AMPLIFICATION, SEQUENCING AND BIOINFORMATIC CHARACTERIZATION OF TRANSMEMBRANE PROTEIN 95 GENE (*TMEM95*) IN MURRAH BUFFALO (*Bubalus bubalis*)

BY

SAMPATHIRAO SHIREESHA
B.V.Sc. & A.H.
GVM/14-002

THESIS SUBMITTED TO THE

SRI VENKATESWARA VETERINARY UNIVERSITY

In partial fulfilment of the requirements

for the award of the degree of

MASTER OF VETERINARY SCIENCE

(ANIMAL GENETICS AND BREEDING)

IN THE FACULTY OF VETERINARY SCIENCE

DEPARTMENT OF ANIMAL GENETICS AND BREEDING
NTR COLLEGE OF VETERINARY SCIENCE, GANNAVARAM
SRI VENKATESWARA VETERINARY UNIVERSITY
TIRUPATI – 517 502

JANUARY, 2017

CERTIFICATE

S. SHIREESHA, GVM/14-002 has satisfactorily prosecuted the course of research and that the thesis entitled "**AMPLIFICATION, SEQUENCING AND BIOINFORMATIC CHARACTERIZATION OF TRANSMEMBRANE PROTEIN 95 GENE (*TMEM95*) IN MURRAH BUFFALO (*Bubalus bubalis*)"** submitted is the result of original research work and is of sufficiently high standard to warrant its presentation to the examination. I also certify that the thesis or part thereof has not been previously submitted by him for a degree of any university.

Date :

Place : Gannavaram

(Dr. K. SUDHAKAR)
Major Advisor
Assistant Professor
Dept. of Animal Genetics & Breeding
N. T. R College of Veterinary Science
GANNAVARAM- 521102 (A.P)

CERTIFICATE

This is to certify that the thesis entitled **"AMPLIFICATION, SEQUENCING AND BIOINFORMATIC CHARACTERIZATION OF TRANSMEMBRANE PROTEIN 95 GENE (*TMEM95*) IN MURRAH BUFFALO (*Bubalus bubalis*)"** submitted in partial fulfillment of the requirements for the degree of **Master of Veterinary Science** of **Sri Venkateswara Veterinary University, Tirupati** is a record of the bonafide research work carried out by **S. SHIREESHA, GVM/14-002** under our guidance and supervision. The subject of the thesis has been approved by the Student's Advisory Committee.

No part of the thesis has been submitted by the student for any other degree or diploma. The published part has been fully acknowledged. All the assistance and help received during the course of investigation have been duly acknowledged by the author of the thesis.

(Dr. K. SUDHAKAR)
Chairman of the Advisory Committee

Thesis approved by the Student Advisory Committee

CHAIRMAN: **(Dr. K. SUDHAKAR)** ______________
Assistant Professor
Dept. of Animal Genetics & Breeding
N.T.R College of Veterinary Science
Gannavaram – 521 102 (A.P)

MEMBER : **(Dr. R. VINOO)** ______________
Professor and Head
Dept. of Animal Genetics & Breeding
N.T.R College of Veterinary Science
Gannavaram – 521 102 (A.P)

MEMBER : **(Dr. Ch. VENKATA SESHAIAH)** ______________
Associate Professor and Head
Instructional Livestock Farm Complex,
N.T.R College of Veterinary Science
Gannavaram – 521 102 (A.P)

CONTENTS

Chapter	Title	Page No.
I	**INTRODUCTION**	1
II	**REVIEW OF LITERATURE**	4
	2.1 GENES INVOLVED IN FEMALE INFERTILTIY	4
	2.2 GENES INVOLVED IN MALE INFERTILTIY	6
	2.3 IDIOPATHIC MALE SUBFERTILTIY/INFERTILITY	13
	2.4 BUFFALO GENOME SEQUENCING	13
	2.5 *TMEM95*	15
III	**MATERIALS AND METHODS**	18
	3.1 MATERIALS	18
	3.1.1 Experimental animals	18
	3.1.2 Chemicals, Oligonucleotide Primers	18
	3.1.3 Consumables	18
	3.1.4 Preparation of Solutions / Reagents for DNA Isolation	18
	3.2 METHODS	20
	3.2.1 Blood sample collection, transport and storage	20
	3.2.2 DNA Extraction	20
	3.2.2.1 Isolation of genomic DNA from whole blood	20
	3.2.2.2 Checking the presence and purity of DNA	22
	3.2.3 Amplification of *TMEM95* Gene by Polymerase chain reaction	22
	3.2.3.1 Oligonucleotide Primer designing	22
	3.2.3.2 Preparation of Primer working solution (10 pmol/µl)	25
	3.2.3.3 Optimisation of PCR conditions	25
	3.2.4 Detection of PCR products	28
	3.2.4.1 Preparation of Reagents for Agarose Gel Electrophoresis	28
	3.2.5 Sequence analysis	30

Chapter	Title	Page No.
	3.2.6 Bioinformatic characterization of *TMEM95*	30
	3.2.7 Estimation of diversity with other species	30
IV	**RESULTS**	31
	4.1 YIELD AND QUALITY OF DNA	31
	4.2 CHARACTERIZATION OF *TMEM95* GENOMIC REGION	31
	4.2.1 Primer designing	33
	4.2.2 PCR amplification of *TMEM95* gene in Murrah buffalo and Gir cattle	33
	4.2.3 DNA sequencing of *TMEM95* gene	33
	4.2.4 Annotation of *TMEM95* gene in Murrah	37
	4.2.5 Functional prediction of *TMEM95* gene in cattle and buffalo	37
	4.2.5.1 Motif prediction	37
	4.2.5.2 Transmembrane topology and signal peptide prediction	46
	4.3 PHYLOGENETIC ANALYSIS	46
V	**DISCUSSION**	51
	5.1 YIELD AND QUALITY CHECK OF DNA	51
	5.2 CHARACTERIZATION OF *TMEM95* GENOMIC REGION	52
	5.2.1 Cross species amplification and sequencing of *TMEM95* region	52
	5.2.2 Annotation of *TMEM95* gene in Murrah and Gir	53
	5.2.3 Functional prediction of *TMEM95* gene in buffalo and cattle	54
	5.3 PHYLOGENETIC ANALYSIS	55
VI	**SUMMARY**	57
	LITERATURE CITED	60
	Appendices	70

LIST OF ILLUSTRATIONS

Fig No.	Title	Page No.
1	Schematic diagram depicting the arrangement of *TMEM95* on chromosome 19 of *Bos taurus*	17
2	Schematic diagram of *TMEM95* organization in *Bos taurus*	17
3	Schematic diagram depicting the arrangement of primers and overlapping regions on *TMEM95*	24
4	Quality and purity check of genomic DNA on 1% Agarose gel	32
5	Photograph of PCR amplification products of five fragments of *TMEM95* gene in Murrah buffalo on agarose gel	35
6	Photograph of PCR amplification products of five fragments of *TMEM95* gene in Gir cattle on agarose gel	35
7a	Quality scores of Murrah buffalo *TMEM95* gene contig	36
7b	Quality scores of Gir cattle *TMEM95* gene contig	36
8a	Chromatogram of *TMEM95* showing heterozygous position at 1284 bp (T and G) in Murrah buffalo	44
8b	Chromatogram of *TMEM95* showing heterozygous position at 1460 bp (C and A) in Murrah buffalo	44
8c	Chromatogram of *TMEM95* showing heterozygous position at 1897 bp (G and A) in Murrah buffalo	44
9	*TMEM95* gene annotation in cattle	45
10	Chromatogram showing two base pair deletion in Murrah *TMEM95* causing a frame shift mutation	45
11	Phylogenetic tree of *TMEM95* genomic region in different species	50

LIST OF TABLES

Table No.	Title	Page No.
1	The effect of various genes on fertility traits reported in different farm animals	10
2	Details of the Primers designed, against different regions on *TMEM95*	23
3	Standardized PCR protocol for *TMEM95* gene in Murrah	29
4	Standardized PCR protocol for *TMEM95* gene in Gir	29
5	Quality and quantity of genomic DNA isolated from blood of Murrah buffalo and Gir cattle	34
6	Phred quality scores	34
7	Predicted protein and coding DNA sequences of *TMEM95* gene in Murrah buffalo and Gir cattle	38
8	Predicted isoform profile of the *TMEM95* gene in different species using GeneWise	42
9	Motif prediction in different isoforms of *TMEM95* gene in cattle, buffalo and yak	43
10	Transmembrane topology and signal peptide predictions in different isoforms of *TMEM95* gene in cattle, buffalo and yak	47
11	Estimates of pair wise evolutionary divergence between *TMEM95* sequences of different species	49

LIST OF SYMBOLS AND ABBREVIATIONS

Base pair	-	bp
Degrees centigrade	-	^{0}C
Deoxyribo Nucleic Acid	-	DNA
Distilled Water	-	dH_20
Ethylene Diamine Tetra Acetic acid	-	EDTA
Gram	-	g
Magnesium Chloride	-	$MgCl_2$
Microgram	-	µg
Microlitre	-	µl
Molarity	-	M
Millimoles	-	mM
Micromoles	-	µM
Milligram	-	mg
Millilitre	-	ml
Million years ago	-	Mya
Minute	-	min
Namely	-	*viz.*
Nanogram	-	ng
Nanometre	-	nm
National Center for Biotechnology Information	-	NCBI
Negative logarithm of hydrogen ion concentration	-	pH
Normality	-	N
Optical density	-	OD

Contd..

Percentage	-	**%**
Picomoles	-	pmol
Polymerase Chain Reaction	-	PCR
Potassium Chloride	-	KCl
Revolutions per minute	-	rpm
Sodium Chloride	-	NaCl
Sodium Dodecyl Sulphate	-	SDS
That is	-	*i.e*
Tris Borate EDTA	-	TBE
Tris EDTA	-	TE
Ultra violet	-	UV
Volts	-	V
Transmembrane protein 95 gene	-	TMEM95

Acknowledgements

I was blessed and deem it my honor to have been guided by ***Dr. K. Sudhakar,*** *Assistant Professor, Department of Animal Genetics and Breeding, N. T. R College of Veterinary Science, Gannavaram. His valuable advice and support, meticulous guidance and patience, endless encouragement and caring, critically going through the manuscript and also spending the precious time in correction, are invaluable because of which I have been able to successfully complete the work assigned to me. I express my wholehearted indebtedness for his support and efforts to embellish this research.*

I would like to place on record my heartfelt thanks to ***Dr. R. Vinoo,*** *Professor and Head, Department of Animal Gentics and Breeding and member of advisory committee for his excellence in pursuing, personal attention and impeccable advices. I am thankful to him for his guidance and help towards completion of this research work.*

I heartily express my sincere thanks to ***Dr. Ch. Venkata Seshaiah,*** *Associate Professor and Head, ILFC, N. T. R College of Veterinary Science, Gannavaram for his generous help and co-operation during the course of study. It gives me immense pleasure and privilege to be associated with him during this course of study.*

I reserve a special bouquet of thanks for ***Dr. M. Muralidhar,*** *Assistant professor, Department of Animal Genetics and Breeding, N. T. R College of Veterinary Science, Gannavaram for his extreme care, personal attention, timely and wise counsel, critical evaluation, wholehearted help and constant encouragement throughout the duration of my course.*

*I express my heartful thanks to **Dr G. Sushma**, Assistant Professor, ILFC, N.T.R College of Veterinary Science, Gannavaram for her friendly treat and valuable suggestions during my study period.*

*I am very much thankful to **Dr. T. SrinivasaRao,** Assistant professor and Head, Department of Veterinary Public health and Epidemiology, **Dr. K. V. Subramanyam,** Associate Professor, Department of Veterinary Microbiology, **Dr. C. Sreedevi,** Associate Professor, Department of Veterinary Parasitology and **Dr. M. Naveen Swaroop,** Assistant Professor, Department of Veterinary Biochemistry, NTR College of Veterinary Science, Gannavaram for their whole hearted help and kind cooperation during my research work.*

*I am also thankful to the PhD scholars, **Dr. H. Siddha linga swamy** and **Dr. Bhaktavatsalam**, my colleagues, **Dr. K. N. S. Kumar** and **Dr. D. V. Praneeth** and my juniors, **Dr. V. Tejaswini, Dr. P. Sushma Devi** and **Dr. A. Bhavya Shanthi** Department of Animal Genetics and Breeding, NTR College of Veterinary Science, Gannavaram for their timely help and making the lab environment memorable, support and wholehearted co-operation during the entire course of study.*

*On my personal note, it gives me immense pleasure to express my sincere gratitude and heartfelt respect to the blessings of my parents **Shri. Viswanadham** and **Smt. Kusuma** and to the love and support of my brother **Seetha Rama Raju**. I reserve special thanks to my aunt **Smt Uma Namballa** and cousins **RaviTeja, Sai Bhargav, Tarun, Bharat, Bhargavi, Sravanthi, Sandhya, Anusha, Supraja, Surendra, Raju, Midhilesh, Madhushree, Lohith** and **Gayatri** for their wholehearted love, affection and encouragement in life.*

*I always fall short of words and felt impossible to describe the support of my best friends **Dr. Amulya Chengalasetty** and **Dr. Vasundhara Priyanka Mudragada**,*

who have been like my soulmates since 7 years and always huge supporters of anything I attempt. I am always thankful for their timely support and help.

I express my profuse thanks to my friends, ***Bhavya Priyanka, Hema, Ramakrishna, Pratapa Reddy, Somasekhar, Sujatha, Satish, Anusha, Satish Padmini*** *and* ***Aruna*** *for their friendship and help during my research work.*

I apologize to all wonderful people whom I have missed and ofcourse I am indebted to all of them who did their best to improve upon my best.

I am extremely thankful to ***Sri Venkateswara Veterinary University*** *for providing the financial assistance in the form of stipend.*

Date: **(S. SHIREESHA)**

DECLARATION

I, **SAMPATHIRAO SHIREESHA** hereby declare that the thesis entitled **"AMPLIFICATION, SEQUENCING AND BIOINFORMATIC CHARACTERIZATION OF TRANSMEMBRANE PROTEIN 95 GENE (*TMEM95*) IN MURRAH BUFFALO (*Bubalus bubalis*)"** submitted to **Sri Venkateswara Veterinary University**, Tirupati for the degree of **MASTER OF VETERINARY SCIENCE** is the result of original research work done by me. I also declare that the materials contained in this thesis have not been published earlier.

Date :

Place : Gannavaram

(S. SHIREESHA)

Name of the Author	**:**	**SAMPATHIRAO SHIREESHA**
Title of the thesis	**:**	**AMPLIFICATION, SEQUENCING AND BIOINFORMATIC CHARACTERIZATION OF TRANSMEMBRANE PROTEIN 95 GENE (*TMEM95*) IN MURRAH BUFFALO (*Bubalus bubalis*)**
Degree to which it is Submitted	**:**	Master of Veterinary Science
Faculty	**:**	Faculty of Veterinary Science
Department	**:**	Department of Aniaml Genetics and Breeding, NTR College of Veterinary Science, Gannavaram - 521102.
Major Advisor	**:**	**Dr. K. SUDHAKAR** Assistant Professor
University	**:**	Sri Venkateswara Veterinary University, Tirupati.
Year	**:**	January, 2017

ABSTRACT

TMEM95 gene is known to be involved in idiopathic male subfertility in cattle. It is required for the integrity of sperm plasma membrane and thus has a role in fertilization. The present study was aimed at bioinformatic characterization of the Transmembrane protein 95 (*TMEM95*) gene in Murrah buffalo.

The *TMEM95* gene was amplified and sequenced in Murrah. Analysis with CodonCode Aligner revealed three heterozygous positions in Murrah sequence at positions 1284 (T and G), 1460 (C and A) and 1897 (G and A) with respect to the 2631 bp reference fragment from the Hereford cattle genome assembly. A 2 bp deletion (at 937 bp) was observed in Murrah buffalo which is causing frame shift mutation. The isoforms were predicted using GeneWise, a comparative gene annotation tool and found that the 2bp deletion resulted in the truncation of isoforms 1, 3 and 4 and were

unlikely to form. The transmembrane topology and signal peptide were predicted using Phobius program, and it was observed that the isoforms 1, 2 and 3 doesn't have any transmembrane domains but the isoform 5 has two transmembrane domains. Analysis of isoforms with MotifScan predicted the presence of Casein kinase II phosphorylation site, which has an important role in sperm morphology. Leucine zipper pattern, N-myristoylation site, protein kinase C phosphorylation site, CHRD domain profile, N-glycosylation site and HIT zinc finger motifs were also predicted. Divergence analysis was carried out for the *TMEM95* region across different mammalian species and less divergence was observed among cattle, bison, yak and mithun. Buffalo and sheep are moderately divergent from cattle. Based on the sequence analysis and functional prediction, it was concluded that the *TMEM95* gene in Murrah buffalo is likely to have function in male fertility. However there may be some species-specific differences with respect to their function between the two species.

CHAPTER I

INTRODUCTION

India is the largest producer of milk in the World and buffaloes contribute the major (56.64%) share to milk production (Dash *et al.,* 2015). Among the Indian breeds of buffalo, Murrah is the most important and efficient producer of milk (Khan *et al.,* 2011).

The intervention of man has altered the reproductive cycle of the farm animals especially that of buffaloes from a free grazing, seasonal mating system to a year round mating system and intensive production. Environmental factors and the demand for economic production may result in partial or complete reproductive failure in terms of fertility.

Fertility is an important multifactorial trait, difficult to study as it affected by several component traits, which in turn are greatly affected by differences in health, nutritional status and environment. The ultimate manifestation of infertility is failure to produce offspring. The causes of infertility may be acquired/environment or congenital/hereditary or may be physiological, anatomical or infectious. Bull fertility is an important trait which has been given less consideration by focusing only on cow fertility. When the other aspects like management of cow, nutrition and environment are good, the problem could be due to male factors rather than females. Bull is regarded as half of the herd hence the fertility of a bull is a crucial factor in determining the reproductive performance of herd.

Male infertility is defined as the inability to cause pregnancy in a fertile female due to inability of the spermatozoa to fertilize and activate the ovum and then support embryonic development. Infertility or subfertility problems are of major concern in buffaloes leading to bull disposal under farm or field conditions (Khatun *et al.*, 2013).The causes may be disturbances of the production, transport or storage of spermatozoa, aberration of libido and partial or complete inability to mate.

Sustainability of dairying can be assured with increased fertility in dairy cattle. Various diagnostic methods have been developed to test the infertility of bulls, which include the signalment (a complete history), clinical examination, breeding soundness evaluation, supported by laboratory tests like hormone assays, microbiology, cytology, serology, cytogenetic examination, genetic testing and semen evaluation. Genomic selection has been successfully used by AI studs to screen potential sires and significantly decrease the generation interval of sires (Sattler, 2013; Schefers and Weigel, 2012).

In general, the genetic factors often overlooked as the most instant factor captures the picture. Thus studying about the genes involved, is important to understand the infertility when the routine approach fails.

Perusal of the literature revealed that genes *viz. STAT1* and *STAT3* (Khatib *et al.*, 2009), *CYP19* (Kumar *et al.*, 2009), *AGRN*, *CD81*, *LGALS3BP*, *LGALS9*, *MEP1B*, *LGMN*, *MMP19*, *TIMP2*, *TGM2*, *MET* and *EPSTI1* (Wolf and Bauersachs 2010), *OPN* and *STAT5A* (Laporta *et al.,* 2011), *TP53* (Yuzawa *et al.*, 2011), *MYL6, NOP10, RNF187, RPS24* and *RPS28* (Zhang *et al.*, 2012), *DGAT1*, *SCD1*, *DECR1*, *CRH*, *CBFA2T1*, *GH*, *LEP, NPY* and *TFAM3* (Wathes *et al.*, 2013) and *NRF1, POLG, POLG2, PPARGC1A, TFAM, BAX*, *ITM2B* and *GDF9* (Ferreira *et al.*, 2016) were associated with female fertility traits.

Apart from these the genes viz. *PRND* (Rondena *et al.*, 2005), *DAZL* (Liu *et al.*, 2007), *P450AROM* (Tiwari *et al.*, 2008), *ITGB5* (Feugang *et al.*, 2009), *DDX3* (Liu *et al.*, 2009), *CDC2* and *CDC25A* (Dong *et al.*, 2009), *BOULE (Zhang et al., 2009)*, *OB* (Abavisani *et al.*, 2011), *H19* (Gui *et al.*, 2012), *PIWIL1* (Yao *et al.*, 2013), *PRDM9* (Lou *et al.*, 2014), *HSFY* and *ZNF280BY* (Yue *et al.*, 2014), *USP9Y*, *BPY2, RBMY, KDM5D, RPS4Y2*, *PRKY, VCY1* and *VCY2 (*Dhanoa *et al.*, 2016) *TSPY*, *HSFY*,

ZNF280BY and *PRAMEY* (Zhang *et al.*, 2016) are few of the genes reported to affect male fertility traits.

In certain cases the male fertility is severely compromised although semen quality is free from any apparent pathological findings and the etiology remains unexplained, such a condition is referred as idiopathic male subfertility. It is often associated with genetic and epigenetic abnormalities, which include chromosome translocations and aneuploidies, Y- chromosome microdeletions and mutations of different genes in human (Carrell *et al*., 2006). In post genomic era, it has been observed that a significant proportion of idiopathic male infertility cases are of genetic etiology (Roy *et al.,* 2007). Pausch *et al.* (2014) observed that in Flekvieh cattle a mutation in transmembrane protein 95 gene (*TMEM95*) made the bulls effectively infertile, with a success rate for insemination less than 2%. It was reported that the integrity of *TMEM95* is required for an undisturbed fertilization. This gene is highly conserved across different mammals including human and hence probably playing similar role in all the species. The availability of sequence from a wide range of species is very important as, it is useful for comparing and contrasting similarities and differences between species. This also makes possible to identify the genes coding for enzymes and the sequence features that regulate their level of expression. Sequencing and characterization of *TMEM95* gene in Murrah buffaloes would help in better understanding of its features, thus providing some insights into infertility or subfertility in bulls. Hence, the present study was undertaken with the following objectives:

1. To amplify the complete *TMEM95* gene region in Murrah buffalo using primers designed using the available cattle genome.
2. To sequence the complete *TMEM95* gene region in Murrah buffalo.
3. To characterize the *TMEM95* gene in Murrah buffalo using bioinformatic tools.
4. To estimate divergence of buffalo *TMEM95* gene from other mammalian species.

CHAPTER II

REVIEW OF LITERATURE

Many economic traits of livestock including fertility are multifactorial. Constant improvements for these quantitative traits in livestock are being carried out based on phenotypic observations on them, evaluation of the progeny and sibs, and by applying the statistical tools. Molecular genetics can overcome some of the limitations of conventional methods.

