michaelis, 1935) and it divided the major wool components into two classes; S-carboxy methyl keratine A (SCMK-A) and S-carboxy methyl keratine B (SCMK-B) as cited by Gong *et al*. (2016). These were classified as lower and higher sulphur content than the average sulphur content in wool. However, with time the SCKM-As were found to be the hair and wool alpha-keratins on the other hand the SCMK-Bs were the KAPs (Powell and Rogers, 1997). Furthermore, keratins also referred to as KRT nomenclature were last reviewed in 2006 whereas the KAPs nomenclature has not been reviewed since the proposition by Rogers and Powell in 1993 and later explained in 1997. Therefore, an updated naming system ought to be done to accommodate the complexity of these Keratins and KAPs (Gong *et al*., 2012). Keratin proteins are a major component of wool and hair fibre which confers their structural and mechanical properties (Zhao *et al*., 2016; Powell, 1997). Hair and wool fibres are produced by the wool/hair follicle bulb which typically consist of three major structural components; cuticle, cortex and sometimes central medulla (Zhao *et al*., 2016).

The protein gene family is characterized into two groups which are the Keratin Intermediate Filament proteins (KRTs) and the Keratin Associated proteins (KAPs) (Zhang *et al.,* 2011). Steinert *et al.* (1994) elucidates that these keratin proteins can be further divided into two major families, the Type I (acidic) keratins and the Type II (basic-neutral) keratins. The KRTs consist of higher ordered copolymers of individual, Type I and Type II family members. Furthermore, protein properties allow an additional division of the proteins of both families into epithelial (cyto-) keratins and hair (trichocytic- or 'hard') keratins (Gong *et al.,* 2016; Langbein and Schweizer, 2005). These epithelial keratins often possess head and tail domains rich in glycine and serine as they exhibit polymorphic variants due to differences in the number of amino acids repeats of the form GGX (Korge *et al.,* 1992a; Korge *et al*., 1992; Hanukoglu and Fuchs, 1983). In the wool fibre, these proteins are linked in a highly organized manner. Approximately 90% of the cortical cells contain longitudinally arrayed KRTs (Gong *et al*., 2012). The KRTs form the skeletal structure of the wool fiber and forms eight to ten nano-meters diameter filaments embedded in a matrix of KAPs through extensive disulphide cross-linkages (Gong *et al*., 2012; Powell, 1997; Rogers *et al*., 1994) between cysteine in the KAPs and in the head and tail domains of the keratins (Powell and Rogers, 1997).

Whilst the KAPs may have little or no distinct effect on KRTs structure, they play an important role in defining the physic-mechanical properties of the micro-fibrils (wool fibre) (Gong *et al*., 2016). Clark *et al.* (2017) illustrated that global analysis of gene expression across multiple tissues has aided genome annotation and supported functional annotation of mammalian genes. Lately our understanding of KAPs has progressed with the advent of the large scale whole genome sequencing of human KAP genes. A widespread genetic variation knowledge in humans and sheep KAP genes is widely discussed (Fuchs and Weber, 1994). The role of KAPs in KRTs is to assembly into arrays considered crucial and therefore most likely affect wool with intercellular connection via desmosomes which gives rise to hair attributes such as strength, inertness and rigidity (Matsunanga *et al*., 2013; Yu *et al*., 2011; Shimomura and Ito, 2005; Powell and Rogers, 1997). The KAPs were originally best understood in sheep reflecting the then economic importance of wool and the majority of wool protein biochemistry undertaken from mid-twentieth century (Gong *et al*., 2012). Wool KRT_{1s} are acidic and vary in size from 392-416 amino acids base pairs, whilst the KRT_{2s} are neutral and range in size from 479-506 amino acids base pairs. Their end domains however, have many cysteine residues to allow them to cross-link to the proteins of the wool matrix (KAPs) through disulphide bonds.

The matrix consists of KAPs typically encoded by a single exon (Powell and Rogers, 1997) and has been divided into three groups based on their amino acid compositions: High-sulphur (HS), Ultra-high sulphur (UHS) and High-glycine-tyrosine (HGT) KAPs. Further, classified into three broad groups according to their amino acid composition the HS (≤30 mol% cysteine content), the UHS (≥30 mol% cysteine content) and the HGT (30-60 mol% glycine and tyrosine) (Gong *et al*., 2016). Furthermore, the KAPs are encoded by small intron-less genes called KRTAPs (Wang *et al*., 2019) and they usually occur in gene clusters (Rogers *et al*., 1994). The HGT keratin associated proteins are primarily found in the orthocortex of wool fibre and first to be expressed in active wool follicles after synthesis of the KIFs (Li *et al*., 2019a; 2019; 2017a; Rogers *et al*., 2006). Gillespie (1990) explains that these HGT-KAPs differ considerable in abundance among and within sheep breeds.

The KAP one family members are quite similar to each other and slightly differ mainly in the number of conserved tandem deca-peptide such as "QTSCCQPXXX" repeats. Often between three to five deca-peptide repeats in KAP1-1, whereas KAP1-2, KAP1-3 and KAP1-4 have three, two, and five repeats respectively (Gong *et al*., 2011; Gong *et al*., 2010; Itenge-Mweza *et al*., 2007). Gong *et al*. (2016) illustrates that the KAP1-1 and KAP1-4 are classified under acidic KAPs whereas KAP1-2 and KAP1-3 are neutral. There are three major (BIIIB2, BIIIB3, BIIIB4) and one minor (BIIIBI) proteins described for ovine KAP3. Zhou *et al*. (2016) illustrates that located on chromosome one are the KAP 6 family genes which are quite diverse and categorized under the HGT-KAP group. Five gene members exist; KRTAP6-1, KRTAP6-2, KRTAP6-3, KRTAP6-4, KRTAP6-5 in sheep. They are further clustered with other HGT-KAP genes (KRTAP7-1, KRTAP8-1, KRTAP8-2, KRTAP20-1 and KRTAP20-2).

2.1.7 Keratin Associated Protein sub-families

To date, 28 KAP families in mammalian species have been identified in over 100 KAP genes (Bai *et al.*, 2018; Bai *et al*., 2019; Ekegbu *et al*., 2018; Gong *et al.*, 2016; Gong *et al*., 2011b; Rogers and Schweizer, 2005), which is based on their amino acid sequence composition (Gong *et al.,* 2016). Khan *et al.* (2014) illustrated that KAP subfamilies were found to exhibit molecular variation within mammals resulting in a variety of hair phenotypes. Nonetheless, sequence homology evaluation between the KAP from different species adds obscurity to KAP naming as evident homologues are at times difficult to find. Although, they may be genetically identical they can also be different in their nucleotide sequences and chromosomal arrangement, in these genes (Gong *et al*., 2019; Wang *et al.*, 2017).

The 27 recognized wool KAPs have been identified in sheep and assigned to 13 families (Li *et al*., 2017; Wang *et al*., 2017; Gong *et al*., 2016). The KAP genes identified have been mapped to and located on three chromosomes. KAP1-n, KAP3-n and KAP4-n are on located on OAR11 (Gong *et al*., 2012; Gong *et al*., 2011a; McLaren *et al*., 1997) the genes for the KAP6-KAP8, KAP11, KAP13 and KAP24 families are located on OAR1 (Li *et al*., 2017; Zhou *et al*., 2016; Gong *et al*., 2014; Zhou *et al*., 2012) and lastly the KAP5-n genes are located on OAR21 (Gong *et al*., 2012; McLaren *et al*., 1997). Three gene loci; KAP1.1 (B2A), KAP1.3 (B2C) and KRT33A formerly known as KRT1.2 have been shown to be tightly linked (Rogers, 1994) and chromosome 11q is the site of location of these three loci (Hediger *et al*., 1991).

2.1.8 Genetic markers

Nicholas (2010) defines polymorphism as the existence of many different forms of alleles which arose due to mutations that may occur when deoxyribonucleic acid replicates. Therefore, Researchers take advantage of these different forms, in order to identify genetic markers. A genetic marker for a particular trait can be defined as a piece of DNA that has genetic variation and affects a phenotype, or a piece of DNA that is closely linked to another piece of DNA that affects a phenotype (Itenge, 2012). Genetic markers are further defined as chromosomal landmarks used in the identification of specific regions in a gene; which have an inheritance pattern that can be traced (Kumar *et al.,* 2006). Commonly used genetic markers include Restriction Fragment Length Polymorphism (RFLPs) (Kumar *et al*., 2006), Polymerase Chain Reaction-Single Strand Conformational Polymorphism (PCR-

SSCP) and Single Nucleotide Polymorphism (SNPs) (Muchadeyi *et al*., 2015). Some reports on application of genetic markers in sheep breeding programmes include; Footrot gene-marker test developed at Lincoln University, New Zealand; Cold tolerance gene-marker test; FecB and FecX(I) for the high prolificacy in sheep carrying the Booroola fecundity gene (FecB) and Inverdale sheep (FecX(I) (Davis *et al.*, 2002). Hence, genetic variation in the keratin genes may be associated with pelt quality traits which are determined by the unique hair formation.

CHAPTER 3

MATERIALS AND METHODS

3.1.0 Research design

The study was conducted to identify genetic variation in the Keratin-associated proteins (KAP1.1, KAP1.3 and KAP3.2) and one Keratin intermediate filament protein (KRT33A) genes. Identification of genetic variation in the four keratin genes was a qualitative approach, whereas genetic variation observed in PCR-Agarose gel electrophoresis for the KAP1.1, and through PCR-Single strand conformational polymorphism account for quantitative data. Figure 3.1 illustrates the research procedure flow chart.

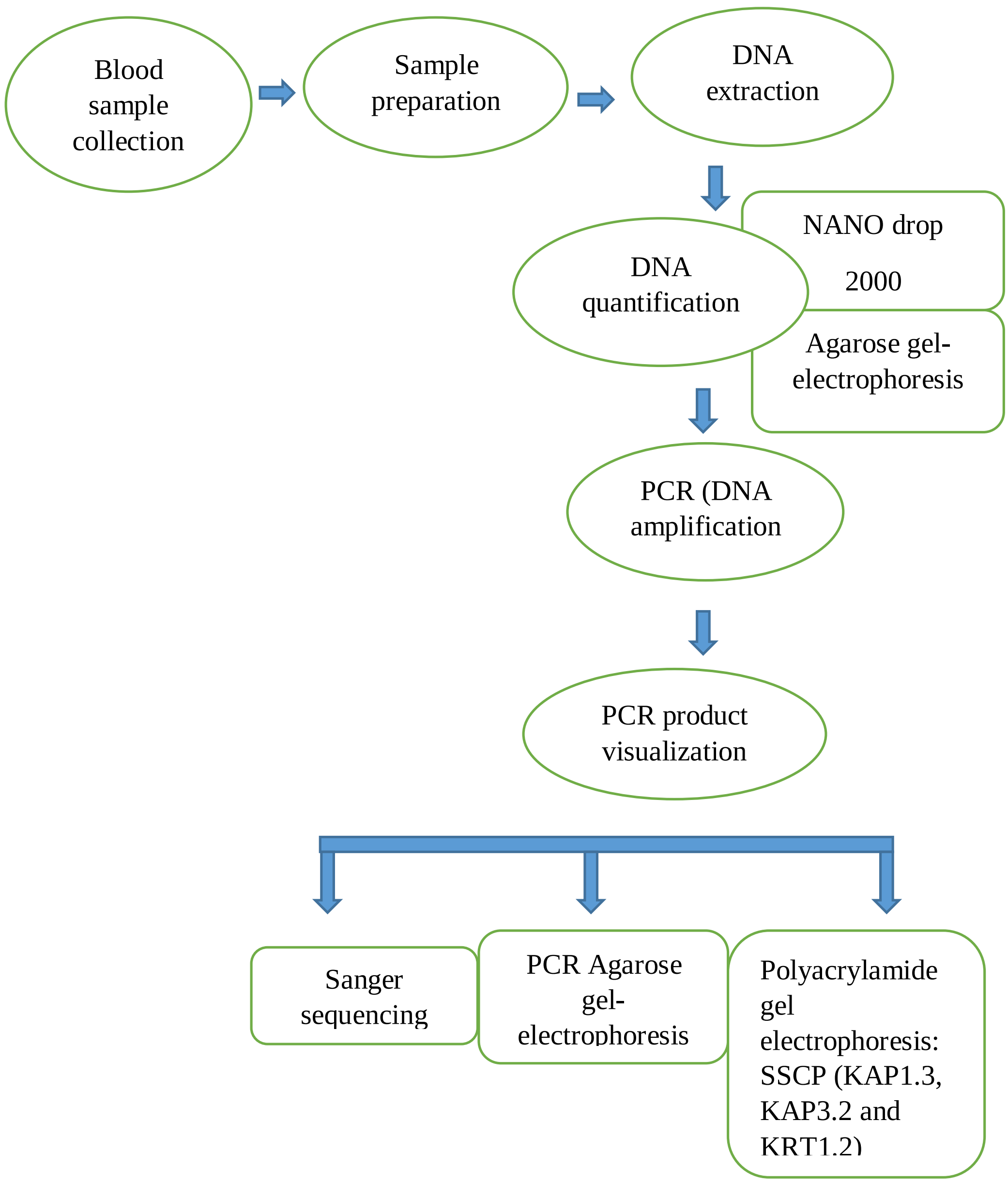


Figure 6.1: Study Procedure Flow Chat

3.1.1 Study area

The study was carried out at Neudamm farm in the Khomas region (22.300 S, 17.22º E), Gellap-Ost research station in Karas region (26.27º S, 18.05º E), Tsumis research station is located in Rehoboth (23.730 S, 17.190E) and Kalahari research station is located in Mariental (25.550S, 16.390E) in Hardap region. The study areas are shown on the map in Figure 3.2.

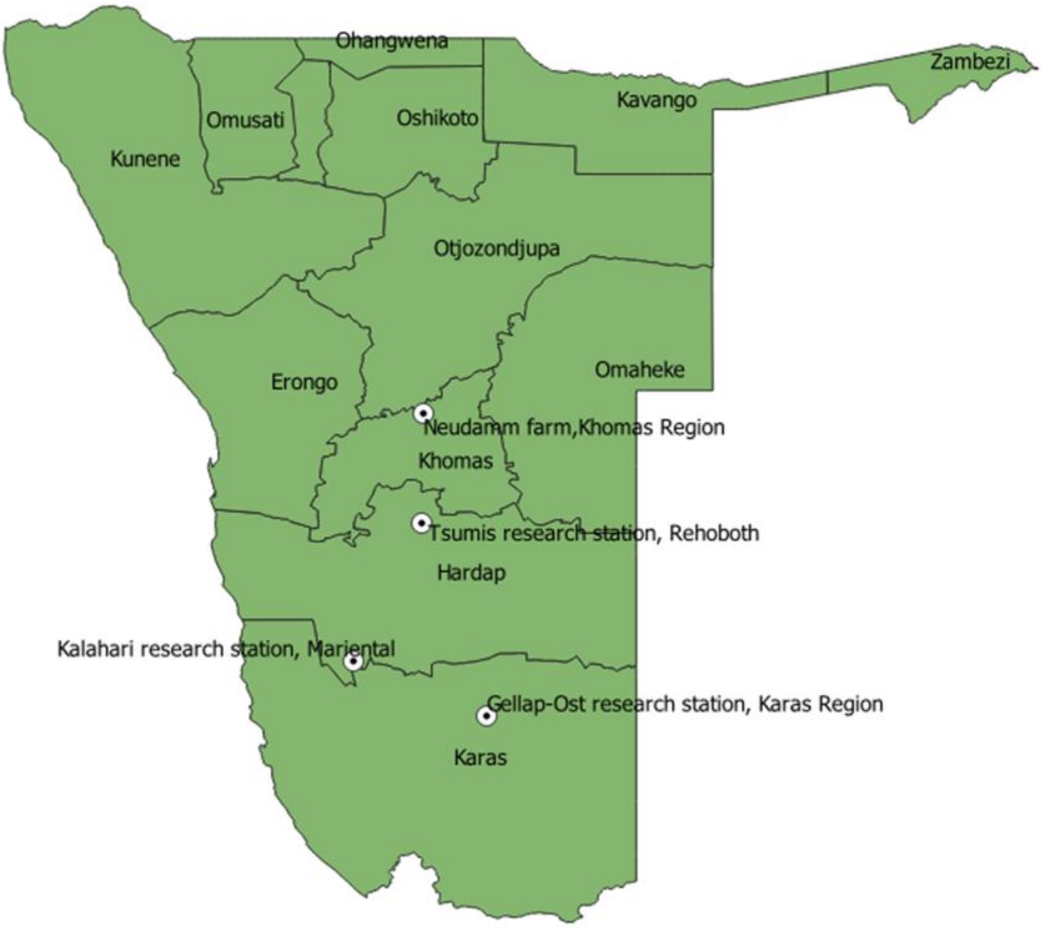


Figure 3.2: Namibian map, depicting the four study sites. The sampling sites map for the Swakara sheep situated in three regions illustrated by a circle with a black dot was drawn using QGIS guidelines.

https://www.qgis.org/en/site/.

3.1.2 Sample size

A total of 205 Swakara sheep were initially randomly sampled and used to access genetic variation in the study. Genomic DNA yield was quantified using a NANO drop and Agarose gel electrophoresis. In addition to that, poorly amplified amplimers were indefinitely disqualified from the study. Hence, the study used a total of 168 Swakara sheep randomly selected from four farms; 46 from Neudamm farm, 46 from Gellap-Ost farm, 45 from Kalahari farm and 31 from Tsumis farm. The herd comprised of 136 ewes, 31 rams and one unknown. Swakara sheep with age range of six months to seven years old rams and ewes were used in the study.

3.1.3 Procedure: Blood sample collection

Approximately four mL of blood sample was withdrawn from each Swakara sheep via the jugular vein puncture in K2E-EDTA anti-coagulant vacutainers tubes. Blood samples were immediately placed in a cooler box with ice packs and transported to the laboratory. One hundred µL blood samples were aliquotted into Eppendorf tubes to minimize contamination prior to refrigeration at 4ºC pending molecular genetics techniques.

3.1.4 Genomic DNA isolation

Genomic deoxyribonucleic acid (DNA) was isolated using the Inqaba biotech-kit protocol. Purity and concentration of genomic DNA was determined by using a Spectrophotometer (Nano-drop 2000). Genomic DNA quality was assessed by using 1% horizontal Agarose gel-electrophoresis stained with 0.1 mg/L ethidium bromide and analyzed at 90 volts for 45 minutes as depicted on Figure 3.3. Agarose gel was then examined on the UV trans-illuminator (Syngene bio imaging, Cambrige, United

Kingdom) for checking of DNA quality. The genomic DNA samples with good quality shown by intact bands were used for further analysis.

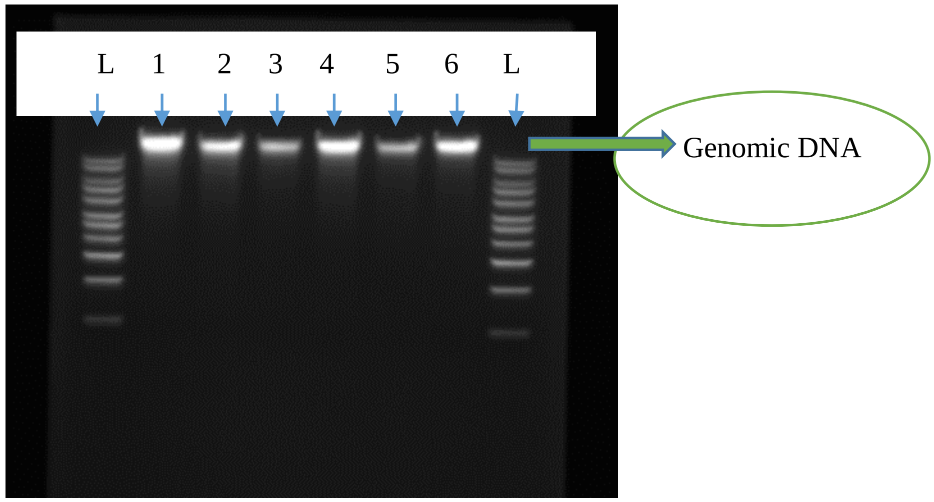


Figure 3.3: DNA quality: Agarose gel-electrophoresis

DNA quality checked on 1% Agarose gel-electrophoresis on six randomly selected Swakara sheep and, analyzed at 90 volts for 45 minutes. The letter L on lane one and lane eight denotes the 100 bp molecular weight marker, and lane two to seven are KAP1.1 genomic DNA samples of the six Swakara sheep sampled at Neudamm farm.

DNA quantification using a spectrophotometer (NANO drop 2000). Genomic DNA samples with NANO drop Optical Density$_{260/280}$ values between 1.8 and 2.0 were subjected to further analysis.

3.1.5 Primers for PCR amplification

The primers used to amplify the specified KAP1.1, KAP1.3, KAP3.2 and KRT33A loci were obtained from the literature, and are summarised in Table 3.1. All primers were synthesised by Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa.

