**UNIVERSITY OF NIGERIA NSUKKA**
**FACULTY OF AGRIUCLTURE**
**DEPARTMENT OF ANIMAL SCIENCE**

**TOPIC**

# IN-SILICO GENETIC ANALYSIS OF THE SEQUENCES OF MYOSTATIN GENE IN BOVIDS (CATTLE, SHEEP AND GOAT)

**BY**

**ODO, KENNETH OGBONNA**
**PG/M.Sc/14/67325**

**SUPERVISOR: DR. H.N. FOLENG**

**SEPTEMBER, 2016**

# TITLE PAGE

# IN-SILICO GENETIC ANALYSIS OF THE SEQUENCES OF MYOSTATIN GENE IN BOVIDS (CATTLE, SHEEP AND GOAT).

**BY**

**ODO KENNETH OGBONNA**
**PG/M.Sc/14/67325**

**A PROJECT SUBMITTED TO THE DEPARTMENT OF ANIMAL SCIENCE, FACULTY OF AGRICULTURE, UNIVERSITY OF NIGERIA, NSUKKA IN PARTIAL FULFILLMENT FOR THE AWARD OF MASTERS DEGREE IN ANIMAL BREEDING AND GENETICS.**

**SUPERVISOR: DR. MRS. H.N. FOLENG**

**SEPTEMBER, 2016**

## CERTIFICATION

Odo Kenneth Ogbonna, a postgraduate student in the department of animal science, faculty of agriculture, university of Nigeria, nsukka with registration number, PG/M.Sc/14/67325 has satisfactorily completed the requirement for course and research work for the degree of master of science in animal science. (animal breeding and genetics).

The work embodied in this thesis is original and has not been submitted in part or full for any other diploma or degree of this or any other university.

é é é é é é é é é ..
Odo Kenneth Ogbonna
(Student)

é é é é é é é é .
Prof. A. G. Ezekwe
(Head of Department)

é é é é é é é é é
Dr. H.N. Foleng
(Supervisor)

é é é é é é é é é
External examiner

## DEDICATION

This work is dedicated to almighty God for his Super Grace which shielded me from the beginning of the program to the end of it.

## ACKNOWLEDGEMENT

My sincere appreciation goes to my dear and ever humane supervisor Dr. Mrs. H.N. Foleng for support, guidance, advice and constructive criticism. To the Head of the Department, Prof. Ezekwe, I want to say that you are really a father to the core. Your constant motivation has made us to be celebrating today. I also appreciate the Dean of the faculty of Agriculture. Special thanks to Emeritus professor, Dr. C.C. Nwosu, Miss Uberu, Dr. Ani, Dr. Onyimonyi, Mr. Ifeanyi Mazi, Mr. Dim, Mr. Bright, Mr. Fred, Mrs. Ohagenyi, and all the lecturers in the department of animal science for their various impact in different styles. It interest me also to express my gratitude to Dr. Yakubu of Nasarawa state university for his guidance throughout the course of the project work.

My profound gratitude to my mother for her constant encouragement and a big thanks to my big family for their concern. My man Engr. CC. Ogbu, you are a friend indeed.

A big thank you to my lovely colleagues; Egom Marvin, Ilekuba Chinenye, Ogwuegbu Mercy, Linus Lawrencia, Omumabuike Jacinta and Ozoke Uche. It was nice meeting you. May God bless you all Amen.