Widespread use of artificial insemination has revealed a great variation among sires (Killan, 1999). Many candidate genes with different functions in metabolism have been identified affecting fertility in both females and males. Infertility may result from molecular defects caused by abnormalities in spermatazoal DNA, RNA or proteins, which impair the ability of spermatozoa to fertilize the ovum (Aitken, 2006). Current advances in animal genome sequencing and associated technologies are providing new insights into genomic study of gametes (Feugang *et al*., 2007).

The literature published on few of the genes involved in female and male infertility is briefly reviewed here.

2.1 Genes involved in female infertility in cattle and buffaloes

Khatib *et al.* (2009) investigated the effects of the interactions between polymorphisms of the *STAT1* and *STAT3* genes on fertilization and early embryonic survival rates using an *in vitro* fertilization system. Statistically significant association was observed between SNP25402 in *STAT3* and fertilization rate. It was concluded that the interactions between SNP in the *STAT3* gene and interactions between the *STAT1* and *STAT3* genes contributed significantly to the phenotypic variation in embryonic survival in cattle.

Kumar *et al*. (2009) analyzed the *CYP19* gene polymorphism by Single Strand Conformational Polymorphism (SSCP) in Murrah heifers of different fertility

performance and observed a significant polymorphism (T/C heterozygote) at position 23 of *CYP19* exon 2 in all late matured and 50% of late maturing animals. Based on these observations they concluded by proposing a hypothesis that the late maturity could be due to methylation of C in the T/C heterozygote condition in regulatory region near to TATA binding protein (TBP) site of *CYP19* gene present on one chromosome.

Yuzawa *et al*. (2011) studied the role of tumor protein p53 gene (*TP53*) polymorphism as a factor in uterine implantation and establishment of pregnancy in four old-aged Japanese Black cows with long-term infertility. It was found that some of the oocytes from individuals having G allele (in exon 6) had developmental competence to the blastocyst stage, but not those having the A allele (in exon 6) had the competence to develop to the blastocyst stage. A polymorphism of *TP53* gene affecting the competence of embryonic development may be the cause of long-term infertility in old-aged Japanese Black cows.

Zhang *et al*. (2012) characterized the novel transcripts and alternatively spliced genes that were associated with embryo quality (loss of which is one of the main factors of infertility) and validated their expression in new biological replications of embryos using quantitative real-time PCR. Quantitative expression of eight novel transcripts differed 2.5- to 90-fold in expression between degenerate embryos and blastocysts. The expression of alternative splicing isoforms of five genes (*MYL6, NOP10, RNF187, RPS24 and RPS28*) differed significantly in the different embryo types. Based on these observations it was concluded that the novel transcripts and alternatively spliced genes, found to be differentially expressed between blastocysts and degenerate embryos, can be used as markers for blastocyst formation and development.

Ferreira *et al*. (2016) conducted a study to know whether the decreased oocyte competence of Repeat Breeding cows (RBs) during summer was associated with an altered gene expression profile and a decrease in mitochondrial DNA (mtDNA) copy number. The mtDNA amounts as well as the expression levels of genes associated with the mitochondria (*MT-CO1, NRF1, POLG, POLG2, PPARGC1A, and TFAM*), apoptosis (*BAX, BCL2, and ITM2B*) and oocyte maturation *(BMP15, FGF8, FGF10, FGF16*, *FGF17* and *GDF9*) from the oocytes collected from heifers, non-RBs in peak lactation (PLs) and RBs were evaluated and was observed that the oocytes retrieved from RBs during winter contained eight times more mtDNA and significantly less mtDNA than those retrieved from RBs during summer. Based on these observations it was concluded that a loss of fertility in RBs during summer might be caused by a possible mitochondrial dysfunction associated with a greater chance of oocytes to undergo apoptosis.

2.2 Genes involved in male infertility in cattle and buffaloes

Tiwari *et al*. (2008) observed a higher ($P < 0.01$) expression of arom*P450* transcript in spermatozoa obtained from the good quality semen (higher mass motility) than that in spermatozoon of poor quality semen (low mass motility). Similarly, higher ($P < 0.01$) expression of arom*P450* mRNA was observed in the motile spermatozoa as compared to non-motile spermatozoa, separated from good quality semen by swim-up.

Dong *et al*. (2009) studied the relationship between the infertility of cattle-yak crosses and the expression levels of *CDC2* and *CDC25A* genes which are the key genes of meiosis. A Real-time quantitative PCR was conducted and it was observed that the expression levels of *CDC2* and *CDC25A* in cattle and yak testes were higher than those in cattle-yak ($P<0.05$) cross. Therefore, it was concluded that the low expression levels of *CDC2* and *CDC25A* genes may have a relationship with the infertility of cattle-yak cross.

Liu *et al*. (2009) characterized *DDX3Y* gene and its homologs *DDX3X* and *PL10* in bovines. The study revealed that the *bDDX3Y-L*, -S, *bDDX3X* and *bPL10* were all widely expressed with predominant expression in testis and brain and revealed that sense and anti-sense RNAs of *bDDX3Y-L*, -S and *bDDX3X* were expressed in interstitial cells. The study was a basis for further investigating the *DDX3* gene function in spermatogenesis, male fertility and gene evolution in mammals.

Zhang *et al*. (2009) studied the association between expression of *bBOULE,* a bovine *DAZ* family gene and male cattle-yak infertility. For this the gene was isolated, and characterized the gene and predicted that it encodes a polypeptide of 295 amino acids with an RNP-type RNA recognition domain. It was observed that bBOULE binds specifically to polypyrimidine RNAs and might act as a nuclear ribonucleoprotein particle auxiliary factor during germ cell formation and morphological changes of germ cells by using tertiary structure analysis. It was also found that the gene expressed in an extremely low level in the testis of sterile male cattle-yaks. Therefore, it was concluded that the *bBOULE* may function in bovine spermatogenesis and the low levels of *bBOULE* expression might lead to male sterility in cattle-yaks.

Abavisani *et al*. (2011) studied the association of leptin which is a hormonal product of the *OB* gene with sperm physiology. The presence of the leptin mRNA transcript in Holstein cattle spermatozoa by reverse transcriptase-polymerase chain reaction (RT-PCR) analysis was investigated and it was concluded that the leptin might be involved in the physiological processes of bovine spermatozoa.

Lin *et al. (*2012) studied quantitation of alternative splicing variants of lactate dehydrogenase C gene (*LDHC*) in testes of adult yak, sexually immature yak calf and sterile male hybrid of yak. Study revealed that the proportions of the *LDHC* variants assayed differed significantly among adult yak, yak calf and cattle-yak and more

LDHC transcripts were spliced in immature or sterile testes. It was opined that the alternative splicing could play a role in the regulation of *LDHC* expression in testes, and could be a factor that plays a role in infertility of yak hybrids.

Gui *et al*. (2012) studied the association between imprinting defects of *H19* and the hybrid sterility of male cattle-yak by examining the methylation patterns of the *H19* imprinting control region (ICR) and *H19* mRNA expression in the testes of cattle-yak, yak and cattle. The results showed that the 3rd CCCTC-binding factor (CTCF) site of the *H19* ICR was significantly hypomethylated in the testes of cattle-yak compared with yak or cattle and the *H19* was expressed at a significantly higher level in cattle-yak than in yak or cattle. Therefore, it was concluded that the disorders in *H19* imprinting might contribute to the sterility of F_1 male hybrids between cattle and yak.

Khatun *et al*. (2013) studied the incidence of the male reproductive anomalies leading to disposal of bovine bulls at GADVASU dairy farm, Ludhiana, Punjab (India). The analysis of data on frequency of various subfertility and disposal pattern of bulls revealed that subfertility traits like poor libido and unacceptable seminal profile were the significant ($p<0.01$) reasons for culling of the breeding bulls. Inadequate sex drive and poor semen quality were the main contributing factors for bull disposal in cattle, whereas poor semen freezability was most frequently observed in buffalo bulls. It was also observed that the occurrence of subfertility problem is at par in crossbred cattle and buffalo bulls and demonstrated the variation in reproductive parameters in the bovine bulls, which could be studied at the molecular level to unveil any genomic markers associated with subfertility and infertility.

Yan *et al*. (2014) conducted a study to know the relationship between the *Dmrt7* deficiency (a member of the DM domain family of genes and essential for male spermatogenesis, between the pachynema and diplonema stages) and male cattle-yak

infertility. *Dmrt7* was cloned by molecular cloning techniques, and the sequence, conserved domains, functional sites and secondary and tertiary structures of the *Dmrt7* encoded protein were predicted and analyzed using bioinformatics methods. It was observed that the *Dmrt7* mRNA and protein expression were significantly higher in testis of cattle and yak than that in cattle-yak hybrid ($p < 0.01$). Results of the study revealed that the low levels of *Dmrt7* expression lead to male sterility in cattle-yak.

Yao *et al*. (2013) conducted a study on association of *PIWIL1* gene with the male sterility of cattle-yaks. The *PIWIL1* and *LINE-1* (long interspersed element-1) retrotransposon mRNA expression and DNA methylation patterns in the testes of cattle, yaks and cattle-yaks were examined and concluded that the higher methylation levels of the *PIWIL1* gene in cattle-yaks coincided with decreased *PIWIL1* mRNA expression, thereby affecting the transposon silencing mechanism and possibly contributing to cattle-yak male sterility.

Yue *et al*. (2014) studied the association of copy number variation of *HSFY* and *ZNF280BY* gene families with testis size and bull fertility in a total of 460 bulls from 15 breeds using a quantitative PCR approach. Bulls in the *Bos taurus* (BTA) lineage had a significantly higher MCN (median copy number) of *HSFY* than bulls in the *Bos indicus* (BIN) lineage, while taurine bulls had a significantly lower MCN of ZNF280BY than indicine bulls. The CNVs of both *HSFY* and *ZNF280BY* were correlated negatively with testis size, while positively with sire conception rate.

The observations of different authors on the effect of various genes on fertility traits in livestock are summarized in Table 1.

Table 1. The Effect of various genes on fertility traits as reported in literature in different farm animals

S.No	Species	Gene (s)	Function (s)	Reference (s)
1.	Bovines	*FSHR*	Spermatogenesis during puberty	Nilsson *et al*. (1986)
		LHCGR	Spermatogenesis in adults	
2.	Bovines	*FUCA1*	Zona pellucid binding/ penetration, sperm egg membrane fusion and post fusion events	Srivastava *et al*. (1986)
3.	Bovines	*TSPY*	Spermatocyte proliferation and differentiation	Vogel *et al*. (1997)
4.	Bovines	*ESR1, ESR2*	Maturation of epididymis, development and function of testes and prostate	Ebling *et al*. (1997), Saunders *et al*. (2001), Stormshak and Bishop (2008)
5.	Horse	*CRISP3*	Equine reproduction	Schambony *et al*. (1998)
		CRISP2,	Interactions between seritoli cells and nascent spermatocytes	Maeda *et al*. (1999)
		CRISP1,	Sperm oocyte fusion	Cohen *et al*. (2000)
6.	Bovines	*PRL*	Inhibit long PRPL activation of JAK2 and transcription via formation of heterodimers.	Bole-Feysot *et al*. (1998)
7.	Pigs	*CD9*	Sperm oocyte fusion	Le Naour *et al*. (2000)
8.	Bovines	*CLGN*	Migration into oviduct, binding to zona pellucida	Ikawa *et al*. (2001)
9.	Humans	*INSL3*	Unilateral cryptorchidism	Klonisch *et al*. (2003)
10.	Bovines	*AR*	Testosterone signaling	Wang *et al*. (2003)
11.	Pigs	*ACTN1* and *ACTG2*	Acrosome reaction	Wimmers *et al*. (2005)
12.	Bovines	*ADAM1* and *ADAM2*	Binding & fusion of sperm-oocyte membrane	Arcelay *et al*. (2008)
13.	Buffalo	*AROMP450*	Sperm motility	Tiwari *et al*. (2008)
14.	Cattle	*STAT1* and *STAT3*	Fertilization and early embryonic pathways	Khatib *et al*. (2009)

Table 1 (contd..)

S.No	Species	Gene (s)	Function (s)	Reference (s)
15.	Buffalo	*CYP19*	Physiology of late maturity	Kumar *et al*. (2009)
16.	Cattle-yak	*CDC25A* and *CDC2*	Meiosis	Dong *et al*. (2009)
17.	Bovines	*DDX3Y*	Spermatogenesis	Wang *et al*. (2009)
18.	Bovines	*BOULE*	Spermatogenesis	Zhang *et al*. (2009)
19.	Bovines	*PLA2*	Sperm capacitation and acrosome reaction	Sato *et al*. (2010)
		CFTR, SHBG, ESR1, ESR2, FSHR, DAZL, MTHFR, INSL3, LGR8 *AR, USP26* and *TAF7L*	Infertility	
20.	Buffalo	*TNP1*	Spermatozoan maturation	Panigrahi and Yadav (2010)
21.	Cattle	*OB*	Physiological process of spermatozoa	Abavisani *et al*. (2011)
22.	Bovines	*RLN1*	Development of male reproductive tract, prostate gland and spermatozoa motility	Miah *et al*. (2011)
23.	Bovines	*GNRH1*	Sperm motility and volume	Sang *et al*. (2011) Yang *et al*. (2011)
24.	Bovines	*INHA*	Acrosome integrity	Sang *et al*. (2011)
		INHBA and *INHBB*	Semen volume per ejaculate	
25.	Cattle	*TP53*	Embryonic development	Yuzawa *et al*. (2011)
26.	Bovines	*SOX9, DMRT1, WNT1, AMH, SF1, DAX1, GATA4, AROMATASE*	Sex determining pathway	Angelopoulou *et al*. (2012)
27.	Bovines	*ACVR2A*	Upregulation of GnRH release from hypothalamus	Mishra *et al.* (2013)
		ACVR2B	Upregulation of GnRH release from hypothalamus	

Table 1 (contd..)

S.No	Species	Gene (s)	Function (s)	Reference (s)
28.	Bovines	*USP9Y*	Spermatocyte development	Bonfiglio *et al*. (2012)
29.	Cattle-yak	*H19*	Sterility	Gui *et al*. (2012)
30.	Yak hybrids	*LDHA*	Sterility	Lin *et al*. (2012)
31.	Bovines	*SRY*	Male sex determination	Mukherjee *et al*. (2013) Winaya *et al*. (2012)
32.	Bovines	*MYL6, NOP10, RNF187, RPS24* and *RPS28*	Embryo quality	Zhang *et al*. (2012)
33.	Bovines	*CLU*	Preventing oxidative damage to sperm, Inhibiting complement induced sperm lysis	Mishra *et al.* (2013)
34.	Bovines	*LDHA, LDHB, LDHC*	Sperm glucose consumption, ATP production and motility	Odet *et al*. (2013)
35.	Cattle-yak	*PIWIL1*	Sterility	Yao *et al*. (2013)
36.	Cattle	*TMEM95*	Sperm plasma membrane integrity	Pausch *et al*. (2014)
37.	Bovines	*HSFY* and *ZNF280BY*	Testis size	Yue *et al*. (2014)
38.	Cattle	*MT-CO1, NRF1, POLG, POLG2, PPARGC1A,TFAM, BAX, BCL2,* and *ITM2B*	Oocyte apoptosis	Ferreira *et al*. (2016)
		BMP15, FGF8, FGF10, FGF16, FGF17, and *GDF9*	Oocyte maturation	

2.3 Idiopathic male sub fertility / infertility

The situation in which the semen analysis is without any apparent pathological findings but the male fertility is compromised may be termed as idiopathic subfertility (Pausch *et al*., 2014). It is often associated with genetic and epigenetic abnormalities, which include chromosome translocations and aneuploidies, Y-chromosome microdeletions and mutations of different genes in human (Carrell *et al*., 2006). In bovines the idiopathic infertility is caused by environmental or genetic factors (Chenoweth, 2005; Arabi and Mohammadpour, 2006; Givens and Marley, 2008)

Parent *et al*. (1999) studied about the subfertility in bulls and its association with levels of sperm proteins like P21b and P25b (having common antigenicity of P34H, which has an important role in idiopathic infertile men) in bovines. It was observed that all bulls with high Non-Return Rates (fertile bulls) demonstrated high amounts of P25b, whereas its levels were decreased in semen from subfertile bulls. Therefore, it was concluded that the protein P25b was a potential fertility marker in the bulls and may be useful for the evaluation of bull fertility.

Lessard *et al.* (2011) studied the causes of idiopathic infertility in a Red Angus bull and reported that the incidence was caused by a failure to complete the capacitation process.

2.4 Buffalo genome sequencing

The availability of sequence from wide range of species is very important. By comparing and contrasting similarities and differences between species, it becomes possible to identify the genes coding for enzymes and the sequence features that regulate their level of expression. Genome sequences make it possible to estimate the effects of each genomic region on a trait of interest. Limited number of studies have

been conducted globally exploring the genetic diversity at molecular level basis in buffaloes in comparison with other farm animal genetic resources.

Parma *et al.* (2004) reported the whole sequence of the water buffalo (*Bubalus bubalis*) mitochondrial genome. It was observed that the mtDNA is 16.355 bp in length and each molecule encodes 37 genes which are responsible for two rRNA (12S and 16S rRNA), 22 tRNAs and 13 protein-coding genes.

Kumar *et al*. (2007) sequenced the mitochondrial region and cytochrome b gene of river buffalo and compared with the Mediterranean and swamp buffalo sequences.

Meignanalakshmi and Mahalinga Nainar (2009) have sequenced the beta-lactoglobulin gene of Murrah buffaloes in a study to observe the polymorphism at the locus.

Michellizi *et al.* (2010) reviewed the buffalo gene sequences available in the NCBI database and observed that 971 buffalo gene sequences are published in the database, which represented 277 functional genes based on orthologous search. *TMEM95* gene is not included in the list.

Tantia *et al.* (2011) sequenced the whole genome of water buffalo genome using Illumina GAIIx technology with paired end and mate pair short read sequencing using the cattle genome (Btau 4.0 assembly) as a reference. The assembly has 185,150 contigs with the median contig length of 2.3 Kb and the largest contig length of 663 Kb.

Zimin *et al*. (2013) started the sequencing of Mediterraniain water buffalo genome using Illumina and 454. The assembly contained 2.76 GB in scaffolds with N50 contig size of 16kb and N50 scaffold size of 1.4Mb.

According to Moaeen-ud-Din (2014) the number of known gene sequences submitted to the GenBank of river buffalo were 1020, which include, 825 entries on

growth and milk production related genes, 307 immune related and 194 genes related to reproduction traits.

Cyp19 gene in Egyptian buffaloes was sequenced and submitted to the NCBI database by Othaman *et al.* (2014).

Perusal of the literature revealed very few studies on *TMEM95* gene in buffaloes. The sequence analysis not only provides the necessary information for understanding the evolutionary pattern but also useful for further study on gene expression regulation and elucidation of molecular mechanisms of gene mutation effects on reproductive traits.

2.5 *TMEM95*

TMEM95 gene is a protein coding gene located on chromosome 19 and 17 in *Bos taurus* and *Homo sapiens* respectively. The *TMEM95* in *Bos taurus* is characterized by six exons (Figures 1 and 2).

It has 6 isoforms and 6 transcript variants. It produces transmembrane protein 95 which is located on the surface of fertile spermatozoa and helps in maintaining the integrity of spermatozoa which is required for an undisturbed fertilization.

A nonsense mutation in *TMEM95* gene may lead to idiopathic subfertility in male animals (Pausch *et al.*, 2014). It was also demonstrated that the deficiency of TMEM95 severely compromises male reproductive performance in cattle and revealed for the first time a phenotypic effect associated with genomic variation in *TMEM95*.

Ramasamy *et al.* (2014) observed that the promoter region of the *TMEM95* gene is hypermethylated in humans suffering from azoospermia, indicating possible role of the gene in fertility.

Zhang *et al.* (2015) have identified two transcripts of *TMEM95* including *TMEM95-SV1* and *TMEM95-SV2* using RT-PCR in cattle that are expressed only in

testis and brain. Bioinformatic predictions of the isoform *TMEM95-SV1* revealed a leucine-rich repeat C-terminal domain and an IZUMO domain. IZUMO domain was involved in acrosome reaction.

The *TMEM95* in *Bubalus bubalis* is not characterized so far. Under these circumstances much scope is left for the researchers in understanding the role of this gene in reproductive pathways.

Perusal of the literature revealed that to understand the role of specific genes affecting fertility requires knowledge on their sequence and expression. Owing to non availability of literature on the *TMEM95* and its involvement in idiopathic male subfertility, there is a necessity to characterize this gene in buffalo. Hence, the present study was undertaken to generate preliminary data on the gene in Murrah buffalo using comparative genomic approach.

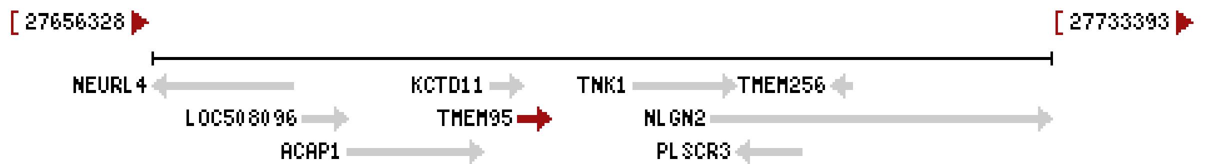

Figure 1. Schematic diagram depicting the arrangement of *TMEM95* on chromosome 19 of *Bos taurus*.

Source: www.ncbi.nlm.nih.gov

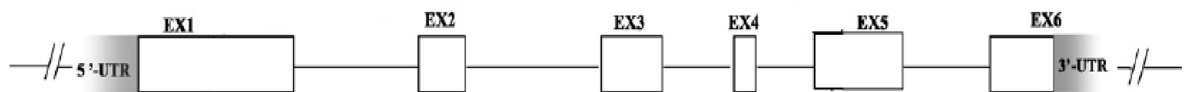

Figure 2. Schematic diagram of *TMEM95* organization in *Bos taurus*.
(Developed by using the TMEM95-AC_000176. Source: Zhang *et al*., 2015)

CHAPTER III

MATERIALS AND METHODS

The present study was carried out on Murrah buffaloes in order to generate the preliminary data on *TMEM95* and on Gir cattle to observe any variation with the published reports of *Bos taurus*.

3.1 MATERIALS

3.1.1 Experimental Animals

The present study does not warrant permission from Institutional Animal Ethics Committee. The blood samples were collected aseptically from the four Murrah buffaloes out of the available pure bred stock at Instructional Livestock Farm Complex, N. T. R College of Veterinary Science, Gannavaram and three pure bred Gir cattle maintained at Veerankilaku dairy farm at random.

3.1.2 Chemicals and Oligonucleotide Primers

All the chemicals used in the present study were procured from Himedia, Mumbai. The oligonucleotide primers were custom synthesized from BioServe Biotechnologies (India) Pvt Ltd, Hyderabad.