Table 3.1: Primers used to amplify the specific regions

Locus	Primer sequence	Source
KAP1.1	5'-CAACCCTCCTCTCAACCCAACTCC-3' 5'-CGCTGCTACCCACCTGGCCATA-3'	Itenge-Mweza *et al.* (2007)
KAP1.3	5'- GGGTGGAACAAGCAGACCAAACTC-3' 5'- TAGTTTGTTGGGACTGTACACTGGC-3'	Rogers *et al.* (1994)
KAP3.2	5'- CGAGACACCAAGACTTCTCTCATC-3' 5'-AGTGAGTGTTGAAGGCCAGATCAC-3'	McLaren *et al.* (1997)
KRT33A	5'- CACAACTCTGGCTTGGTGAACTTG-3' 5'- CTTAGCCATATCTCGGATTCCCTC-3'	Rogers *et al.* (1993)

3.1.6 PCR Amplification

The protocols for amplification were adapted from Itenge-Mweza *et al.* (2007). Amplification consisted of an automated preheating Thermo cycler (Thermo Scientific, Arktik thermal cycler) lid at 110°C for two minutes, an initial denaturation of one minute at 95ºC, followed by 30 cycles of denaturation at 95ºC for one minute, annealing at the temperature specified in Table 3.2 for 1 min and extension at 72 ºC for 1 min, with a final extension of 72 ºC for 7 min. After the execution of PCR program the PCR products were stored at 4ºC.

Table 3.2: Optimized PCR annealing temperatures and predicted amplimer sizes for each locus investigated

Locus	Annealing temperature (ºC)	Amplimer size (base pairs)
KAP1.1	65	311
KAP1.3	65	598
KAP3.2	58	424
KRT33A	65	480

3.1.7 Amplification of the loci using PCR

PCR amplifications were performed in 25 µL reactions containing 50 ng genomic DNA from whole blood, 1X One Taq Master Mix (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa), 200 mM of forward primer, 200 mM of reverse primer, and nuclease free water (New England BioLabs@ Inc.).

3.1.8 Agarose gel electrophoresis

Amplimers were analysed in 1.0% w/v Agarose gels (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa) prepared with 0.5X Tris-Acetate EDTA (TAE) buffer (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa), containing 0.1 mg/L of ethidium bromide. Seven µL of PCR product was mixed with 1.8X loading dye (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa). A 100 bp molecular weight marker (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa) was included for fragment size comparison, and DNA bands were visualized under the UV trans-illuminator (Syngene bio imaging, Cambrige, United Kingdom).

3.1.9 Detection of length variation in the KAP1.1 gene

Amplimers were analysed in 2.0% w/v Agarose gels (New England BioLabs@ Inc.) prepared with with 0.5X Tris-Acetate EDTA (TAE) buffer, containing 0.1 mg/L of ethidium bromide. Seven µL of PCR product was mixed with 1.8X loading dye (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa) and the gels were electrophoresed at a constant voltage of 90 V for 90 minutes, prior to visualisation with UV trans-illuminator (Syngene bio imaging, Cambrige, United Kingdom).

3.2.0 Detection of sequence variation using Polymerase chain reaction-Single Strand Conformational Polymorphism (PCR-SSCP)

In exploring genetic variation in the KRT33A, KAP3.2 and KAP1.3 loci, amplified PCR products were subjected to PCR-Single Strand Conformational Polymorphism (SSCP) analysis. Polyacrylamide gels (38:2 acrylamide/ bis-acrylamide, (Bio-Rad Laboratories, Ltd, Johannesburg, South Africa) vertical gels (Protean II 16 x 16 cm, 1.0 mm thick spacers, 24 well comb, Bio-Rad Laboratories, Ltd, Johanesburg, South Africa) were prepared containing 1X TBE, 7 M of Urea, 30% ammonium persulphate (APS) and 30 µL TEMED (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa).

The polyacrylamide gels were pre-run for an hour prior to loading the PCR amplimers to remove any polar impurity, in the vertical gel electrophoresis tank. The PCR products were mixed with 50 µL loading dye (95% formamide, 10 mM NA_2EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol) (Inqaba Biotechnical Industries (Pty)

Ltd, Pretoria, South Africa), denatured by firstly preheating the Thermo-cycler lid (Thermo Scientific, Arktik thermal cycler) for two minutes followed by denaturing at 95ºC for five minutes and immediately placed samples on wet ice before loading 10 µL aliquots. The polyacrylamide gels were run on 1X TBE (Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa) running buffer at 200 V between 16 and 17 hours (Table 3.3). Furthermore the DNA bands on PAGE were visualized by the silver staining technique according to Itenge (2007) method adapted from Sanguinetti *et al*. (1994) with minor modifications like inclusion of urea in PAGE preparation and use of ‘stop’ solution is optional as the developer solution does not darken PAGE, and illustrated on Appendix E.

Table 3.3: Optimised SSCP conditions for the loci investigated

Locus	Polyacrylamide gel % (38:2)*	Time (Hours)	Voltage (V)
KAP1.3	10	16 hours 30 minutes	200
KAP3.2	10	17 hours	200
KRT33A	10	16 hours 30 minutes	200

* (38:2) the ratio of Acrylamide/ Bis-acrylamide

The polyacrylamide gels in the study had a better resolution and visibility when run between 16-17 hours. However, PAGE run below or above the standard time had poor visibility, mostly smearing was observed when samples were subjected to prolonged hours and poor fragment separation with less time.

3.2.1 Genotyping for each locus

KAP1.1

Each animal was genotyped by comparison to allele standards generated on agarose gels that also had a 100 bp molecular weight marker inorder to show the 311 bp fragment of the KAP1.1 gene, which was labelled as the B allele. Animals with the A allele could be observed through the fragment size of 341 bp, while the animals with the C allele had a fragment size of 281 bp.

KAP1.3, KAP3.2 and KRT33A

Each animal was genotyped by firstly confirming the presence or absence of polymorphism prior to grading the samples. As a result of PAGE gels being fragile during handling all grading was done while in the 'Stop' solution as illustrated in Appendix E, with respect to each gene under study.

3.2.2 Deoxyribonucleic acid (DNA) Sequencing

Allelic variation was confirmed through sequencing twenty-six randomly selected PCR products in both forward and reverse direction using primers on Table 3.1, at the University of Porto in Portugal. For the KAP1.1 locus only six samples were sequenced to determine PCR amplimer length variation and 20 PCR products were sequenced at the specified loci (KAP1.3, KAP3.2 and KRT33A). PCR products of each reaction were sequenced following the ABI Prism BigDye^{TM} Terminator v3.1 Sequencing Kit protocol on an ABI3130xl DNA Analyser (Applied Biosystems, Foster City, California, USA).

3.2.3 Statistical analysis

All data was gathered and consolidated on Microsoft excel 13 prior to data analysis. Chi-square test of association at 5% level of significance was used to test for allele frequency distribution across all four farms; Neudamm, Kalahari, Tsumis and Gellap-Ost. The Spearman's Rank correlation formulae was used to calculate genotype frequencies of the four loci under study; KRT33A, KAP1.1, KAP3.2 and KAP1.3. Frequency distribution of the different alleles in Swakara sheep were tested for Hardy Weinberg Equilibrium.

The Chi square test of association equation used:

$$x_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where; $_c$ is the degrees of freedom

O_i is the observed allele frequencies

E_i is the expected allele frequencies

$\sum$ is the summation of each data entry

χ^2 is the critical Chi-square value

Spearman's Rank correlation: Allele frequency distribution:

A= [2 (AA) + (AB) + (AC)]/ 2(sample size)

B= [2 (BB) + (AB) + (BC)]/ 2(sample size)

C= [2 (CC) + (AC) + (BC)]/ 2(sample size)

Hardy Weinberg Equilibrium equations

$p^2 + 2pq + q^2 = 1$

$p^2 + q^2 + r^2 + 2pq + 2qr + 2pr = 1$

Where p, q, r denote the different alleles per gene

(https://www.icalcu.com/stat/chisqtest.html)

3.2.4 Research ethics

An ethical clearance with FANR/003/2019 as a reference number was obtained from the University of Namibia. Veterinarians and animal health technicians collected blood samples from the Swakara sheep on the four farms under study. All animals were handled humanly.

CHAPTER 4

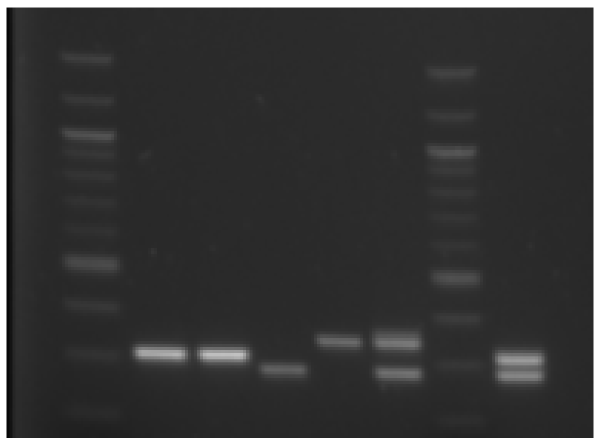
RESULTS

4.1.0 Keratin Associated Protein (KAP1.1) gene

Three unique banding patterns that varied in length were observed for the KAP1.1 amplimer using PCR-Agarose gel electrophoresis typing methods. These were designated as A, B and C (Figure 4.1). Sequence analysis revealed that the length of each of these AFLP bands were 341 bp, 311 bp and 281 bp, respectively. Sequencing analysis confirmed that the length variation was the result of an insertion or deletion of a 30 bp region of the sequence. Table 4.1 depicts the genetic diversity across the four farms under study within the KAP1.1 gene of Swakara sheep. The Chi-square calculated value 32.28 was statistically significant (P=0.004). Hence it was inferred that the Swakara sheep population was in Hardy Weinberg Equilibrium with respect to the KAP1.1 gene. In this study, the B allele had a higher allele frequency (0.57%) in comparison to the A and C allele frequencies (Table 4.1).

Table 4.1: Genotype and allele frequencies of the KAP1.1 gene in four geographic locations

Locus	**Breed**	**Geographic location (Farm)**	**Farm total**	**Genotype frequency**	**Allele frequency %**	**Chi-square**
KAP1.1	Swakara	Gellap-Ost	46	AA = 2	A = 0.13	32.28
		Kalahari	45	AB = 25	B = 0.57	
		Neudamm	46	AC = 15	C = 0.3	
		Tsumis	31	BB = 58		
				BC = 50		
				CC = 18		

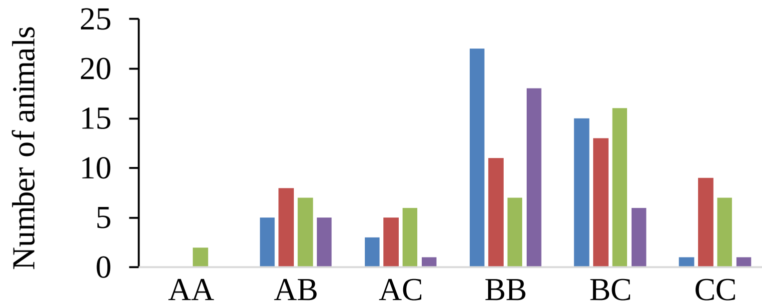


1 2 3 4 5 6 7 8

L BB BB CC AA BC L BC

Figure 4.1: Length polymorphism within the KAP1.1 amplimer. Separated in a 2% Agarose-gel electrophoresis at 90 volts for 95 minutes. Lanes one and seven contain a 100bp molecular weight marker (L), lanes two to six and eight shows possible allelic variation within the KAP1.1 locus among Swakara sheep, where lane two and three denote the homozygous B allele, lane four denotes the homozygous C allele, lane five the homozygous A allele, lane six and eight denote the heterozygous BC allele.

A graphical representation and distribution of the three alleles identified in the study with respect to each farm as shown on Figure 4.2. The AA genotype was found only at Kalahari farm and it was observed that the B allele was highly distributed across the four farms.



	AA	AB	AC	BB	BC	CC
Farm Neudamm	0	5	3	22	15	1
Farm Gellap-Ost	0	8	5	11	13	9
Farm Kalahari	2	7	6	7	16	7
Farm Tsumis	0	5	1	18	6	1

Figure 4.2: KAP1.1 Allele frequency distribution on four farms Gellap-Ost, Kalahari, Neudamm and Tsumis

Figure 4.3 depicts the phylogenetic tree for the KAP1.1 locus identified in sheep. The tree was constructed using the amino acid consensus sequences. The Swakara sheep are indicated as ‘Sample and a number assigned afterward’ whereas ‘FJ; L; MG; XM and a number assigned afterward’ denote the predicted reference sequences for comparison obtained from the NCBI. KAP1.1 Sheep sequence (XM024980561.1) was used as an out-group. The sequences were consolidated and drawn using MEGA software (Version 6).

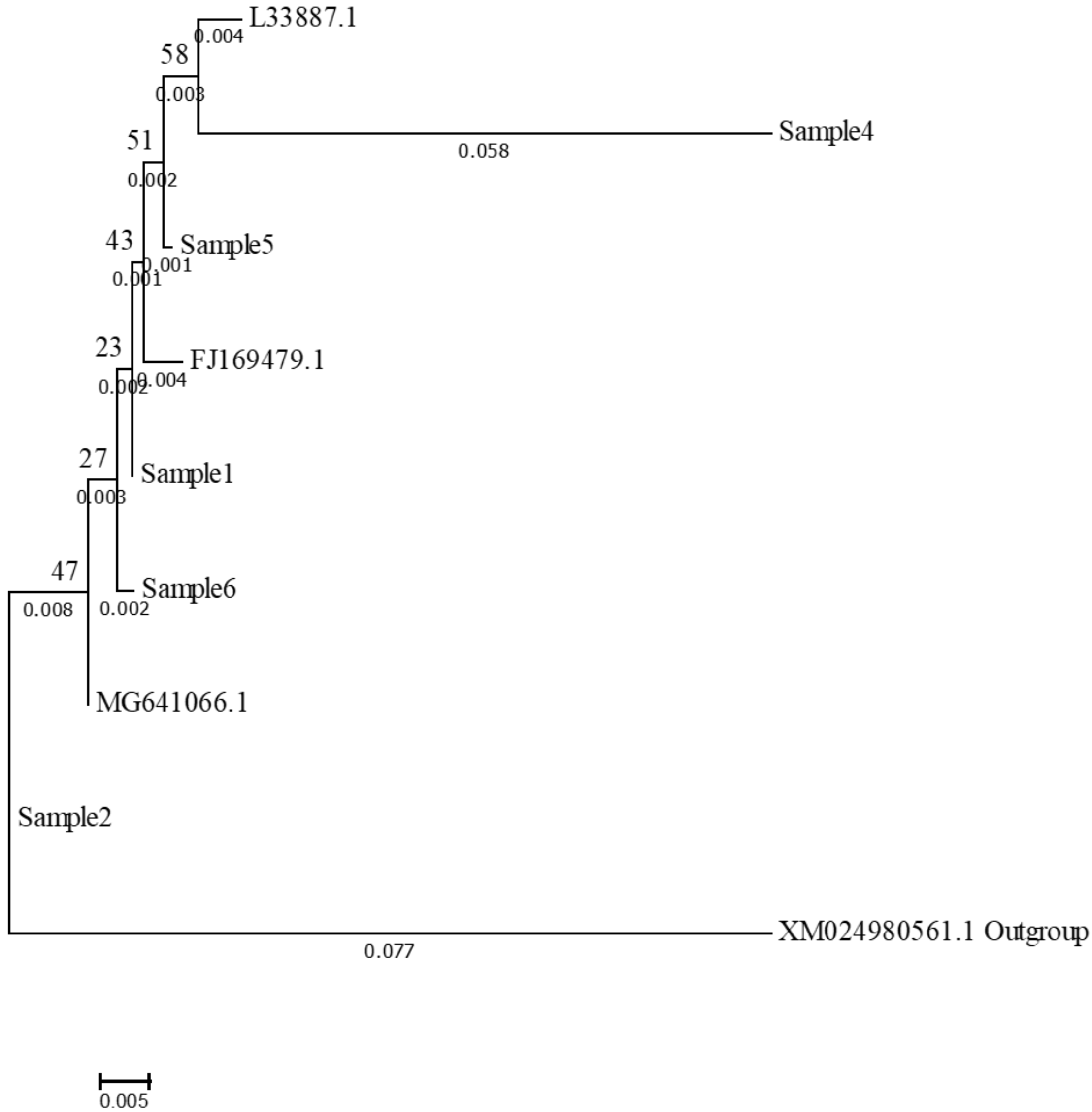
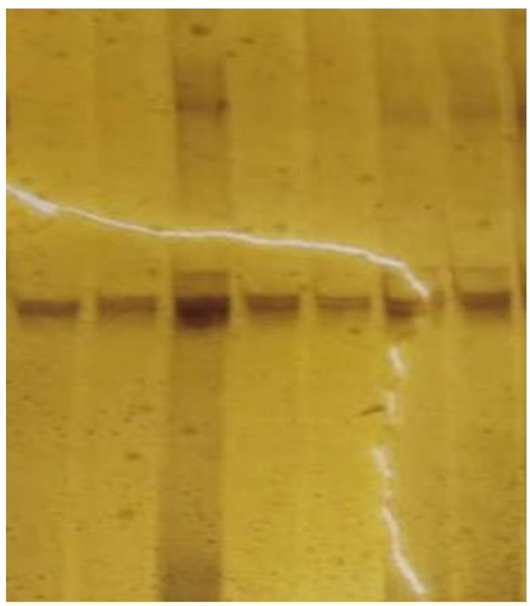


Figure 4.3: The evolutionary history was inferred using the Neighbor-Joining method, based on a 322 bp sequence and NCBI sequences were used for comparison in the KAP1.1 locus, in Swakara sheep. The optimal tree with the sum of branch length = 0.14911027 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The analysis involved 9 nucleotide sequences. Evolutionary analyses were conducted in MEGA version 6.

4.1.1 Keratin Associated Protein (KAP1.3) gene

Figure 4.4 depicts polymorphism of ovine KAP1.3 gene and three variants (AA, CC and AB) in both homozygous and heterozygous forms were detected by PCR-SSCP. An overall sample size of 20 was used to identify polymorphism. Three KAP1.3 alleles were found 10 (AA), 7 (AB) and 3 (CC) with allele frequencies of 0.5, 0.35 and 0.15, respectively. The Chi-square value was statistically significant at *P*=0.160, and was in Hardy Weinberg Equilibrium.



1	2	3	4	5	6	7
AA	AA	AB	CC	CC	AB	AB

Figure 4.4: SSCP analysis of the 570 bp amplimer of the KAP1.3 gene, in Swakara sheep and electrophoresed on 10% acrylamide/bis-acrylamide gel that contained 7 M concentration of urea run for 16 hours 30 minutes at 200V at room temperature. Lane one to lane seven are different genotypes at the KAP1.3 gene.

Figure 4.5 depicts the phylogenetic tree for the KAP1.3 locus identified in Swakara sheep. The tree was constructed using the amino acid consensus sequences. The Swakara sheep are indicated as 'Sample and a number assigned afterward' whereas AY; and a number assigned afterward' denote the predicted reference sequence for comparison obtained from the NCBI. KAP1.3 Sheep sequence (AY835593.1) was used as an out-group. The sequences were consolidated and drawn using MEGA software (Version 6).

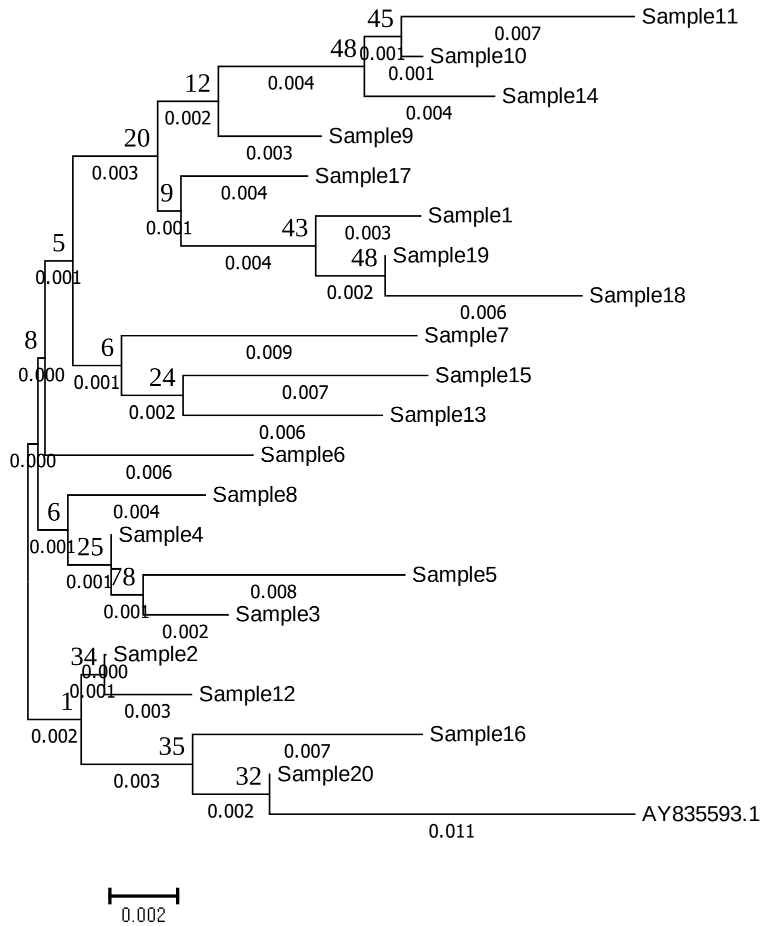


Figure 4.5: The evolutionary history was inferred using the Neighbor-Joining method, based on a 570 bp sequence and NCBI sequences were used for comparison in the KAP1.3 locus, in Swakara sheep. The optimal tree with the sum of branch length = 0.11929587 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The analysis involved 21 nucleotide sequences. KAP1.3 (AY835593.1) Sheep was used as an out-group/root of the tree. Evolutionary analyses were conducted in MEGA version 6.