# TABLE OF CONTENTS.

## 3.0 CHAPTER THREE:MATERIALS AND METHODS

# LISTS OF TABLES

**Table**

# LISTS OF FIGURES

**Figures**

# LISTS OF GRAPHS

## Graphs

# LISTS OF PLATES

**Plates**

## ABSTRACT

Double muscling is a heritable trait. It has been revealed that myostatin (MSTN) or growth differentiation factor 8(GDF8) is the genetic agent of this trait. The gene is a myokine, a protein that inhibits myogenesis (muscle cell growth and differentiation). In-Silico genetic analysis was done to analyze the sequences of Myostatin gene in cattle, sheep and goat. A total of thirty seven (37) nucleotides with their corresponding amino acid sequences comprising of (26 for goat, 9 for cattle and 2 for sheep) were retrieved from the Genebank. The genetic polymorphism with three variants (M180L, S276N, S279K), five (K178N, V151L, E247D, Q329L, G355I) and six (R98H, I119T, S125M, G133S, T165N, H328T) for goat, sheep and cattle respectively, appeared not to impair the gene function while three variants (K153F, T240K, L270Q), six (S191P, W203L, S205C, N222Q, D231C, R303G), seven (D110L, I158A, R175V, K193V, S205G, P301L, F353N) were deleterious. The results revealed beneficial amino acid variants which can be used as possible markers for growth and development in goats, sheep and cattle. Although Capra and Ovis family had more propinquity and organized branch in the phylogenetic tree, the Neigbour-joining showed that sequences of the three species are similar. Also, goats and sheep appeared more similar in their amino acid contents compared to cattle. However, the distribution pattern was the same for the three species in respect of (aspartate 6.1, cysteine 3.5, glutamate 6.7, methionine 2.1, phenylalanine 3.7, proline 6.4 and tyrosine 3.2 with leucine 9.9, as the highest). The results also showed that this gene has a high degree of conservation during evolution of various species, which implies that MSTN, is an essential factor in biological muscle control. Physicochemical properties also showed extinction coefficient =51630 for sheep and cattle, half life=30hours for goats, sheep and cattle and aliphatic index =84.45 for goats and sheep, other parameters varied from one species to another. The secondary protein structure prediction in the bovine myostatin protein showed highest alpha helix (23.20%) and random coil (44.00%) with caprine and ovine at 22.67% and 43.20%. However, the extended strand (25.87%) and beta turn (8.27%) predictions were higher in both caprine and ovine species with bovine at 25.33% and 7.47%. Tertiary protein structure prediction of goat and sheep are the same while that of cattle differed. Furthermore, the results showed that capra and ovis family are much similar in function compared to cattle. Finally, the comparative inferences of myostatin gene sequences of the species studied conferred similarity in goat and sheep than cattle.

# 1.0 CHAPTER ONE: GENERAL INTRODUCTION

## 1.1 INTRODUCTION

Improving agricultural production and human food supply are the major human concern all over the world especially in developing countries. Meat is a very important source of food to human as it supplies protein and energy (Aleriza *et al*., 2014). The first step in animal breeding was selection of the best animals by ranchers, but today scientists consider recognizing genetic aspects of major genes affecting meat production. In recent years, the tools of new molecular techniques developed and caused the discovery of new growth factors that are involved in the regulation of muscle mass (Diel *et al*., 2010). It has already been pointed out that one of the interesting aspects of hypertrophy or double muscle is the dramatically increased muscle as a result of a combination of muscle fibre hyperplasia and hypertrophy (Mcpheron and Lee, 1997). The gene encoding myostatin was discovered by geneticists Se-Jin Lee and Mcpherron Alexander who also produced a strain of mutant mice that lack the gene. These Myostatin ñKnockoutò mice have approximately twice as much as normal mice (Mc Pherron *et al*., 1997) which were subsequently named ñmighty miceò. The term myostatin also known as (growth differentiation factor (GDF8) or MSTN gene is a myokine, a protein produced by muscle cells that acts on muscle cells (autocrine function) to inhibit myogenesis; muscle cell growth and differentiation. Myostatin is a secreted growth differentiation factor (GDF8) that is a member of the Transforming growth factor (TGF) beta protein family (Joulia-Ekaza and Cabello, 2007). Animal with mutant genotypes in GDF8 gene not only produce more meat, the quality of the meat in their products is also different.

In these animals, the posterior limbs are round and prominent, the muscle are in protrusion mood and clear lines under the skin are visible. Prominent examples of mammals are Belgium blue and pre-montese, which show significant characteristics of double-muscle (Kambadur *et al*., 1997).