3.1.3 Consumables

All the consumables such as glassware and plasticware used for the present study were obtained from Borosil and Tarson respectively.

3.1.4 Preparation of Solutions / Reagents for DNA Isolation

The compositions of different solutions/reagents used in DNA isolation were as follows:

EDTA (10 per cent) solution

Ethylene Diamine Hexa Acetic Acid (EDTA)	:	1 g
Triple glass distilled water to make up to	:	10 ml

RBC Lysis Buffer

Ammonium chloride	:	8.0235 g
Potassium chloride (10 mM)	:	0.7455 g
EDTA	:	0.0372 g

Triple glass distilled water to make up to 1000 ml and autoclaved and stored at 4^0C

Tris Buffer Saline

Potassium chloride	:	0.0373 g
Tris (hydroxymethyl) aminomethane HCl	:	0.0303 g
Magnesium chloride (10 mM)	:	1 ml
Sodium chloride (0.4 M)	:	8.1800 g
Triple glass distilled water to make up	:	1000 ml

pH adjusted to 7.4 and autoclaved and stored at 4^0C

Sodium Dodecyl Sulphate (10 per cent SDS)

Sodium dodecyl sulphate	:	10 g
Triple glass distilled water to make up to	:	100 ml

Proteinase – k

Proteinase – k	:	20 mg
Triple glass distilled water to make up to	:	1 ml

Stored at - 20°C

Saturated Sodium Chloride

Sodium chloride	:	29.22 g
Triple glass distilled water to make up to	:	100 ml

TE buffer (DNA Storage buffer)

Tris (1M, pH 8.0)	:	1 ml
EDTA (0.5M, pH 8.0)	:	200 µl

Triple glass distilled water to make up to 100 ml and autoclaved and stored at room temperature

Tris-EDTA buffer (TE buffer pH 8.0)

Tris (hydroxymethyl) aminomethane HCl (10 mM), pH 7.6	:	1.2114 g
EDTA (0.1 mM)	:	0.3722 g

Triple glass distilled water to make up to 1000 ml, pH adjusted to 8.0
Autoclaved and stored at 4^0C

Ethanol (70 per cent)

Ethanol 99.9 per cent	:	70 ml
Distilled water	:	30 ml

<u>0.5 M EDTA</u>

EDTA	:	18.612 g
Triple glass distilled water to make up to	:	100 ml

<u>Chloroform and Isoamyl Alcohol Solution</u>

Chloroform	:	24 ml
Isoamyl alcohol	:	1 ml

<u>3.0 M Sodium acetate</u>

Sodium acetate	:	12.3045 g
Triple glass distilled water to make up to	:	50 ml

3.2 METHODS

3.2.1 Blood sample collection, transport and storage

Nine ml of whole blood was collected from jugular vein of each animal using 10 ml K2E (EDTA) BD Vacutainer tubes. Immediately after collection, the samples were labeled and transported to the laboratory in an ice packed container and stored at -20 ^{0}C until use.

3.2.2 DNA Extraction

3.2.2.1 Isolation of genomic DNA from whole blood

DNA was isolated from blood samples using a modified high salt method (Miller *et al*., 1988).

The steps followed for isolation of DNA from blood samples are given below.

i. Three ml of blood and nine ml of RBC lysis buffer were taken in a 15 ml centrifuge tube. This tube was inverted several times and incubated on ice by shaking for 10 minutes for complete lyses of red blood cells. Nuclear material was pelleted by centrifugation at 4000 rpm at room temperature for 10 minutes and discarded the supernatant.

ii. Nuclear pellet was washed with 10 ml of RBC lysis buffer and centrifuged once again. This step was repeated for three to four times till the clear nuclear pellet was obtained.

iii. Nuclear pellet was resuspended in 10 ml Tris buffer saline (pH 7.4) and centrifuged at 3000 rpm for 10 minutes. The pellet was collected and the procedure was repeated for 3 times.

iv. 250 µl of TE buffer (pH 8.0) was added to the pellet and vortexing was performed. To this tube 5 µl of proteinase-k and 12 µl of 0.5M EDTA were added and mixed thoroughly. To this solution 500 µl of SDS was added with gentle mixing and incubated overnight at 55 ^{0}C. After incubation 150 µl of saturated NaCl was added and shaken vigorously. To this mixture equal volume of Chloroform: Isoamylalcohol (24:1) was added, mixed and centrifuged at 13,000 rpm for 15 minutes at 24 ^{0}C.

v. Aqueous upper phase was transferred to fresh tube without disturbing the interphase.

vi. Finally, DNA was precipitated by adding double the volume of absolute ethanol and 1/3 rd volume of 3M Sodium acetate to the supernatant collected and inverted several times until the DNA was precipitated and kept at 4 ^{0}C for 2 hours and centrifuged at 13,000 rpm for 30 minutes at 4 ^{0}C.

vii. The supernatant was discarded and 300 µl of 70 per cent ethanol was added and centrifuged at 13,000 rpm for 5 minutes at room temperature.

viii. The supernatant was discarded and the pellet was air dried.

ix. The dried DNA pellet was resuspended in 300 µl of TE buffer (pre warmed) and kept at 37 ^{0}C for 10 minutes for dissolution.

x. The resuspended pellet was stored at -20 ^{0}C until further use.

3.2.2.2 Checking the presence and purity of DNA

On extraction of DNA, concentration was checked by agarose gel electrophoresis. Agarosoe gel of 1 per cent was prepared by boiling agarose in 1X TBE buffer. The prepared gel was poured into an electrophoresis plate and left at room temperature for about 30 minutes for polymerization. One µl of 6x loading buffer containing bromophenol blue dye and six µl of dH_2O were mixed and then loaded into the wells of the gel. The DNAs were run on agarose gel at 90 V for about 40 minutes in 1x TBE buffer, stained in 0.5 µl/ml ethidium bromide solution and was visualized under UV light. DNA was quantified by using NanoDrop™ 2000/2000c (Thermo Fisher Scientific) using the convention that 1 optical absorbance unit at 260 nm equals 40 µg per ml. Purity of DNA was judged on the basis of OD ratio at 260:280 and samples having the acceptable purity *i.e.* 1.7 to 2.0 were used for further analysis. The concentration of the DNA in samples for Murrah and Gir were 109.7 ng/µl and 50.0 ng/µl, respectively.

3.2.3 Amplification of *TMEM95* Gene by Polymerase chain reaction

The basis for PCR comes from natural DNA replication and so requires similar molecular reagents and follows the same basic principles. During thermocycling, DNA is denatured, primers are annealed to the flanking region of the target region and new strands are extended by the heat stable Taq DNA polymerase (isolated from the thermophilic bacterium, *Thermus aquaticus*) (Coen and Scharf, 1991).

The *TMEM95* gene sequence of *B. taurus* (Accession No. AC_000176.1) was taken from the available GenBank database, NCBI and subjected to BLASTN to find the regions of local similarity between the nucleotide sequences.

3.2.3.1 Oligonucleotide Primer designing

In the present study, to design primers for sequencing the *TMEM95* region in Murrah buffalo, comparative genomics approach was used. Since cattle is

evolutionarily a closely related species to buffalo, the genomic region in cattle is primarily used for primer designing. The *TMEM95* gene sequence of *B. taurus* was subjected to BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat) to obtain additional 100 bp on to the 5' and 3' ends of the gene. The resultant sequence was given as input to GenScript, (http://www.genscript.com/cgi-bin/site2/sequencing_primer_design), an online DNA sequencing primer designing tool and primers were designed so as to get overlapping products by keeping the distance between at 600 bp and the melting temperature at 60 °C ± 5 °C. The obtained primers from GenScript were having variants at few positions, which were observed based on multiple sequence alignment of catlle and buffalo. Therefore, the primers were redesigned manually by taking the general principles of primer designing into consideration. The conditions were that primers were a minimum of 20 bp in length, about 50% GC content, ending with either G or C base, both forward and reverse primers have similar Tm, and there should be overlapping of a minimum 100 bp between the adjacent primer pairs. Five different primer pairs were designed (Table 2) that cover entire 2.6 kb region of the *TMEM95* gene and the overlapping of primers was shown in Figure 3.

Table 2. **Details of the Primers designed, against different regions on *TMEM95***

Fragment	**Primer Sequence**	**Tm**	**GC%**	**Length (bp)**	**Product (bp)**	**Start**	**Stop**
1	**F:** CTGTTTTGGTACCAGGTGTG	56.56	50	20	617	38	57
	R: AGGTTCTAGAGCATTCTAGCAC	57.34	45.45	22		654	633
2	**F:** TCCATCTCCCACTAGAAGAATC	56.43	45.45	22	696	547	568
	R: CTTCCAAGCCCCTAGGTTGT	59.3	55	20		1242	1223
3	**F:** TCCAGGTCAGTGAGGGAATC	58.43	55	20	689	1101	1120
	R: CAACAGTACACCTCGAAGCC	58.57	55	20		1789	1770
4	**F:** CAGAGAGGAACGGAGGCTTC	59.83	60	20	687	1679	1698
	R: GACCATCTGACACTGGGACT	58.73	55	20		2365	2346
5	**F:** GCCTAGCCTCCAGTTCTAAG	57.1	55	20	621	2191	2210
	R: GTTGAGAACCACTGTGCTAC	56.39	50	20		2811	2792

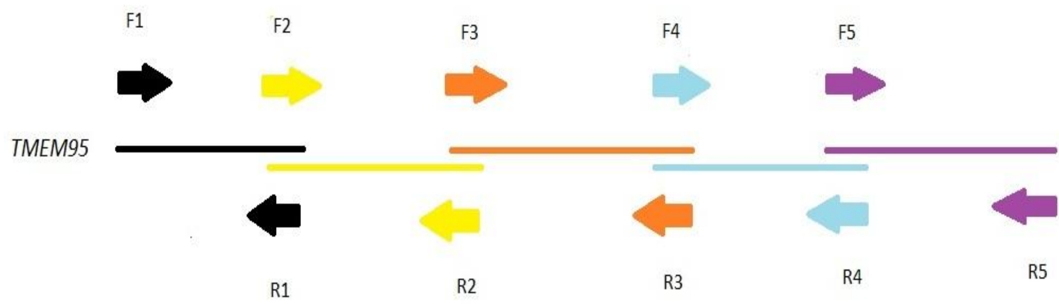

Figure 3. Schematic diagram depicting the arrangement of primers and overlapping regions on *TMEM95*

F1to F5: forward primers; R1 to R5: reverse primers for respective forward primers

3.2.3.2 Preparation of Primer working solution (10 pmol/µl)

The oligonucleotide primers were custom synthesized by BioServe Biotechnologies (India) Pvt Ltd, Hyderabad. Oligos supplied in freeze dried powder form were reconstituted in nuclease and protease free water (Himedia, Mumbai, India) to a volume (in µl) equivalent to the mass (µg) of primer and further diluted to give a final concentration of 10 pmol/µl.

3.2.3.3 Optimisation of PCR conditions

PCR was set up for two reactions, one each for Murrah and Gir and the total volume of each reaction was 15 µl. Fourteen µl of reaction volume comprising 2x master mix, nuclease free water and forward and reverse primers which was aliquoted in each PCR tube and one µl sample Murrah DNA and one µl of Gir DNA were added to each tube to make the final volume. Each reaction volume contained.

PCR Components		Volume
DNA Template	-	1.0 µl
Primer Forward (10 pmol/µl)	-	0.5 µl
Primer Reverse (10 pmol/µl)	-	0.5 µl
2 x PCR Master Mix	-	7.5 µl
Nuclease free water	-	5.5 µl
Total	-	15.0 µl

PCR mix was prepared for one additional sample to cover pipetting error. DNA template from Gir cattle amplified is used as positive control in the present study.

The PCR conditions were optimized by setting different time-temperature combinations for annealing and extension process. The PCR was carried out in total volume of 15µl in a 200 µl PCR tube. The combination that gave best result, in terms of yield and specificity of the product was further used for amplification. The PCR was carried out in a thermal cycler (Eppendorf, Master cycler TM, Germany).

The conditions that employed were initial denaturation at 94 ^{0}C / 2 min, denaturation at 94 ^{0}C / 30 sec, annealing at 55 ^{0}C / 30 sec, extension at 72 ^{0}C / 1 min and final extension at 72 ^{0}C / 10 min. On observation of non specific amplification, the denaturation and extension duration was increased along with an increment of annealing temperature. The primer working solution was also diluted up to 1 pmol/µl and the PCR was repeated with the same conditions and amplicons of specific length (Table 2) were obtained both in Murrah and Gir with very low concentration of primer dimers and the protocol was standardized for the first pair of primers.

PCR was carried out for amplification of the second fragment with second set of primers (1 pmol/µl) by using same protocol. On gel electrophoresis no bands were observed both in Murrah and Gir samples. The annealing temperature was decreased to 50 ^{0}C and primer working solution concentration was increased up to10 pmol/µl and the PCR was repeated. Unambiguious bands of specific length (Table 2) were observed and this protocol was standardized for second pair of primers.

PCR was carried out with third set of primers (1 pmol/µl) and kept at 55 ^{0}C annealing temperature. No bands were obtained both in Murrah and Gir samples. The annealing temperature was decreased to 48 ^{0}C and PCR was performed. The concentration of primer working solution was increased to10 pmol/µl and the PCR was done with the same protocol. Very faint band was observed in Murrah and no band was obtained in Gir. To increase the specificity, 5–10 per cent Glycerol (PCR adjuvant) was added at 2 µl for 15 µl reation and PCR was done. No bands were observed both in Murrah and Gir samples. Subsequently the annealing temperature was set for a gradient of 52 ^{0}C -58 ^{0}C and faint bands were observed at 52.3 ^{0}C and 53 ^{0}C and clear bands of specific length (Table 2) were observed at remaining all temperatures. Hence, PCR was repeated with the annealing temperature 56 ^{0}C and again clear band was observed. This protocol was standardized for third pair of

primers for Murrah. For Gir samples also a gradient was set from 55.8 ^{0}C, 56 ^{0}C, 56.2 ^{0}C. Clear bands were observed both at 55.8 ^{0}C and 56 ^{0}C and this protocol (at 56 ^{0}C) was standardized for third pair of primers for Gir.

PCR was carried out with fourth set of primers (1 pmol/µl) and kept at 55 ^{0}C annealing temperature. On gel electrophoresis no bands were observed. The concentration of primer working solution was increased up to 5 pmol/µl and the PCR were repeated with same cycling conditions. Clear bands of specific length (Table 2) were obtained along with no primer dimers. This protocol was standardized for fourth pair of primers.

PCR was carried out with fifth pair of primers (1 pmol/µl) and kept at 55 ^{0}C annealing temperature. No bands were observed both in Murrah and Gir samples. The concentration of primer working solution was increased up to 5 pmol/µl. The annealing temperature was decreased up to 50 ^{0}C and faint band was observed only in Murrah sample. Then the concentration of primer working solution was increased up to 10 pmol/µl and PCR was done with the same protocol. No bands were observed. The annealing temperature was set for a gradient of 52 ^{0}C -58 ^{0}C. Amplicon of specific length (Table 2) was obtained only at 57.1 ^{0}C. The PCR was repeated with the annealing temperature 58 ^{0}C and again clear band was observed. Therefore, this protocol was standardized for fifth pair of primers for Murrah. As the Murrah sample got amplified at 58 ^{0}C, the gradient for Gir sample was set from 56.4 ^{0}C to 57.8 ^{0}C and reactions were kept at 56.1 ^{0}C, 57.1 ^{0}C, 57.3 ^{0}C, 57.6 ^{0}C along with negative control. Faint bands were obtained at all temperatures. Then the same gradient was repeated and two reactions were kept at 57.3 ^{0}C and 57.8 ^{0}C. In both samples, clear bands at around 600 bp were obtained. The reaction was repeated with 57.3 ^{0}C and clear band was obtained and this protocol was standardized for fifth pair of primers for Gir.

A negative control was run along with the samples at every PCR setup. A negative control is a reagent blank which consists of all reagents used during sample processing but contains no sample / DNA template. The final PCR protocols followed for amplification of *TMEM95* in Murrah and Gir are presented in Tables 3 and 4 respectively.

3.2.4 Detection of PCR products

3.2.4.1 Preparation of Reagents for Agarose Gel Electrophoresis

1.5 % Agarose gel

Agarose	- 0.6 g
1x TBE working solution	- 40 ml

Preparation of 10X TBE buffer (stock) 500 ml

Tris base	- 54 g
EDTA	- 4.65 g
Boric acid	- 27.5 g

Make up to 500 ml with triple distilled water.

Preparation of 1X TBE buffer (working) 500 ml

10 ml of stock was mixed with 90 ml of triple distilled water to prepare 100 ml of 10x TBE working solution.

Five µl of each PCR amplicons were analyzed on 1.5 per cent (w/v) agarose gel containing 0.5 µg/mL ethidium bromide in 1x TBE buffer and visualized on UV transilluminator (Gel Doc ™ XR+, Biorad, U.S.A). The horizontal gel electrophoresis was carried out at 90 volts for 40 minutes. The sizes and quantities of PCR products were verified by comparison with 100 bp DNA ladder (Genei Merck, Banglore, India). The photograph of the gel was obtained by a gel image system.

Table 3. **Standardized PCR protocol for *TMEM95* gene in Murrah**

Step No.	Step	Primer1 (1 pmol/µl)	Primer2 (10 pmol/µl)	Primer3 (10 pmol/µl)	Primer4 (5 pmol/µl)	Primer5 (10 pmol/µl)
1	Initial denaturation	94 ^{0}C / 2 min	94 ^{0}C / 2 min	94 ^{0}C / 2 min	94 ^{0}C / 2 min	94 ^{0}C / 2 min
2	Denaturation	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec
3	Annealing	55 ^{0}C / 30 sec	50 ^{0}C / 30 sec	56 ^{0}C / 45 sec	55 ^{0}C / 30 sec	58 ^{0}C / 45 sec
4	Extension	72 ^{0}C / 2 min	72 ^{0}C / 2 min	72 ^{0}C / 2 min	72 ^{0}C / 2 min	72 ^{0}C / 2 min
5	No. of cycles from step 2 – 4	30	30	30	30	30
6	Final Extension	72 ^{0}C / 10 min	72 ^{0}C / 10 min	72 ^{0}C / 10 min	72 ^{0}C / 10 min	72 ^{0}C / 10 min

Table 4. **Standardized PCR protocol for *TMEM95* gene in Gir**

Step No.	Step	Primer1 (1 pmol/µl)	Primer2 (10 pmol/µl)	Primer3 (10 pmol/µl)	Primer4 (5 pmol/µl)	Primer5 (10 pmol/µl)
1	Initial denaturation	94 ^{0}C / 2 min	94 ^{0}C / 2 min	94 ^{0}C / 2 min	94 ^{0}C / 2 min	94 ^{0}C / 2 min
2	Denaturation	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec	94 ^{0}C / 45 sec
3	Annealing	56.2 ^{0}C / 30 sec	50 ^{0}C / 45 sec	56 ^{0}C / 45 sec	55 ^{0}C / 30 sec	57.3 ^{0}C / 45 sec
4	Extension	72 ^{0}C / 2 min	72 ^{0}C / 2 min	72 ^{0}C / 2 min	72 ^{0}C / 2 min	72 ^{0}C / 2 min
5	No. of cycles from step 2 – 4	30	30	30	30	30
6	Final Extension	72 ^{0}C / 10 min	72 ^{0}C / 10 min	72 ^{0}C / 10 min	72 ^{0}C / 10 min	72 ^{0}C / 10 min

3.2.5 Sequence analysis

The PCR products of Murrah and Gir samples were kept on ice and sent for sequencing along with primers. The sequencing was performed on automated sequencer (ABI prism) using Sanger's dideoxy chain termination method at SciGenome Labs Pvt. Ltd. Cochin.

3.2.6 Bioinformatic characterization of *TMEM95*

The sequences obtained were interpreted and aligned by using CodonCode Aligner 6.0.2 software (LI-COR, Inc., Lincoln, USA) and analysed by the Basic Local Alignment Search Tool (BLAST).

The GeneWise (http://www.ebi.ac.uk/Tools/psa/genewise/) online tool was used to predict the isoforms for the sequenced *TMEM95* gene allowing for introns and frameshifting errors.

The transmembrane topology and signal peptide prediction was performed with Phobius (Kall *et al*., 2004) bioinformatic tool. The predictor is based on a hidden Markov model (HMM) that models the different sequence regions of a signal peptide and the different regions of a transmembrane protein in a series of interconnected states.

Motif Scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan), an online tool was used to identify the motifs present in the protein sequence.

3.2.7 Estimation of diversity with other species

MEGA7 (Molecular Evolutionary Genetic Analysis) (Kumar *et al*., 2016) software was used to estimate the divergence of buffalo *TMEM95* gene from other mammalian species. ModelTest (Posada and Crandall, 1998) is performed to check the best suitable model to explain the sequence divergence. It selects the best-fit nucleotide substitution model for a set of aligned sequences.

CHAPTER IV

RESULTS

A total of four blood samples were collected from Murrah buffaloes and three from Gir cattle for isolation of genomic DNA for the present study.

4.1 Yield and Quality of DNA

Agarose gel electrophoresis was used to check the presence of DNA (Figure 4). The concentration of DNA in the samples from Murrah ranged from 109.7 to 249.7 ng/µl and that of Gir were 50.0 to 128.7 ng/µl. The nanodrop results of the genomic DNA of Murrah buffalo and Gir cattle are presented in Table 5. For most of the samples the ratio of optical densities were more than 1.7 indicating good deproteinisation.

4.2 Characterization of *TMEM95* genomic region

Searching the NCBI gene database indicated that *TMEM95* is located on chromosome 19 in *Bos taurus* (Accession No. AC_000176.1). The region is located between the co-ordinates 27687749-27690379 (including UTR regions) on plus strands flanking *KCTD11* gene and *TNK1* gene respectively on either side (Figure 1). The length of the region is 2631 bp. The available *Bos indicus* (Ongole breed) (Canavez *et al.,* 2012) homologous genomic region is downloaded from NCBI database (Accession No. CM003039.1; Co-ordinates 27432672- 27435302). The obtained sequences are given in Appendix I (A and B). Using the sequence alignment tool BLAST, it was observed that there were no differences in the genomic regions of taurine and indicine cattle with respect to this region.

The orthologous regions of *TMEM95* gene from the genome assemblies of *Bubalus bubalis* (Mediterranian buffalo) and *Bos mutus* (Yak) were obtained from NCBI database using the DNA sequence of cattle as reference. The obtained sequences are given in Appendix I (C and D).