Figure 4.6 illustrates the allele frequency distribution with the KAP1.3. Overall the AA genotype was highly distributed when compared to the AB and CC across the four farms under study. Gellap-Ost farm only had an A allele frequency distribution whilst Neudamm, Kalahari and Tsumis had allelic diversity.

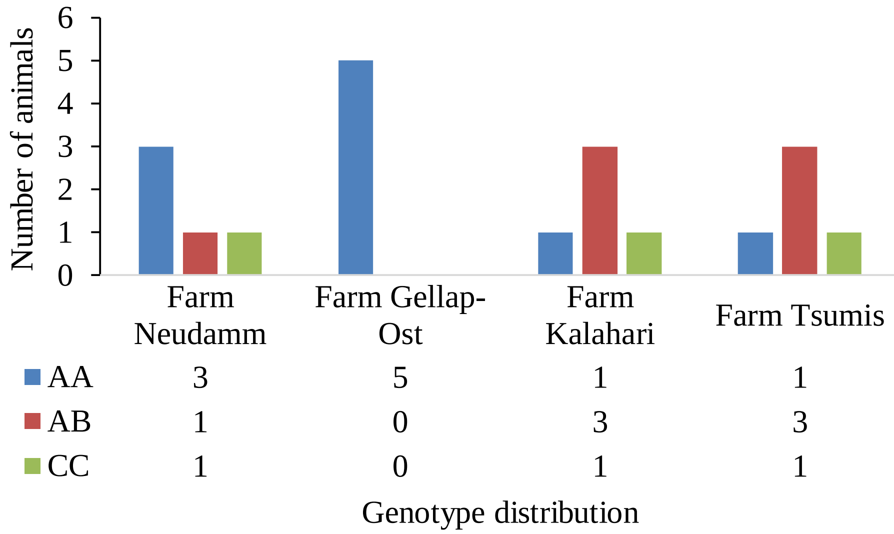
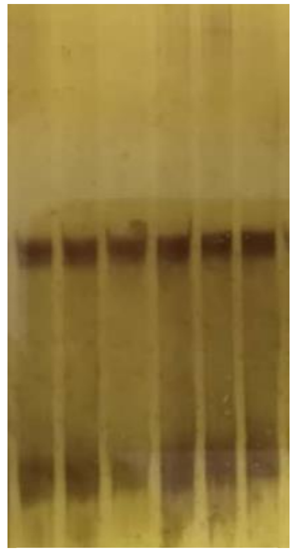


Figure 4.6: KAP1.3 Allele frequency distribution

4.1.2 Keratin Associated Protein (KAP3.2) gene

Figure 4.7 depicts a monomorphic distribution of ovine KAP3.2 across all four farms under study and one homozygous variant (AA) form was detected by PCR-SSCP. An overall sample size of 20 was used to identify the monomorphic pattern with a homogenous allele frequency of 100%.



AA genotype

Figure 4.7: All samples genotyped as AA, and indicated from lane one to six. SSCP analysis of the 426 bp amplimer of the KAP3.2 gene, in Swakara sheep and electrophoresed on 10% acrylamide/bis-acrylamide gel that contained 7 M concentration of urea run for 17 hours at 200V at room temperature.

Figure 4.8 depicts the phylogenetic tree for the KAP3.2 locus identified in sheep. The tree was constructed using the amino acid consensus sequences. The Swakara sheep are indicated as ‘Sample and a number assigned afterward.’ Whereas AY, and number assigned afterward’ denote the predicted reference sequences for comparison obtained from NCBI. KAP3.2 (M21099.1) Sheep sequence was used as the root of the tree. The sequences were consolidated and drawn using MEGA 6 software.

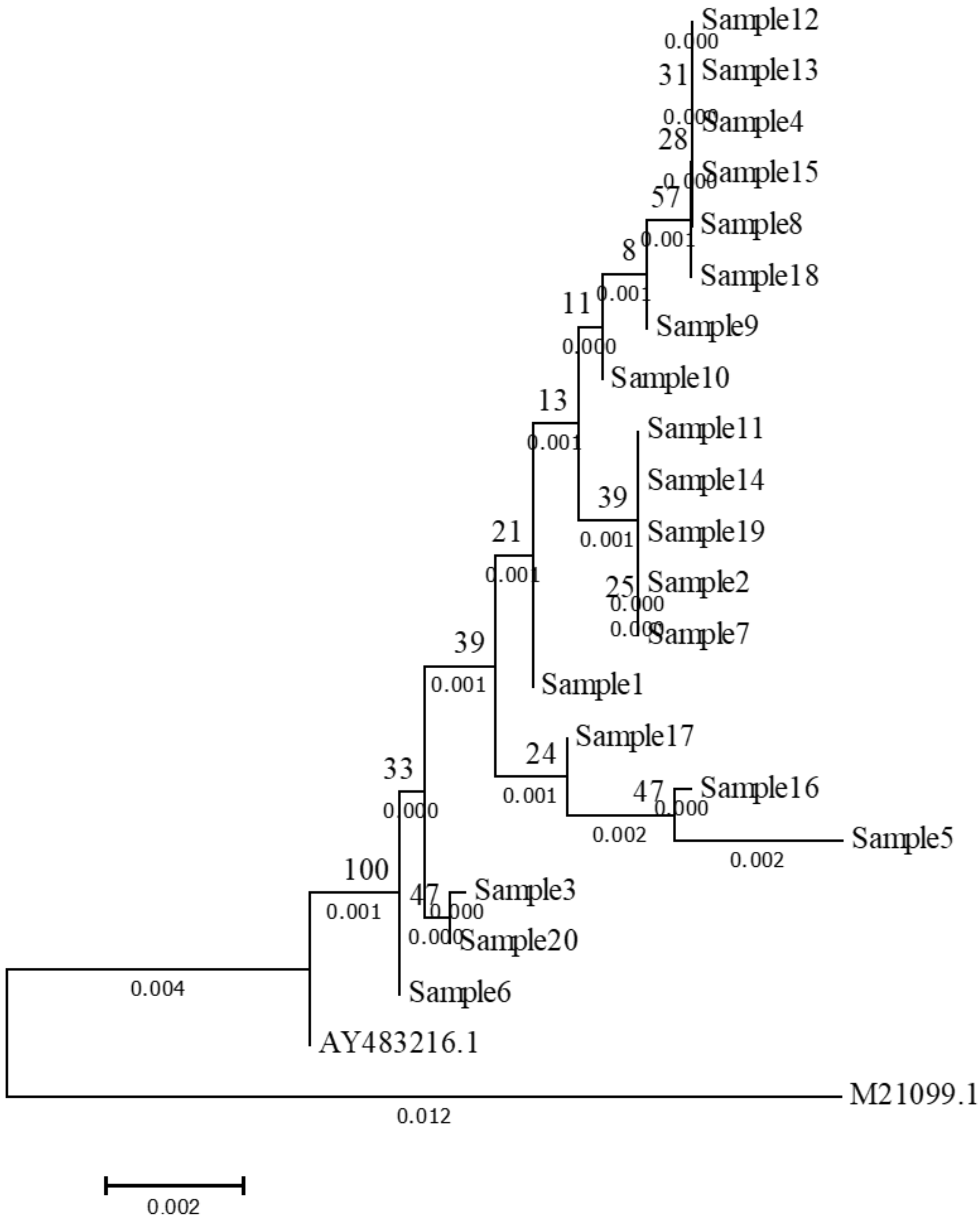


Figure 4.8: The evolutionary history was inferred using the Neighbor-Joining method, based on a 426 bp sequence and NCBI sequences were used for comparison in the KAP3.2 locus, in Swakara sheep. The optimal tree with the sum of branch length = 0.02063384 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The analysis involved 22 nucleotide sequences. Evolutionary analyses were conducted in MEGA version 6.

Figure 4.9 demonstrates the allele frequency distribution with the KAP3.2 gene in Swakara. Uniquely the monomorphic pattern denoted as AA genotype was mostly distributed across the four farms under study.

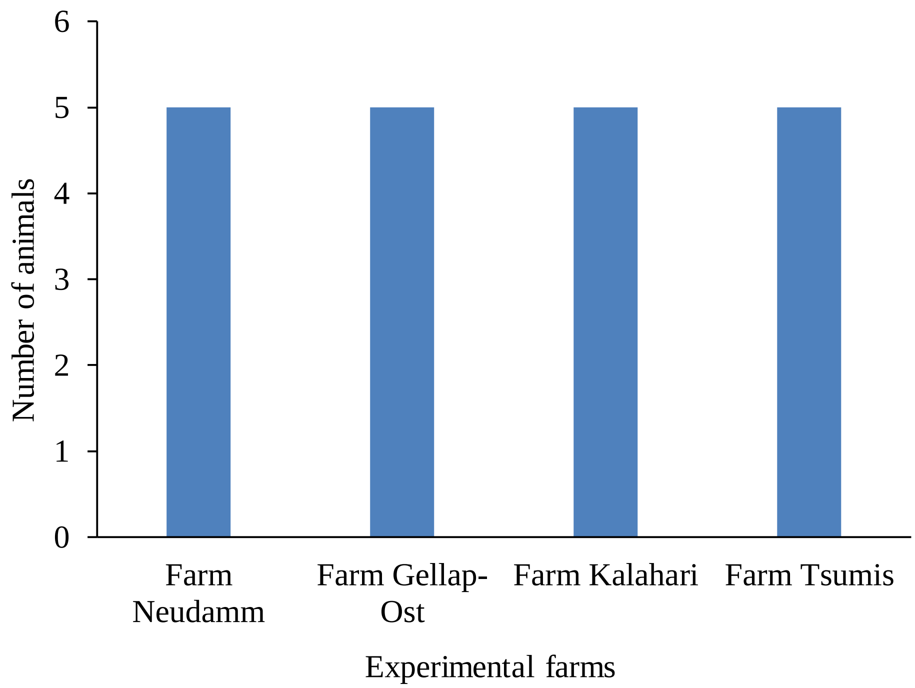


Figure 4.9: KAP3.2 allele frequency distribution in Swakara sheep from four farms. A unique monomorphic pattern was observed.

4.1.3 Keratin Intermediate filament protein (KRT33A) gene

Figure 4.10 exemplifies the allele frequency distribution observed in the KRT33A gene of the Swakara sheep. Moreover, Figure 4.11 depicts polymorphism of ovine KRT33A and two variants (AA and BB) with a homozygous forms were detected by PCR-SSCP. An overall sample size of 20 was used to identify polymorphism. Two KRT33A alleles were found 5 (AA) and 15 (BB) with allele frequencies of 0.25 and 0.75, respectively. There were no differences between the A and B alleles (P=0.402).

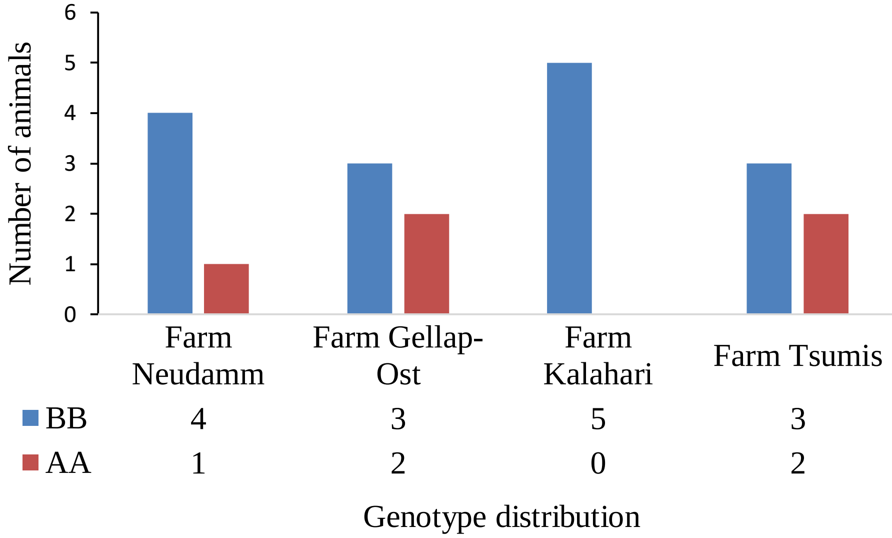


Figure 4.10: KRT33A allele frequency distribution of Swakara sheep situated on four farms

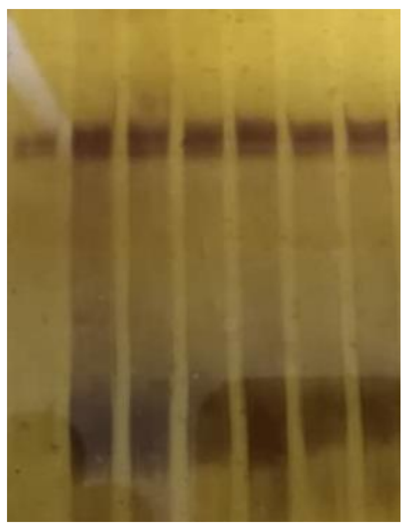


Figure 4.11: SSCP analysis of the 480 bp amplimer of the KRT33A gene, in Swakara sheep electrophoresed on 10% acrylamide/bis-acrylamide gel that contained 7 M concentration of urea run for 16 hours 30 minutes at 200V at room temperature. Lane one to lane seven are different genotypes at the KRT33A gene.

Figure 4.12 depicts the phylogenetic tree for the KRT33A locus identified in sheep. The tree was constructed using the amino acid consensus sequences. The Swakara sheep are indicated as ‘Sample and a number assigned afterward.’ Whereas KY, M and KF, and number assigned afterward’ denote the predicted reference sequences for comparison obtained from NCBI. KRT33A (M23912.2) Sheep sequence was used as the root of the tree. The sequences were consolidated and drawn using MEGA 6 software.

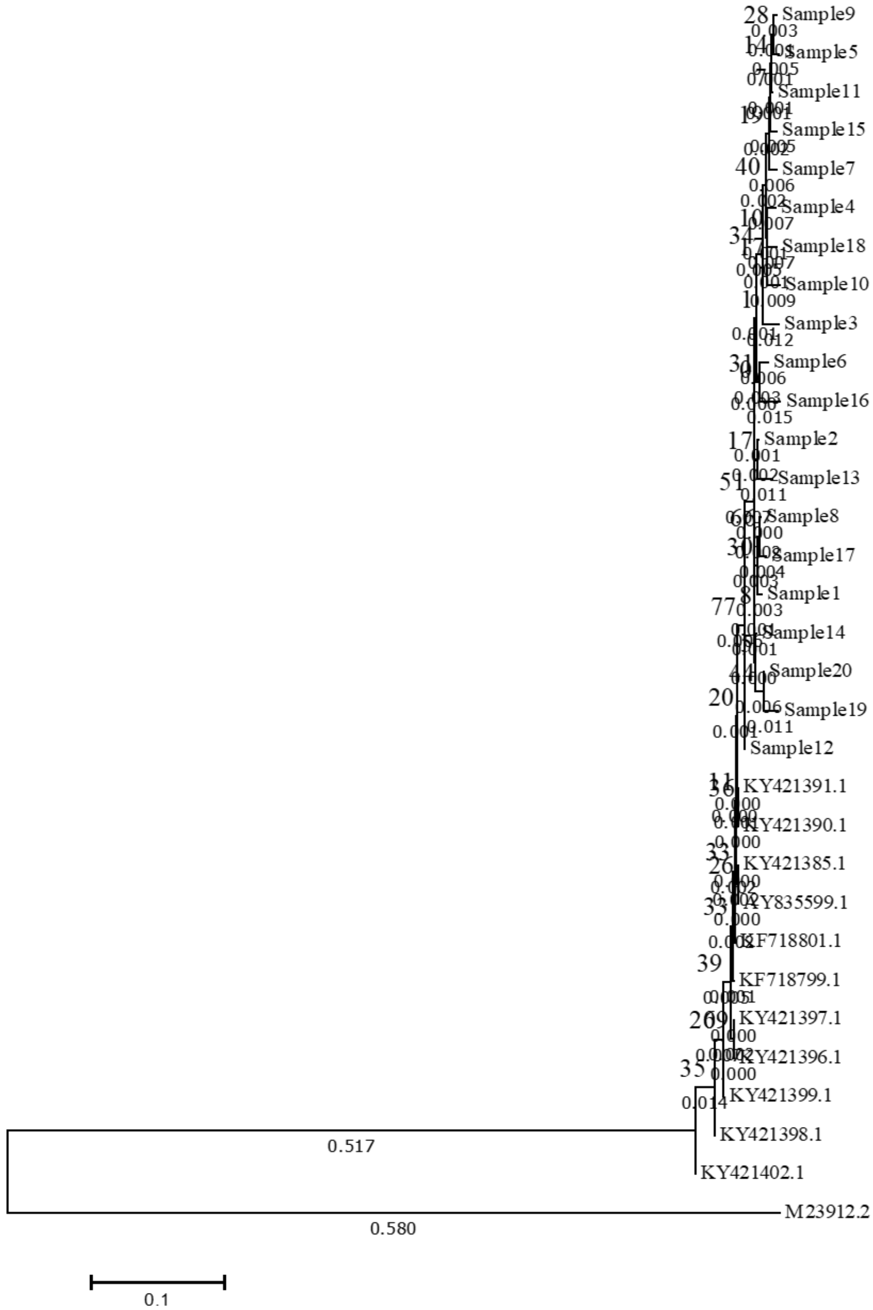


Figure 4.12: The evolutionary history was inferred using the Neighbor-Joining method, based on a 480 bp sequence and NCBI sequences were used for comparison in the KRT33A locus, in Swakara sheep. The optimal tree with the sum of branch length = 1.23245998 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The analysis involved 32 nucleotide sequences. KRT33A (M23912.2) Sheep was used as an out-group/root of the tree Evolutionary analyses were conducted in MEGA version 6.

4.1.4 DNA sequence polymorphism software version 5.10 output results

A total of 26 samples were selected for sequencing at the University of Porto in Portugal following the ABI Prism BigDye^{TM} Terminator v3.1 Sequencing Kit protocol on an ABI3130xl DNA Analyzer (Applied Biosystems, Foster City, California, USA). The sequence output on Table 4.2 depicts a summary of the keratin gene, amplimer length, number of variable site, number of haplotypes, haplotype diversity, standard deviation of haplotype diversity and the Guanine-cytosine content. The KAP1.1, KAP1.3 and KRT33A loci had a haplotype diversity of 0.778, 0.883 and 0.733, respectively. There was no haplotype diversity observed in KAP 3.2 gene. GenBank sequences used for comparison in the study shared 98-100% similarity with the KAP1.1, KAP1.3, KAP3.2 and KRT33A genes in the study. Furthermore, sequence alignments are illustrated in text and adapted as a text file from BioEdit software version 7.0.5.