Myostatin referred to as growth differentiation factor-8 (GDF8) is a member of the mammalian growth transforming family (TGF-B Super family), which is expressed specifically in developing an adult skeletal muscle (Gonzalez-Cadavid and Bhasin, 2004). Mice that lack myostatin indicated a widespread increase in skeletal muscle mass due to an increase in both myofiber size (hypertrophy) and myofiber number (hyperplasia): (Mcpherron *et al.*, 1997). Muscular

hypertrophy (mh) also called double muscling has been intensified during a study in cattle as a heritable physiological character and is found in Austriana de loss valles, Belgian blue and Redmontese breeds of cattle (Smith *et al.*, 1997). Use of double muscle beef breeds has been encouraged as a result of their high meat yield and superior meat quality associated with high proportion of white, glycolytic muscle fibres (Shah *et al.,* 2006). Double-muscled cattle also deposit much less fat than other breeds (Potts *et al.,* 2003).

MSTN is synthesized as a biologically inactive precursor molecule (Full length MSTN protein) comprising 3 domains, the signal peptide, the propeptide (N-Terminus), and the C- Terminal domain. MSTN is composed of 375 amino acid precursors, and has same C-terminal fragments of about 109 amino acid residue in mice, rats, human, swine, fowl and turkey and only 3 amino acid residue in C- terminal region thereof are not the same in monkeys, cows, and sheep, The C-terminal regions are expected to include physiologically active portions of MSTN (Thomas *et al.*, 2000). Major effect of a single gene on processing yields opened a potential channel for improving processing yields of animals using knockout technology (Arif *et al.*, 2002). Therefore, in present study, in-silicon genetic analysis of sequence of myostatin gene in mammalian species is important in understanding the evolution, differentiation and the effects of polymorphism on the myostatin gene (MSTN ) and how they are related within and among the mammalian species under study.

## 1.2 Objectives of the Study

The general objective of the study was to analyze the myostatin gene in selected bovids (cattle, sheep and goat) with a view to providing relevant genetic information for breeding and selection programmes in the studied species in Nigeria.

**The Specific objectives were to;**

1. Examine the attendant effects of various amino acids substitutions of the myostatin gene in the selected species.
2. Examine the genetic diversity of myostatin *in silico* on their evolution and differentiation within and among species.
3. Study the various physicochemical properties of myostatin gene and;
4. Predict the secondary and tertiary structures of the myostatin gene in the selected species.

## 1.3 **Justification**

Myostatin genetic polymorphisms have evoked considerable research interest in recent years because of its possible association with growth. The major challenge that faces molecular geneticists is to identify markers for genes that control the phenotypic variation in the target traits. Recent advances in high-through put technologies have generated massive amounts of genome sequence and genotype data for a number of species. The method to identify functional SNPs from a pool, containing both functional and neutral SNPs is challenging by experimental protocols. Therefore, computational predictions have become indispensable for evaluating the impact of nonsynonymous single-nucleotide variants discovered in exome sequencing.

A good knowledge of the sequences of myostatin gene will help in identifying the variants responsible for various factors attributed to the gene. This will help breeders to plan breeding programmes more easily and in turn reduce the problem of low supply of high quality protein facing the increasing populations especially in a developing economy such as Nigeria.