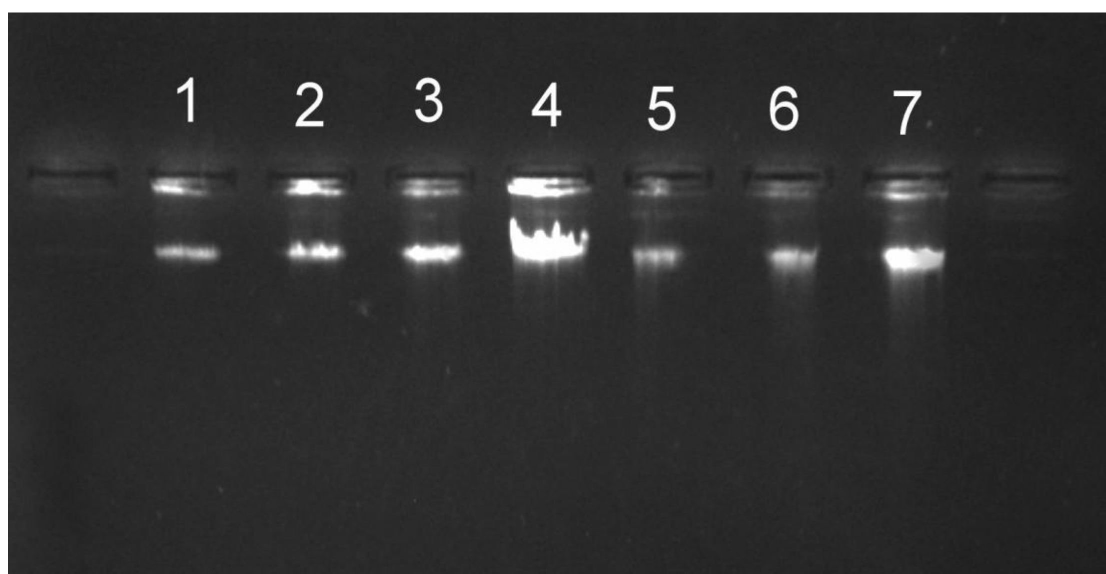

Figure 4. **Quality and purity check of genomic DNA on 1% Agarose gel**
1-Murrah_5M; 2-Murrah_8M; 3-Murrah_10M; 4-Murrah_12M; 5-Gir_1G; 6-Gir_9G; 7-Gir_12G;

4.2.1 Primer designing

To design primers for sequencing the Murrah and Gir *TMEM95*, comparative genomics approach is used. Five different primer pairs were designed that cover the entire 2.6 kb region of the *TMEM95* gene.

4.2.2 PCR amplification of *TMEM95* gene in Murrah buffalo and Gir cattle

The PCR conditions were optimized by setting different time-temperature combinations for annealing and extension processes. The PCR conditions were optimized to get a single desired PCR amplicon. In the present study the DNA five fragments across *TMEM95* from both the genetic groups were amplified with five sets of primers having overlapping regions. The corresponding PCR products were 617, 696, 689, 687 and 621bp in length respectively in both Murrah buffalo and Gir cattle (Figure 5 and 6). None of the oligonucleotide primers yielded PCR products with negative control.

In the present study genomic DNA of Gir cattle is used as positive control for PCR amplification and DNA sequencing. Genomic DNA of Murrah (sample-5M) and Gir (sample-12G) was used for PCR amplification and sequencing.

4.2.3 DNA sequencing of *TMEM95* gene

The amplicons were subjected to purification and DNA sequencing was done by Sanger sequencing method (SciGenome Labs Pvt. Ltd. Cochin). Quality of the sequences is assessed by Phred scores. Phred quality scores are logarithmically linked to error probabilities. The phred quality score is the negative ratio of the error probability to the reference probability level of 1 expressed in Decibel (dB). Phred scores of more than 30 are considered as good quality (Table 6). High quality sequences were obtained for both Murrah buffalo and Gir cattle for the *TMEM95* gene region (Figure 7a and 7b). The poor quality ends were trimmed and base calling was manually verified. There were three heterozygous positions observed in the Murrah

Table 5. Quality and quantity of genomic DNA isolated from blood of Murrah buffalo and Gir cattle

Sample No.	λ_{260}	$\lambda_{260}/\lambda_{280}$	Concentration of the DNA (ng/µl)	Total Yield (µg)
Murrah_5M	2.194	1.80	109.7	32.91
Murrah_8M	4.490	1.88	224.5	67.35
Murrah_10M	2.307	1.82	115.4	34.62
Murrah_12M	4.994	1.82	249.7	74.91
Gir_1G	1.575	1.89	78.8	23.64
Gir_9G	2.574	1.84	128.7	38.61
Gir_12G	0.999	1.79	50.0	15.00

Table 6. Phred quality scores

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

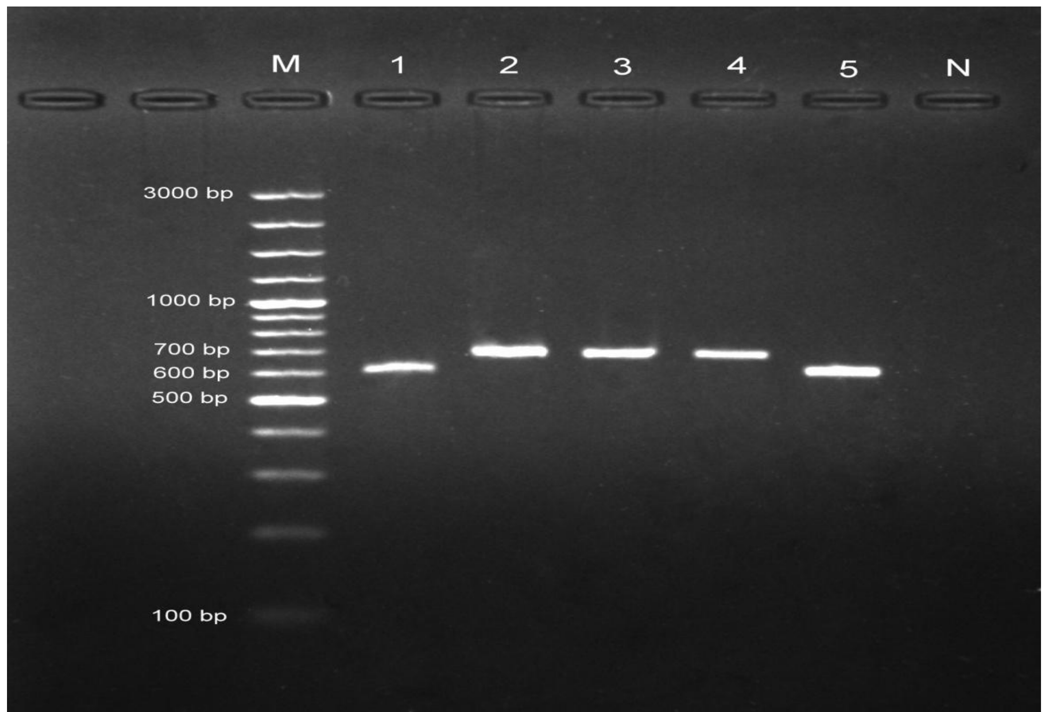

Figure 5. **Photograph of PCR amplification products of five fragments (1-5) of *TMEM95* gene in Murrah buffalo on agarose gel.**

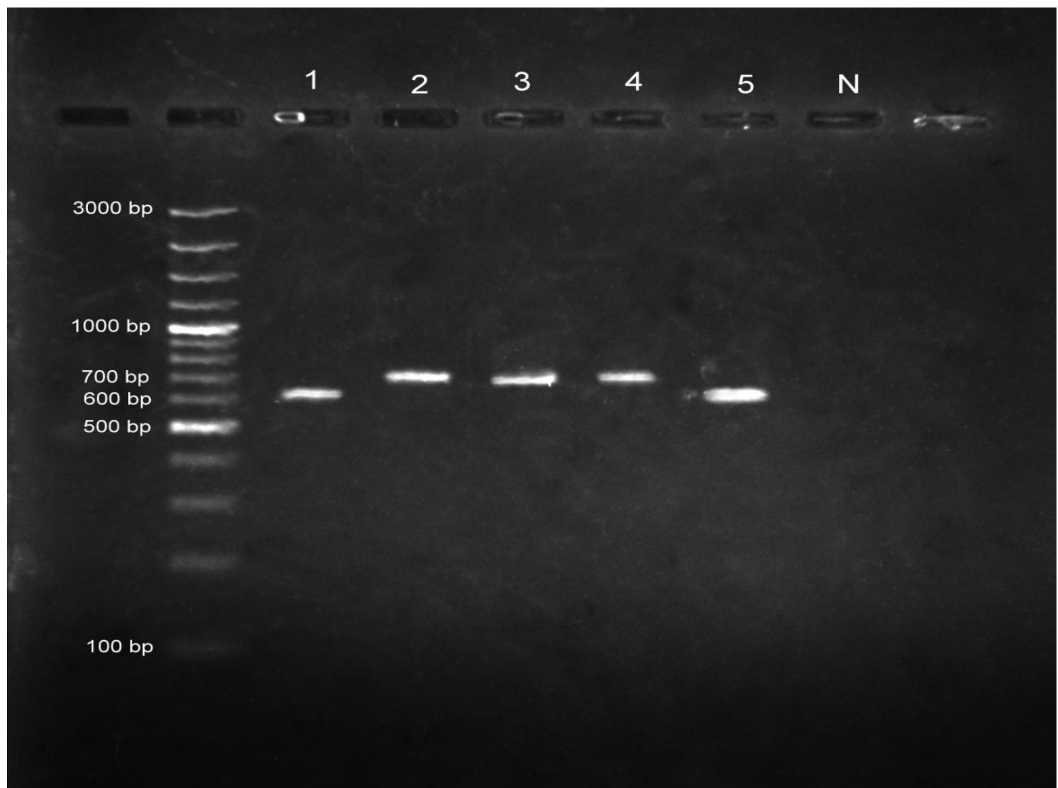

Figure 6. **Photograph of PCR amplification products of five fragments (1-5) of *TMEM95* gene in Gir cattle on agarose gel.**

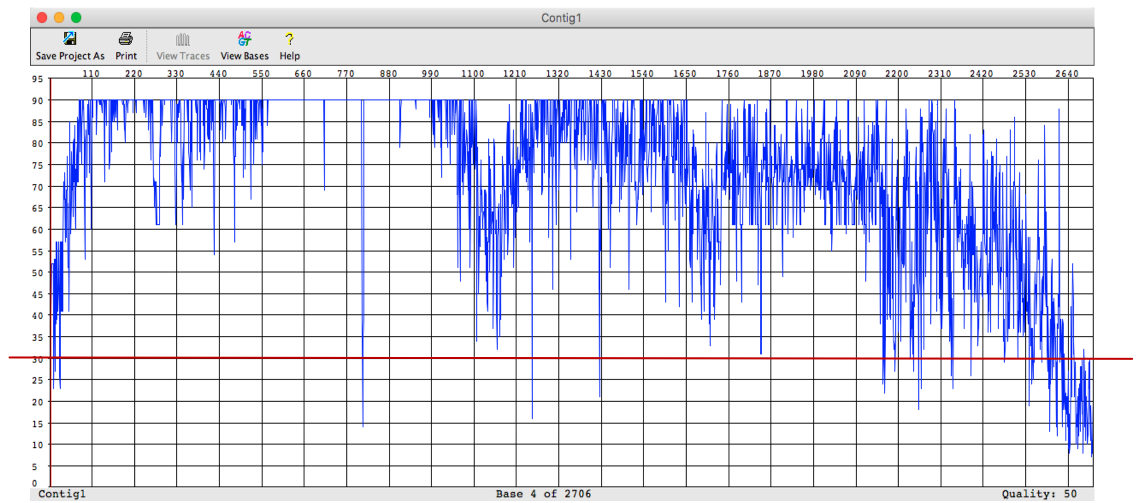

Figure 7a. **Quality scores of Murrah buffalo *TMEM95* gene contig**

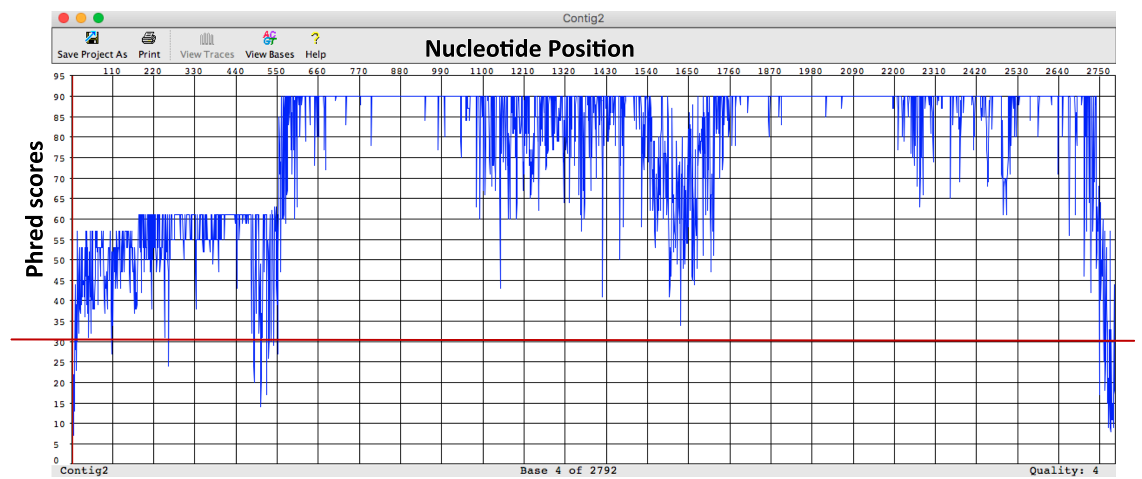

Figure 7b. **Quality scores of Gir cattle *TMEM95* gene contig**

sequence. These are located at positions 1284 (T and G), 1460 (C and A) and 1897 (G and A) (Figure 8a, 8b and 8c) with reference to the 2631 bp reference fragment from the Hereford cattle genome assembly (Accession No. AC_000176.1). Whereas, analysis of the sequences obtained from Gir cattle in the present study revealed three mutations at 1527 (A→C), 1581 (C→G) and 2195 (T→G).

4.2.4 Annotation of *TMEM95* gene in Murrah

There are six isoforms reported in the NCBI for the *TMEM95* gene in cattle (Figure 9). In the present study, a two base pair deletion at 937 bp (Figure 10) with reference to the 2631 bp reference fragment from the Hereford cattle genome assembly (Accession No. AC_000176.1) causing frame shift was observed in Murrah *TMEM95*. The predicted protein and coding DNA sequences using GeneWise for Murrah buffalo and Gir cattle are shown in Table 7. In Murrah as a result of the frameshift mutation three isoforms 1, 3 and 4 (Table 8) are unlikely to be formed.

4.2.5 Functional prediction of *TMEM95* gene in cattle and buffalo

4.2.5.1 Motif prediction

To understand the function of *TMEM95* in cattle, buffalo and yak, the polypeptide sequences of different isoforms were subjected to motif prediction using Motif Scan program (http://myhits.isb-sib.ch/cgi-bin/motif_scan). The motifs predicted from the sequence of both Murrah and Gir obtained in the present study were Casein kinase II phosphorylation site, leucine zipper pattern, N-myristoylation site, protein kinase C phosphorylation site, CHRD domain profile, N-glycosylation site and HIT zinc finger. The numbers of these motifs in different isoforms are given in Table 9. There are clear differences with respect to the number of these motifs between cattle and buffalo.

Table 7. Predicted protein and coding DNA sequences of *TMEM95* gene Murrah buffalo and Gir cattle.

Isoform-1

>Gir_isoform1

MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKIRHPGMGLTTIT
PVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEIKGSLYSLPSYWQWLRKTKLR
EYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLPRKSRSLGSSDPAPLCVRNCPAPGCPEPRGGV
SWGIRAGDSLTTSHQYPLILPPALVALPSSLPRPPTLCFPSGTTTSKQKVDL

>Gir_Cds_1

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCC
AGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTG
TGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAAATCAGACATCCAGGGATGGGCCTAACAA
CAATCACCCCTGTGCCCCAAGGGGCCTGCAAGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGT
CCACAGCCCTCTCTGGACACACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCC
CAGAACCTTCTTCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGCAGCACCATCC
TGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGA
AGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTC
CCGAGCCTCGCGGTGGAGTGAGTTGGGGGATCAGGGCGGGAGACAGCCTCACCACTTCCCACCAGT
ACCCCCTCATCCTTCCTCCTGCCCTCGTGGCCCTTCCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTT
CCCTCAGGTACAACTACTTCCAAGCAAAAAGTTGACTTG

Isoform-2

>Murrah_isoform2

MWTLALGGIFLAAVEACVFCRFPDRELSGRLAQLCSQMEVQWKDCEVSWTFSAFALDDASLNKITEKTHR
VLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPACREGQHHPVQLFHLPGLRGVLLAPKALLPRKSR
SLGSSDPAPLCVRNCPAPGCSEPRGGVSWGIRAGDSLITSHQYPLILPPALAALSSSLPRPPTLCFPSGTTTSK
QKVDL

>Murrah_Cds_2

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCC
AGACCGTGAGTTGTCGGGCCGCCTGGCTCAGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTGT
GAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGCATCCTTGAACAAAATCACAGAGAAGAC
TCACAGAGTCCTGAGAGTCATGGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGC
TTCGAAAGACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCGCCTGCCGTGAGGGGCA
GCACCATCCTGTACAACTGTTCCACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGC
TTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTC
CTGGGTGTTCCGAGCCTCGCGGTGGAGTGAGTTGGGGGATCAGGGCGGGAGACAGCCTCATCACTT
CCCACCAGTACCCCCTCATCCTTCCTCCTGCCCTCGCGGCCCTTTCCTCCTCCCTTCCCAGGCCTCCCACC
CTCTGTTTTCCCTCAGGTACAACTACTTCCAAGCAAAAAGTTGACTTG

>Gir_isoform2

MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDASLNKITEKTHR
VLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLPRKSR
SLGSSDPAPLCVRNCPAPGCPEPRGGVSWGIRAGDSLTTSHQYPLILPPALVALPSSLPRPPTLCFPSGTTTS
KQKVDL

>Gir_Cds_2

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCC
AGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTG

TGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGCATCCTTGAACAAAATCACAGAGAAGA
CTCACAGAGTCCTGAGAGTCATGGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGG
CTTCGAAAGACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGC
AGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTG
CTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCT
CCTGGGTGTCCCGAGCCTCGCGGTGGAGTGAGTTGGGGGATCAGGGCGGGAGACAGCCTCACCACT
TCCCACCAGTACCCCCTCATCCTTCCTCCTGCCCTCGTGGCCCTTCCCTCCTCCCTTCCCAGGCCTCCCAC
CCTCTGTTTTCCCTCAGGTACAACTACTTCCAAGCAAAAAGTTGACTTG

Isoform-3

>Gir_isoform3

MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKIRHPGMGLTTIT
PVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEIKGSLYSLPSYWQWLRKTKLR
EYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLPRSLLPTSASPPPRSLSSEASPASMGMVPGITH
CFLGPL

>Gir_Cds_3

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCC
AGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTG
TGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAAATCAGACATCCAGGGATGGGCCTAACAA
CAATCACCCCTGTGCCCCAAGGGGCCTGCAAGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGT
CCACAGCCCTCTCTGGACACACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCC
CAGAACCTTCTTCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGCAGCACCATCC
TGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGT
CCTTACTCCCCACCTCTGCCTCACCCCCACCCAGATCTCTGAGCTCTGAGGCTTCCCCAGCATCGATGG
GTATGGTCCCCGGGATCACTCACTGCTTCCTGGGTCCCCTA

Isoform-4

>Gir_isoform4

MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKIRHPGMGLTTIT
PVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEIKGSLYSLPSYWQWLRKTKLR
EYNREALCPPSCRGSTILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVPSLAVEYNYF
QAKS

>Gir_Cds_4

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCC
AGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTG
TGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAAATCAGACATCCAGGGATGGGCCTAACAA
CAATCACCCCTGTGCCCCAAGGGGCCTGCAAGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGT
CCACAGCCCTCTCTGGACACACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCC
CAGAACCTTCTTCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCACCATCCTGTA
CAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTC
ACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCG
AGCCTCGCGGTGGAGTACAACTACTTCCAAGCAAAAAGT

Isoform-5

>Murrah_isoform5

MPSQPRWGQPSLQQPPAGLMWTLALGGIFLAAVEACVFCRFPDRELSGRLAQLCSQMEVQWKDCEVS
WTFSAFALDDASLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPACRGSTILYNCST

CQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVPSLAVE

>Murrah_Cds_5

ATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAGCCCCCAGCAGGCCTCATGTGGACACT
GGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCCAGACCGTGAGT
TGTCGGGCCGCCTGGCTCAGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTGTGAGGTCTCCTG
GACATTCTCGGCCTTTGCCTTAGATGATGCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCC
TGAGAGTCATGGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACC
AAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCGCCTGCCGGGGCAGCACCATCCTGTACA
ACTGTTCCACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCAC
GATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTTCCGAG
CCTCGCGGTGGAG

>Gir_isoform5

MPSQPRWGQPSLQQPPAGLMWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVS
WTFSAFALDDASLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPSCRGSTILYNCST
CQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVPSLAVE

>Gir_Cds_5

ATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAGCCCCCAGCAGGCCTCATGTGGACACT
GGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCCAGACCGTGAGT
TGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTGTGAGGTCTCCTG
GACATTCTCGGCCTTTGCCTTAGATGATGCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCC
TGAGAGTCATGGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACC
AAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCACCATCCTGTACAA
CTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCACG
ATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCC
TCGCGGTGGAG

Isoform-6

>Murrah_isoform6

MWTLALGGIFLAAVEACVFCRFPDRELSGRLAQLCSQMEVQWKDCEVSWTFSAFALDDASLNKITEKTHR
VLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPACPLPALLYPAGGSTILYNCSTCQGFEVYCWPRKR
CFPGSHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQAKS

>Murrah_Cds_6

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCC
AGACCGTGAGTTGTCGGGCCGCCTGGCTCAGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTGT
GAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGCATCCTTGAACAAAATCACAGAGAAGAC
TCACAGAGTCCTGAGAGTCATGGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGC
TTCGAAAGACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCGCCTGCCCACTGCCGGCC
CTACTGTATCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCCACCTGCCAGGGCTTCGAGGTGTA
CTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCT
TTGTGTGCGGAACTGCCCTGCTCCTGGGTGTTCCGAGCCTCGCGGTGGAGTACAACTACTTCCAAGCA
AAAAGT

>Gir_isoform6

MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDASLNKITEKTHR
VLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPSCPLPALLSPAGGSTILYNCSTCQGFEVYCWPRKR
CFPGSHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQAKS

>Gir_Cds_6

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGCCGCTTCCC

AGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGGTCCAGTGGAAGGACTG
TGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGCATCCTTGAACAAAATCACAGAGAAGA
CTCACAGAGTCCTGAGAGTCATGGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGG
CTTCGAAAGACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCCACTGCCGGC
CCTACTGTCTCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGT
ACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTC
TTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTACAACTACTTCCAAGC
AAAAAGT

Table 8. Predicted isoform profile of the *TMEM95* gene in deferent species using GeneWise

Species	Breed	Isoforms																	
		1			2			3			4			5			6		
		Amino acid length	Exons	Introns	Amino acid length	Exons	Introns	Amino acid length	Exons	Introns	Amino acid length	Exons	Introns	Amino acid length	Exons	Introns	Amino acid length	Exons	Introns
B. taurus	Hereford	263	5	4	220	6	5	218	4	3	217	6	5	185	6	5	183	7	6
B. indicus	Ongole	263	5	4	220	6	5	218	4	3	217	6	5	185	6	5	183	7	6
B. indicus	Gir	263	5	4	220	6	5	218	4	3	217	6	5	185	6	5	183	7	6
B. bubalis	Mediterranean	**x**	**x**	**x**	220	6	5	**x**	**x**	**x**	**x**	**x**	**x**	185	6	5	183	7	6
B. bubalis	Murrah	**x**	**x**	**x**	220	6	5	**x**	**x**	**x**	**x**	**x**	**x**	185	6	5	183	7	6
B. mutus		263	5	4	220	6	5	218	4	3	217	6	5	185	6	5	183	7	6

Table 9. **Motif prediction in different isoforms of *TMEM95* gene in cattle, buffalo and yak**

Isoform	Organism	Motif						
		Casein kinase II phosphorylation site	Leucine zipper pattern	N-myristoylation site	Protein kinase C phosphorylation site	CHRD domain profile	N-glycosylation site	HIT zinc finger
1	Cattle	3	1	5	3	1	×	×
	Buffalo	×	×	×	×	×	×	×
	Yak	4	1	5	3	1	×	×
2	Cattle	2	1	4	5	1	×	×
	Buffalo	1	1	4	4	1	×	×
	Yak	3	1	4	5	1	×	×
3	Cattle	3	×	4	2	1	×	×
	Buffalo	×	×	×	×	×	×	×
	Yak	3	×	4	2	1	×	×
4	Cattle	2	×	3	2	1	1	1
	Buffalo	×	×	×	×	×	×	×
	Yak	2	×	3	2	1	1	1
5	Cattle	1	×	3	4	×	1	1
	Buffalo	1	×	3	3	×	1	×
	Yak	1	×	3	4	×	1	1
6	Cattle	1	×	2	3	1	1	×
	Buffalo	1	×	2	3	1	1	×
	Yak	1	×	2	3	1	1	×

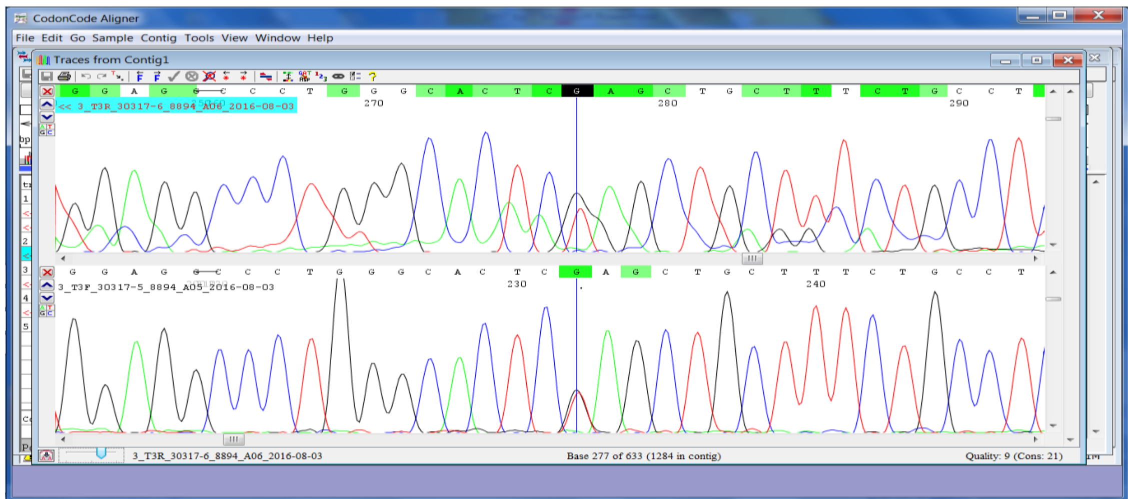

Figure 8a. Chromatogram of *TMEM95* showing heterozygous position at 1284 bp (T and G) in Murrah buffalo

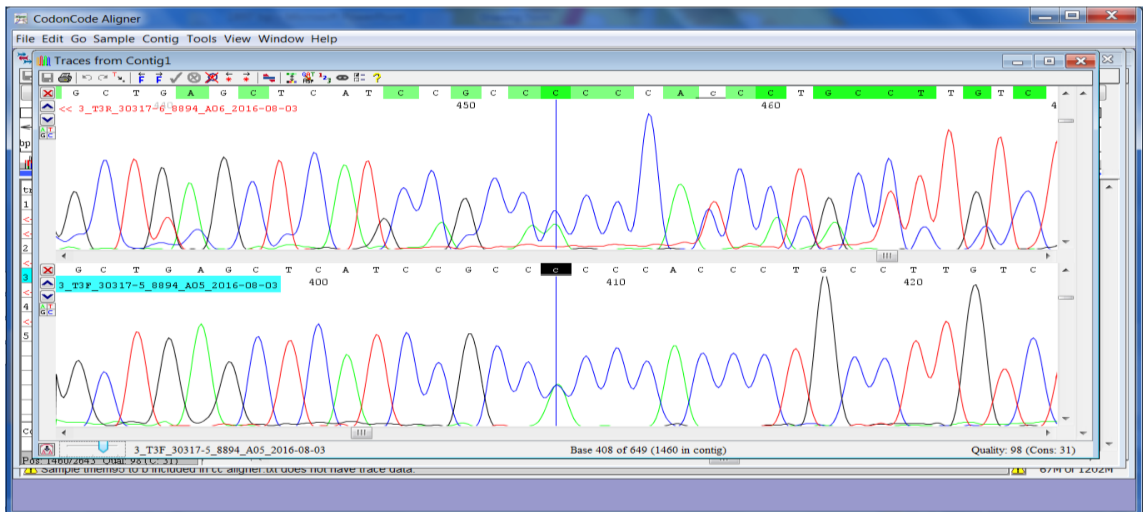

Figure 8b. Chromatogram of *TMEM95* showing heterozygous position at 1460 bp (C and A) in Murrah buffalo

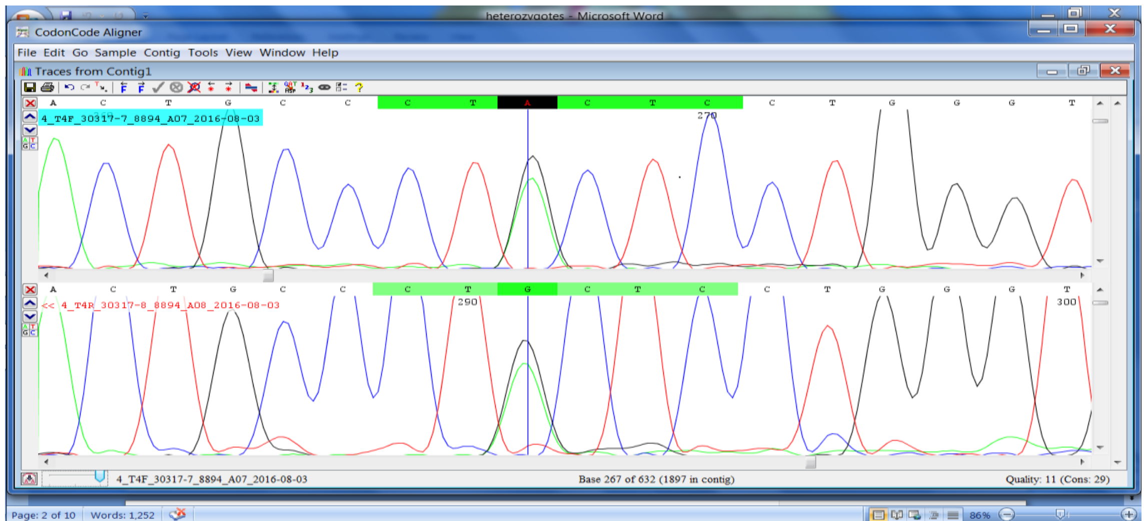

Figure 8c. Chromatogram of *TMEM95* showing heterozygous position at 1897 bp (G and A) in Murrah buffalo

Figure 9. ***TMEM95*** **gene annotation in cattle**

Source: www.ncbi.nlm.nih.gov

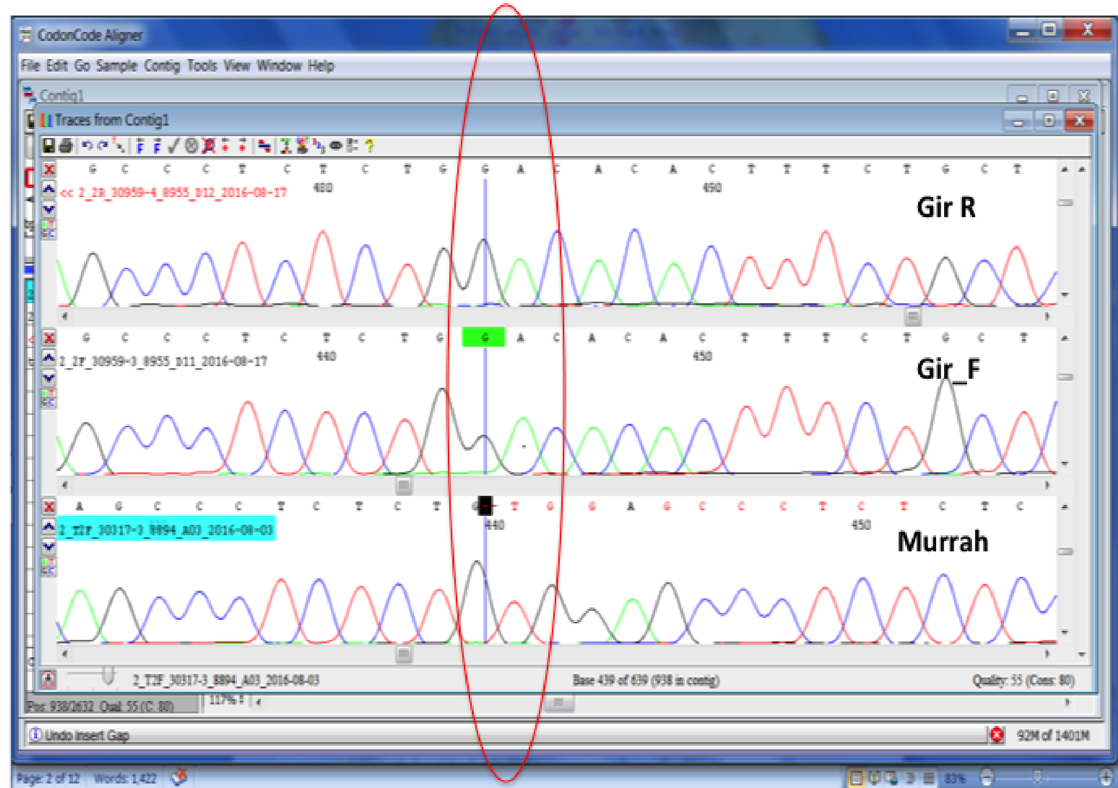

Figure 10. Chromatogram showing two base pair deletion at 937 bp in Murrah *TMEM95* causing a frame shift mutation.

4.2.5.2 Transmembrane topology and signal peptide prediction

The transmembrane topology and signal peptide were predicted using Phobius program (Kall *et al*., 2004). The location of transmembrane domain and signal peptide are presented in Table 10. Functional analysis of various isoforms in Murrah and Gir indicated that isoforms 1, 2 and 3 doesn’t have any transmembrane domains. Whereas, the isoform 5 possesses two transmembrane domains one each on N-terminal side and C-terminal side of the protein along with no signal peptide. All other isoforms contained signal peptide on the N-terminal side.

4.3 Phylogenetic analysis

Homologous regions of the *TMEM95* gene in other species viz. bison, wild yak, mithun, sheep, pig, horse, cat, dog, human and mouse were obtained using BLAT search against the corresponding genomes. All the sequences were aligned using ClustalW program as implemented in MEGA7 (Kumar *et al*, 2016). The analysis involved 15 nucleotide sequences. All positions containing gaps and missing data were eliminated. All the ambiguous positions were removed for each sequence pair and a total of 3160 positions were analyzed. Evolutionary analyses were conducted using MEGA7 and all the individual sequences were classified into ruminant and monogastric. The mean diversity within the ruminants and monogastric species was calculated as the number of base differences per sequence from mean diversity calculations within each group (Nei and Kumar 2000). The mean distance within the groups found to be 318.26. The number of base differences per sequence from the mean diversity between the groups is expressed as mean diversity between the groups and is 88.38. The number of base differences per sequence from averaging overall sequence pairs is expressed as overall mean distance and is found to be 406.64. The coefficient of differentiation between ruminants and monogastrics is 0.22. The

Table 10. **Transmembrane topology and signal peptide predictions in different isoforms of *TMEM95* gene in cattle, buffalo and yak**

Isoform	Peptide length	Species	Signal peptide	N-Region	H-Region	C-Region	Non-cytoplasmic	Transmembrane	Cytoplasmic
I	263	Cattle	1-16	1-2	3-11	12-16	17- 263		
		Yak	1-16	1-2	3-11	12-16	17- 263		
II	220	Cattle	1-16	1-2	3-11	12-16	17-220		
		Buffalo	1-16	1-2	3-11	12-16	17-220		
		Yak	1-16	1-2	3-11	12-16	17-220		
III	218	Cattle	1-16	1-2	3-11	12-16	17-218		
		Yak	1-16	1-2	3-11	12-16	17-218		
IV	217	Cattle	1-16	1-2	3-11	12-16	17-188	189-213	214-217
		Yak	1-16	1-2	3-11	12-16	17-188	189-213	214-217
V	185	Cattle					1-19	20- 41; 165-184	42-164; 185-185
		Buffalo					1-19	20- 41; 165-184	42-164; 185-185
		Yak					1-19	20- 41; 165-184	42-164; 185-185
VI	183	Cattle	1-16	1-2	3-11	12-16	17-154	155-179	180-183
		Buffalo	1-16	1-2	3-11	12-16	17-154	155-179	180-183
		Yak	1-16	1-2	3-11	12-16	17-154	155-179	180-183

The numbers indicate position of the amino acid in each of the isoform.

pairwise distances in terms of number of base differences per sequence are shown in Table 11.

ModalTest (Posada and Crandall, 1998) was performed to check the best suitable model to explain the sequence divergence. It selects the best-fit nucleotide substitution model, General Time Reversal with Gamma distribution and variant site (GTR +G+I) (lnL-10894.8221) for a set of aligned sequences. Phylogenetic tree was constructed using maximum likelihood method with 1000 bootstrap iterations. The phylogenetic tree is shown in Figure 11. Phylogeny and divergence of *TMEM95* gene is in accordance with the evolutionary divergence of these species.

Table 11. Estimates of pair wise evolutionary divergence between *TMEM95* sequences of different species

Species	Herford cattle	Ongole cattle	Gir cattle	Bison	Wild yak	Mithun	Murrah buffalo	Mediterranean buffalo	Sheep	Pig	Horse	Dog	Cat	Human	Mouse
Herford cattle	0.0														
Ongole cattle	0.0														
Gir cattle	2.0	0.0													
Bison	7.0	6.0	7.0												
Wild yak	6.0	5.0	6.0	7.0											
Mithun	6.0	5.0	6.0	7.0	0.0										
Murrah buffalo	43.0	42.0	43.0	44.0	43.0	43.0									
Mediterranean buffalo	43.0	42.0	43.0	44.0	43.0	43.0	2.0								
Sheep	85.0	84.0	85.0	86.0	84.0	84.0	93.0	92.0							
Pig	435.0	434.0	435.0	435.0	432.0	432.0	435.0	434.0	434.0						
Horse	493.0	492.0	494.0	494.0	494.0	494.0	492.0	492.0	482.0	463.0					
Dog	652.0	650.0	652.0	652.0	650.0	650.0	648.0	648.0	650.0	617.0	594.0				
Cat	592.0	591.0	593.0	591.0	590.0	590.0	584.0	585.0	577.0	572.0	513.0	514.0			
Human	638.0	636.0	637.0	638.0	657.0	657.0	633.0	635.0	622.0	582.0	532.0	682.0	640.0		
Mouse	753.0	752.0	754.0	754.0	754.0	754.0	752.0	751.0	740.0	753.0	705.0	764.0	749.0	741.0	0.0

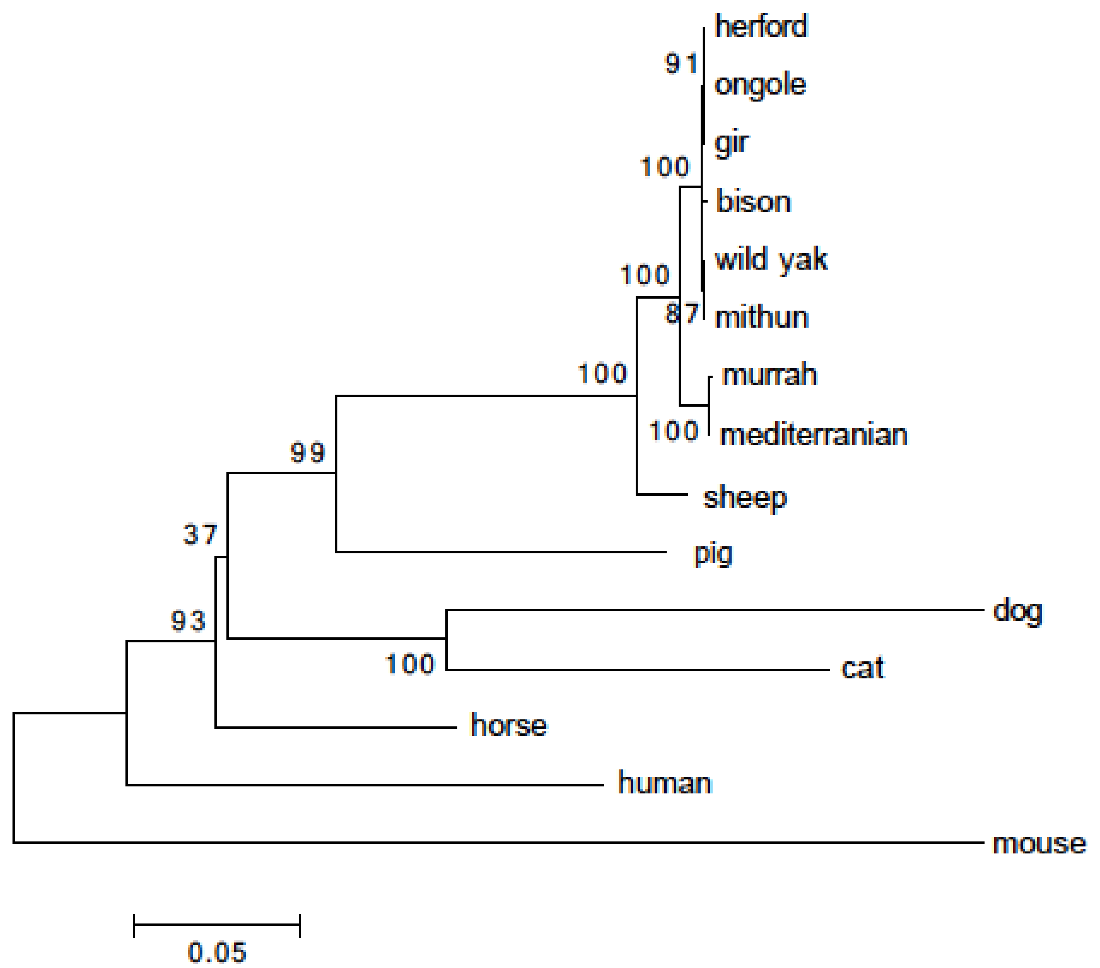

Figure 11. Phylogenetic tree of *TMEM95* genomic region in different species. Bootstrap values for 1000 iterations is shown on the branches.

CHAPTER V

DISCUSSION

The buffalo genome sequencing is in its primitive stage. Two independent groups started sequencing buffalo genome, one by Zimin *et al*. (2013) from University of Maryland, USA, sequencing the Mediterraniain water buffalo (NCBI Accession No. AWWX01488834.1) and another one by Tantia *et al*. (2011) from Anand Agricultural University, India, sequencing the Jaffarabadi buffalo breed (NCBI Accession No. ACZF020000000). The former is at scaffold stage and the latter is at the contig level. The annotation and analysis of the buffalo genome is not yet complete.

The present study is aimed at characterizing the gene *TMEM95* in Murrah buffalo, which is known to be involved in idiopathic male subfertility in cattle and conserved across taxa including human (Pausch *et al*., 2014). The gene-gene and gene environment interactions in the regulation of male fertility have been poorly characterized (Roy *et al*., 2007). Comparative genomic approach is a useful tool for characterizing the sequences in unknown species (Hardison, 2003). Hence in the present study it was attempted to characterize the *TMEM95* gene in Murrah buffalo and understand the functional differences between cattle and buffalo.

5.1 Yield and Quality of DNA

In the present study, DNA was isolated from blood samples using modified high salt method (Miller *et al*., 1988). The high salt method was reported to be very efficient for isolating DNA from cattle and buffalo blood (Aravindakshan *et al*., 1997). The ratio of optical density at 260/280 nm for majority of the samples lies within a range of 1.79 to 1.89, indicating a good deproteinization. The DNA obtained in the present study was of good quality and suitable for further analysis.

5.2 Characterization of *TMEM95* genomic region

The *TMEM95* sequences of taurine and indicine were obtained from NCBI and were compared with that of Mediterranean buffalo and *B. mutus (*yak*)*. It was observed that there were no differences in the genomic regions of taurine and indicine cattle with respect to this region.