Table 4.2: Summary of sequence output from DNA Sequence Polymorphism software version 5.10

Keratin gene	**Base pair**	**Variable site(s)**	**No. of Haplotype(s)**	**Haplotype diversity (Hd.)**	**SD of Hd.**	**Variance of Hd.**	**G+C content**
KAP1.1	322	7	4	0.778	0.091	0.00822	0.617
KAP1.3	570	8	8	0.883	0.017	0.00030	0.637
KAP3.2	426	0	1	0.000	0.000	0.0000	0.574
KRT33A	480	6	4	0.733	0.028	0.00078	0.643

Alignment:\Text file sequences\KAP1.1

```
              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  5         15         25         35         45         55
Sample1       ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1    ---------- ---------- ---------- ---------- --aaccctcc tctcaaccca
MG641066.1    ---------- ---------- ---------- ---------- ---------- ----------
Sample5       ---------- ---------- ---------- ---------- ---------- ----------
Sample6       ---------- ---------- ---------- ---------- ---------- ----------
Sample2       ---------- ---------- ---------- ---------- ---------- ----------
L33887.1      agatgcagaa ggtggagcca aaactcaaaa acttctctta acaaccctcc tctcaaccca
Sample4       ---------- ---------- ---------- ---------- ---------- ----------
XM02498056    ---------- ---gcagacc aaactcagaa acttctccaa gcatcccagc tctcagccta

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  65         75         85         95        105        115
Sample1       ---------- ---------- ---------c agcttctgtg gatttcccat ctgttccact
FJ169479.1    actcctgaca ccatggcctg ctgttccacc agcttctgtg gatttcccat ctgttccact
MG641066.1    ---------- --atggcctg ctgttccacc agcttctgtg gatttcccat ctgttccact
Sample5       ---------- ---------- ------cacc agcttctgtg gatttcccat ctgttccact
Sample6       ---------- ---------- ------cacc agcttctgtg gatttcccat ctgttccact
Sample2       ---------- ---------- ---------c agcttctgtg gattt-ccat ctgttccact
L33887.1      actcctgaca ccatggcctg ctgttccacc agcttctgtg gatttcccat ctgttccact
Sample4       ---------- ---------- --------cc agcttctgtg gatttcccat ctgttccact
XM02498056    acccctgaca ccatggcctg ctgttctact agcttctgtg gctttcccat ctgttccact

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 125        135        145        155        165        175
Sample1       ggtgggacct gtggctccag tccctgccag ccgacctgct gccagaccag ctgctgccag
FJ169479.1    ggtgggacct gtggctccag tccctgccag ccgacctgct gccagaccag ctgctgccag
MG641066.1    ggtgggacct gtggctccag tccctgccag ccgacctgct gccagaccag ctgctgccag
Sample5       ggtgggacct gtggctccag tccctgccag ccgacctgct gccagaccag ctgctgccag
Sample6       ggtgggacct gtggctccag tccctgccag cagacctgct gccagaccag ctgctgccag
Sample2       ggt-ggacct gt-gctccag tccctgccag ccgacctgct gccagaccag ctgctgccag
L33887.1      ggtgggacct gtggctccag tccctgccag ccgacctgct gccagaccag ctgctgccag
Sample4       gtgga----c tgtgctccag tccctgccag ccgacctgct gccagaccag ctgctgc--a
XM02498056    gctgggacct gtggct---- ---------- ---------- ---------- ----------

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 185        195        205        215        225        235
Sample1       ccaacc---- ---------- ---------- ------tcca tccagaccag ctgctgccaa
FJ169479.1    ccaacc---- ---------- ---------- ------tcca tccagactag ctgctgccaa
MG641066.1    ccaacc---- ---------- ---------- ------tcca tccagaccag ctgctgccaa
Sample5       ccaacc---- ---------- ---------- ------tcca tccagaccag ctgctgccaa
Sample6       ccaacctcca tccagaccag ctgctgccag ccaacttcca tccagaccag ctgctgccaa
Sample2       ccaacc---- ---------- ---------- ------tcca tccagaccag ctgctgccaa
L33887.1      ccaacc---- ---------- ---------- ------tcca tccagaccag ctgctgccaa
Sample4       gcaacc---- ---------- ---------- ------tcca tccagaccag ctgctgccaa
XM02498056    ---------- ---------- ---------- ---------- ------ccag ctgctgccga

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 245        255        265        275        285        295
Sample1       ccgatctcca tccagaccag ctgctgccag ccaacctcca tccagaccag ctgctgccag
FJ169479.1    ccgatctcca tccagaccag ctgctgccag ccaacctcca tccagaccag ctgctgccag
MG641066.1    ccgatctcca tccagaccag ctgctgccag ccaacctcca tccagaccag ctgctgccag
Sample5       ccgatctcca tccagaccag ctgctgccag ccaacctcca tccagaccag ctgctgccag
Sample6       ccgatctcca tccagaccag ctgctgccag ccaacctcca tccagaccag ctgctgccag
Sample2       ccgatctcca tccagaccag ctgctgccag ccaacctcca tccagaccag ctgctgccag
L33887.1      ccgat----- ---------- ---------- -----ctcca tccagaccag ctgctgccag
Sample4       ccgat----- ---------- ---------- -----ctcca tccagaccag ctgctgccag
XM02498056    tcaac----- ---------- ---------- -----ctgca gtcagaccag ctgctgccag

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 305        315        325        335        345        355
Sample1       ccaacctgcc tccagaccag tggctgtgag acgggctgtg gcattggtgg cagcattggc
FJ169479.1    ccaacctgcc tccagaccag tggctgtgag acgggctgtg gcattggtgg cagcattggc
MG641066.1    ccaacctgcc tccagaccag tggctgtgag acgggctgtg gcattggtgg cagcattggc
Sample5       ccaacctgcc tccagaccag tggctgtgag acgggctgtg gcattggtgg cagcattggc
Sample6       ccaacctgcc tccagaccag tggctgtgag acgggctgtg gcattggtgg cagcattggc
```

```
Sample2         ccaacctgcc tccagaccag tggctgtgag accggctgtg gcattggtgg cagcattggc
L33887.1        ccaacctgcc tccagaccag tggctgtgag accggctgtg gcattggtgg cagcattggc
Sample4         ccaacctgcc tccagaccag agg------- ---------- ---------- ----------
XM02498056      ccaacctgcc tccagaccag tggctgtgag accggctgtg gcattggtgg cagcattggc

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     365        375        385        395        405        415
Sample1         tatggccagg tgggtagcag cga------- ---------- ---------- ----------
FJ169479.1      tatggccagg tgggtagcag cga------- ---------- ---------- ----------
MG641066.1      tatggccagg tgggtagcag cg-------- ---------- ---------- ----------
Sample5         tatggccagg tgggtaaca- ---------- ---------- ---------- ----------
Sample6         tatggccagg tgggtagca- ---------- ---------- ---------- ----------
Sample2         tatggccagt tgggtagcag cg-------- ---------- ---------- ----------
L33887.1        tatggccagg tgggtagcag cggagctgtg agcagccgca ccaggtggtg ccgccctgac
Sample4         ---------- ---------- ---------- ---------- ---------- ----------
XM02498056      tatggccagg tgggtagcag cggagctgtg agcagccgca ccaggtggtg ccgccctgac

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     425        435        445        455        465        475
Sample1         ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1      ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1      ---------- ---------- ---------- ---------- ---------- ----------
Sample5         ---------- ---------- ---------- ---------- ---------- ----------
Sample6         ---------- ---------- ---------- ---------- ---------- ----------
Sample2         ---------- ---------- ---------- ---------- ---------- ----------
L33887.1        tgccgcgtgg agggcaccag cctgccaccc tgctgtgtgg tgagctgcac acccccgtcc
Sample4         ---------- ---------- ---------- ---------- ---------- ----------
XM02498056      tgccgtgtgg agggcaccag cctgcctccc tgctgtgtgg tgagctgcac acccccctcc

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     485        495        505        515        525        535
Sample1         ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1      ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1      ---------- ---------- ---------- ---------- ---------- ----------
Sample5         ---------- ---------- ---------- ---------- ---------- ----------
Sample6         ---------- ---------- ---------- ---------- ---------- ----------
Sample2         ---------- ---------- ---------- ---------- ---------- ----------
L33887.1        tgctgccagc tgtactatgc ccaggcctcc tgctgccgcc catcctactg tggacagtcc
Sample4         ---------- ---------- ---------- ---------- ---------- ----------
XM02498056      tgctgccagc tgtactatgc ccaggcctcc tgctgccgcc catcctactg tggacagtcc

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     545        555        565        575        585        595
Sample1         ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1      ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1      ---------- ---------- ---------- ---------- ---------- ----------
Sample5         ---------- ---------- ---------- ---------- ---------- ----------
Sample6         ---------- ---------- ---------- ---------- ---------- ----------
Sample2         ---------- ---------- ---------- ---------- ---------- ----------
L33887.1        tgctgccgcc cagcctgctg ctgccagccc acctgcattg agcccatctg tgagcccagc
Sample4         ---------- ---------- ---------- ---------- ---------- ----------
XM02498056      tgctgccgcc cagcctgctg ctgccagccc acctgcattg agcccatctg tgagcccacc

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     605        615        625        635        645        655
Sample1         ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1      ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1      ---------- ---------- ---------- ---------- ---------- ----------
Sample5         ---------- ---------- ---------- ---------- ---------- ----------
Sample6         ---------- ---------- ---------- ---------- ---------- ----------
Sample2         ---------- ---------- ---------- ---------- ---------- ----------
L33887.1        tgctgtgagc ccacctgctg aaagcaatgc tgctgattgc ttaag----- ----------
Sample4         ---------- ---------- ---------- ---------- ---------- ----------
XM02498056      tgctgccaac ccacctgtta aaaatcttct gatggaaatc ctaagacaat ggcacttcaa

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                     665        675        685        695        705        715
Sample1         ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1      ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1      ---------- ---------- ---------- ---------- ---------- ----------
Sample5         ---------- ---------- ---------- ---------- ---------- ----------
```

```
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
L33887.1     ---------- ---------- ---------- ---------- ---------- ----------
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
XM02498056   agttagccag tgtagagtcc ---------c aacaaacttc taaccactag cacccataac

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 725        735        745        755        765        775
Sample1      ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1   ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1   ---------- ---------- ---------- ---------- ---------- ----------
Sample5      ---------- ---------- ---------- ---------- ---------- ----------
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
L33887.1     ---------- ---------- ---------- ---------- ---------- ----------
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
XM02498056   ccagtctacc acctcatgtc aaaagctaca taattttcct aacatggaaa ttgatttata

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 785        795        805        815        825        835
Sample1      ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1   ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1   ---------- ---------- ---------- ---------- ---------- ----------
Sample5      ---------- ---------- ---------- ---------- ---------- ----------
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
L33887.1     ---------- ---------- ---------- ---------- ---------- --aatgaagg
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
XM02498056   ataaaaaaac aaaaatattt catagatctg gatatgaaca gaagtcctac agtgtgaaaa

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 845        855        865        875        885        895
Sample1      ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1   ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1   ---------- ---------- ---------- ---------- ---------- ----------
Sample5      ---------- ---------- ---------- ---------- ---------- ----------
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
L33887.1     gggaagcaa- ---------- ---------- -aacaaacct gcgctgatct ct--------
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
XM02498056   gaaa------ --tgtgtaat actttactat aaatatcttc tgaaatgttt caacaactca

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 905        915        925        935        945        955
Sample1      ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1   ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1   ---------- ---------- ---------- ---------- ---------- ----------
Sample5      ---------- ---------- ---------- ---------- ---------- ----------
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
L33887.1     ---------- -------gaa gaaccattc- ---------- ---------- ----------
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
XM02498056   tcggcctttg gatttaataa aatttttcat tcttcagtga ---------- ----------

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 965        975        985        995       1005       1015
Sample1      ---------- ---------- ---------- ---------- ---------- ----------
FJ169479.1   ---------- ---------- ---------- ---------- ---------- ----------
MG641066.1   ---------- ---------- ---------- ---------- ---------- ----------
Sample5      ---------- ---------- ---------- ---------- ---------- ----------
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
L33887.1     ---------- ---------- ---------- ---------- ---------- ----------
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
XM02498056   ---------- ---------- ---------- ---------- ---------- ----------

             ....|....| ....|....| ...
                1025       1035
Sample1      ---------- ---------- ---
FJ169479.1   ---------- ---------- ---
MG641066.1   ---------- ---------- ---
```

```
Sample5      ---------- ---------- ---
Sample6      ---------- ---------- ---
Sample2      ---------- ---------- ---
L33887.1     ---------- ---------- ---
Sample4      ---------- ---------- ---
XM02498056   ---------- ---------- ---
```

Alignment:\Text file sequences\KAP1.3

```
             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 5          15         25         35         45         55
Sample1      ---------- ---------- ---------- ---------- ---------- tcag-cctac
Sample19     ---------- ---------- ---------- ---------- ---------- --------ac
Sample18     ---------- ---------- ---------- ---------- ---------- -cag-cctac
Sample11     ---------- ---------- ---------- ---------- ---------- -cag-cctac
AY835593.1   gggtggaaca agcagaccaa actcagaaac ttctccaagc atcccagctc tcggcctaac
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
Sample12     ---------- ---------- ---------- ---------- ---------- ---cgcctac
Sample20     ---------- ---------- ---------- ---------- ---------- --------ac
Sample8      ---------- ---------- ---------- ---------- ---------- --------ac
Sample13     ---------- ---------- ---------- ---------- ---------- --------ac
Sample3      a--------- ---------- ---------- ---------- atccaagctc tcag-cctac
Sample16     ---------- ---------- ---------- ---------- ---------- tcag-cctac
Sample6      ---------- ---------- ---------- ---------- ---------- -cag-cctac
Sample7      ---------- ---------- ---------- ---------- ---------- ------ctac
Sample5      ---------- ---------- ---------- ---------- -------ctc tcag-cctac
Sample15     ---------- ---------- ---------- ---------- ---------- -----cctac
Sample4      ---------- ---------- ---------- ---------- ---------- -----cctac
Sample14     ---------- ---------- ---------- ---------- ---------- -----cctac
Sample9      ---------- ---------- ---------- ---------- ---------- ------ctac
Sample10     ---------- ---------- ---------- ---------- ---------- ------ctac
Sample17     ---------- ---------- ---------- ---------- ---------- --------ac

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 65         75         85         95         105        115
Sample1      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample19     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample18     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample11     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
AY835593.1   ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample2      -cctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample12     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample20     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample8      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample13     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample3      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample16     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccacagc
Sample6      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample7      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample5      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample15     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample4      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample14     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample9      ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample10     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc
Sample17     ccctgacacc atggcctgct gttccaccag cttctgtgga tttcccatct gttccactgc

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 125        135        145        155        165        175
Sample1      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample19     tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample18     tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample11     tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccaacc
AY835593.1   tgggacctgt ggctccagct gctgccgatc aacctgcagc cagaccagct gctgccagcc
Sample2      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample12     tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample20     tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample8      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
```

```
Sample13      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccaacc
Sample3       tgggacctgt ggctccagct gctgccgatc aacctgcagt caaaccagct gctgccagcc
Sample16      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample6       tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample7       tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccaacc
Sample5       tgggacctgt ggctccagct gctgccgatc accctgcagt cagaccagct gctgccagcc
Sample15      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccaacc
Sample4       tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample14      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccaacc
Sample9       tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc
Sample10      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccaacc
Sample17      tgggacctgt ggctccagct gctgccgatc aacctgcagt cagaccagct gctgccagcc

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 185        195        205        215        225        235
Sample1       aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample19      aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgagac
Sample18      aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgagac
Sample11      aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgagac
AY835593.1    aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgagac
Sample2       aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample12      aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample20      aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgagac
Sample8       aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample13      aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample3       aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample16      aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgaaac
Sample6       aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample7       aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgaaac
Sample5       aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample15      aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgaaac
Sample4       aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample14      aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample9       aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgagac
Sample10      aacctccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gctgtgagac
Sample17      aacttccatc cagaccagct gctgccagcc aacctgcctc cagaccagtg gttgtgagac

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 245        255        265        275        285        295
Sample1       cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample19      cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample18      cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample11      cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
AY835593.1    cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample2       cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample12      cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample20      cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample8       cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample13      cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample3       cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample16      cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample6       cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample7       cggctgtggc attggtggca gcaatggcta tggccaggtg ggtagcagcg gaactgtgag
Sample5       cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample15      cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample4       cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag
Sample14      cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample9       cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample10      cggctgtggc attggtggca gcattggcta tggccaggtg ggtagcagcg gagctgtgag
Sample17      cggctgtggc attggtggca gcactggcta tggccaggtg ggtagcagcg gagctgtgag

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 305        315        325        335        345        355
Sample1       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample19      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample18      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample11      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
AY835593.1    cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample2       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample12      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample20      cagccgcacc agatggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
```

```
Sample8       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample13      cagccgcacc agatggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample3       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample16      cagccgcacc agatggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample6       cagccgcacc agatggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample7       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample5       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample15      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample4       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample14      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample9       cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample10      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg
Sample17      cagccgcacc aggtggtgcc gccctgactg ccgcgtggag ggcaccagcc tgcctccctg

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   365        375        385        395        405        415
Sample1       ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample19      ctgtgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample18      ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample11      ctgtgtggtg agctgcacac ccccgtcctg ctgccagctg tactatgccc aggcctcctg
AY835593.1    ctgtgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample2       ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample12      ctgtgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample20      ctgtgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample8       ctgcgtggtg agctgcacaa ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample13      ctgggtggtg agctgcacac ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample3       ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample16      ctgtgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample6       ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample7       ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample5       ctgcgtggtg agctgcacat ccccgtcctg ctgccaactg tactatgccc aggcctcctg
Sample15      ctgggtggtg agctgcacac ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample4       ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample14      ctgtgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample9       ctgtgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample10      ctgtgtggtg agctgcacac ccccgtcctg ctgccagctg tactatgccc aggcctcctg
Sample17      ctgcgtggtg agctgcacat ccccgtcctg ctgccagctg tactatgccc aggcctcctg

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   425        435        445        455        465        475
Sample1       ctgccgccca tccactg-gt ggacaa---- ---------- ---------- ----------
Sample19      ctgccgccca tccactg-gt ggacaa---- ---------- ---------- ----------
Sample18      ctgccgccca tccactt-gt ggacaaacct gcgaccgcat ctagtt---- ----------
Sample11      ctgccgccca tccactt-gt ggacaatcct gctgccgc-- ccactc---- ----------
AY835593.1    ctgccgccca tcctact-gt ggacagtcct gctgccgc-- ccagcctgct gctgccagcc
Sample2       ctgccgccca tcctact-gt ggacaa---- ---------- ---------- ----------
Sample12      ctgccgccca tcctact-gt ggacaa---- ---------- ---------- ----------
Sample20      ctgccgccca tcctact-gt ggacaa---- ---------- ---------- ----------
Sample8       ctgccgccca tcctact-gt gg-------- ---------- ---------- ----------
Sample13      ctgccgccca tcctact-gt ggaca----- ---------- ---------- ----------
Sample3       ctgccgccca tcctact-gt ggaca----- ---------- ---------- ----------
Sample16      ctgccgccca tcctact-gt ggacaa---- ---------- ---------- ----------
Sample6       ctgccgccca tccaact-gt ggacaatcct gcgaccg--- ---------- ----------
Sample7       ctgccgccca tccaact-gt ggaca----- ---------- ---------- ----------
Sample5       ctgccgccca tcctact-gg ggacaa---- ---------- ---------- ----------
Sample15      ctgccgccca tccaact-gt ggacaatcct gctgccgc-- ccatcct--- ----------
Sample4       ctgccgccca tcc-act-gt ggacaa---- ---------- ---------- ----------
Sample14      ctgccgccca tccactt-gt ggacaa-cct gctgccgc-- ccatccttct ttgttcaaat
Sample9       ctgccgccca tccaatttgt ggacaa---- ---------- ---------- ----------
Sample10      ctgccgccca tccaattggt ggacaatcct gctgccgc-- cc-------- ----------
Sample17      ctgccgccca tccaattggt ggacaaacct gctgccgc-- ccatccta-- ----------

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   485        495        505        515        525        535
Sample1       ---------- ---------- ---------- ---------- ---------- ----------
Sample19      ---------- ---------- ---------- ---------- ---------- ----------
Sample18      ---------- ---------- ---------- ---------- ---------- ----------
Sample11      ---------- ---------- ---------- ---------- ---------- ----------
AY835593.1    cacctgcact gagcccgtct gtgagcccac ctgctcccaa cccatctgtt aaaaacctac
Sample2       ---------- ---------- ---------- ---------- ---------- ----------
Sample12      ---------- ---------- ---------- ---------- ---------- ----------
```

```
Sample20     ---------- ---------- ---------- ---------- ---------- ----------
Sample8      ---------- ---------- ---------- ---------- ---------- ----------
Sample13     ---------- ---------- ---------- ---------- ---------- ----------
Sample3      ---------- ---------- ---------- ---------- ---------- ----------
Sample16     ---------- ---------- ---------- ---------- ---------- ----------
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample7      ---------- ---------- ---------- ---------- ---------- ----------
Sample5      ---------- ---------- ---------- ---------- ---------- ----------
Sample15     ---------- ---------- ---------- ---------- ---------- ----------
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
Sample14     aaaatccact ga-------- ---------- ---------- ---------- ----------
Sample9      ---------- ---------- ---------- ---------- ---------- ----------
Sample10     ---------- ---------- ---------- ---------- ---------- ----------
Sample17     ---------- ---------- ---------- ---------- ---------- ----------

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                545        555        565        575        585        595
Sample1      ---------- ---------- ---------- ---------- ---------- ----------
Sample19     ---------- ---------- ---------- ---------- ---------- ----------
Sample18     --gaggcgcc cttgacacaa aaaaaa---- ---------- ---------- ----------
Sample11     --gtcgtgtt tttaaaagga cgctttttt- ---------- ---------- ----------
AY835593.1   tgatggaaat tctaagacaa tggcacttca aaattagcca gtgtacagtc ccaacaaact
Sample2      ---------- ---------- ---------- ---------- ---------- ----------
Sample12     ---------- ---------- ---------- ---------- ---------- ----------
Sample20     ---------- ---------- ---------- ---------- ---------- ----------
Sample8      ---------- ---------- ---------- ---------- ---------- ----------
Sample13     ---------- ---------- ---------- ---------- ---------- ----------
Sample3      ---------- ---------- ---------- ---------- ---------- ----------
Sample16     ---------- ---------- ---------- ---------- ---------- ----------
Sample6      ---------- ---------- ---------- ---------- ---------- ----------
Sample7      ---------- ---------- ---------- ---------- ---------- ----------
Sample5      ---------- ---------- ---------- ---------- ---------- ----------
Sample15     ---------- ---------- ---------- ---------- ---------- ----------
Sample4      ---------- ---------- ---------- ---------- ---------- ----------
Sample14     ---------- ---------- ---------- ---------- ---------- ----------
Sample9      ---------- ---------- ---------- ---------- ---------- ----------
Sample10     ---------- ---------- ---------- ---------- ---------- ----------
Sample17     ---------- ---------- ---------- ---------- ---------- ----------

             .

Sample1      -
Sample19     -
Sample18     -
Sample11     -
AY835593.1   a
Sample2      -
Sample12     -
Sample20     -
Sample8      -
Sample13     -
Sample3      -
Sample16     -
Sample6      -
Sample7      -
Sample5      -
Sample15     -
Sample4      -
Sample14     -
Sample9      -
Sample10     -
Sample17     -
```