# 2.0 CHAPTER TWO: LITERATURE REVIEW

## 2.1 Myostatin growth factor.

Myostatin, or growth differentiation factor 8 (GDF8), is a protein that is encoded by the MSTN gene and is a secreted growth differentiation factor that is a member of the transforming growth factor (TGF) beta protein family. The TGF beta protein family controls primarily proliferation and cellular differentiation; myostatin inhibits muscle differentiation and growth in the process known as myogenesis and is produced primarily in skeletal muscle cells, circulates in the blood and acts on muscle tissue. Myostatin was discovered in (1997) by geneticists Dr. Se-Jin Lee and Alexandra McPherron who also produced a strain of mutant mice that lack the gene, which had twice as much muscle as normal mice. Like other TGF-ß family members, myostatin is synthesized as a precursor protein, and contain 244 amino acid residues of the human myostatin Propeptide; the propeptide undergoes proteolytic processing at a dibasic site to generate an N-terminal propeptide and a disulfide-linked C-terminal dimer, which is the biologically active molecule. The circulating form of myostatin consists of a latent complex of the myostatin C-terminal dimer and other proteins, including the myostatin propeptide, which inhibit the biological activity of the C-terminal dimer (Filipo, 2014). The enzyme that cleavages the myostatin propeptide is unknown, but some researchers suggest that members of the bone morphogenetic protein-1/tolloid (BMP-1/TLD) family of metalloproteinases may be involved in activating latent myostatin in vivo., myostatin displays the traditional TGF-ß family hand-shaped architecture, with each monomer consisting of four curved beta strands or ófingers', a cystine knot motif in the ópalm' region, and a major helix or ówrist'.

## 2.2 Myostatin Protein

Myostatin actively inhibits skeletal muscle development (Bellinge *et al.,* 2005). Myostatin is a member of the transforming growth factor (TGF)-ß superfamily and cannot be classified into the existing TGF-ß subfamilies, such as inhibins or bone morphogenic proteins. This deviation from the typical TGF-ß family is particularly evident in the C-terminal region (Mc Pherron *et al.,* 1997). Myostatin, like other members of the transforming growth factor-ß (TGF-ß) family, is synthesized by a 376 amino acid precursor protein including three domains namely, a C-terminal

domain or active molecule, an N-terminal propeptide domain which will be cleaved at the RSRR site during maturation, and a signal sequence (Fig.1).

Proteasic digestion processing between the propeptide domain and the C-terminal domain results in an N-terminal propeptide and the mature form of myostatin, a 12-kDa carboxy-terminal fragment. Both mature and unprocessed myostatin form disulfide linked dimers. Moreover, the only active form of the protein is the processed myostatin dimer (Joulia-Ekasa and Cabello, 2006).