5.2.1 Cross species amplification and sequencing of *TMEM95* region

Since cattle and buffalo genome sequences are highly conserved (Iannuzzi *et al.,* 2003), the primers designed for cattle genome were used for sequencing the *TMEM95* region in Murrah buffalo. Further, by BLAST search against Mediterranean buffalo genome sequence, homologous region is parsed. Similarly homologous region for *Bos indicus* (Ongole breed) and *Bos mutus* (Yak) are parsed. All the four sequences were aligned using clustalW to design conserved primers. Five pairs of primers were designed to obtain full-length sequence of *TMEM95* region. The primers produced unambiguous amplicons for both cattle and buffalo genomic DNA. However, there is slight difference in annealing temperature for two of the fragments between Murrah and Gir (Tables 3 and 4). Theoretically, annealing temperature depends on primer composition but not on sequence composition (Innis *et al*., 1990). It is difficult to attribute the reason to any factor for the differences in annealing temperature for the two fragments but there may be hitherto unidentified factors that influence annealing temperatures specific to a template and need to be explored.

The overlapping of at least 100 bp between the adjacent fragments to be able to obtain unambiguous full-length sequence has produced good quality final product (Figures 5 and 6).

The genomic sequences obtained from Murrah buffalo and Gir cattle were found to be having Phred scores of more than 30, which is considered as good quality.

Sequence alignment of the *TMEM95* of Murrah and Mediterranean buffalo revealed that the region is completely conserved between the two breeds. However, there are three heterozygous positions observed both in Murrah and Mediterranean sequences (Figure 8a, 8b and 8c) suggesting there is possibility of limited variation of this region in buffalo. Probably the gene might be involved in an important function in buffalo and hence there is selective pressure in both coding and intronic regions of the gene that lead to high conservation. In the present study the sequences obtained from Gir cattle revealed mutations at 1527 (A→C), 1581 (C→G) and 2195 (T→G) compared to Hereford *TMEM95* sequence. On analyzing the Gir sequence with that of *B. indicus* (Ongole) available in NCBI database (Accession No.CM003039.1) at two of the above positions *i.e.* 1527 and 2195 degeneracy of bases was observed, which were referred as M (representing A or C) and K (G or T) respectively according to IUPAC nucleotide code. Further, it was observed that at position 562 another degenerate base R (representing A or G) was present in Ongole cattle. No difference at position 1581 was observed between Gir and Ongole. Population studies of this region will provide better understanding of the natural variation in buffaloes and cattle.

5.2.2 Annotation of *TMEM95* gene in Murrah and Gir

There are six isoforms reported in cattle for the *TMEM95* region in NCBI database. All these isoforms are supported by RefSeq database. The function of these isoforms is not yet annotated except that the gene is involved in male fertility and expressed on the surface of sperm (Pausch *et al*., 2014). Earlier research indicated that the transmembrane proteins are involved in sperm-egg interactions (Suri, 2004 and Wang *et al*., 2009). Comparative gene annotation tool GeneWise was used to predict the isoforms of *TMEM95* in buffalo. The six isoforms are conserved between cattle and yak but only three are conserved in buffalo. The difference is attributed to a two base pair deletion at 937 bp in buffalo (both in Murrah and Mediterranian), which is

leading to frame shift and disruption of the three missing isoforms in buffalo. Pausch *et al.* (2014) reported that a putative mutation (rs378652941, c.483C.A, p.Cys161X) in the isoforms 4, 5 and 6, leading to premature stop codon in cattle. In the present study, interestingly the isoform 4 is affected in buffalo due to the two base pair deletion (at 937 bp), which might lead to a frame shift mutation in *TMEM95* of Murrah resulting in a truncated protein which could be due to a non-sense mediated mRNA decay (Chang *et al*., 2007). However it is not clear which isoform exactly affects the male fertility. It was opined that the *TMEM95* mutants doesn't show any abnormalities with respect to sperm quality but affect fertility (Pausch *et al*., 2014). Zhang *et al*. (2015) have identified two novel alternatively splicing transcripts of *TMEM95* in cattle. Both the forms are expressing only in testis and brain. Further, it was shown that *TMEM95* gene might have role in brain function. Similarly, there are species-specific differences in expression of number of isoforms. Hence spatial expression profiling of various isoforms in cattle and buffalo would help in unraveling the specific function of the gene in these species.

5.2.3 Functional prediction of *TMEM95* gene in buffalo and cattle

To understand functional implications of various isoforms in cattle and buffalo, the peptide sequences of each isoform were subjected to prediction of transmembrane topology and signal peptide. Results indicated that the isoforms 1, 2 and 3 doesn't have any transmembrane domains and the isoform 5 possesses two transmembrane domains one each on N-terminal side and the C-terminal side of the protein along with no signal peptide. All other isoforms contain signal peptide on the N-terminal side. Signal peptide is essential for localization of the protein (Blobel and Dobberstein 1975). It would be interesting to verify localization of these various isoforms in testis and other tissues to understand precise function of each of the isoform.

To further understand the biological function of *TMEM95* gene and its isoforms, the protein sequences were scanned for various motifs located in them. Motifs are indicators of their biological function. The sequences showed casein kinase II phosphorylation site, leucine zipper pattern, N-myristoylation site, protein kinase C phosphorylation site and CHRD domain profile. It was shown that lack of casein kinase II in mice led to morphological defects in sperm function (Xu *et al*., 1999). Leucine zipper is involved in regulatory function, Myristoylation is involved in signal transduction (Hyashi and Titani 2010) while Protein kinases were involved in modifying function of target proteins. CHRD domain function is not yet known. There are species-specific differences in the presence or absence and in number of these motifs if present. Summarizing the domain architectures of these isoforms, *TMEM95* plausibly involved in sperm morphology and regulatory function and there may exist functional differences between different isoforms.

5.3 Phylogenetic analysis

To understand the diversity of *TMEM95* genomic region in different species, pairwise differences between sequences from different species that are closely and distantly related to cattle were analysed. Less divergence is observed among cattle, bison, yak and mithun. Buffalo and sheep are moderately divergent from cattle. Overall, all the ruminants are relatively more closer compared to the monogastric species including human. Further phylogenetic analysis (Figure 10) indicated clear clustering of ruminants and monogastric species. Further, more variation is observed among monogastric species. It is too primitive to conclude functional differences of the gene between ruminants and monogastrics based on the divergence. Initial divergence of the Laurasiatheres (includes pigs, dog, cat, ruminants etc) and Euarchontoglires (primates and rodents) occured about 95 million years ago (mya). Perissodactyla (Horse etc), Carnivora (Dog, cat etc.) and Pholidota (Pangolin) were

diverged from Cetartiodactyla (Pig and ruminants) about 80 mya. Pigs and ruminants split about 60 mya (Murphy *et al.,* 2004) Phylogeny and divergence of *TMEM95* gene is in accordance with the evolutionary divergence of these species.

In the present study the *TMEM95* gene which is likely to affect the male fertility was characterized using bioinformatic tools. Three mutations in Gir cattle and a two base pair deletion and three heterozygous positions in Murrah were significant observations. Three of the isoforms were unlikely to be formed in Murrah. However, there may be some species-specific differences with respect to their function between the cattle and buffalo. Further, spatial expression profiling of various isoforms and analysis of natural variation would help in better understanding of the function of the gene and its role in male fertility. As *TMEM95* gene has a role in sperm integrity, which is required for undisturbed fertilization and their transcripts could be mirrors of sperm health this area warrants much more research.

CHAPTER VI

SUMMARY

Fertility is an important multifactorial trait in livestock, difficult to study as it affected by several component traits. Infertility or subfertility problems are major concern in buffaloes leading to bull disposal. Thus studying about the genes involved, is important to understand the infertility when the routine approach fails. It was reported that the integrity of *TMEM95* is required for an undisturbed fertilization. Sequencing and characterization of *TMEM95* gene in Murrah buffaloes would help in better understanding of its features thus providing some insights into infertility or subfertility in bulls. Hence, the present work was undertaken to amplify, sequence and characterize *TMEM95* gene in Murrah buffalo and Gir cattle using bioiformatic tools.

Genomic DNA was isolated by high salt method from the blood samples of four Murrah and three Gir cattle of ILFC, N T R College of Veterinary Science, Gannavaram and Veerankilaku Dairy farm respectively. One Sample each from the two breeds, having high quality DNA were used for the present study. Five pairs of primers were designed for the sequencing of *TMEM95* region in Murrah buffalo by using comparative genomics approach.

The sequence of *TMEM95* gene of Murrah and Gir were interpreted and aligned by CodonCode Aligner software 6.0.2 . Three heterozygous positions were observed in the Murrah sequence at positions 1284 (T and G), 1460 (C and A) and 1897 (G and A) with reference to the 2631 bp reference fragment from the Hereford cattle genome assembly.

The isoforms were predicted in Murrah buffalo by using GeneWise, a comparative gene annotation tool and a 2 bp deletion (at 937 bp) was observed in both Mediterranian and Murrah buffalo which is causing frame shift mutation leading to malformation of the

the isoforms 1, 3 and 4. Whereas, three mutations were observed in Gir with reference to Hereford cattle sequence available in NCBI. The six isoforms were found to be conserved between cattle and yak but only three were conserved in buffalo.

To understand functional implications of various isoforms in cattle and buffalo, the peptide sequences of each isoform were subjected to prediction of transmembrane topology and signal peptide by Motifscan and Phobius programs. The Phobius results revealed that isoforms 1, 2 and 3 doesn't have any transmembrane domains both in buffalo and cattle. The Isoform 5 had two transmembrane domains, one on N-terminal side and the other on C-terminal side without signal peptide along with no signal peptide. All the isoforms contain signal peptide on the N-terminal side. Signal peptide is essential for localization of the protein. Motif scan results revelead that the gene has casein kinase II phosphorylation site, leucine zipper pattern, N-myristoylation site, protein kinase phosphorylation site and CHRD domains. It was shown that lack of casein kinase II in mice led to morphological defects in sperm function.

The diversity of *TMEM95* genomic region in different species was studied using the information obtained on pairwise differences between sequences from different species that are closely and distantly related to cattle. Less divergence was observed among cattle, bison, yak and mithun. Buffalo and sheep were moderately divergent from cattle. Overall, all the ruminants were found to be relatively more closer, compared to the monogastric species, including human.

The results of the present study indicated that *TMEM95* gene in Murrah buffalo is likely to have function in male fertility like in cattle. However, there may be some species-specific differences with respect to their function between the two species. Further, spatial expression profiling of various isoforms and analysis of natural variation

would help in better understanding of the function of the gene and its role in male fertility. As *TMEM95* gene has a role in sperm integrity which is required for undisturbed fertilization and their transcripts could be mirrors of sperm health, there is much scope for research to validate the present findings.

LITERATURE CITED

Abavisani, A., Baghbanzadeh, A., Shayan, P. and Dehghani, H. 2011. Leptin mRNA in bovine spermatozoa. Research in Veterinary Science, **90**:439–442.

Aitken, R. J. 2006. Sperm function tests and fertility. International Journal of Andrology, **29**:69-75.

Angelopoulou, R., Lavranos, G. and Manolakou, P. 2012. Sex determination strategies in 2012: towards a common regulatory model? Reproductive Biology and Endocrinology, **10**:13.

Arabi, M. and Mohammadpour, A. A. 2006. Adverse effects of cadmium on bull spermatozoa. Veterinary research communications, **30**:943-951.

Aravindakshan, T. V., Nainar, A. M and Nachimuthu, K. (1997). High salt method: A simple and rapid procedure for isolation of genomic DNA from buffalo (*Bubalus bubalis*) white blood cells. Indian Journal of Experimantal Biology, **35**:903-905.

Arcelay, E., Salicioni, A. M., Wertheimer, E. and Visconti. P.E. 2008. Identification of proteins undergoing tyrosine phosphorylation during mouse sperm capacitation. International Journal of Developmental Biology, **52**:463-472.

Blobel, G. and Dobberstein, B. 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. The Journal of Cell Biology, **67**:835–51.

Bole-Feysot, C. H., Goffin, V., Edery, M., Binart, N. and Kelly, P. A. 1998. Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. Endocrine Reviews, **19**:225-268.

Bonfiglio, S., De Gaetano, A., Tesfaye, K., Grugni, V., Semino, O. and Ferretti, L. 2012. A novel USP9Y polymorphism allowing a rapid and unambiguous classification of *Bos taurus* Y chromosomes into haplogroups. Animal Genetics, **43**:611-613.

Canavez, F. C., Luche, D. D., Stothard, P., Leite, K. R. ., Sousa-Canavez, J. M., Plastow, G., Meidanis, J., Souza, M. A., Feijao, P., Moore, S. S. and Camara-Lopes, L. H. 2012. Genome Sequence and Assembly of *Bos indicus*. Journal of Heredity, **103**:342-348.

Carrell, D. T., Jonge, C. D. and Lamb, D. J. 2006. The Genetics of male infertility: a field of study Whose time is now. Archives of Andrology, **52**:269–274.

Chang, Y. F., Imam, J. S. and Wilkinson, M. F. 2007. The nonsense-mediated decay RNA surveillance pathway. Anuual Review of Biochemistry, **76**:51–74.

Chenoweth, P. 2005. Genetic sperm defects. Theriogenology, **64**:457–68.

Coen, D, M. and Scharf, S. J. 1991. The polymerase chain reaction. Current protocols in Molecular Biology, vol.2 .F.M Ausubel. New York,Greene publishing Associates and Wiley Interscience.

Cohen, D. J., Ellerman, D. A. and Cuasnicu, P. S. 2000. Mammalian sperm-egg fusion: evidence that epididymal protein DE plays a role in mouse gamete fusion. Biology of Reproduction, **63:**462-468.

Dash, S., Chakravarty, A. K., Singh, A., Shivahre, P. R., Upadhyay, A., Sah, V. and Singh, K. M. 2015. Assessment of expected breeding values for fertility traits of Murrah buffaloes under subtropical climate. Veterinary World **8**:320-325.

Dhanoa, J. K., Mukhopadhyay, C. S. and Arora, J. S. 2016. Y-chromosomal genes affecting male fertility: A review. Veterinary World, **9**:783-791.

Dong, L. Y., Li, Q. F., Qu, X. G., Li, Y. X., Li, X. F., Hu, H. T., and Xie, Z. 2009. Expression levels of Cdc2 and Cdc25A mRNA in cattle, yak and cattle-yak testis. Yi chuan= Hereditas/Zhongguo yi chuan xue hui bian ji, **31**:495-499.

Ebling, D. W., Ruffer, J., Whittington, R., Vanarsdalen, K., Broderick, G. A., Malkowicz, S. B. and Wein, A. J. 1997. Development of prostate cancer after pituitary dysfunction: a report of 8 patients. Urology, **49**:564-568.

Ferreira, R. M., Chiaratti, M. R., Macabelli, C. H., Rodrigues, C. A., Ferraz, M. L., Watanabe, Y. F., Smith, L. C., Meirelles, F. V. and Baruselli, P. S. 2016. The infertility of repeat-breeder cows during summer is associated with decreased mitochondrial DNA and increased expression of mitochondrial and apoptotic genes in oocytes. Biology of Reproduction, **94**:1-10.

Feugang, J. M., Kaya, M., Page, G. P., Chen, L., Mehta, T., Hirani, K., Nazareth, L., Topper, E., Gibbs, R. and Memili, E. 2009. Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. BMC Genomics, **10**:176.

Feugang, J. M., Rodriguez-Osorio, N. and Memili, E. 2007. Advances and applications of genomics and proteomics in gamete and embryo science. Acta Scientiae Veterinariae, **35**:847-856.

Givens, M. and Marley, M. 2008. Pathogens that cause infertility of bulls or transmission via semen. Theriogenology, **70**:504 -7.

Gui, L. M., Shan, L. Z. and Xiang, P. Z. 2012. The mRNA expression and methylation Status in imprinting control region of H19 Gene between Cattle-Yak and their Parents. Journal of Integrative Agriculture, **11**:1691-1699.

Hardison, R.C 2003. Comparative Genomics. PLOS Biology **1**:e58.

Hayashi, N. and Titani, K. 2010. N-myristoylated proteins, key components in intracellular signal transduction systems enabling rapid and flexible cell responses. Proceedings of the Japan Academy. Series B, Physical and Bilologiacl Sciences, **86:**494–508.

Iannuzzi, L., Di Meo G. P., Perucatti, A., Schibler, L., Incarnato, D. and Gallagher, D. 2003. The river buffalo (*Bubalus bubalis*, 2n = 50) cytogenetic map: assignment of 64 loci by fluorescence in situ hybridization and R-banding. Cytogenet Genome Res. **102**:65–75.

Ikawa, M., Nakanishi, T., Yamada, S., Wada, I., Kominami, K., Tanaka, H., Nozaki, M., Nishimune, Y. and Okabe, M. 2001. Calmegin is required for fertilin α/β heterodimerization and sperm fertility. Developmental Biology, **240**:254-261.

Innis, M, A., Gelfand, D. H., Sninsky, J. J. and White, T.J. 1990. PCR protocols: A guide to methods and applications. San Diego, CA: Academic Press.

Kall, L., Krogh, A. and Sonnhammer, E. L. 2004. A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology, **338**:1027-1036.

Khan, H. M., Mohanty, T. K., Bhakat, M., Raina, V. S. and Gupta, A. K. 2011. Relationship of Blood Metabolites with Reproductive Parameters during Various Seasons in Murrah Buffaloes. Asian-Australasian Journal of Animal Sciences, **24**:1192-1198.

Khatib, H., Huang, W., Mikheil, D., Schutzkus, V. and Monson, R.L. 2009. Effects of signal transducer and activator of transcription (*STAT*) genes *STAT1* and *STAT3*

genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. Journal of Dairy Science, **92:**6186–6191.

Khatun, M., Kaur, S., Kanchan and Mukhopadhyay, C. S. 2013. Subfertility problems leading to disposal of breeding bulls. Asian-Australasian Journal of Animal Sciences, **26**:303-308.

Killian, G. 1999. High fertility proteins enhance reproduction rates in Dairy cattle. NRI Reasearch Highlights, Competitiive Grants Program, Factsheet, USDA.

Klonisch, T., Steger, K., Kehlen, A., Allen, W.R., Froehlich, C., Kauffold, J., Bergmann M. and Hombach-Klonisch, S. 2003. INSL3 ligand-receptor system in the equine testis. Biology of Reproduction. **68:**1975-1981.

Kumar, O. S., Sharma., D., Singh., D. and Sharma, M. K. 2009. CYP19 (cytochrome P450 aromatase) gene polymorphism in Murrah buffalo heifers of different fertility performance. Research in Veterinary Science, **86**:427-437

Kumar, S., Nagarajan, N., Sandhu, J. S., Kumar, N., Behl, V. and Nishanth, G. 2007. Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. Animal Genetics **38**:227-232.

Kumar, S., Stecher, G. and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 For Bigger Datasets. Molecualr Biology and Evolution, **33**:1870-1874.

Laporta, J., Driver, A. and Khatib, H. 2011. Expression and alternative splicing of POU1F1 pathway genes in pre implantation bovine embryos. Journal of Dairy Science, **94**:4220-4223.

Le Naour, F., Rubinstein, E., Jasmin, C., Prenant, M. and Boucheix, C. 2000. Severely reduced female fertility in CD9-deficient mice. Science, **287**:319-321.

Lessard, C., Siqueira, L. G., Amours, O. D., Sullivan, R., Leclerc, P. and Palmer,C. 2011. Infertility in a beef bull due to a failure in the capacitation process. Theriogenology, **76**:891-899.

Lin, H., SuYu, J., YaOu, X., YuPing, L., YaQiu, L. and YuCai, Z. 2012. Quantitation of alternative splicing variants of lactate dehydrogenase C gene in testes of adult yak, sexually immature yak calf and sterile male hybrid of yak. Canadian Journal of Animal Science, **92**:291-296.

Liu, W. S., Wang, A., Uno, Y., Galitz, D., Beattie, C. W. and Ponce de Leon, F. A. 2007. Genomic structure and transcript variants of the bovine DAZL gene. Cytogenetic and Genome Research, **116**:65-71.

Liu, W. S., Wang, A., Yang, Y., Chang, T. C., Landrito, E. and Yasue, H. 2009. Molecular characterization of the DDX3Y gene and its homologs in cattle. Cytogenetic and Genome Research, **126**:318-328.

Lou, Y. N., Liu, W. J., Wang, C. L., Huang, L., Jin, S. Y., Lin, Y. Q. and Zheng, Y. C. 2014. Histological evaluation and PRDM9 expression level in the testis of sterile male cattle-yaks. Livestock Science, **160**:208-213.

Maeda, T., Nishida, J. and Nakanishi, Y. 1999. Expression pattern, subcellular localization and structure-function relationship of rat Tpx-1, a spermatogenic cell adhesion molecule responsible for association with Sertoli cells. Development, Growth and Differentiation, **41:**715-722.

Meignanalakshmi, S. and Mahalinga Nainar, A. 2009. PCR-RFLP analysis of beta-lactoglobulin gene in Murrah buffaloes Tamilnadu. Journal of Veterinary and Animal Sciences, **5**:194-197.

Miah, A. G., Salma, U., Sinha, P. B., Holker, M., Tesfaye, D., Cinar, M. U., Tsujii, H. and Schellander, K. 2011. Intracellular signaling cascades induced by relaxin in the stimulation of capacitation and acrosome reaction in fresh and frozen-thawed bovine spermatozoa. Animal Reproduction Science, **125**:30-41.

Michelizzi, V. N., Dodson, M. V., Pan, Z., Amaral, M. E. J., Michal, J. J., McLean, D. J., Womack, J. E. and Jiang, Z. 2010. Water Buffalo Genome Science Comes of Age. International Journal of Biological Sciences, **6**:333-349.

Miller, S. A., Dykes, D. D. and Polesky, H. F. 1988. A Simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research, **16:**1215.

Mishra, C., Palai, T. K., Sarangi, L. N., Prusty, B. R. and Maharana, B. R. 2013. Candidate gene markers for sperm quality and fertility in bulls. Veterinary World, **6**:905-910.

Moaeen-ud-Din, M. 2014. Buffalo genome research - a review. Animal Science Papers and Reports, **32**:187-199.

Mukherjee, A., Dass, G., Mohana rao, G. J., Gohain, M., Brahma, B., Datta, T. K. and De, S. 2013. Absolute copy number differences of Y chromosomal genes between

crossbred (*Bos taurus* x *Bos indicus*) and Indicine bulls. Journal of Animal Science and Biotechnology, **4**:15.

Murphy, W. J., Pevzner, P. A. and O'Brien S. J. 2004. Mammalian phylogenomics comes of age. Trends in Genetics, **20**:631-639.

Nei, M. and Kumar, S. 2000. Molecular evolution and phylogenetics. Oxford university press.

Nilsson, B., Rosen, S. W., Weintraub, B. D. and Zopf, D. A. 1986. Differences in the carbohydrate moieties of the common alpha-subunits of human chorionic gonadotropin, luteinizing hormone, follicle-stimulating hormone and thyrotropin: preliminary structural inferences from direct methylation analysis. Endocrinology, **119**:2737-2743.

Odet, F., Gabel, S., London, R. E., Goldberg, E. and Eddy, E.M. 2013. Glycolysis and mitochondrial respiration in mouse LDHC-null sperm. Biology of Reproduction, **88**:1-7.