Alignment:\Text file sequences\KAP3.2

```
               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   5         15         25         35         45         55
Sample1        ---------- ---------- ---------- ---------- ---------- --catggctt
AY483216.1     -cgagacacc aagacttctc tcatcaaccc aacaaaaccc agctcctgac accatggctt
Sample6        ---------- ---------- ---------- ---------- ---------- --catggctt
Sample9        ---------- ---------- ---------- ---------- ---------- --catggctt
Sample17       ---------- ---------- ---------- ---------- ---------- accatggctt
Sample2        ---------- ---------- ---------- ---------- -----ctga- accatggctt
Sample7        ---------- ---------- ---------- ---------- ---------- ---atggctt
Sample10       ---------- ---------- ---------- ---------- ---------- ---atggctt
Sample11       ---------- ---------- ---------- ---------- ---------- --catggctt
Sample14       ---------- ---------- ---------- ---------- ---------- --catggctt
Sample19       ---------- ---------- ---------- ---------- ---------- --catggctt
Sample3        ---------- ---------- ---------- ---------- ---------- cccatggctt
Sample4        ---------- ---------- ---------- ---------- ---------- -ccatggctt
Sample8        ---------- ---------- ---------- ---------- ---------- ---atggctt
Sample12       ---------- ---------- ---------- ---------- ---------- ---atggctt
Sample13       ---------- ---------- ---------- ---------- ---------- -ccatggctt
Sample15       ---------- ---------- ---------- ---------- ---------- --catggctt
Sample18       ---------- ---------- ---------- ---------- ---------- --catggctt
Sample16       ---------- ---------- ---------- ---------- ---------- -ccatggctt
Sample5        ---------- ---------- ---------- ---------- ---------- -ccatggctt
Sample20       ---------- ---------- ---------- ---------- ---------- ----tggctt
M21099.1       tcgagacacc aagacttctc tcatcaaccc aacaaaaccc agctcctgac accatggctt

               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   65         75         85         95        105        115
Sample1        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
AY483216.1     gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample6        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample9        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample17       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample2        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample7        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample10       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample11       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample14       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample19       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample3        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample4        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample8        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample12       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample13       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample15       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample18       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample16       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample5        gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct
Sample20       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccactcat ctgctcctct
M21099.1       gctgcgctcc ccgctgctgc agcgtccgca ctggtcctgc caccac-cat ctgctcctct

               ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  125        135        145        155        165        175
Sample1        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
AY483216.1     gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample6        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample9        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample17       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample2        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample7        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample10       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample11       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample14       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample19       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample3        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample4        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample8        gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample12       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample13       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample15       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
```

```
Sample18      gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample16      gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample5       gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
Sample20      gacaaattct gtcggtgtgg agtctgcctg cccagcacct gcccacacga catcagcctc
M21099.1      gacaaattct gtcggtgtgg agtctgcctg cccagcagct gcccacacga catcagcctc

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 185        195        205        215        225        235
Sample1       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
AY483216.1    ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample6       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample9       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample17      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample2       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample7       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample10      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample11      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample14      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample19      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample3       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample4       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample8       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample12      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample13      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample15      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample18      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample16      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample5       ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
Sample20      ctccagccca cttgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg
M21099.1      ctccagccca cctgctgtga caactccccc gtgccctgct atgtgcctga cacctatgtg

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 245        255        265        275        285        295
Sample1       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
AY483216.1    ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample6       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample9       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample17      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample2       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample7       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample10      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample11      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample14      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample19      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample3       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample4       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample8       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample12      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample13      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample15      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample18      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample16      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample5       ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
Sample20      ccaacttgct ttctgctcaa ctcttcccac cccactcctg gactgagcgg gatcaacctg
M21099.1      ccaacttgct ttctgctcaa ctcttcccac cccactcctc gactgagcgg gatcaacctg

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 305        315        325        335        345        355
Sample1       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
AY483216.1    acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample6       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample9       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample17      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample2       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample7       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample10      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample11      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample14      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample19      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample3       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample4       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
```

```
Sample8       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample12      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample13      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample15      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample18      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample16      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample5       acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
Sample20      acgaccttca ttcagcctgg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca
M21099.1      acgaccttca ttcagcctcg ctgtgaaaat gtctgcgagc cccgctgtta aacagccaca

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 365        375        385        395        405        415
Sample1       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggc------
AY483216.1    tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttcaa
Sample6       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttcaa
Sample9       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tgg-------
Sample17      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggc------
Sample2       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaaa
Sample7       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaaa
Sample10      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggc------
Sample11      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaaa
Sample14      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaaa
Sample19      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaaa
Sample3       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttcaa
Sample4       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaac
Sample8       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaac
Sample12      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaac
Sample13      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaac
Sample15      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaac
Sample18      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttaac
Sample16      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttcaa
Sample5       tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttcaa
Sample20      tctttgcacg ggttcagtga tgagctgctc --aagtctta atgcgtgatc tggccttcaa
M21099.1      tctttgcacg ggcttagtga tgagctgctc ctaagtctta atgcgtgatc tggccttcaa

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 425        435        445        455        465        475
Sample1       ---------- ---------- ---------- ---------- ---------- ----------
AY483216.1    ca-ctcact- ---------- ---------- ---------- ---------- ----------
Sample6       ca-ctcac-- ---------- ---------- ---------- ---------- ----------
Sample9       ---------- ---------- ---------- ---------- ---------- ----------
Sample17      ---------- ---------- ---------- ---------- ---------- ----------
Sample2       caactcacta ---------- ---------- ---------- ---------- ----------
Sample7       ca-------- ---------- ---------- ---------- ---------- ----------
Sample10      ---------- ---------- ---------- ---------- ---------- ----------
Sample11      ca-ctcacta ---------- ---------- ---------- ---------- ----------
Sample14      ca-ctcacta ---------- ---------- ---------- ---------- ----------
Sample19      ca-ctc---- ---------- ---------- ---------- ---------- ----------
Sample3       ca-ctc---- ---------- ---------- ---------- ---------- ----------
Sample4       ca-------- ---------- ---------- ---------- ---------- ----------
Sample8       c--------- ---------- ---------- ---------- ---------- ----------
Sample12      ca-------- ---------- ---------- ---------- ---------- ----------
Sample13      ca-------- ---------- ---------- ---------- ---------- ----------
Sample15      ca-c------ ---------- ---------- ---------- ---------- ----------
Sample18      c--------- ---------- ---------- ---------- ---------- ----------
Sample16      ccactc---- ---------- ---------- ---------- ---------- ----------
Sample5       acactc---- ---------- ---------- ---------- ---------- ----------
Sample20      ca-ct----- ---------- ---------- ---------- ---------- ----------
M21099.1      ca-ctcacta ttacctacat caaattaaat caatcccaat agcttgggca gtgttttcat

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 485        495        505        515        525        535
Sample1       ---------- ---------- ---------- ---------- ---------- ----------
AY483216.1    ---------- ---------- ---------- ---------- ---------- ----------
Sample6       ---------- ---------- ---------- ---------- ---------- ----------
Sample9       ---------- ---------- ---------- ---------- ---------- ----------
Sample17      ---------- ---------- ---------- ---------- ---------- ----------
Sample2       ---------- ---------- ---------- ---------- ---------- ----------
Sample7       ---------- ---------- ---------- ---------- ---------- ----------
Sample10      ---------- ---------- ---------- ---------- ---------- ----------
Sample11      ---------- ---------- ---------- ---------- ---------- ----------
```

```
Sample14      ---------- ---------- ---------- ---------- ---------- ----------
Sample19      ---------- ---------- ---------- ---------- ---------- ----------
Sample3       ---------- ---------- ---------- ---------- ---------- ----------
Sample4       ---------- ---------- ---------- ---------- ---------- ----------
Sample8       ---------- ---------- ---------- ---------- ---------- ----------
Sample12      ---------- ---------- ---------- ---------- ---------- ----------
Sample13      ---------- ---------- ---------- ---------- ---------- ----------
Sample15      ---------- ---------- ---------- ---------- ---------- ----------
Sample18      ---------- ---------- ---------- ---------- ---------- ----------
Sample16      ---------- ---------- ---------- ---------- ---------- ----------
Sample5       ---------- ---------- ---------- ---------- ---------- ----------
Sample20      ---------- ---------- ---------- ---------- ---------- ----------
M21099.1      gcatttgccc aacttcttgg attctttctt cttttgcata gaattggaga ccttcttcct

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 545        555        565        575        585        595
Sample1       ---------- ---------- ---------- ---------- ---------- ----------
AY483216.1    ---------- ---------- ---------- ---------- ---------- ----------
Sample6       ---------- ---------- ---------- ---------- ---------- ----------
Sample9       ---------- ---------- ---------- ---------- ---------- ----------
Sample17      ---------- ---------- ---------- ---------- ---------- ----------
Sample2       ---------- ---------- ---------- ---------- ---------- ----------
Sample7       ---------- ---------- ---------- ---------- ---------- ----------
Sample10      ---------- ---------- ---------- ---------- ---------- ----------
Sample11      ---------- ---------- ---------- ---------- ---------- ----------
Sample14      ---------- ---------- ---------- ---------- ---------- ----------
Sample19      ---------- ---------- ---------- ---------- ---------- ----------
Sample3       ---------- ---------- ---------- ---------- ---------- ----------
Sample4       ---------- ---------- ---------- ---------- ---------- ----------
Sample8       ---------- ---------- ---------- ---------- ---------- ----------
Sample12      ---------- ---------- ---------- ---------- ---------- ----------
Sample13      ---------- ---------- ---------- ---------- ---------- ----------
Sample15      ---------- ---------- ---------- ---------- ---------- ----------
Sample18      ---------- ---------- ---------- ---------- ---------- ----------
Sample16      ---------- ---------- ---------- ---------- ---------- ----------
Sample5       ---------- ---------- ---------- ---------- ---------- ----------
Sample20      ---------- ---------- ---------- ---------- ---------- ----------
M21099.1      gtcttttaga gtctatcact attctttgga aatgaaccat tttgactatt caacaataaa

              ....|....| ....|.
                 605        615
Sample1       ---------- ------
AY483216.1    ---------- ------
Sample6       ---------- ------
Sample9       ---------- ------
Sample17      ---------- ------
Sample2       ---------- ------
Sample7       ---------- ------
Sample10      ---------- ------
Sample11      ---------- ------
Sample14      ---------- ------
Sample19      ---------- ------
Sample3       ---------- ------
Sample4       ---------- ------
Sample8       ---------- ------
Sample12      ---------- ------
Sample13      ---------- ------
Sample15      ---------- ------
Sample18      ---------- ------
Sample16      ---------- ------
Sample5       ---------- ------
Sample20      ---------- ------
M21099.1      cttcattctt acctag
```

Alignment:\Text file sequences\KRT33A

```
              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 5          15         25         35         45         55
Sample1       ---------- ---------- ---------- ---------- ---------- ----------
```

```
Sample14      ---------- ---------- ---------- ---------- ---------- ----------
Sample2       ---------- ---------- ---------- ---------- ---------- ----------
KY421385.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
AY835599.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KF718801.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KY421399.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KY421398.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KY421391.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KY421390.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
Sample12      ---------- ---------- ---------- ---------- ---------- ----------
KY421402.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KY421397.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KY421396.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
KF718799.1    ---------- ---------- ---------- --cacaactc tggcttggtg aact------
Sample8       ---------- ---------- ---------- ---------- ---------- ----------
Sample17      ---------- ---------- ---------- ---------- ---------- ----------
Sample13      ---------- ---------- ---------- ---------- ---------- ----------
Sample3       ---------- ---------- ---------- ---------- ---------- ----------
Sample4       ---------- ---------- ---------- ---------- ---------- ----------
Sample11      ---------- ---------- ---------- ---------- ---------- ----------
Sample15      ---------- ---------- ---------- ---------- ---------- ----------
Sample9       ---------- ---------- ---------- ---------- ---------- ----------
Sample5       ---------- ---------- ---------- ---------- ---------- ----------
Sample18      ---------- ---------- ---------- ---------- ---------- ----------
Sample7       ---------- ---------- ---------- ---------- ---------- ----------
Sample6       ---------- ---------- ---------- ---------- ---------- ----------
Sample20      ---------- ---------- ---------- ---------- ---------- ----------
Sample16      ---------- ---------- ---------- ---------- ---------- ----------
Sample10      ---------- ---------- ---------- ---------- ---------- ----------
Sample19      ---------- ---------- ---------- ---------- ---------- ----------
M23912.2      gatcccctgg aggaaggcat ggcaacccac tccagtattc ttgcctggag aatcccatgg

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   65         75         85         95        105        115
Sample1       ---------- ---------- ---------- ---------- ---------- ----------
Sample14      ---------- ---------- ---------- ---------- ---------- ----------
Sample2       ---------- ---------- ---------- ---------- ---------- ----------
KY421385.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
AY835599.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KF718801.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KY421399.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KY421398.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KY421391.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KY421390.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
Sample12      ---------- ---------- ---------- ---------- ---------- ----------
KY421402.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KY421397.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KY421396.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
KF718799.1    ---------- ----tggact ---------- ---------- ---------- ----ctgtgt
Sample8       ---------- ---------- ---------- ---------- ---------- ----------
Sample17      ---------- ---------- ---------- ---------- ---------- ----------
Sample13      ---------- ---------- ---------- ---------- ---------- ----ccagtg
Sample3       ---------- ---------- ---------- ---------- ---------- ----------
Sample4       ---------- ---------- ---------- ---------- ---------- ----------
Sample11      ---------- ---------- ---------- ---------- ---------- ----------
Sample15      ---------- ---------- ---------- ---------- ---------- ----------
Sample9       ---------- ---------- ---------- ---------- ---------- ----------
Sample5       ---------- ---------- ---------- ---------- ---------- ----gcttgt
Sample18      ---------- ---------- ---------- ---------- ---------- ----gttggt
Sample7       ---------- ---------- ---------- ---------- ---------- ----------
Sample6       ---------- ---------- ---------- ---------- ---------- ----------
Sample20      ---------- ---------- ---------- ---------- ---------- ----------
Sample16      ---------- ---------- ---------- ---------- ---------- ----gcagat
Sample10      ---------- ---------- ---------- ---------- ---------- ----------
Sample19      ---------- ---------- ---------- ---------- ---------- ----gttggt
M23912.2      ccaggggagc ctggtgggct acagtccatg gggtcgcaaa gtcggacaca actgagtgac

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  125        135        145        155        165        175
Sample1       ---------- ---------- ---------- ---------c agcaccatgt ctttc-actt
Sample14      ---------- ---------- ---------- ---------- -gcaccatgt ctttc-actt
```

```
Sample2         ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttc-actt
KY421385.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
AY835599.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KF718801.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KY421399.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KY421398.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KY421391.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KY421390.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
Sample12        ---------- ---------- ---------- ---------- -gcaccatgt ctttcaactt
KY421402.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KY421397.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KY421396.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
KF718799.1      tcagctgggc acc------- ---------t cgctccctcc agcaccatgt ctttcaactt
Sample8         ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttcaactt
Sample17        -------gtc cc-------- ---------- tcgctcctcc agcaccatgt ctttcaactt
Sample13        tatgctgggc ac-------- ---------- tcgctcctcc agcaccatgt ctttcaactt
Sample3         ---------- ---------- ---------- cgctccctcc agcaccatgt ctttcaactt
Sample4         ---------- ---------- ---------- ---------- --caccatgt ctttcaactt
Sample11        ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttcaactt
Sample15        ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttcaactt
Sample9         ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttcaactt
Sample5         tcg--ctggc ac-------- ---------- tcgctcctcc agcaccatgt ctttcaactt
Sample18        acagctggga ac-------- ---------- tcgctcctcc agcaccatgt ctttcaactt
Sample7         ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttcaactt
Sample6         ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttcaactt
Sample20        ---------- ---------- ---------- ---------c agcaccatgt ctttcaactt
Sample16        ttag-ctgcg ac-------- ---------- tcgctcctcc agcaccatgt ctttcaactt
Sample10        ---------- ---------- ---------- -cgctcctcc agcaccatgt ctttcaactt
Sample19        tcggcctggc ac-------- ---------- -tcgctcctc agcaccatgt ctttcaactt
M23912.2        taagcacagc acagcacaga aagaatctat tcttctttct aaaaccatga ctttggaatt

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    185        195        205        215        225        235
Sample1         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgtg tgccctccag
Sample14        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample2         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KY421385.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
AY835599.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KF718801.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KY421399.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KY421398.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KY421391.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KY421390.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample12        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KY421402.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
KY421397.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgtg tgccctccag
KY421396.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgtg tgccctccag
KF718799.1      ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample8         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample17        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample13        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample3         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample4         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgtg tgccctccag
Sample11        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample15        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample9         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample5         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample18        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample7         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctggg tgccctccag
Sample6         ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctggg tgccctccag
Sample20        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctggg tgccctccag
Sample16        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample10        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
Sample19        ctgcctgccc aacctgagct tccgctccag ctgctcctcc aggccctgcg tgccctccag
M23912.2        tgacttagca aaca------ ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    245        255        265        275        285        295
Sample1         ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample14        ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
Sample2         ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
```

```
KY421385.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
AY835599.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
KF718801.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
KY421399.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
KY421398.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
KY421391.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
KY421390.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
Sample12     ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
KY421402.1   ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
KY421397.1   ctgctgtggc accaccctgc ctggggcctg caacatcccc gccagcgtgg gcagctgcaa
KY421396.1   ctgctgtggc accaccctgc ctggggcctg caacatcccc gccagcgtgg gcagctgcaa
KF718799.1   ctgctgtggc accaccctgc ctggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample8      ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagggtgg gcagctgcaa
Sample17     ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample13     ctgctgtggc accaccctgc ctggggcctg caacatcccc gccagggtgg gcagctgcaa
Sample3      ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample4      ctgctgtggc accaccctgc ctggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample11     ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
Sample15     ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagggtgg gcagctgcaa
Sample9      ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagtgtgg gcagctgcaa
Sample5      ctgctgtggc accaccctgc cgggggcctg caacatcccc gccagtgtgg gcagctgcaa
Sample18     ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample7      ctgctgtggc accaccctgc cgggggcctg caacatcccc gccagggtgg gcagctgcaa
Sample6      ctgctgtggc accaccctgc cgggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample20     ctgctgtggc accaccctgc cgggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample16     ctgctgtggc accaccctgc cgggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample10     ctgctgtggc accaccctgc ctggggcctg caacatcccc gccagcgtgg gcagctgcaa
Sample19     ctgctgtggc accaccctgc ccggggcctg caacatcccc gccagcgtgg gcagctgcaa
M23912.2     ---------- actatatggg ccagggttat gtttgtccac ctcagaaaga tgagccagag

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                305        315        325        335        345        355
Sample1      ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample14     ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample2      ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
KY421385.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
AY835599.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
KF718801.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
KY421399.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
KY421398.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
KY421391.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
KY421390.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample12     ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
KY421402.1   ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
KY421397.1   ctggttctgt gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
KY421396.1   ctggttctgt gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
KF718799.1   ctggttctgt gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample8      ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample17     ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample13     ctggttctgg gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample3      ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample4      ctggttctgt gag---ggct ccttcaacgg caacgaaaag gagaccatgc agttcttgaa
Sample11     ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample15     ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcatgaa
Sample9      ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample5      ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample18     ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
Sample7      ctggttctgg gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample6      ctggttctgg gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample20     ctggttctgg gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
Sample16     ctggttctgg gag---ggct ccttcaacgg caacgaaaaa gagaccatgc agttcttgaa
Sample10     ctggttctgt gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcctgaa
Sample19     ctggttctgc gag---ggct ccttcaacgg caacgagaag gagaccatgc agttcttgaa
M23912.2     acttgccacc gagtcagggc tgtcaagccg aaaaagagaa gaagcaatag tgtccttgga

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                365        375        385        395        405        415
Sample1      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga atgcggagct
Sample14     cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample2      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421385.1   cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga atgcggagct
```

```
AY835599.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga atgcggagct
KF718801.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421399.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421398.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421391.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421390.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample12      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421402.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421397.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KY421396.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
KF718799.1    cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample8       cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga atgcggagct
Sample17      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga atgcggagct
Sample13      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample3       cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample4       cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample11      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample15      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample9       cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample5       cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample18      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample7       cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample6       cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga atgcggagct
Sample20      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample16      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample10      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
Sample19      cgaccggctg gccagctacc tggagaaggt gcggcagctg gagcgggaga acgcggagct
M23912.2      ctgaccacac cccagtaat- ---gaaggtc agcgctttca tgccacatga gtccagccct

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   425        435        445        455        465        475
Sample1       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample14      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample2       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421385.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
AY835599.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KF718801.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421399.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421398.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421391.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421390.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample12      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421402.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421397.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KY421396.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
KF718799.1    ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample8       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample17      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample13      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample3       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample4       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample11      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample15      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample9       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample5       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample18      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample7       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample6       ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample20      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample16      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample10      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
Sample19      ggagagacgc atcctggagc gcagccagca gcaggagccc ctcgtgtgcc ccaactacca
M23912.2      ttgaagccag accatgaagc tcttatttca -ctgatgccc agtgtggaag aggagttctg

              ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   485        495        505        515        525        535
Sample1       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample14      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample2       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421385.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
AY835599.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
```

```
KF718801.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421399.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421398.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421391.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421390.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample12      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421402.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421397.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KY421396.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
KF718799.1    gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample8       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample17      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample13      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample3       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample4       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample11      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample15      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample9       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample5       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample18      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample7       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample6       gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample20      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample16      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample10      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
Sample19      gtcctacttc cggaccatcg aggagctcca gcagaaggtg aggggctggg cgctccattg
M23912.2      acttagcaag aagcccaact aaagtcttt- -tatgagttg agggact--- ----------

              ....|....| ....|....| ....|....| ....|....| ....|....
                 545        555        565        575        585
Sample1       cctccagcag gaagttgtag ggagggaatc cgagtattgg --ctaaga-
Sample14      cctccagcag gaagttgtag ggagggaatc cgagtattgg --ctaag--
Sample2       cctccagcag gaagttgtag ggagggaatc cgagaattgg ggctaag--
KY421385.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
AY835599.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KF718801.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KY421399.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KY421398.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KY421391.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KY421390.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
Sample12      cctccagcag gaagttgtag ggagggaatc cgagatatgg --cta----
KY421402.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KY421397.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KY421396.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
KF718799.1    cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaag--
Sample8       cctccagcag gaagttgtag ggagggaatc cgagtaatgg --ctaag--
Sample17      cctccagcag gaagttgtag ggagggaatc cgagtaaggg --ctaaga-
Sample13      cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaaga-
Sample3       cctccagcag gaagttgtag ggagggaatc cgaaaattgg --gct----
Sample4       cctccagcag gaagttgtag ggagggaatc cgagtaatgg --gctaaga
Sample11      cctccagcag gaagttgtag ggagggaatc cgagtaatgg --gctaaga
Sample15      cctccagcag gaagttgtag ggagggaatc cgagtaattg --gctaaga
Sample9       cctccagcag gaagttgtag ggagggaatc cgagaatttg --gctaa--
Sample5       cctccagcag gaagttgtag ggagggaatc cgagtattgg --gcta---
Sample18      cctccagcag gaagttgtag ggagggaatc cgagatatgg --gctaag-
Sample7       cctccagcag gaagttgtag ggagggaatc cgagtaatgg --gctaag-
Sample6       cctccagcag gaagttgtag ggagggaatc cgagaattgg --ctaaag-
Sample20      cctccagcag gaagttgtag ggagggaatc cgagtaatgg --ctaag--
Sample16      cctccagcag gaagttgtag ggagggaatc cgagatatgg --ctaaga-
Sample10      cctccagcag gaagttgtag ggagggaatc cgagaaaagt --ggtaag-
Sample19      cctccagcag gaagttgtag ggagggaatc cgagtaattg ggcta----
M23912.2      -cttctgagg aaacata--- ---------- ---------- ---------
```