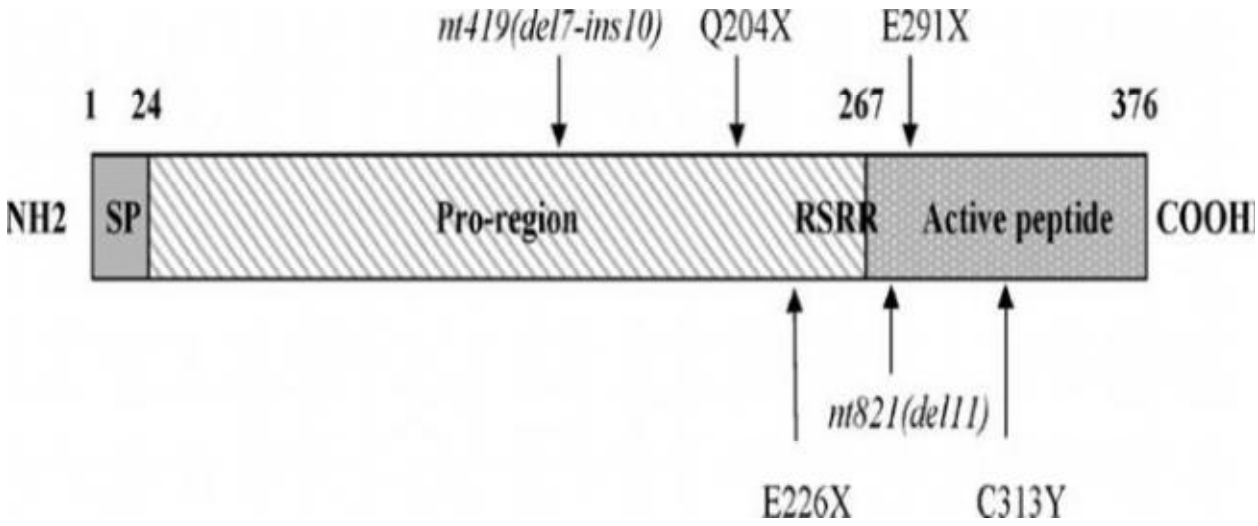



**Figure1; Myostatin protein structure and natural mutations in the bovine myostatin gene.**

The three domains are the active peptide at the C-terminal part, the pro-region and the signal peptide (SP). The arrows show the position of mutations that are responsible for the increased muscle growth in some cattle breeds (Mc Pherron *et al.,* 1997).

In mice, myostatin is predominantly present in both developing muscles, (even as early as 9.5 days postcoitum), and adult skeletal muscles (Mc Pherron *et al.,* 1997). However, there are several reports of various animal species having the occurrence of myostatin mRNA or protein in their other tissues and plasma (Gonzalez-Cadavid *et al.,* 1998).

## 2.3 The myostatin pathway

While myostatin is bound to the follistatin-related gene (FLRG), and the growth and differentiation factor-associated serum protein-1 (GASP-1), human small glutamin-rich tetratricopeptide (hSGT) repeat-containing protein is reproduced with permission of (hSGT), T-cap, follistatin or the myostatin propeptide. It can then be found either in the serum or in an inactive local state. The active myostatin dimer gets attached to the activine type II receptor (ActRIIB), which then activates the type I receptor (ALK4 or ALK5) by transphosphorylation. Smad2 and Smad3 are then activated as a result of the previous process. Smad4 joins them afterwards. Finally, they translocate to the nucleus, activating target gene transcription. So far, two inhibitors of this signalization, namely Smad7 and Smurf1, have been identified. Smurf1 is an E3 ubiquitin ligase that mediates ubiquitination and the consequent degradation of the R-Smads (Fig.2.). Expression of Smad7 is induced by the myostatin expression. This could express the existence of a negative regulatory feedback loop mechanism (Zhu *et al.,* 2004).

In vitro studies show that myostatin causes C2C12 myoblasts to be accumulated in the G0/G1 and G2 cell-cycle phases, consequently diminishing the number of S-phase cells. Moreover, myostatin causes failure in myoblast differentiation, which is related to a strong decrease in the expression of differentiation markers (Joulia-Ekaza and Cabello, 2006). Furthermore, under proliferation and differentiation conditions, myostatin expression diminishes the apoptotic rate of cells (Joulia *et al.,* 2003). Using antisense myostatin mRNA, the opposite results were obtained by preventing endogenous myostatin expression. This approach suggests that myogenin and p21 cyclin-dependent kinase inhibitors might probably be the main physiological targets of myostatin (Joulia *et al,* 2003).