Othaman, O. E., Ahmed, W. M., Balabel, E. A., Zaabal, M. M., El Khadrawy H. H. and Hanafy, E. M. 2014. Genetic polymorphism of CYP19 Gene and its association with ovarian activity in Egyptian Buffaloes. Global Veterinaria, **12**:768-773.

Panigrahi, S. K. and Yadav, B. R. 2010. Polymorphism in *TNP-1* gene of Murrah buffalo bulls. African Journal of Biotechnology, **9**:7224-7229.

Parent, S., Lefievre, L., Brindle, Y. and Sullivan, R. 1999. Bull subfertility is associated with low levels of a sperm membrane antigen. Molecular Reproduction and Development, **52**:57-65.

Parma, P., Pujada, M.E., Feligini, M., Greppi, G. and Enne, G. 2004. Water Buffalo (*Bubalus bubalis*): Complete Nucleotide Mitochondrial Genome Sequence. DNA Sequence, **15**:369-373.

Pausch, H., Kolle, S., Wurmser, C., Schwarzenbacher, H. and Emmerling, R. 2014. A Nonsense Mutation in *TMEM95* Encoding a Nondescript Transmembrane Protein Causes Idiopathic Male Subfertility in Cattle. PLOS Genetics, **10**:e1004044.

Posada, D. and Crandall, K. A. 1998. MODEL TEST: Testing the model of DNA substitution. Bioinformatics, **14**:817-818.

Ramasamy, R., Ridgeway, A., Lipshultz, L. I. and Lamb, D. J. 2014. Integrative DNA

methylation and gene expression analysis identifies discoidin domain receptor 1

association with idiopathic nonobstructive azoospermia. Fertility and Sterility, **102**:968-973.

Rondena, M., Ceciliani, F., Comazzi, S., Pocacqua, V., Bazzocchi, C., Luvoni, C., Chigioni, S. and Paltrinieri, S. 2005. Identification of bovine doppel protein in testis, ovary and ejaculated spermatozoa. Theriogenology, **63**:1195-1206.

Roy, A., Lin, Y. N. and Matzuk, M. M. 2007. Genetics of Idiopathic Male Infertility:The Power of a Cross-Species Approach. The Genetics of Male Infertility, edited by D. T. Carrell, Chapter **6**:99-114.

Sang, L., Du, Q. Z., Yang, W. C., Tang, K. Q., Yu, J. N., Hua, G., Zhang, X. X. and Yang, L.G. 2011. Polymorphisms in follicle stimulation hormone receptor, inhibin alpha, inhibin bata A, and prolactin genes and their association with sperm quality in Chinese Holstein bulls. Animal Reproduction Science, **126**:151-156.

Sato, H., Taketomi, Y., Isogai, Y., Miki, Y., Yamamoto, K., Masuda, S., Hosono, T., Arata, S., Ishikawa, Y., Ishii, T., Kobayashi, T., Nakanishi, H., Ikeda, K., Taguchi, R., Hara, S., Kudo, I. and Murakami, M. 2010. Group III secreted phospholipase A2 regulates epididymal sperm maturation and fertility in mice. Journal of Clinical Investigation, **120**:1400-1414.

Sattler, C.G., 2013. Progeny testing and genomics: Where are we and where are we going? In: Proc. Dairy Cattle Reprod. Council Annu. Conf., Indianapolis, IN, pp 98– 103.

Saunders, P. T., Sharpe, R. M., Williams, K., Macpherson, S., Urquart, H., Irvine, D. S. and Millar, M. R. 2001. Differential expression of oestrogen receptor alpha and beta proteins in the testes and male reproductive system of human and non human primates. Molecular Human Reproduction, **7**:227-236.

Schambony, A., Hess, O., Gentzel, M. and Topfer-Petersen, E. 1998. Expression of CRISP proteins in the male equine genital tract. Journal of Reproduction and Fertility, **53**:67-72.

Schefers, J. M., and K. A. Weigel. 2012. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Anim. Frontiers **2**:4–9.

Srivastava, P. N., Arbtan, K., Takei, G. H., Huang, T. F. and Yanagimachi, R. 1986. α-L-fucosidase from bull seminal plasma: Its purification and acrosome reaction

promoting property. Biochemical and Biophysical Research Communications, **137**: 1061-1068.

Stormshak, F. and Bishop, C.V. 2008. Estrogen and progesterone signaling: Genomic and nongenomic actions in domestic ruminants. Journal of Animal Science, **86**:299-315.

Suri, A. 2004. Sperm specific proteins-potential candidate molecules for fertility control. Reproductive Biology and Endocrinology, **2**:10.

Sutovsky, P., Terada, Y. and Schatten, G. 2001. Ubiquitin-based sperm assay for the diagnosis of male factor infertility. Human Reproduction, **16**:250-258.

Tantia, M. S., Vijh, R. K., Bhasin, V., Poonam Sikka, Vij, P. K., Kataria, R. S., Mishra, B. P., Yadav, S. P., Pandey, A. K., Sethi, R. K., Joshi, B. K., Gupta, S. C. and Pathak, K. M. L. 2011. Whole-genome sequence assembly of the water buffalo (*Bubalus bubalis*). The Indian Journal of Animal Sciences, **81**:38-46.

Tiwari, A., Singh, D., Kumar, O. S. and Sharma, M.K. 2008. Expression of cytochrome P450 aromatase transcripts in buffalo (*Bubalus bubalis*)-ejaculated spermatozoa and its relationship with sperm motility. Domestic Animal Endocrinology, **34**:238-249.

Vogel, T., Dechend, F., Manz, E., Jung, C., Jakubiczka, S., Fehr, S., Schmidtke, J. and Schnieders, F. 1997. Organization and expression of bovine TSPY. Mammalian Genome, **8**:491-496.

Wang, H., Liu, J., Cho, K. H. and Ren, D. 2009. A Novel, Single, transmembrane protein CATSPERG is associated with CATSPER1 Channel Protein. Biology of Reproduction, **81**:539-544.

Wang, R. S., Yeh, S., Tzeng, C. R. and Chang, C. 2003. Androgen receptor roles in spermatogenesis and fertility: lessons from testicular cell specific androgen receptor knockout mice, **30**:119-132.

Wathes, D. C., Clempson, A. M. and Pollott, G. E. 2013. Associations between lipid metabolism and fertility in the dairy cow. Reproduction, Fertility and Development, **25**:48-61.

Wimmers, K., Lin, C. L., Tholen, E., Jennen, D. G., Schellander, K. and Ponsuksili, S. 2005. Polymorphisms in candidate genes as markers for sperm quality and boar fertility. Animal Genetics, **36:**152-155.

Winaya, A., Rahayu, I. D., Amin, M. and Herliantin. 2012. The genetic variation of Bali cattle (*Bos javanicus*) based on sex related Y-chromosome gene. Animal Production, **13**:50-55.

Wolf, E. and Bauersachs, S. 2010. Functional genome research in reproductive biology and biotechnology − a mini review. Animal Science Papers and Reports, **28**:123-131.

Xu, X., Toselli, P. A, Russell, L. D. and Seldin, D. C. 1999. Globozoospermia in mice lacking the casein kinase II alpha' catalytic subunit. Nature Genetics, **23**:118-21.

Yan, P., Xiang, L., Guo, X., Bao, P. J., Jin, S. and Wu, X. Y. 2014. The low expression of Dmrt7 is associated with spermatogenic arrest in cattle-yak. Molecular Biology Reports, **41**:7255-7263.

Yang, W. C., Tang, K. Q., Yu, J. N., Zhang, C. Y., Zhang, X. X. and Yang, L. G. 2011. Effects of MboII and BspMI polymorphisms in the gonadotropin releasing hormone receptor (GnRHR) gene on sperm quality in Holstein bulls. Molecular Biology Reports, **38**:3411-3415.

Yao, G., QiFa, L., ZengXiang, P., MingGui, L., Hua, L. and Zhuang, X. 2013. Molecular cloning, gene expression and methylation status analysis of PIWIL1 in cattle-yaks and the parental generation. Animal Reproduction Science, **140**:131-137.

Yue, X. P., Dechow, C., Chang, T. C., DeJarnette, J. M., Marshall, C. E., Lei, C. Z. and Liu, W. S. 2014. Copy number variations of the extensively amplified Y-linked genes, HSFY and ZNF280BY, in cattle and their association with male reproductive traits in Holstein bulls. BMC Genomics, **15**:113.

Yuzawa, T., Fukui, E., Matsumoto, H., Kawada, T., Kawanobe, A., Niira, K., Ochi, M. and Yoshizawa, M. 2011. Tumor protein p53 gene polymorphism and developmental competence of embryos derived from *in vitro* fertilization in old-aged and long-term infertile Japanese black cows. Journal of Mammalian Ova Research, **28**:75-78.

Zhang, B., Penagaricano, F., Chen, H. and Khatib, H. 2012. Novel transcripts and alternatively spliced genes are associated with early development in bovine embryos. Animal, **6**:1199-1205.

Zhang, G. W., Guan, J. Q., Luo, Z. G., Zhang, W. X., Wang, L., Luo, X. L. and Zuo, F. Y. 2016. A tremendous expansion of TSPY copy number in crossbred bulls (*Bos taurus* x *Bos grunniens*). Journal of Animal Science, **94**:1398-1407.

Zhang, Q., Li, J., Li, Q., Li, X., Liu, Z., Song, D. and Xie, Z. 2009. Cloning and characterization of the gene encoding the bovine BOULE protein. Molecular Genetics and Genomics, **281**:67-75.

Zhang, S., Cai, H., Yang, Q., Shi, T., Pan, C., Lei, C., Dang, R., Chen, H. and Lan, X. 2015. Identification of novel alternative splicing transcript and expression analysis of bovine *TMEM95* gene. Gene, Available online 16 September 2015, ISSN 0378-1119.

Zimin, A. M. G, Ferre, F., Biagini, T. and Shroeder, S. 2013. The Buffalo Reference Genome Assembly. In Plant and Animal Genome XXI conference, Plant and Animal Genome.

Appendix I

DNA sequences used for Primer designing

(A) Bos taurus breed Hereford chromosome 19, Bos_taurus_UMD_3.1.1, whole genome shotgun sequence

NCBI Reference Sequence: AC_000176.1
>gi|258513348:27687749-27690379 Bos taurus breed Hereford chromosome 19, Bos_taurus_UMD_3.1.1, whole genome shotgun sequence
AAGGGGTGTGTGTGCACATGTGCTTCACACATCCACGTAGGCAGGAGGGGCCTGCTGA
GCTTTGAGTCACTAGGATCTTCTAATGAGAGGTAAGAGAAGTCAACTGGGCTTAGCTA
GGCCTTGGAGGCCTGTATAGAACACTCTGTTGCTAAGGCAACCATGAGCCTGTTGCTA
GGAGATAGCTGGGGAAGGCCCAAGGCTGCTCAGGGCGAAGAAGAGATGAACAGAAG
AGTTTGAAAGAGTGGGGGAGGATGTGGACAGGCATGGGTTTCTGAGTCCCTGGGGGT
ACGGAATACTCACAAACAGCTTCTTCACTGGGGAGTCACTTACTATTTATCACTGGTCA
TGATTTACAGGAAGAAAAATAAAATTGCTTTTGGAATCACCAACTCATGAAAGCTGCT
GAGGGCTAGGTTGGGGGTATAGCAGAAGCCCTGCTTTCTCTATCCATCTCCCACTAGA
AGAATCTTCTTCCAGTCCTCAATTCACTGGGCCCCAAGTTTACCCCAGCCAGAGATGG
GGCAGTGTCGCTGTGCTAGAATGCTCTAGAACCTCCAGACAAAAGGCTGGAACCAGG
ATCACAGAGCATTCCAATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAG
CCCCCAGCAGGCCTCATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTG
AGGCCTGTGTCTTCTGCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTT
TGCAGCCAGATGGAGGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCT
TTGCCTTAGGTAAAATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCC
CCAAGGGGCCTGCAAGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACA
GCCCTCTCTGGACACACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAG
GGCCCAGAACCTTCTTCCAGGTCAGTGAGGGAATCACCTGGTCCTTACCCATCTCCAC
AGATGATGCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCAT
GGGTGAAGTCAAGGGGGAAGGCACAACCTAGGGGCTTGGAAGGCACAACCAAGGAG
GGGGCTGGGGAGGTGGAGTAGGGACGGGTGGAGGGCAGATGGAGACAGGAGAGTGT
GATGGGGTGGTGCTGGGTGAAGGGGTGGAGGTGAACCATCCCTGCTGAGGCCCTGGG
CACTCTAGCTGCTTTCTGCCTCCCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCA
TATTGGCAATGGCTTCGAAAGACCAAGCTCCGGGAGTACAACAGAGAAGGTACGGAT
GAGGGGAAGGACCAAGGGGTGCCTAGGTCCTTGGGCAAAGGGGTGTATGGCTGAGCT
CATCCGCCCCCCACCCTGCCTTGTCTTTCATTTCAGCTCTCTGCCCTCCCTCCTGCCGTG
AGTAAGAAAGGGAAGGGGTAGAGAAAGCCGGGGGCTTGGGGCACTTGAGGAGAGCA
GCCACGGAGGGCGGTGGGAGCAGAGAGGAACGGAGGCTTCCCGCCGCACTGCCGGCC
CTACTGTCTCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCT
TCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGTCCTTACTCCCCACCTCTG
CCTCACCCCCACCCAGATCTCTGAGCTCTGAGGCTTCCCCAGCATCGATGGGTATGGTC
CCCGGGATCACTCACTGCTTCCTGGGTCCCCTATAGGAAGTCACGATCTTTGGGAAGC
TCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCG
CGGTGGAGTGAGTTGGGGGATCAGGGCGGGAGACAGCCTCACCACTTCCCACCAGTA
CCCCCTCATCCTTCCTCCTGCCCTCGTGGCCCTTCCCTCCTCCCTTCCCAGGCCTCCCAC
CCTCTGTTTTCCCTCAGGTACAACTACTTCCAAGCAAAAAGTTGACTTGTGAAAACGA
GGACAGAGCCTAGCCTCCAGTTCTAAGGAGCTCGTGTGCCCACAACATCCACTTCTCT
CTTGGGGTTGGGGGTGGGGTGGGGAGGGCGGGGGACAGCCAGACCCTTAGTTCCAGC
CCCAAGTGGTTCAATCGTCCAGACCCTGATACTTAGATATACCCCATAGTCCCAGTGTC
AGATGGTCCAGGGACCCAGGCACCCACAGCTGGAAAGCTCCTCCCCTTCATCCCTCAA
CCAATCAGACAGGGGCAGTCCCCGTGCCAGGAAAATATCTACAGAAGGAAAGAATCC
CCTCTGCTCACCACCACTCCCACATCTCCTTCTGCCTGTTCCCGGAACGTGGGGGCTGC
CAGGCCCCAGAAACTCCTCTTGGCCCTCCTGTGACTGATTGGGAATCCGGGAATGAGT

GTTCTGGAAAAGTCCCTCCCCCTAGAATGAGACCACAGCAGTCAGCCTGCTGGCATGC
TCCCCTTCCATTTAACATTGTGAAGCCTCCCACAGAGAGAGGCCTCCCTCCTGCCAGGC
CCAACTAAACGTGCTGTCAAATGGC

(B) Bos indicus isolate QUIL7308 breed Nelore chromosome 19, whole genome shotgun sequence

GenBank: CM003039.1

>gb|CM003039.1|:27432672-27435302 Bos indicus isolate QUIL7308 breed Nelore chromosome 19, whole genome shotgun sequence

AAGGGGTGTGTGTGCACATGTGCTTCACACATCCACGTAGGCAGGAGGGGCCTGCTGA
GCTTTGAGTCACTAGGATCTTCTAATGAGAGGTAAGAGAAGTCAACTGGGCTTAGCTA
GGCCTTGGAGGCCTGTATAGAACACTCTGTTGCTAAGGCAACCATGAGCCTGTTGCTA
GGAGATAGCTGGGGAAGGCCCAAGGCTGCTCAGGGCGAAGAAGAGATGAACAGAAG
AGTTTGAAAGAGTGGGGGAGGATGTGGACAGGCATGGGTTTCTGAGTCCCTGGGGGT
ACGGAATACTCACAAACAGCTTCTTCACTGGGGAGTCACTTACTATTTATCACTGGTCA
TGATTTACAGGAAGAAAAATAAAATTGCTTTTGGAATCACCAACTCATGAAAGCTGCT
GAGGGCTAGGTTGGGGGTATAGCAGAAGCCCTGCTTTCTCTATCCATCTCCCACTAGA
AGAATCTTCTTCCAGTCCTCAATTCACTGGGCCCCAAGTTTACCCCAGCCAGAGATGG
GGCAGTGTCGCTGTGCTAGAATGCTCTAGAACCTCCAGACARAAGGCTGGAACCAGG
ATCACAGAGCATTCCAATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAG
CCCCCAGCAGGCCTCATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTG
AGGCCTGTGTCTTCTGCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTT
TGCAGCCAGATGGAGGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCT
TTGCCTTAGGTAAAATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCC
CCAAGGGGCCTGCAAGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACA
GCCCTCTCTGGACACACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAG
GGCCCAGAACCTTCTTCCAGGTCAGTGAGGGAATCACCTGGTCCTTACCCATCTCCAC
AGATGATGCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCAY
GGGTGAAGTCAAGGGGGAAGGCACAACCTAGGGGCTTGGAAGGCACAACCAAGGAG
GGGGCTGGGGAGGTGGAGTAGGGACGGGTGGAGGGCAGATGGAGACAGGAGAGTGT
GATGGGGTGGTGCTGGGTGAAGGGGTGGAGGTGAACCATCCCTGCTGAGGCCCTGGG
CACTCTAGCTGCTTTCTGCCTCCCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCA
TATTGGCAATGGCTTCGAAAGACCAAGCTCCGGGAGTACAACAGAGAAGGTACGGAT
GAGGGGAAGGACCAAGGGGTGCCTAGGTCCTTGGGCAAAGGGGTGTATGGCTGAGCT
CATCCGCCCCCCACCCTGCCTTGTCTTTCATTTCAGCTCTCTGCCCTCCCTCCTGCCGTG
AGTAAGAAAGGGAAGGGGTAGMGAAAGCCGGGGGCTTGGGGCACTTGAGGAGAGCA
GCCACGGAGGGCGGTGGGAGCAGAGAGGAACGGAGGCTTCCCGCCGCACTGCCGGCC
CTACTGTCTCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCT
TCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGTCCTTACTCCCCACCTCTG
CCTCACCCCCACCCAGATCTCTGAGCTCTGAGGCTTCCCCAGCATCGATGGGTATGGTC
CCCGGGATCACTCACTGCTTCCTGGGTCCCCTATAGGAAGTCACGATCTTTGGGAAGC
TCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCG
CGGTGGAGTGAGTTGGGGGATCAGGGCGGGAGACAGCCTCACCACTTCCCACCAGTA
CCCCCTCATCCTTCCTCCTGCCCTCGTGGCCCTTCCCTCCTCCCTTCCCAGGCCTCCCAC
CCTCTGTTTTCCCTCAGGTACAACTACTTCCAAGCAAAAAGTTGACTTGTGAAAACGA
GGACAGAGCCTAGCCTCCAGTTCTAAGGAGCTCGTGTGCCCACAACATCCACTTCTCT
CTTGGGGTTGGGGGTGGGGTGGGGAGGGCGGGGGACAGCCAGACCCTTAGKTCCAGC
CCCAAGTGGTTCAATCGTCCAGACCCTGATACTTAGATATACCCCATAGTCCCAGTGTC
AGATGGTCCAGGGACCCAGGCACCCACAGCTGGAAAGCTCCTCCCCTTCATCCCTCAA
CCAATCAGACAGGGGCAGTCCCCGTGCCAGGAAAATATCTACAGAAGGAAAGAATCC
CCTCTGCTCACCACCACTCCCACATCTCCTTCTGCCTGTTCCCGGAACGTGGGGGCTGC
CAGGCCCCAGAAACTCCTCTTGGCCCTCCTGTGACTGATTGGGAATCCGGGAATGAGT
GTTCTGGAAAAGTCCCTCCCCCTAGAATGAGACCACAGCAGTCAGCCTGCTGGCATGC
TCCCCTTCCATTTAACATTGTGAAGCCTCCCACAGAGAGAGGCCTCCCTCCTGCCAGGC

CCAACTAAACGTGCTGTCAAATGGC

(C) Bubalus bubalis breed Mediterranean contig488833, whole genome shotgun sequence

GenBank: AWWX01488834.1

>gb|AWWX01488834.1|:19601-22209 Bubalus bubalis breed Mediterranean contig488833, whole genome shotgun sequence