CHAPTER 5

DISCUSSION

5.1.0 Genetic variation in Swakara sheep of Namibia

Genetic variation has become an objective tool traditionally used for improving animal species (Melus *et al*., 2009). Numerous studies showed that there has been an increase in the number of KAP genes defined in humans and sheep species and progressive accounts of variation in these genes (Hua *et al*., 2012). Rogers *et al.* (2005) illustrated that keratin associated proteins are encoded by a large number of genes which are polymorphic. Between two and nine alleles have been identified among the KAP genes (Gong *et al*., 2010). Gene expression can be affected by the structure and function of the encoded proteins (Elmaci *et al*., 2013), and variation in KAP genes has greatly influenced wool traits (Yardibi *et al*., 2015) in most ovine species.

5.1.1 Length polymorphism analysis of the KAP1.1 amplimer

In this study, three length polymorphism alleles were observed within the KAP1.1 gene, separated on 2% Agarose-gel electrophoresis and designated A, B and C (Figure 4.1). The length polymorphism of the KAP 1.1 locus determined in this study in the Swakara sheep was previously reported in Romney sheep of New Zealand (Roger *et al.*, 1994) and Merino sheep of New Zealand (Itenge-Mweza *et al*., 2007). Sequencing results of the three alleles were in agreement with the previously published KAP1.1 alleles. Upon alignment, the sequencing results revealed that the A, B and C alleles had 100% homology with reference to GenBank accession number FJ169479, GenBank accession number FJ169479.1 and GenBank accession number MG641066, respectively.

Results obtained showed that the B allele frequency was highly distributed in comparison to the A and C allele frequencies. A haplotype diversity of 0.778 was observed confirming the variation in allele distribution. The predominant B allele could be due to adaption of the Swakara sheep in Namibia since their conception in 1907. Yardibi *et al*. (2015) obtained a similar distribution allele frequency in Kivircik, Chios and Awassi wool sheep breeds in Turkey. In Namibia, Swakara sheep are primarily kept for pelt production which are sought for in the fashion industry. The study showed that the KAP1.1 locus is polymorphic, and this may contribute to an improved understanding of the organization of the keratin genes found on chromosomes 3 and 11. Polymorphism observed at the KAP1.1 locus may prompt further studies on KAP genes in Swakara sheep, which could lead to the identification of genetic markers linked to superior pelts.

5.1.2 Polymorphism identified at the KAP1.3 gene

In this study, three alleles were observed within the KAP1.3 locus with amplimer length of 570 bp run on 10% polyacrylamide gel-electrophoresis in Swakara sheep. The study results obtained differed from other published literature; Mahajan *et al*. (2017) used primers published by Chen *et al*. (2011) to detect a 598 bp amplimer using PCR-RFLP *Bsr* 1 restriction enzyme in Rambouillet sheep. Furthermore, the PCR product of 598 bp was reported for KAP1.3 by Rogers *et al*. (2007) in six sheep breeds, and Rogers *et al*. (1994) and Itenge-Mweza *et al*. (2007) also reported it in Merino sheep. In agreement, Xu *et al*. (2008) observed 598 bp amplimer in five sheep breeds of Xinjiang area, Chen *et al*. (2011) among Chinese Merino sheep and Kumar *et al*. (2016) in ten sheep breeds of India. Therefore, indicating the two common methods used to identify polymorphism within the KAP1.3 gene.

In this study three different band patterns were obtained and designated as A, B and C with an allele frequency of 0.5, 0.35 and 0.15, respectively. Overall the A allele was highly distributed in comparison to the B and C alleles. It was observed that the A allele with the highest frequency was inherent mostly in ewes in comparison to rams. Mahajan *et al*. (2017) observed three diverse patterns and genotype using RFLP at two restriction sites which resembled AA (0.41), AB (0.47) and BB (0.12), with an overall 0.645 (A) and 0.355 (B) gene frequencies. Meena *et al*. (2018) reported polymorphism in Magra sheep at the KAP1.3 locus, were three genotypes identified were XX, XY and YY; with a genotype frequency of 0.12, 0.17, 0.71, respectively. Furthermore, with an allele frequency of 0.20 (X) and 0.80 (Y). The YY genotype was predominantly distributed in comparison to XX. Arora *et al*. (2011) study concurs with Meena *et al*. (2018). In contrast Kumar *et al*. (2016) had a higher frequency of XY genotype in comparison to the YY genotype in Indian sheep breeds. Darwish *et al*. (2018) reported seven alleles (A, B, C, D, F, G, T) in the KAP1.3 locus of Egyptian sheep and Powell *et al*. (1983) studies had a similar distribution. Furthermore, Gong *et al.* (2015); Mahajan *et al.* (2017) and Sulayman *et al.* (2017); reported six alleles (α, β, g, V, S, t) and Itenge-Mweza *et al.* (2007) identified 10 alleles designated A-J in New Zealand Merino sheep. However, different RFLP pattern using Bsr 1 restriction enzyme were reported by Rogers *et al*. (1993) in six sheep breeds, Xu *et al*. (2008) in five sheep breeds of Xinjiang.

Mahajan *et al*. (2017) asserts that differences in several studies on sheep could be as a result of selective breeding schemes, breed differences and mere evolutionary pressure.There is genetic variation that exists within the KAP1.3 gene in the Swakara sheep of Namibia.

Sequenced DNA output was used to infer and construct phylogenetic trees between the 20 Swakara sheep located in four geographic locations using MEGA software version 6.0. A haplotype diversity of 0.883 was observed in the KAP 1.3 gene as illustrasted in Table 4.2. Altshuler *et al*. (2005) and Stumpf (2004) define a haplotype as a set of alleles in an organism closely linked on a chromosome and inherited together from a single parent. In agreement, Itenge-Mweza (2007) alluded that tight linkage between closely linked alleles (KAP1.1, KAP1.3 and KRT33A) on a chromosome could be associated with a particular trait of interest. Therefore, the KAP1.3, 0.883 haplotype diversity could be associated to superior pelt quality. Furthermore, breed difference, geographical location, animal husbandry practices and vegetation in an area may influence the allele frequency concentration in a population pool. The KAP1.3 gene was found polymorphic with AA, AB and CC genotypes using KAP1.3 primers and revealed by PCR-SSCP analysis. Polymorphism in KAP1.3 might be a potential genetic marker in Swakara sheep.

5.1.3 Identification of genetic variation in the KAP3.2 gene

The KAP3.2 gene was amplified from genomic DNA of Swakara sheep. One unique monomorphic genotype denoted as AA was observed with amplimer length of 426 bp. A 10% polyacrylamide gel electrophoresis was used to determine the presence of variation within four geographic locations. Itenge (2012) identified three alleles, designated A, B and C at the KAP3.2 locus with the sire used in a half-sib family being homozygous, and was hence deemed uninformative. Furthermore, Wang *et al*. (2011) observed three genotypes; AA, AB and BB with AA genotype highly distributed in comparison to the others and overall with $P>0.05$. In contrary, McLaren *et al*. (1997) and Wang *et al*. (2011) reported two alleles in the KAP3.2 locus. Moreover, the monomorphic distribution may be as a result of species difference and adaptation of the Swakara sheep to the Namibian environment, as studies by Itenge (2012) and Wang *et al*. (2011) had similar findings of the sire used in the half-sib family. The KAP3.2 gene findings in this study could further be as a result of the crossbreeding of the Swakara sheep (then Karakul) with Namaqua Afrikaner and Black head Persian sheep resulting in a uniform inherent pattern of the AA genotype which was mostly observed and confirmed by the sequence haplotype diversity of zero. Based on the samples in the study it may be alluded that the Swakara sheep in Namibia and other sheep breeds may or may not have a monomorphic pattern at the KAP3.2 gene hence the need to optimize the PCR-SSCP protocol.

5.1.4 Polymorphism identified at the KRT33A amplimer

In this study, two length polymorphism alleles were identified at the KRT33A locus with amplimer length of 480 base pairs using PCR-SSCP typing methods. Ahlawat *et al*. (2014) observed three alleles in the Dumba, Marwari and Patanwadi sheep breeds in India. Itenge-Mweza *et al.* (2007) further identified six alleles (A, B, C, D, E, F) in the KRT33A gene, (previously referred to as KRT1.2 gene), of the Merino sheep in New Zealand. The two alleles identified in this study were designated A and B with an allele frequency of 0.25 and 0.75, respectively. Sequence data output revealed a haplotype diversity of 0.733 in the KRT33A gene, which may be used to exploit traits of importance such as pelt quality in Swakara sheep. Kumar *et al.* (2016) reported three genotypes (MM, MN, NN) in Patanwadi and Nali sheep breed with an allele frequency distribution of 0.778 (M) and 0.222 (N), with genotype frequencies of 0.648 (MM), 0.260 (MN) and 0.092 (NN). On the contrary, the Nali breed had allele frequency distribution of 0.643 (M) and 0.357 (N), with genotype frequencies of 0.464 (MM), 0.366 (MN) and 0.170 (NN). Meena *et al*. (2018) observed three genotypes (MM, MN, NN) in the Magra sheep. The MM genotype was predominant in comparison to the NN genotype in Meena *et al.* (2018); which concurs to Kumar *et al*. (2016) studies. However, Meena *et al*. (2018) denoted that MM genotype was 0.56 whilst NN genotype was 0.07 which was lower than that observed among Nali and Patanwadi sheep breeds. A similar pattern was found in a study done by Rogers *et al.* (1993) in Romney, Merino, Dorset Down, Border Leicester and Coop worth sheep. The predominant B allele observed in this study could be attributed to integral pattern of inheritance in Swakara sheep and its adaptation to the Namibian environment. It is also possible that there are more than two alleles within the KRT33A gene of the Swakara sheep, but the PCR-SSCP genotyping method used may not have detected all the variations.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

In Namibia, Swakara sheep are primarily kept for pelt production which are sought after in the fashion industry. Three Keratin genes under study were polymorphic (KAP1.1, KAP1.3 and KRT33A) and one monomorphic (KAP3.2). Polymorphism in the KAP1.1 (3 alleles), KAP1.3 (3 alleles) and KRT33A (2 alleles) positively correlates with the indicated haplotype diversity, illustrated in Table 4.2. Results obtained may contribute to an improved understanding of the organization of the keratin genes (KAP1.1, KAP1.3, KAP3.2 and KRT33A) found on chromosomes 11. Studies have shown that a group of alleles tightly linked on a chromosome are likely to be inherited together from a single parent to the progeny. Hence, with the advent of the three keratin genes found to exhibit high genetic diversity in the study. Expressed as haplotype diversity in Table 4.2 (KAP1.1=0.778; KAP1.3=0.883; KRT33A=0.733), may be further exploited in possible identification of genetic markers associated with exceptional pelt quality in Swakara sheep.

The monomorphic distribution of the KAP3.2 gene could be attributed to homogeneity of all Swakara sheep sampled from the four study sites; Neudamm, Gellap-Ost, Kalahari and Tsumis. Moreover, the study results showed that most ewes had the genotype AB in the KAP1.3 gene. This predominant AB could probably be inherent in most females in comparison to males. Primarily Hardap region had a higher frequency distribution of six in comparison to Khomas with one and Karas region with zero in the KAP1.3 gene.

Studies have shown that during DNA replication different forms of alleles may arise due to mutation (Nicholas, 2010). Therefore, as a result of these different forms, researchers develop genetic markers linked to traits of economic importance in domesticated animals. Genetic markers are chromosomal landmarks used in the identification of specific regions in a gene; which have an inheritance pattern that can be traced. When identified, these genetic markers may be used to complement the currently used visual appraisal technique used during pelting sorting and grading in the Swakara industry. To our knowledge, no genetic markers that are associated with pelt quality traits in Swakara sheep have been identified so far. Polymorphism observed at the KAP1.1, KAP1.3, and KRT33A loci may prompt further studies on KAP genes in Swakara sheep, which may help with the identification of genetic markers linked to superior pelts and strategic selection of breeding stock. It may also potentially improve the accuracy of selection for exceptional lustrous pelt pattern. Campbell (2007) asserts that pattern and curl types improves significantly with the increase in Swakara (then Karakul) genetic contribution.

This work has the potential to provide fundamental baseline knowledge that other researchers can use in better utilization and improvement of the Swakara sheep breeding programs in developing superior pelt quality and overall Swakara sheep sustainability in Namibia. Distribution of allele frequency across the four farms under study slightly differs with what other researchers have reported in literature which could be attributed to the distinctive nature of the Swakara sheep of Namibia.

As a result of improved breeding schemes Namibia's Swakara sheep pelts have been perceived to be different from those of Bukhara, Uzbekistan where the Karakul sheep originated. Furthermore, the study is a crucial baseline platform for other researchers to build on in better utilization and upgrading of the Swakara sheep in Namibia despite that current conclusion were drawn from Swakara sheep from only four farms. Furthermore, the distribution of allele frequency across the four farms under study concurs with what other researchers have reported in literature. Irrespective, of the erratic and harsh climatic conditions in Namibia the Swakara sheep industry has thrived well as seen in the remarkable exceptional pelts produced and sold in Copenhagen, Denmark.

The study recommends that a larger sample size be used that includes both commercial and communal Swakara sheep flocks in order to exploit genetic diversity and inclusion of the breeding herd from commercial farmers. It is further recommended that another study on a half-sib family be conducted using the KAP1.1 gene. Hence, genetic variation would allow successful development of genetic markers linked to superior pelt quality.

The PCR-SSCP protocol needs to be optimized in order to confirm the presence of genetic variation obtained in this study in the three genes (KAP1.1, KAP1.3 and KRT33A), and the monomorphic pattern in KAP3.2 gene in the Swakara sheep of Namibia. It is further recommended that urea be used when preparing polyacrylamide gels, which maintains the denatured state of the PCR product when run on the vertical gels. Furthermore,

Moreover, the study did not perform cloning of allele standards for each locus. However, to improve the genotyping process, it is recommended that for each locus, genomic DNA be selected from various sheep animals such that all the alleles detected are represented in their collective genotypes. This DNA can then be amplified and amplimers subsequently cloned in future studies. The genotype of each animal would then be confirmed accurately by comparing allele standards generated through cloning.

Furthermore, it is advisable to re-sequence the genes in order to validate the results and prior to depositing on the NCBI GenBank

REFERENCES

Ahlawat, A. R, P. U. Gajbhiye, L. L. L. Prince, A. S. Meena, V. B. Dongre, and S. G. Gajjar. 2014. Genetic polymorphism in growth hormone gene in Patanwadi, Marwari and Dumba breeds of sheep. *Indian Journal of Animal Science.* 84(9): 98-99.

Altshuler, D, P. Donnelly, and International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature*. 437(7063): 1299.

Amareswari, P, M. G. Prakash, B. Ekambaram, M. Mahendar, and C. H. Krishna. 2018. Molecular genetics studies on Nellore and Deccani sheep using microsatellite markers. *Indian Journal of Animimal Resolution*. 52(6): 805-810. DOI: 10.18805/ijar.B-3300.

Anonymous. 1998. Agricultural investment opportunities in Namibia. Website on internet. 1-14 pages.

Anonymous. 2005. Swakara product guide. Karakul board of Namibia. Formeset, Cape Town.

Arora, R, H. S. Yadav, S. Bhatia, and B. P. Mishra. 2011. Genetic polymorphism in the KAP1.3, CYHR1 and BMP 15 genes in Indian Sheep. *Indian Journal of Animal Sciences*. 64: 219-225.

Awgichew, K, T. A. Gipson, R. C. Merkel, and A. Yami. 2007. Selecting breeding stock for sheep production. Tech. Bull. No. 4. ESGPIP. https://issuu.com/esgpip/docs/esgpip_sheep_and_goat_handbook

Bai, L, H. Gong, H. Zhou, J. Tao, and J. G. H. Hickford. 2018. A nucleotide substitution in the ovine KAP20-2 gene leads to a premature stop codon that affects wool fibre curvature. *Animal Genetics*. 49: 357-358. https://doi.org/10.1111/age.12668. (Accessed 12 August 2019.)

Bai, L, J. Wang, H. Zhou, H. Gong, J. Tao, and J. G. H. Hickford. 2019. Identification of ovine KRTAP28-1 and its association with wool weight and mean fibre diameter-associated traits. *Animals*. 9:142.

Benbouza, H, J. M. Jacquemin, J. P. Baudoin, and G. Mergeai. 2006. Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. *Biotechnologie, agronomie, société et environnement*. 10(2): 77-81.

Bravenboer, B. 2007. *Karakul-Gift from the Arid Land Namibia 1907-2007*. Karakul Board of Namibia and the Karakul Breeders' Society of Namibia. Windhoek, Namibia.

Campbell, L. J. 2007. Evaluation of two indigenous South African sheep breeds as pelt producers. Doctoral dissertation, University of Pretoria.

Clark, E. L, S. J. Bush, M. E. P. McCulloch, I. L. Farquhar, R. Y. L. Lefevre, C. Pridans, H. G. Tsang, C. Wu, C. Afrasiabi, M. Watson, C. B. Whitelaw, T. C. Freeman, K. M. Summers, A. L. Archibald, and D. A. Hume. 2017. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLOS Genet*. 13(9): e1006997. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5626511/ (Accessed 14 April 2018.)

Chen, H. Y, X. C. Zeng, W. Q. Hui, Z. S. Zhao, and B. Jia. 2011. Developmental expression patterns and association analysis of sheep KAP8.1 genes in Chinese Merino sheep. *Indian Journal of Animal Sciences*. 81(4): 391-396. ISSN: 0367-8318.

Darwish, H. R, I. M. Farag, A. M. Darwish, M. G. Eshak, and R. W. Ahmed. 2018. Genetic polymorphism of KRT1.2 gene and its association with improving of some wool characteristics in Egyptian sheep. *Asian Journal of Scientific Research*. 11: 295-300.

Davis, G. H, S. M. Galloway, I. K. Ross, S. M. Gregan, J. Ward, B. V. Nimbkar, P. M. Ghalsasi, C. Nimbkar, G. D. Gray, I. I. Subandriyo, B. Tiesnamurti, E. Martyniuk, E. Eythorsdottir, P. Mulsant, F. Lecerf, J. P. Hanrahan, G. E. Bradford, and T. Wilson. 2002. DNA tests in prolific sheep from eight countries provide new evidence on origin of the Booroola (FecB) mutation. Biology of Reproduction. 66(6): 1869-74.

"Development of a Sector Growth Strategy for the Namibian Agro-processing Industry: Swakara Wool Processing-In-depth Value Chain Analysis and Strategy" by Optimal Agricultural Business Systems (OABS), April 2016.

http://www.mti.gov.na/downloads/Swakara%20Wool_Strategies_Web.pdf

Devandra, C, and G. B. McLeroy. 1982. *Goat and Sheep production in the tropics*. Longman.

Duffield-Harding, J. 2005. Swakara product guide: The Karakul Board of Namibia. Former-set Printers. Cape Town. ISBN 99916-68-04-7.

Ekegbu, U. J., I. S. Haruna, G. Mahmoud, H. Zhou, and J. G. H. Hickford. 2018.

Genetic polymorphisms in New Zealand sheep breeds. *World Journal of Agriculture and Soil Science*. 1(2): 1-4.

Elmaci, C, S. Sahin, and Y. Oner. 2013. Distribution of different alleles of aromatase cytochrome P450 (CYP19) and melatonin receptor 1A (MTRN1A) genes among native Turkish sheep breeds. *Kafkas Univ Vet Fak Derg*. 19: 929-933. DOI: 10. 9775/kvfd.2013.8900.

Ensminger, M. E. 1991. Animal Science: Animal Agriculture Series. Interstate Publishers, INC. Danville, Illinois.

FAO. 2009. Sheep and goats for diverse products and profits, by Pollott, G and R. T. Wilson. FAO Diversification booklet No. 9. Rome.

Farmers Weekly. 2012. www.swakara.net. (Accessed 10 July 2017.)

Fuchs, E, and K. Weber. 1994. Intermediate filaments: Structure, dynamics, function and disease. In Richardson, C. C, (Ed). *Annual Review of Biochemistry*, Annual Reviews. pp. 345-382.

Gillespie, J. M. 1990. The proteins of hair and other hard a-keratins. In *Cellular and molecular biology of intermediate filaments*. Springer. pp. 95-128.

Goddard, D. R, and L. Michaelis. 1935. Derivatives of keratin. *J. Biol. Chem*. 112(1): 361-371.

Gong, H, H. Zhou, J. M. Dyer, and J. G. Hickford. 2011. Identification of the ovine KAP11-1 gene (KRTAP11-1) and genetic variation in its coding sequence. *Mol Biol Rep*. 38:5429-33.

Gong, H, H. Zhou, J. M. Dyer, and J. G. H. Hickford. 2014. The sheep KAP8-2 gene, a new KAP8 family member that is absent in humans. *SpringerPlus*. 3: 528. https://doi.org/10.1186/2193-1801-3-528 (Accessed 18 August 2019.)

Gong, H, H. Zhou, J. M. Dyer, J. E. Plowman, and J. G. Hickford. 2011a. Identification of the Keratin associated protein 13-3 (KAP13-3) gene in sheep. *Open J Genet*. 1: 60-4.

Gong, H, H. Zhou, Z. Yu, J. Dyer, J. E. Plowman, and J. G. H. Hickford. 2011b. Identification of the ovine keratin-associated protein KAP1-2 (KRTAP1-2). *Exp. Dermatol*. 20: 815-819.

Gong, H, H. Zhou, and J. G. Hickford. Polymorphism of the ovine keratin-associated protein 1-4 gene (KRTAP1-4). 2010. *Mol. Biol. Rep.* 37: 3377-3380.

Gong, H, H. Zhou, S. Hodge, J. M. Dyer, and J. G. Hickford. 2015. Association of wool traits with variation in the ovine KAP1-2 gene in Merino cross lambs. *Small Rumin. Res*. 124: 24-29.

Gong, H, H. Zhou, G. W. McKenzie, Z. Yu, S. Clerens, J. M. Dyer, J. E. Plowman, M. W. Wright, R. Arora, C. S. Bawden, Y. Chen, J. Li, and J. G. H. Hickford. 2012. An updated nomenclature for Keratin-Associated Proteins. *International Journal of Biological Sciences*. 8(2): 258-264. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3269608/ (Accessed 14 April 2018.)

Gong, H, H. Zhou, R. H. Forrest, S. Li, J. Wang, J. M. Dyer, Y. Luo, and J. G. H. Hickford. 2016. Wool keratin-associated protein genes in sheep-A review. *Genes*. 7:24. DOI: 10.3390/genes7060024.

Gong, H, H. Zhou, J. Wang, S. Li, Y. Luo, and J. G. H. Hickford. 2019. Characterization of an Ovine Keratin Associated Protein (KAP) gene, which would produce a protein rich in glycine and tyrosine but lacking in cysteine. *Genes*. 10: 849. DOI: 103390/genes10110848.

Hanukoglu, I, and E. Fuchs.1983. The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins. *Cell*. 33: 915-924.

Hediger, R, H. A. Ansari, and G. F. Stranziger. 1991. Chromosome banding and gene localizations support extensive conservation of chromosome structure between cattle and sheep. *Cytogenetic and Genome Research*. 57(2-3): 127-134.

Heren, R. 2000. The Science of Animal Agriculture. Delmar Publishers.

https://www.icalcu.com/stat/chisqtest.html (Accessed 14 February 2020.)

Hua, G, Z. Huitong, W. Grant, Y. Zhidong, C. Stefan, M. D. Jolon, E. P. Jeffery, W. W. Matthew, A. Reena, B. Simon, C. Yulin, L. Jinquan, and G. H. H. Jonathan. 2012. An updated nomenclature for Keratin Associated Proteins (KAPs). *Int. J. Biol. Sci*. 8: 258-264. DOI: 107150/ijbs. 3278.

Hua, G, Z. Huitong, and J. G. H. Hickford. 2011. Diversity of the glycine/tyrosine-rich Keratin Associate Protein 6 gene (KAP 6) family in sheep. *Mol. Biol. Rep*. 38: 31-38. DOI: 10. 1007/s11033-010-0074-6.

Itenge, T. O. 2002. The search for genetic markers affecting wool characteristics. *Science in Africa*.

Itenge-Mweza, T. 2007. Identification of genetic markers associated with wool quality traits in merino sheep. Doctoral dissertation. University of Lincoln.

Itenge, T. O, and M. N. T. Shipandeni. 2015. Sale Trends of Swakara Pelt Offered at the Copenhagen Fur Auction from 1994-2013. *Applied Animal Husbandry & Rural Development*. 8:1-5.

Itenge-Mweza, T. O, R. H. J. Forrest, G. W. McKenzie, A. Hogan, J. Abbott, O. Amoafo, and J. G. H. Hickford. 2007. Polymorphism of the KAP1.1, KAP1.3 and K33 genes in Merino sheep. *Molecular and cellular probes*. 21(5): 338-342.

Itenge, T. O. 2012. Identification of Polymorphism in the Keratin Genes (KAP3. 2, KAP6.1, KAP7, KAP8) and Microsatellite BfMS in Merino Sheep Using Polymerase Chain Reaction-Single Strand Conformational Polymorphism (PCR-SSCP) Analysi. In *Electrophoresis*. IntechOpen. pp. 193-220. ISBN 980-953-307-117-1.

Karakul Board Annual Report. 2008/2009. Swakara.

Karakul Annual Report. 2010. (50th edit). Karakul Breeders' Society and Karakul Board of Namibia.

Khan, I, E. Maldonado, V. Vasconcelos, J. Stephen, W. E. Johnson, and A. Antunes. 2014. Mammalian keratin associated proteins (KRTAPs) Sub-genomes: Disentangling hair diversity and adaptation to terrestrial and aquatic environments. *BMC genomics*. 15(1): 779. DOI: 10.1186/1471-2164-15-779 PMID: 25208914.

Kirsten, G. J. C. 1966. The Production and Marketing of Karakul Pelts. *Agrekon*. 5(3):

22-29.

Korge, B. P, J. G. Compton, P. M. Steinert, and D. Mischke. 1992a. The two size alleles of human keratin 1 are due to a deletion in the glycine-rich carboxyl-terminal V2 subdomain. *J. Invest. Dermatol*. 99: 697-702.

Korge, B. P, S. Q. Gran, W. McBride, D. Mischke, and P. M. Steinert. 1992. Extensive size polymorphism of the human keratin 10 chain residues in the C-terminal V2 subdomain due to variable numbers and sizes of glycine loops. *Proc. Nat. Acad. Sci*. 89: 910-914.

Kumar, P, V. Choudhary, K. G. Kumar, T. K. Bhattacharya, B. Bhushan, A. Sharma, and A. Mishra. 2006. Nucleotide sequencing and DNA polymorphism studies on IGFBP-3 gene in sheep and its comparison with cattle and buffalo. *Small Ruminant Research*. 64(3): 285-292.

Kumar, R, A. S. Meena, R. Kumari, B. Jyotsana, L. L. L. Prince, and S. Kumar. 2016. Polymorphism of KRT1.2 and KAP1.3 genes in Indian sheep breeds. *Indian Journal of Small Ruminants*. 22(1): 28-31.

Kevorkian, S. E. M, M. Zaulet, M. A. Manea, S. E. Georgescu, and M. Costache. 2011. Analysis of ORF region of the prion protein gene in the Botosani Karakul sheep breed from Romania. Turk. *J. Vet. Anim. Sci*. 35(2): 105-109.

Kunchi, C, C. K. Venkateshan, and R. B. Adusumalli. 2019. Nano-indentation of hair cortex and medulla regions. *Fibers and polymers*. 20(7): 1538-1545.

Kruger, B, P. Hugo, and J. van Zyl. 2013. The importance of the Namibian Swakara

Industry and opportunities for growth. Agra Professional Services, Windhoek, Namibia.

Langbein, L, and J. Schweizer. 2005. Keratins of the Human Hair Follicle. *Int. Rev. Cytol*. 243: 1-78.

Li, W, H. Gong, H. Zhou, and J. Wang. 2018. Variation in the ovine keratin-associated protein 15-1 gene affects wool yield. *The Journal of Agricultural Science*. Doi.org/10.1017/S0021859618000953.

Li, S, H. Zhou, H. Gong, F. Zhao, J. Wang, X. Liu, J. Hu, Y. Luo, and J. G. H. Hickford. 2019. Identification of the ovine Keratin-Associated Protein 21-1 gene and its association with variation in wool traits. *Animals*. 9(450). DOI: 10.3390/ani9070450.

Li, W, H. Gong, H. Zhou, J. Wang, S. Li, X. Liu, Y. Luo, and J. G. H. Hickford. 2019. Variation in KRTAP6-1 affects wool fibre diameter in New Zealand Romney ewes. *Arch. Anim. Breed*. 62: 509-515. https://doi.org/10.5194/aab-62-509-2019. (Accessed 18 August 2019.)

Li, S, H. Zhou, H. Gong, F. Zhao, J. Wang, X. Lui, Y. Luo, and J. G. Hickford. 2017a. Identification of the ovine keratin-associated protein 22-1 (KAP22-1) gene and its effect on wool traits. *Genes*. 8: 27.

Li, S, H. Zhou, H. Gong, F. Zhao, J. Wang, Y. Luo, and J. G. H. Hickford. 2017. Variation in the ovine KAP6-3 gene (KRTAP6-3) is associated with variation in mean fibre diameter-associated wool traits. *Genes*. 8(204). DOI: 10.3390/genes8080204.

Lundie, R. S. 2011. The genetics of color in fat-tailed sheep: A review. *Tropical Animal*

Health Production. 43: 1245 - 65. DOI: 10.1007/s11250-011-9850-0.

Mahajan, V, A. K. Das, R. K. Taggar, D. Kumar, and R. Sharma. 2017. Association of polymorphic variants of KAP1.3 gene with wool traits in Rambouillet sheep. *Indian J. Anim. Sci.* 87: 1237-1242.

Malesa, M.T. 2015. Population genetics of Swakara sheep inferred using genome-wide SNP genotyping. MSc Thesis. University of KwaZulu-Natal. https://researchspace.ukzn.ac.za/xmlui/bitstream/handle/10413/13979/Malesa_Masoko_Tshenakaho_2015.pdf?sequence=1&isAllowed=y

Marshall, R. C, D. F. G. Orwin, and J. M. Gillespie 1991. Structure and biochemistry of mammalian hard keratin. *Electron microscopy reviews*. 4(1): 47-83. DOI: 10.1016/0892-0354 (91) 90016-6.

Martins, C, and K. J. Peters. 1992. Alternative use of Karakul sheep for pelt and lamb production in Botswana. I. Reproduction and growth performance. *Small Ruminant Research*. 9: 1-10.

Matsunanga, R, R. Abe, D. Ishii, S. I. Watanabe, M. Kiyoshi, B. Nocker, M. Tsichiya, and K. Tsumoto. 2013. Bidirectional binding property of high glycine-tyrosine keratin-associated protein contributes to the mechanical strength and shape of hair. *J Struct Biol*. 183(3): 484-494. DOI: 10.1016/j.jsb.2013.06.005. PMID: 23791804.

McLaren, R. J, G. R. Rogers, K. P. Davies, J. F. Maddox, and G. W. Montgomery. 1997.

Linkage mapping of wool keratin and keratin-associated protein genes in sheep. *Mamm. Genome*. 8: 938-940.

Meena, A. S, R. Kumar, B. Jyotsana, H. K. Narula, and S. Kumar. 2018. Genetic polymorphism of KRT1.2, KAP1.3 and THH gene in Magra sheep. *Indian Journal of Small Ruminants*. 24(1): 27-30.

Melus, V, R. Kasarda, O. Kadlecik, and A. Trakovicka. 2009. Breeding potential of the Slovak Pinzgau cattle seeking biochemical and molecular biological traits. *Archiva Zootechnica*. 12: 34-37.

Ministry of Agriculture Water and Forestry (MAWF): Namibia livestock catalogue. 2013. John Meinert Print, Windhoek. ISBN: 978-99945-0-067-3.

Muchadeyi, F. C, M. T. Malesa, P. Soma, and E. F. Dzomba. 2015. Runs of Homozygosity in Swakara pelt producing sheep: implications on sub-vital performance. Proceedings for Association for the Advancement of Animal Breeding and Genetics. 21: 310-313.

Nambahu, H. N. Personal communication. 4 July 2019.

Namibia's 5th National Development Plan. 2017. National Planning Commission of Namibia: www.npc.gov.na/ (Accessed 13 March 2018.)

Nel, J. A. 1966. Genetic studies in Karakul sheep. Doctoral dissertation. University of Stellenbosch.

Nicholas, F. W. 2010. (3rd edit). *Introduction to Veterinary Genetics*. Wiley-Blackwell Publishers.

Neubert, L. 1989. The Karakul Industry: Policy Options for Independent Namibia.

United Nations Institute for Namibia.

Nsoso, S. J, and M. J. Madimabe. 2003. A survey of karakul sheep farmers in southern Kalahari, Botswana: management practices and constraints to improving production. *South African Society of Animal Science.* 4: 23-7. https://www.nda.agric.za/docs/AAPS/Articles/Sheep/Karakul.pdf

Nsoso, S. J, and M. J. Madimabe. 1999. The sheep industry in Botswana: promoting the Karakul sheep industry. *South African Journal of Animal Science*. 29(3).

Onions, W. J. 1962. Wool- an introduction to its properties, varieties, uses and production. Wool. An introduction to its properties, varieties, uses and production.

Parry, D, and P. M. Steinert. 1992. Intermediate filament structure. *Cur Opin Cell Biology*. 4: 94-98. DOI: 10.1016/51570-0232 (02) 00211-8.

Plowman, J. E. 2003. Proteomic database of wool components. *J. Chromatogr*. 787: 63-76.

Powell, B. C, J. Arthur, and A. Nesci. 1995. Characterisation of s gene encoding a cysteine-rich keratin associated protein synthesized late in rabbit hair follicle differentiation. *Differentiation.* 58(3): 227-232.

Powell, B. C. 1997. Molecular Genetics of Sheep. In Piper, L. and A. Ruvinsky. (Eds.). *The Genetics of Sheep.* CAB International. pp. 149-161.

Powell, B. C, and G. E. Rogers. 1986. Hair keratin: Composition, structure and biogenesis. In Biology of the integument. *Springer*. pp. 695-721.

Powell, B. C, M. J. Sleigh, K. A. Ward, and G. E. Rogers. 1983. Mammalian keratin gene families: Organisation of genes coding for the B2 high-sulphur proteins of sheep wool. *Nucleic Acids Research*. 11: 5327-5346.

Powell, B. C. 1996. The keratin proteins and genes of wool and hair. *Wool Technology and Sheep Breeding*. 44:100-118.

Powell, B. C, and G. Rogers. 1997. The role of keratin proteins and their genes in the growth, structure and properties of hair. *Exs.* 78: 59-148.

Powell, B. C, and G. E. Rogers. 1994. Differentiation in hard keratin tissues: Hair and related structures. In: Keratinocyte Handbook. Leigh I. M., Lane E. B. and Watt F. M, (Eds.). Cambridge University Press. pp. 401-436.

Rogers, G. E. 2004. Hair follicle differentiation and regulation. *Int. J. Dev. Biol*. 48: 163-170.

Rogers, G. R, J. G. H. Hickford, and R. Bickerstaffe. 1993. Msp 1 RFLP in the gene for a Type I intermediate filament wool keratin. *Animal Genetics*. 24: 218.

Rogers, M. A, and J. Schweizer. 2005. Human KAP genes, only the half of it? Extensive size polymorphisms in hair keratin-associated protein genes. *J. Investig*. Dermatol. 124: 7-9.

Rogers, G. R. 1994. Identification of a quantitative trait loci for wool strength in Romney sheep. Doctoral dissertation. University of Lincoln.

Rogers, M. A, H. Winter, L. Langbein, A. Wollschlager, S. Praetzel-Wunder, L. F. Jave-Suarez, and J. Schweizer. 2007. Characterization of human KAP24.1, a cuticular hair keratin associated protein with unusual amino-acid composition and repeat structure. *J. Invest. Dermatol*. 127: 1197-1204.

Rogers, M. A, L. Langbein, S. Praetzel-Wunder, H. Winter, and J. Schweizer. 2006. Human hair keratin-associated proteins (KAPs). *Int. Rev. Cytol*. 251: 209-263.

https://doi.org/10.1016/S0074-7696(06)51006-X. (Accessed 18 August 2019.)

Rogers, M. A, L. Edler, H. Winter, L. Langbein, I. Beckmann, and J. Schweizer. 2005. Characterization of new members of the human type II keratin gene family and a general evaluation of the keratin gene domain on chromosome 12q13. 13. *Journal of investigative dermatology*. 124(3): 536-544.

Rogers, G. R, J. G. H, Hickford, and R. Bickerstaff. 1994. Polymorphism in two genes for B2 high sulphur proteins of wool. *Animal Genetics* 25: 407-415.

Sanguinetti, C. J, N. E. Dias, A. J. G. Simpson. 1994. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques*. 17: 915-919.

Schoeman, S. J. 1998. Genetic and environmental factors influencing the quality of pelt traits in karakul sheep. *South African Journal of Animal Science*. 38(3/4): 125-135.

Schoeman, S. J. 1968. Phenotypic parameters in Karakul Sheep.

Shimomura, Y, and M. Ito. (Eds.). 2005. Human hair keratin-associated proteins. *J Investiq Dermatol Symp Proc*. 10(3): 230-3. PMID: 16382671.

Steinert, P. M, and D. R. Roop. 1988. Molecular and cellular biology of intermediate filaments. *Ann. Rev. Biochem*. 57: 593-625.

Steinert, P. M, A. C. T. North, and D. A. D. Parry. 1994. Structural features of keratin intermediate filaments. *J. Invest*. Dermatol. 103: S19-S24.

Stumpf, M. P. 2004. Haplotype diversity and SNP frequency dependence in the description of genetic variation. European journal of human genetics. 12(6): 469.

Sulayman, A, A. Mamat, M. Taursun, X. X. Huang, K. Tian, X. Xu, and X. Fu. 2017. Identification of polymorphisms and association of five KAP genes with sheep wool traits. *Asian Aust. J. Anim. Sci*. DOI 10.5713/ajas.16.06.32.

Swakara Annual Report. 2015/2016. (54th edit). Annual Report 2016 Swakara Breeders Society and Swakara Board of Namibia. ISSN: 2026-7835.

Swakara Annual Report. 2011. (51st edit). Swakara Breeders' Society and Karakul Board of Namibia.

The Karakul Board of Namibia. 2006. Code of Practice for the Care and Handling of Karakul Sheep in Namibia. The Karakul Board of Namibia. Windhoek, Namibia.

Visser, W. H, P.H. Hugo, and Krommenhoek. 1998. Spotlight on Agriculture: Karakul-Black Diamond. MAWRD. Capital press. 10: 3059. www.swakara.net (Accessed 20 August 2017.)

Wang, Z. Y, Y. L. Chen, Q. L. Xu, Z. R. Ma, and Q. Q. Qi. 2011. Polymorphism of KAP3.2 gene and its effect on partial economics traits in Tibetan sheep (J). *Chinese journal of Animal and Veterinary Sciences*. 2.

Wang, J, H. Zhou, J. Zhu, J. Hu, X. Liu, S. Li, Y. Luo, and J. G. Hickford. 2017. Identification of the ovine keratin-associated protein 15-1 gene (KRTAP15-1) and genetic variation in its coding sequence. *Small Rumin. Res*. 153: 131-136.