AAGGGGTGTGTGTGCACATGTGCTTCACACATCCACGTAGGCAGGAGGGGCCTGCTGA
GCTTGAGTCACTAGGATCTTCTAATGAGAGGTAAGAGAAGTCAACTGGGCTTAGCTGG
GCCTTGGAGGCCTGTATAGAACTCTCTGTTGCTAAGGCAACCATGAGCCTGTTGCTAG
GAGATAGCTGGGAAAGGCCCAAGGCTGCTCAGGGCGAAGAAGAGATGAACAGAAGA
GTTTGAAAGAGTGGGGGAGGATGTGGACAGGCATGGGTTTCTGAGTCCCTGGGGGTA
CGGAATACTCACAAACAGCTTCTTCACTGGGGAGTCACTTACTGTTTATCACTGGTCAT
GATTTACAGGAAGAAAAATAAAATTGCTTTCGGAATCACCAACTCATGAAAGCCGCTG
AGGGCTAGGTTGGGGGTATAGCAGAAGCCCTGCTTTCTCTATCCATCTCCCACTAGAA
GAATCTTCTTCCAGTCCTCAATTCACTGGGCCCCAAGTTTACCCCAGCCAGAGATGGG
GCAGTGTCACTGTGCTAGAATGCTCTAGAACCTCCAGACAAAAGGCTGGAACCAGGAT
CACAGAGCATTCCAATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAGCC
CCCAGCAGGCCTCATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAG
GCCTGTGTCTTCTGCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCTCAGCTTTG
CAGCCAGATGGAGGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTT
GCCTTAGGTAAAATCAGACATCCAGGGATGGGCCTAGCAACAATCACCCCTGTGCCCC
AAGGGGCCTGCAAGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGC
CCTCTCTGTGGAGCCCTCTCTCCCTCTGGGAGAGCTTCTTTGGACCCCGGGAAAGGGCC
CAGAACCTTTTTCCAGGTCAGTGAGGGAATCACCCGGTCCTTACCCATCTTCACAGAT
GATGCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGGT
GAAGTCAAGGGGGAAGGCACAACCTAGGGGCTTGGAAGGCACAACCAAGGAGGGGG
CTGGGGAGGTGGAGTAGGGACGGGTGGGGGGCAGATGGAGACAGGAGAGTGTGATG
GGGTGGTGCTGGGCGAAGGGGTGGAGGCCCTGGGCACTCTAGCTGCTTTCTGCCTCCC
CAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGAC
CAAGCTCCGGGAGTACAACAGAGAAGGTACGGATGAGGGGAAGGACCAAGGGGTGC
CTAGGTCCTTGGGCAAAGGGGTGTATGGCTGAGCTCATCCGCCACCCACCCTGCCTTG
TCTCTCATTTCAGCTCTCTGCCCTCCCGCCTGCCGTGAGTAAGAAAGGGAAGGGGTAG
CGAGAGCCGGGGGCTTGGGGCACTTGAGGAGAGCAGCCACGGAGGGCGGTGGGAGC
AGAGAGGAACGGAGGCTTCCCGCCGCACTGCCGGCCCTACTGTATCCTGCAGGGGGC
AGCACCATCCTGTACAACTGTTCCACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCC
GAAAGCGCTGCTTCCCAGGTCCTTACTCCCCACCTCTGCCTCACCCCCACCCAGATCTC
TGAGCTCTGAGGCTTCCCCAGCATCGATGGGTATGGTCCCCGGGATCACTCACTGCCT
CCTGGGTCCCTTATAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGT
GTGCGGAACTGCCCTGCTCCTGGGTGTTCCGAGCCTCGCGGTGGAGTGAGTTGGGGGA
TCAGGGCGGGAGACAGCCTCATCACTTCCCACCAGTACCCCCTCATCCTTCCTCCTGCC
CTCGCGGCCCTTTCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTTCCCTCAGGTACA
ACTACTTCCAAGCAAAAAGTTGACTTGTGAAAACGAGAACAGAGCCTAGCCTCCAGTT
CTAAGGAGCTCGTGTGCTCACAACATCCACTTCTCTCTTGGGGTTGGGGGTGGGGTGG
GGAGGGCGGGGGACAGCCCAACCCTTAGTTCCAGCCCCAAGTGGTTCAATCTTCCAGA
CCCTGATACTCAGACATACCCCATAGTCCCAGTGTCAGATGGTCCAGGGACCCAGGCA
CCCACAGCTGGAAAGCTCCTCCCCTTCATCCCTCAACCAATCAGACGGGGGCAGTCCC
CGTGCCAGGAAAATATCTACAGAAGGAAAGAATCCCCTCTGCTCACCACCACTCCCAC
ATCTCCTTCTGCCTGTTCCCGGAACGTGGGGGCTGCCAGGCCCCAGAAACTCCTCTTGG
CCCTCCTGTGACTGATTGGGAATCCGGGAATGAGTGTTCTGGAAAAGTCCCTCCCCCT
AGAATGAGACCACAGCAGTCAGCCTGCTGGCATGCTCCCCTTCCATTTAACATTGTGA
AGCCTCCCACAGAGAGAGGCCTCCCTCCTGCCAGGCCCAACTAAACGTGCTGTCAAAT
GGC

(D) Bos mutus isolate yakQH1 unplaced genomic scaffold, BosGru_v2.0 scaffold348_1, whole genome shotgun sequence

NCBI Reference Sequence: NW_005393148.1

>gi|550711427:119575-121770 Bos mutus isolate yakQH1 unplaced genomic scaffold, BosGru_v2.0 scaffold348_1, whole genome shotgun sequence

TGGACCTACGGGTCTGGCTGTCCCCGCCCTCCCCCCCCCACCCCCAACCCCAAGAGA
GAAGTGGATGTTGTGGGCACACGAGCTCCTTAGAACTGGAGGCTAGGCTCTGTCCTCG
TTTTCACAAGTCAACTTTTTGCTTGGAAGTAGTTGTACCTGAGGGAAAACAGAGGGTG
GGAGGCCTGGGAAGGGAGGAGGGAAGGGCCACGAGGGCAGGAGGAAGGATGAGGGG
GTACTCGTGGGAAGTGGTGAGGCTGTCTCCCGCCCTGATCCCCCAACTCACTCCACCG
CGAGGCTCGGGACACCCAGGAGCAGGGCAGTTCCGCACACAAAGAGGAGCAGGATCC
GAGCTTCCCAAAGATCGTGACTTCCTATAGGGGACCCAGGAAGCAGTGAGTGATCCCG
GGGACCATACCCATCGATGCTGGGGAAGCCTCAGAGCTCAGAGATCTGGGTGGGGGT
GAGGCAGAGGTGGGGAGTAAGGACCTGGGAAGCAGCGCTTTCGGGGCCAACAGTACA
CCTCGAAGCCCTGGCAGGTTGAACAGTTGTACAGGATGGTGCTGCCCCCTGCAGGAGA
CAGTAGGGCCGGCAGTGCGGCGGGAAGCCTCCGTTCCTCTCTGCTCCCACCGCCCTCC
GTGGCTGCTCTCCTCAAGTGCCCCAAGCCCCCGGCTTTCGCTACCCCTTCCCTTTCTTA
CTCACGGCAGGAGGGAGGGCAGAGAGCTGAAATGAAAGACAAGGCAGGGTGGGGGG
CGGATGAGCTCAGCCATACACCCCTTTGCCCAAGGACCTAGGCACCCCTTGGTCCTTC
CCCTCATCCGTACCTTCTCTGTTGTACTCCTGGAGCTTGGTCTTTCGAAGCCATTGCCA
ATATGAAGGGAGTGAGTAGAGAGACCCTTTGATCTCTGGGGAGGCAGAAAGCAGCTA
GAGTGCCCAGGGCCTCAGCAGGGATGGTTCACCTCCACCCCTTCACCCAGCACCACCC
CATCACACTCTCCTGTCTCCATCTGCCCTCCACCCGTCCCTACTCCACCTCCCCAGCCC
CCTCCTTGGTTGTGCCTTCCAAGCCCCTAGGTTGTGCCTTCCCCCTTGACTTCACCCATG
ACTCTCAGGACTCTGTGAGTCTTCTCTGTGATTTTGTTCAAGGATGCATCATCTGTGGA
GATGGGTAAGGACCGGGTGATTCCCTCACTGACCTGGAAGAAGGTTCTGGGCCCTTTC
CCAGGGTCCAAAGAAGCTCTCCCAGAGGGAGCAGAAAGTGTGTCCAGAGAGGGCTGT
GGACTGAGGAGATGGCTGAGAGGCAGAGCTGTGAGGCCACTTGCAGGCCCCTTGGGG
CACAGGGGTGATTGTTGTTAGGCCCATCCCTGGATGTCTGATTTTACCTAAGGCAAAG
GCCGAGAATGTCCAGGAGACCTCACAGTCCTTCCACTGGACCTCCATCTGGCTGCAAA
GCCGGGCCAGGCGGCCCGACAACTCACGGTCTGGGAAGCGGCAGAAGACACAGGCCT
CAACGGCTGCCAGGAAGATCCCACCTAGTGCCAGTGTCCACATGAGGCCTGCTGGGGG
CTGCTGGAGGGAGGGCTGGCCCCACCTAGGCTGGGAAGGCATTGGAATGCTCTGTGAT
CCTGGTTCCAGCCTTTTGTCTGGAGGTTCTAGAGCATTCTAGCACAGCGACACTGCCCC
ATCTCTGGCTGGGGTAAACTTGGGGCCCAGTGAATTGAGGACTGGAAGAAGATTCTTC
TAGTGGGAGATGGATAGAGAAAGCAGGGCTTCTGCTATACCCCCAACCTAGCCCTCAG
CAGCTTTCATGAGTTGGTGATTCCAAAAGCAATTTTATTTTTCTTCCTGTAAATCATGA
CCAGTGATAAATAGTAAGTGACTCCCCAGTGAAGAAGCTGTTTGTGAGTATTCCGTAC
CCCCAGGGACTCAGAAACCCATGCCTGTCCACATCCTCCCCCACTCTTTCAAACTCTTC
TGTTCATCTCTTCTTCGCCCTGAGCAGCCTTGGGCCTTCCCCAGCTATCTCCTAGCAAC
AGGCTCATGGTTGCCTTAGCAACAGAGTGTTCTATACAGGCCTCCAAGGCCTAGCTGA
GCCCAGTTGACTTCTCTTACCTCTCATTAGAAGATCCTAGTGACTCAAAGCTCAGCAGG
CCCCTCCTGCCTACGTGGATGTGTGAAGCACATGTGCACACACACCCCTT

APPENDIX II

Genewise predictions of different isoforms in cattle, buffalo and yak

Isoform1
>Btaurus_isoform1
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLREYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLP
RKSRSLGSSDPAPLCVRNCPAPGCPEPRGGVSWGIRAGDSLTTSHQYPLILPPALVALPSSL
PRPPTLCFPSGTTTSKQKVDL
>Btaurus_Cds_1
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGC
AGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCC
GAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTT
GTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTGAGTTGGGG
GATCAGGGCGGGAGACAGCCTCACCACTTCCCACCAGTACCCCCTCATCCTTCCTCCT
GCCCTCGTGGCCCTTCCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTTCCCTCAGGT
ACAACTACTTCCAAGCAAAAAGTTGACTTG
>Ongole_Isoform1
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLREYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLP
RKSRSLGSSDPAPLCVRNCPAPGCPEPRGGVSWGIRAGDSLTTSHQYPLILPPALVALPSSL
PRPPTLCFPSGTTTSKQKVDL
>Ongole_Cds_1
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGC
AGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCC
GAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTT
GTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTGAGTTGGGG
GATCAGGGCGGGAGACAGCCTCACCACTTCCCACCAGTACCCCCTCATCCTTCCTCCT
GCCCTCGTGGCCCTTCCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTTCCCTCAGGT
ACAACTACTTCCAAGCAAAAAGTTGACTTG
>Bmutus_isoform1
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLQEYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLP
RKSRSLGSSDPAPLCVRNCPAPGCPEPRGGVSWGIRAGDSLTTSHEYPLILPPALVALPSSLP
RPPTLCFPSGTTTSKQKVDL
>Bmutus_Cds_1

ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCAGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGC
AGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCC
GAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTT
GTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTGAGTTGGGG
GATCAGGGCGGGAGACAGCCTCACCACTTCCCACGAGTACCCCCTCATCCTTCCTCCT
GCCCTCGTGGCCCTTCCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTCCCTCAGGT
ACAACTACTTCCAAGCAAAAAGTTGACTTG

Isoform2

>Btaurus_isoform2
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDA
SLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPSCREGQHHPVQL
FNLPGLRGVLLAPKALLPRKSRSLGSSDPAPLCVRNCPAPGCPEPRGGVSWGIRAGDSLTT
SHQYPLILPPALVALPSSLPRPPTLCFPSGTTTSKQKVDL

>Btaurus_Cds_2
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
GGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGCAGCACCATCC
TGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTG
CTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAA
CTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTGAGTTGGGGGATCAGGGCG
GGAGACAGCCTCACCACTTCCCACCAGTACCCCCTCATCCTTCCTCCTGCCCTCGTGGC
CCTTCCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTTCCCTCAGGTACAACTACTTC
CAAGCAAAAAGTTGACTTG

>Ongole_isoform2
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDA
SLNKITEKTHRVLRVxEIKGSLYSLPSYWQWLRKTKLREYNREALCPPSCREGQHHPVQLF
NLPGLRGVLLAPKALLPRKSRSLGSSDPAPLCVRNCPAPGCPEPRGGVSWGIRAGDSLTTS
HQYPLILPPALVALPSSLPRPPTLCFPSGTTTSKQKVDL

>Ongole_Cds_2
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCANGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
GGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGCAGCACCATCC
TGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTG
CTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAA
CTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTGAGTTGGGGGATCAGGGCG
GGAGACAGCCTCACCACTTCCCACCAGTACCCCCTCATCCTTCCTCCTGCCCTCGTGGC
CCTTCCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTTCCCTCAGGTACAACTACTTC
CAAGCAAAAAGTTGACTTG

>Meditarian_isoform2
MWTLALGGIFLAAVEACVFCRFPDRELSGRLAQLCSQMEVQWKDCEVSWTFSAFALDDA

SLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPACREGQHHPVQL
FHLPGLRGVLLAPKALLPRKSRSLGSSDPAPLCVRNCPAPGCSEPRGGVSWGIRAGDSLITS
HQYPLILPPALAALSSSLPRPPTLCFPSGTTTSKQKVDL
>Meditarian_cds_2
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCTCAGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
GGGAGTACAACAGAGAAGCTCTCTGCCCTCCCGCCTGCCGTGAGGGGCAGCACCATCC
TGTACAACTGTTCCACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTG
CTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAA
CTGCCCTGCTCCTGGGTGTTCCGAGCCTCGCGGTGGAGTGAGTTGGGGGATCAGGGCG
GGAGACAGCCTCATCACTTCCCACCAGTACCCCCTCATCCTTCCTCCTGCCCTCGCGGC
CCTTTCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTCCCTCAGGTACAACTACTTC
CAAGCAAAAAGTTGACTTG
>Bmutus_isoform2
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDA
SLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLQEYNREALCPPSCREGQHHPVQL
FNLPGLRGVLLAPKALLPRKSRSLGSSDPAPLCVRNCPAPGCPEPRGGVSWGIRAGDSLTT
SHEYPLILPPALVALPSSLPRPPTLCFPSGTTTSKQKVDL
>Bmutus_Cds_2
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
AGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGCAGCACCATCC
TGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTG
CTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAA
CTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTGAGTTGGGGGATCAGGGCG
GGAGACAGCCTCACCACTTCCCACGAGTACCCCCTCATCCTTCCTCCTGCCCTCGTGGC
CCTTCCCTCCTCCCTTCCCAGGCCTCCCACCCTCTGTTTCCCTCAGGTACAACTACTTC
CAAGCAAAAAGTTGACTTG
Isoform-3
>Btaurus_isoform3
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLREYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLP
RSLLPTSASPPPRSLSSEASPASMGMVPGITHCFLGPL
>Btaurus_Cds_3
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGC
AGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCC
GAAAGCGCTGCTTCCCAGGTCCTTACTCCCCACCTCTGCCTCACCCCCACCCAGATCTC
TGAGCTCTGAGGCTTCCCCAGCATCGATGGGTATGGTCCCCGGGATCACTCACTGCTTC
CTGGGTCCCCTA

>Ongole_isoform3
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLREYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLP
RSLLPTSASPPPRSLSSEASPASMGMVPGITHCFLGPL
>Ongole_Cds_3
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGC
AGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCC
GAAAGCGCTGCTTCCCAGGTCCTTACTCCCCACCTCTGCCTCACCCCCACCCAGATCTC
TGAGCTCTGAGGCTTCCCCAGCATCGATGGGTATGGTCCCCGGGATCACTCACTGCTTC
CTGGGTCCCCTA
>Bmutus_isoform3
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLQEYNREALCPPSCREGQHHPVQLFNLPGLRGVLLAPKALLP
RSLLPTSASPPPRSLSSEASPASMGMVPGITHCFLGPL
>Bmutus_Cds_3
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCAGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGTGAGGGGC
AGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCC
GAAAGCGCTGCTTCCCAGGTCCTTACTCCCCACCTCTGCCTCACCCCCACCCAGATCTC
TGAGCTCTGAGGCTTCCCCAGCATCGATGGGTATGGTCCCCGGGATCACTCACTGCTTC
CTGGGTCCCCTA

Isoform4

>Btaurus_isoform4
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLREYNREALCPPSCRGSTILYNCSTCQGFEVYCWPRKRCFPG
SHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQAKS
>Btaurus_Cds_4
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCA
CCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAA
GCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGT

GCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTACAACTACTTCCA
AGCAAAAAGT
>Ongole_isoform4
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLREYNREALCPPSCRGSTILYNCSTCQGFEVYCWPRKRCFPG
SHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQAKS
>Ongole_cds_4
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCGGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCA
CCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAA
GCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGT
GCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTACAACTACTTCCA
AGCAAAAAGT
>Bmutus_isoform4
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALGKI
RHPGMGLTTITPVPQGACKWPHSSASQPSPQSTALSGHTFCSLWESFFGPWERAQNLLPEI
KGSLYSLPSYWQWLRKTKLQEYNREALCPPSCRGSTILYNCSTCQGFEVYCWPRKRCFPG
SHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQAKS
>Bmutus_Cds_4
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGGTAAA
ATCAGACATCCAGGGATGGGCCTAACAACAATCACCCCTGTGCCCCAAGGGGCCTGCA
AGTGGCCTCACAGCTCTGCCTCTCAGCCATCTCCTCAGTCCACAGCCCTCTCTGGACAC
ACTTTCTGCTCCCTCTGGGAGAGCTTCTTTGGACCCTGGGAAAGGGCCCAGAACCTTCT
TCCAGAGATCAAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAG
ACCAAGCTCCAGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCA
CCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAA
GCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGT
GCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAGTACAACTACTTCCA
AGCAAAAAGT

Isoform-5

>Btaurus_isoform5
MPSQPRWGQPSLQQPPAGLMWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEV
QWKDCEVSWTFSAFALDDASLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREY
NREALCPPSCRGSTILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGV
PSLAVE
>Btaurus_Cds_5
ATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAGCCCCCAGCAGGCCTCA
TGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGC
CGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGG
TCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGC
ATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATCAA
AGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCCGG
GAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCACCATCCTGTACA

ACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCC
AGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCC
CTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAG
>Ongole_isoform5
MPSQPRWGQPSLQQPPAGLMWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEV
QWKDCEVSWTFSAFALDDASLNKITEKTHRVLRVxEIKGSLYSLPSYWQWLRKTKLREYN
REALCPPSCRGSTILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVP
SLAVE
>Ongole_Cds_5
ATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAGCCCCCAGCAGGCCTCA
TGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGC
CGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGG
TCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGC
ATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCANGGAGATCAA
AGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCCGG
GAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCACCATCCTGTACA
ACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCC
AGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCC
CTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAG
>Meditarian_isoform5
MPSQPRWGQPSLQQPPAGLMWTLALGGIFLAAVEACVFCRFPDRELSGRLAQLCSQMEV
QWKDCEVSWTFSAFALDDASLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREY
NREALCPPACRGSTILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGV
PSLAVE
>Meditarian_Cds_5
ATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAGCCCCCAGCAGGCCTCA
TGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGC
CGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCTCAGCTTTGCAGCCAGATGGAGG
TCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGC
ATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATCAA
AGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCCGG
GAGTACAACAGAGAAGCTCTCTGCCCTCCCGCCTGCCGGGGCAGCACCATCCTGTACA
ACTGTTCCACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCC
AGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCC
CTGCTCCTGGGTGTTCCGAGCCTCGCGGTGGAG
>Bmutus_isoform5
MPSQPRWGQPSLQQPPAGLMWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEV
QWKDCEVSWTFSAFALDDASLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLQEY
NREALCPPSCRGSTILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGV
PSLAVE
>Bmutus_Cds_5
ATGCCTTCCCAGCCTAGGTGGGGCCAGCCCTCCCTCCAGCAGCCCCCAGCAGGCCTCA
TGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCTGC
CGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGAGG
TCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGATGC
ATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATCAA
AGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCCAG
GAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCGGGGCAGCACCATCCTGTACA
ACTGTTCAACCTGCCAGGGCTTCGAGGTGTACTGTTGGCCCCGAAAGCGCTGCTTCCC
AGGAAGTCACGATCTTTGGGAAGCTCGGATCCTGCTCCTCTTTGTGTGCGGAACTGCC
CTGCTCCTGGGTGTCCCGAGCCTCGCGGTGGAG
Isoform-6
>Btaurus_isoform6

MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDA
SLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPSCPLPALLSPAGG
STILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQA
KS
>Btaurus_Cds_6
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
GGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCCACTGCCGGCCCTACTGTC
TCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTG
TACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGA
TCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTG
GAGTACAACTACTTCCAAGCAAAAAGT
>Ongole_isoform6
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDA
SLNKITEKTHRVLRVxEIKGSLYSLPSYWQWLRKTKLREYNREALCPPSCPLPALLSPAGGS
TILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQA
KS
>Ongole_Cds_6
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCANGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
GGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCCACTGCCGGCCCTACTGTC
TCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTG
TACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGA
TCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTG
GAGTACAACTACTTCCAAGCAAAAAGT
>Meditarian_isoform6
MWTLALGGIFLAAVEACVFCRFPDRELSGRLAQLCSQMEVQWKDCEVSWTFSAFALDDA
SLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLREYNREALCPPACPLPALLYPAG
GSTILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQ
AKS
>Meditarian_Cds_6
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCTCAGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
GGGAGTACAACAGAGAAGCTCTCTGCCCTCCCGCCTGCCCACTGCCGGCCCTACTGTA
TCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCCACCTGCCAGGGCTTCGAGGTG
TACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGA
TCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTTCCGAGCCTCGCGGTG
GAGTACAACTACTTCCAAGCAAAAAGT
>Bmutus_isoform6
MWTLALGGIFLAAVEACVFCRFPDRELSGRLARLCSQMEVQWKDCEVSWTFSAFALDDA
SLNKITEKTHRVLRVMEIKGSLYSLPSYWQWLRKTKLQEYNREALCPPSCPLPALLSPAGG
STILYNCSTCQGFEVYCWPRKRCFPGSHDLWEARILLLFVCGTALLLGVPSLAVEYNYFQA
KS

>Bmutus_Cds_6
ATGTGGACACTGGCACTAGGTGGGATCTTCCTGGCAGCCGTTGAGGCCTGTGTCTTCT
GCCGCTTCCCAGACCGTGAGTTGTCGGGCCGCCTGGCCCGGCTTTGCAGCCAGATGGA
GGTCCAGTGGAAGGACTGTGAGGTCTCCTGGACATTCTCGGCCTTTGCCTTAGATGAT
GCATCCTTGAACAAAATCACAGAGAAGACTCACAGAGTCCTGAGAGTCATGGAGATC
AAAGGGTCTCTCTACTCACTCCCTTCATATTGGCAATGGCTTCGAAAGACCAAGCTCC
AGGAGTACAACAGAGAAGCTCTCTGCCCTCCCTCCTGCCCACTGCCGGCCCTACTGTC
TCCTGCAGGGGGCAGCACCATCCTGTACAACTGTTCAACCTGCCAGGGCTTCGAGGTG
TACTGTTGGCCCCGAAAGCGCTGCTTCCCAGGAAGTCACGATCTTTGGGAAGCTCGGA
TCCTGCTCCTCTTTGTGTGCGGAACTGCCCTGCTCCTGGGTGTCCCGAGCCTCGCGGTG
GAGTACAACTACTTCCAAGCAAAAAGT