Wang, J, H. Zhou, Y. Luo, M. Zhao, H. Gong, Z. Hao, J. Hu, and J. G. Hickford. 2019. Variation in the caprine KAP24-1 gene affects cashmere fibre diameter. *Animals*. 9(1): 5.

www.swakara.net (Accessed 5 October 2017.)

Xu, H. F, Z. S. Zhao, A. Y. Xue, L. G. L. Amina, Q. Ban, J. Zhang. 2008. Studies on the relation between KAP1.3 gene and 5 breeds of sheep, and with some wool traits. *Journal of Shihezi University (Natural Science).* 26(1): 60-63.

Yardibi, H, F. E. Gursel, and A. Ates. 2015. Polymorphism of the Kap 1.1, Kap 1.3 and K33 Genes in Chios, Kivircik and Awassi. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi.* 21(4): 535-538.

Yu, Z. D, J. E. Wildermoth, O. A. M. Wallace, S. W. Gordon, N. J. Maqbool, P. H. Maclean, A. J. Nixon, and A. J. Pearson. 2011. Annotation of sheep keratin intermediate filament genes and their patterns of expression. *Exp Dermatol*. 20(7): 582-8.
DOI: 10.1111/j.1600-0625.2011.01274.x PMID: 21554405.

Zhang, Y, He, Y. Xue, and Y. Chen. 2011. Sequence variant in the KAP1.1 gene associate with Cashmere trait in two Cashmere goat breeds. *African Journal of Agricultural Research*. 6(12): 2889-2894.

Zhao, Z, G. Liu, X. Li, J. Huang, Y. Xiao, X. Du, and M. Yu. 2016. Characterization of the promoter regions of two sheep Keratin-Associated Protein Genes for hair cortex-specific expression. *Plos One*. 11(4): e0153936.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4839604/
(Accessed 7 May 2017.)

Zhou, H, H. Gong, J. Wang, J. M. Dyer, Y. Luo, and J. G. H. Hickford. 2016. Identification of four new gene members of the KAP6 gene family in sheep. *Sci. Rep*. 6: 24074. https://doi.org/10.1038/srep24074 (Accessed 19 August 2019.)

Zhou, H, H. Gong, W. Yan, Y. Luo, and J. G. Hickford. 2012. Identification and sequence analysis of the keratin associated protein 24-1 (KAP24-1) gene homologue in sheep. *Gene*. 511: 62-5.

APPENDICES

APPENDIX A: Ethical clearance certificate

UNAM
UNIVERSITY OF NAMIBIA

ETHICAL CLEARANCE CERTIFICATE

Ethical Clearance Reference Number: FANR/003/2019 **Date: 28 October 2019**

This Ethical Clearance Certificate is issued by the University of Namibia Animal Research Ethics Committee (AREC) in accordance with the University of Namibia's Research Ethics Policy and Guidelines. Ethical approval is given in respect of undertakings contained in the Research Project outlined below. This Certificate is issued on the recommendations of the ethical evaluation done by the Research & Publications Committee sitting with the Postgraduate Studies Committee.

Title of Project: Identification of genetic variation in keratin genes of the Swakara sheep in Namibia

Nature/Level of Project: Master by Thesis

Researcher: Nellia Felicity Nyoni

Student Number: 201212522

Faculty: Faculty of Agriculture and Natural Resources

Supervisors:

Take note of the following:

(a) Any significant changes in the conditions or undertakings outlined in the approved Proposal must be communicated to the AREC. An application to make amendments may be necessary.

(b) Any breaches of ethical undertakings or practices that have an impact on ethical conduct of the research must be reported to the AREC.

(c) The Principal Researcher must report issues of ethical compliance to the AREC (through the Chairperson of the Faculty/Centre/Campus Research & Publications Committee) at the end of the Project or as may be requested by AREC.

(d) The AREC retains the right to:

(i) Withdraw or amend this Ethical Clearance if any unethical practices (as outlined in the Research Ethics Policy) have been detected or suspected,

(ii) Request for an ethical compliance report at any point during the course of the research.

AREC wishes you the best in your research.

AREC Chairperson (Dr M Janik)

APPENDIX B: DNA extraction protocol one
ZR Genomic DNATM – Tissue Mini-Prep: Whole Blood, Serum and Plasma

The following is for the purification of DNA from 100 µL whole blood, serum or plasma (the volume can be adjusted to a maximum of 200 µL depending on your requirements. Fresh, frozen or preserved blood (in EDTA, citrate, or heparin) can be used. If the material cannot be processed immediately, the sample can be "stabilized" for later processing (as noted below) although the immediate processing is of blood samples is recommended.

1) Adjust total volume of sample (blood) to 100 µL with water in a micro-centrifuge tube and then add the following:

2X Digestion buffer 95 µL

Proteinase K 5 µL

*Vortex

2) Incubate the tubes at 55ºC for 20 minutes.

3) Add 700 µL of Genomic Lysis Buffer to the tube

*Vortex

4) Transfer the mixture to a Zymo-SpinTM 11C Column in a collection tube

*Centrifuge at 10 000 X g for a minute

5) Add 200 µL of DNA Pre-Wash Buffer to the spin column in a new collection tube

*Centrifuge at 10 000 X g for a minute

6) Add 400 µL of g-DNA Wash Buffer to the spin column

*Centrifuge at 10 000 X g for a minute.

7) Transfer the spin column to a clean micro-centrifuge tube (Eppendorf). Add 50 µL DNA Elution Buffer or water to the spin column. Incubate 2-5 minutes at room temperature and then centrifuge at top speed for 30 seconds to elude the DNA. The eluded DNA can be used immediately for molecular based applications or stored at -20ºC for future use.

APPENDIX C: DNA extraction protocol two

1. Add 400 µL of Genomic Lysis Buffer to 100 µL of blood, serum, or plasma (4:1). Thoroughly mix by vortexing 4-6 seconds, then let stand for 5-10 minutes at room temperature.

Note: Add 200 µL of Genomic Lysis Buffer to all samples less than 50 µL. For samples larger than 50 µL add proportional amount (4:1) Genomic Lysis Buffer (e.g. add 800 µL Genomic Lysis Buffer to 200 µL of blood sample)

2. Transfer the mixture to Zymo-Spin IIC™ column in a collection tube. Centrifuge at 10, 000 x g for 1 minute. Discard the collection tube with the flow trough.

3. Transfer the Zymo-Spin IIC™ column to a new collection tube. Add 200 µL of DNA Pre-Wash Buffer to the spin column. Centrifuge at 10,000 x g for one minute.

4. Add 500 µL of g-DNA Wash Buffer to the spin column. Centrifuge at 10,000 x g for one minute.

5. Transfer the spin column to a clean micro-centrifuge tube. Add 50 µL DNA Elution Buffer or water to the spin column. Incubate 2-5 minutes at room temperature and then centrifuge at top speed for 30 seconds to elude the DNA. The eluded DNA can be used immediately for molecular based applications or stored -20ºC for future use.

APPENDIX D: Polyacrylamide gel electrophoresis (PAGE) preparation

Prepare PAGE electrophoresis running buffer for the tank

1) 1 X TBE in a 7 litre PAGE tank

2) $C_1 V_1 = C_2 V_2$

10* X = 1 * 7 Litres

X= 0.7 Litres

Therefore 7 – 0.7 litres = 6.3 litres (Distilled water)

* = Multiplication sign

C = Concentration

V = Volume

Polyacrylamide gel electrophoresis methods were adapted from Sanguinetti *et al*. (1994) and, the DNA bands on PAGE were visualized by the silver staining technique described by (Benbouza *et al*., 2006; Itenge 2007) with minor modifications facilitated by a team from the Ministry of Agriculture and the student.

Steps involved:

1. Though-roughly wash glass plates with soap detergent using a soft sponge. Repeat step to ensure that the glass plates are clean
2. Let it air dry for at-least 30 minutes
3. Use 5ml of absolute ethanol to clean the dry glass sandwich with smooth paper towels
4. Let it air dry for a minimum of two minutes
5. With the aid of a stand, let the two glasses lean on a stand, spray gentle WD40 silicon free spray (often used in mechanics) and is readily available in the market
6. Let it air dry
7. Place the glass sandwich together and cast on a PAGE stand
8. Firmly clip the sides of the stand to minimize leakage of PAGE solution when placed in between the sandwich (the two glasses)
9. PAGE gel preparation: Prepare 40% Acrylamide solution

a) Add 38g of acrylamide powder, add 2g of methylene bis-acrylamide and make up to100 mL with distilled water.
b) Prepare 30% Ammonium persulphate

bi.) Weigh 3 grams of APS and top up with distilled water to make up a 10 mL of working stock

* APS can be prepared once a week

How to prepare PAGE gel (30mL)

a) From the above 40% Acrylamide solution use 7.5 mL

b) For each polyacrylamide gel, use 12.6 grams of Urea (7 M)

c) Add 3 mL of 10X TBE

d) Top up with distilled water to make 30 mL

e) From the prepared APS, add 300 µL of APS

f) Lastly, add 30 µL of TEMED

*TEMED is added lastly as it quickly crystalizes the PAGE solution

10. Gentle load your gel quickly using a 5 mL pipette, delay in this may result in PAGE gel crystalizing while in the beaker. Hence, focus is needed in this phase. Let your PAGE gel crystalize upon successful loading in between the glass sandwich for at-least an hour
11. Remove the casting stand and place your glass sandwich in the PAGE tank, pre-run to remove excess urea
12. Clean each well using a 5 mL pipette, so as to flush out urea or any gel blocking your wells prior to loading your rapidly denatured PCR products
13. Load the PCR products, set up the machine, ensure that all codes are appropriately connected
14. Run the PAGE protocol

Challenges encountered:

1. Failure to split the glass sandwich as the silicon agents firstly, clamped together, until WD40 was sourced. The PAGE protocol in this study was standardized in 6 weeks after numerous trials
2. Failure of glass sandwich to split resulted in loss of PCR amplimers

APPENDIX E: Gel staining procedure
Method adapted from Itenge, 2007.

Fixation stock solution

1) In a 1L add 100 mL of Absolute ethanol add 5 mL of Acetic acid (100%)

Fixation working solution

2) Use 500 mL from the Fixation stock solution

Staining solution

3) From the stock solution above (Fixation stock solution) prepare a staining solution

4) Use 500 mL of the fixing solution and add 1g of Silver nitrate ($AgNO3$) with a final concentration of 1X.

*Use a brown bottle

*Work at room temperature

Developer solution

5) Use 500 mL of distilled water

6) Add 15g of Sodium hydroxide (NaOH)

7) And lastly add 600 µL of Formaldehyde

*Add distilled water, NaOH and a magnetic stirrer in a bottle. Heat on a hot plate to completely dissolve the NaOH pellets at about 30ºC. Add formaldehyde, keep the solution on the hot plate until the formaldehyde is mixed, and remove from the hot plate.

Rinse phase

8) Use 500 mL of distilled water to rinse

Stop solution

1) From the Fixing stock solution, use 500mL as a stop solution

*Stop solution step can be optional as the developer solution does not often darken your polyacrylamide gel.

Gel staining procedure

1) Fix with 500 mL fixing solution for 5 minutes, remove the solution

2) Stain with 500 mL staining solution for 7 minutes, remove the solution

3) Rinse with 500 mL of distilled water for 3 minutes, remove the water

4) Develop with 500 mL developer solution; leave until yellow colour develops and bands are clearly seen, remove the solution

5) Fix with 500 mL fixing solution for 3 minutes, remove the solution and score

*In the study gels were developed for approximately 35 minutes

*The fixing solution on the last phase is optional as it was observed the yellow gel background remains as is overnight

APPENDIX F: Sequences used to construct phylogenetic trees

KAP1.1

```
>Sample1
----------------CAACTCCTGACACCATGGCCtGCTGTTCCACCAGCTTCTGTGGA
TTTCCCATCTGTTCCACTGGTGGGACCTGTGGCTCCAGTCCCTGCCAGCCGACCTGCTGC
CAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGC-----------CAAC
CGATCTCCATCCAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGCCAGC
CAACCTGCCTCCAGACCAGTGGCTGTGAGACGGGCTGTGGCATTGGTGGCAGCATTGGCT
ATGG------------------

>Sample2
------------------------------------tgctgtTCCACCAGCTTCTGTGGA
TTTCCCATCTGTTCCACTGGTGG-ACCTGTG-CTCCAGTCCCTGCCAGCCGACCTGCTGC
CAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGC-----------CAAC
CGATCTCCATCCAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGCCAGC
CAACCTGCCTCCAGACCAGTGGCTGTGAGACYGGCTGTGGCATTGGTGGCAGCATTGGCT
ATG-------------------

>Sample4
--CCCTCCTCTCAACCCAACTCCTGACACCATGGCCTGCTGTTCCACCAGCTTCTGTGGA
TTTCCCATCTGTTCCACTGGTGGGACCTGTGGCTCCAGTCCCTGCCAGCCGACCTGCTGC
CAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGC-----------CAAC
CGATCTCCATCCAGACCAGCTGCTGCCAGCCAACCT------------------------
------GCCTCCAGACCAGTGGCTGTGAGACCGGCT------------------------
```

>Sample5
-ACCCTCCTcTCAACCCAACTCCTGACACCATGGCCTGCTGTTCCACCAGCTTCTGTGGA
TTTCCCATCTGTTCCACTGGTGGGACCTGTGGCTCCAGTCCCTGCCAGCCGACCTGCTGC
CAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGC-----------CAAC
CGATCTCCATCCAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGCCAGC
CAACCTGCCTCCAGACCAGTGGCTGTGAGACGGGCTGTGGCATTGGTGGCAGCATTGGCT
ATGGCCAGGTGGGTAgCAGC—

>Sample6
AACCCTCCTCTCAACCCAACTCCTGACACCATGGCCTGCTGTTCCACCAGCTTCTGTGGA
TTTCCCATCTGTTCCACTGGTGGGACCTGTGGCTCCAGTCCCTGCCAGCAGACCTGCTGC
CAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGCACTTCTCGCTCCAGC
CAATCTCCATCCAGACCAGCTGCTGCCAACCAACCTCCATCCAGACCAGCTGCTGCCAGC
CAACCTGCCTCCAGACCAGTGGCTGTCAGCCAGGCTGTGGCATTGGTGGCAGCATTGGCT
ATGGCCAGGTGGGTAGCAGC—

KAP1.3
>Sample1
ACTTcTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGCGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC---------------

>Sample10
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCARCCAACCTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGRTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACAYCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATC----------------

>Sample11
ACTTCTCCAAGCATCCCAGCTYTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCARCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACAYCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC---------------

>Sample12
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGTGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTAcTG----------

>Sample13
ACTTCTCCAAGCATCCCAGCTYTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCARCCAACCTCCATCCAGACCAGCTGCTGCCAG

CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGRTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACAYCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTACTG----------

>Sample14
ACTTCTCCAAGCATCCCAGCTYTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCARCCAACCTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACAYCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC--------------

>Sample15
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCARCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACAYCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC--------------

>Sample16
AcTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGRTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGTGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTAC-----------

>Sample17
ACTTCTCCAAGCATCCCAGCTYTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATC---------------

>Sample18
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATC---------------

>Sample19
ACTTCTCCAAGCATCCCAGCTYTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA

```
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACTTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGTTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGTGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------

>Sample2
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTACTGTGGACA----
------------------------------------------------------------
------------------------------------------------------------
------------------------------

>Sample20
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGRTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGTGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTACTG----------
------------------------------------------------------------
------------------------------------------------------------
------------------------------

>Sample3
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGCGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTACTG----------
------------------------------------------------------------
------------------------------------------------------------
------------------------------

>Sample4
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACTTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGCGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC---------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------

>Sample5
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGCGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAgCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTACTGtGGACA----
------------------------------------------------------------
------------------------------------------------------------
------------------------------

>Sample6
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
```

AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACCTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGRTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC---------------

>Sample7
ACTTCTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCARCCAACCTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGRTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGYGTGGTGAGCTGCACAYCCCCGTCC
TGcTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCC---------------

>Sample8
ACTTcTCCAAGCATCCCAGCTCTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCARCCAACCTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGCTGTGAGACCGGCTGTGGCATTGGTGGCAGCACTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGCGTGGTGAGCTGCACAYCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATCCTAC------------

>Sample9
ACTTCTCCAAGCATCCCAGCTYTCAGCCTAACCCCTGACACCATGGCCTGCTGTTCCACC
AGCTTCTGTGGATTTCCCATCTGTTCCACTGCTGGGACCTGTGGCTCCAGCTGCTGCCGA
TCAACCTGCAGTCAGACCAGCTGCTGCCAGCCAACYTCCATCCAGACCAGCTGCTGCCAG
CCAACCTGCCTCCAGACCAGTGGYTGTGAGACCGGCTGTGGCATTGGTGGCAGCAYTGGC
TATGGCCAGGTGGGTAGCAGCGGAGCTGTGAGCAGCCGCACCAGGTGGTGCCGCCCTGAC
TGCCGCGTGGAGGGCACCAGCCTGCCTCCCTGCTGTGTGGTGAGCTGCACATCCCCGTCC
TGCTGCCAGCTGTACTATGCCCAGGCCTCCTGCTGCCGCCCATC----------------

KAP3.2
>Sample1
--AGACACCCAgACTTcTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample10
CGAGACACCAAGACTTcTcTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCT------

>Sample11
CGAGACACCAAGACTTcTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT

CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample12
--------CAAGACTTcTcTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCaCCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample13
CGAGACACCAAGACTTCTcTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample14
--AGACACCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample15
CGAGACACCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCC-------

>Sample16
CGAGACACCAAGACTTcTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTTCAAC-

>Sample17
CGAGACACCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample18
---GACACCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC

AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCC-------

>Sample19
-------CCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample2
-------CCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample20
CGAGACACCAAGACTTcTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCaCCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTTCAACA
CTC---

>Sample3
CGAGACACCAAGAcTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTTCAACA
CTCA--

>Sample4
CGAGACACCAAGACTTCTcTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample5
CGAGACACCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTTCAA--

>Sample6
CGAGACACCAAGACTTcTcTCaTCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC

GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTTcAaCA
CTC---

>Sample7
--aGACACCAAGACTTcTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample8
-------CCAAGACTTCTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCTT-----

>Sample9
----------AGACTTcTCTCATCAACCCAACAAAACCCAGCTCCTGACACCATGGCTTG
CTGCGCTCCCCGCTGCTGCAGCGTCCGCACTGGTCCTGCCACCACCATCTGCTCCTCTGA
CAAATTCTGTCGGTGTGGAGTCTGCCTGCCCAGCACCTGCCCACACGACATCAGCCTCCT
CCAGCCCCACTTGCTGTGACAACTCCCCCGTGCCCTGCTATGTGCCTGACACCTATGTGCC
AACTTGCTTTCTGCTCAACTCTTCCCACCCCACTCCTGGACTGAGCGGGATCAACCTGAC
GACCTTCATTCAGCCTGGCTGTGAAAATGTCTGCGAGCCCCGCTGTTAAACAGCCACATC
TTTGCACGGGTTCAGTGATGAGCTGCTC--AAGTCTTAATGCGTGATCTGGCCT------

KRT1.2
>Sample1
-------gtggCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAAYGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample10
-------gtgGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCC--------------

>Sample11
------TGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample12
--CaACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAGAT---------

>Sample13
-------gtggcTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC

CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAGATATGGCTAAG

>Sample14
CaCAACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGTGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCCTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample15
---AACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample16
-aCAACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCTTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAGATAWGGCTAAG

>Sample17
-------gtggctTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAAYGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample18
--CAACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCTTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAg-----------

>Sample19
--acCTCGCTCCctcCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCTTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample2
-------gtGgCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCC-------------

>Sample20
CaCAACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCTTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample3
-------gtgGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC

```
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAg-----------

>Sample4
-------gTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGTGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCT
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGTGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCTTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample5
CACaACTGTGGCTTGGTGAACTTGGACTCTGTGTTcagcTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample6
------tgtgGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGCGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAAYGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample7
CaCAACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGYGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCY
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGYGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCYTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample8
CaCAACTGTGGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGYGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCCTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAAYGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGAG-----------

>Sample9
-------gtgGCTTGGTGAACTTGGACTCTGTGTTCAGCTGGGCACCTCGCTCCCTCCAGCACCATGTCTTTCAACTTCT
GCCTGCCCAACCTGAGCTTCCGCTCCAGCTGCTCCTCCAGGCCCTGCGTGCCCTCCAGCTGCTGTGGCACCACCCTGCCC
GGGGCCTGCAACATCCCCGCCAGTGTGGGCAGCTGCAACTGGTTCTGCGAGGGCTCCTTCAACGGCAACGAGAAGGAGAC
CATGCAGTTCCTGAACGACCGGCTGGCCAGCTACCTGGAGAAGGTGCGGCAGCTGGAGCGGGAGAACGCGGAGCTGGAGA
GACGCATCCTGGAGCGCAGCCAGCAGCAGGAGCCCCTCGTGTGCCCCAACTACCAGTCCTACTTCCGGACCATCGAGGAG
CTCCAGCAGAAGGTGAGGGGCTGGGCGCTCCATTGCCTCCAGCAGGAAGTTGTAGGGAGGGAATCCGA------------
```