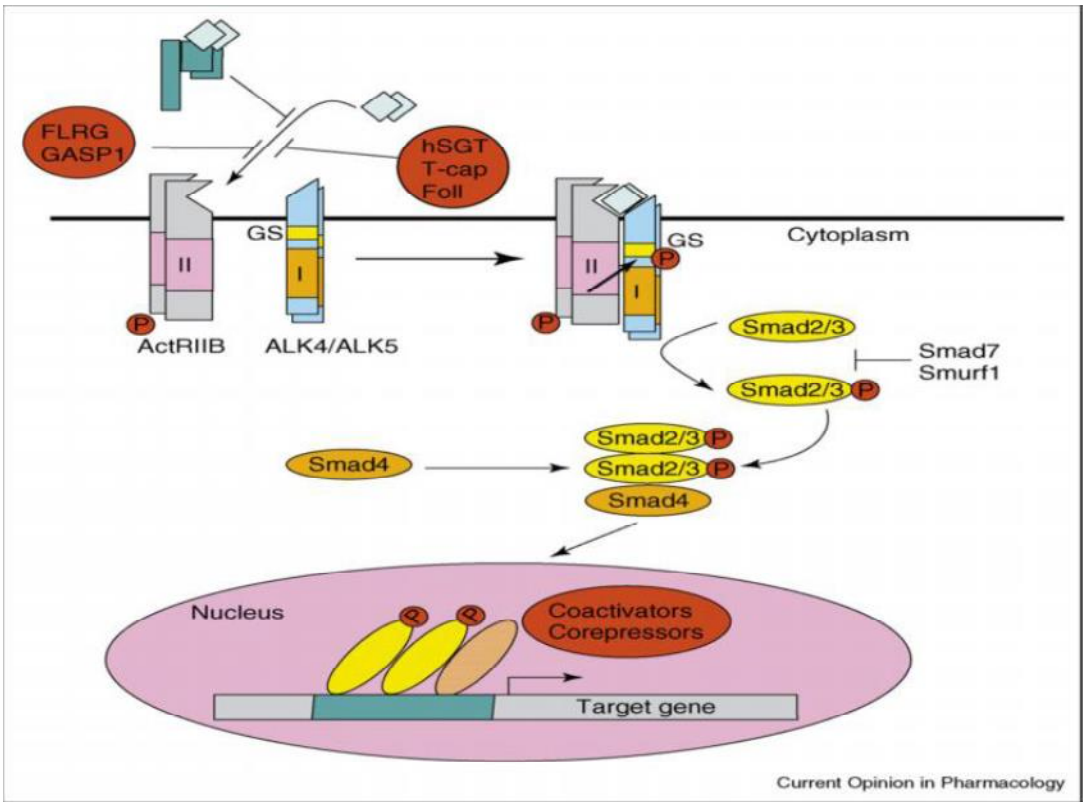



**Figure 2. Famous elements of the myostatin pathway ((Joulia-Ekasa and Cabello, 2006).**

## 2.4 Cellular Actions of Myostatin

### 2.4.1 Proteolytic processing and regulated bioactivity

Several myostatin binding proteins have been identified and include follistatin (Lee and McPherron, 2001), follistatin-like related gene (FLRG; also known as follistatin like-3, FSTL-3, and follistatin-related peptide or FLRP) (Hill *et al.,* 2002), growth/differentiation factor-associated serum protein (GASP)-1 (Hill *et al.,* 2003) and titin (T)-cap (Nicholas *et al.,* 2002) T-cap is a sarcomeric protein that binds the N-terminal domain of titin where it helps regulate the cytoskeletal organization of muscle cells. T-cap also binds myostatin presumably in the golgi and prevents its secretion. The N-terminal peptide that results from proteolytic processing of pro-myostatin also binds myostatin with high affinity, and like follistatin, FLRG and GASP-1, can prevent receptor binding and activation (Lee and McPherron, 2001), Such interactions commonly occur between proteolyzed fragments of TGFϐ superfamily members (Koli *et al.,* 2001) and although the exact mechanism of ligand activation has yet to be determined for myostatin, it appears to strongly resemble that of its superfamily siblings. This includes removal

of the signal peptide from prepro-myostatin and proteolysis of pro-myostatin at the furin/PACE cleavage site. This separates the bioactive domain from the N-terminal latency-associated peptide (LAP), which binds to the disulfide-linked myostatin dimer. The two proteins are then secreted as a small latent complex where they enter the circulation or potentially bind to the extracellular matrix forming a large latent complex. Proteolytic cleavage of LAP then releases myostatin from circulating and extracellular complexes. Thus, myostatin bioactivity is not mediated *per se* by increased synthesis or release from skeletal muscle, but by three independent proteolytic events, of which the latter two may be regulated.

The nullifying effects of LAP and follistatin on myostatin bioactivity have been demonstrated quite conclusively using *in vitro* and/or *in vivo* systems, with LAP receiving the most attention. Both proteins bind myostatin and prevent receptor binding and activation *in vitro*, whereas transgenic mice overexpressing LAP or follistatin develop hypermuscularity similar to that seen in the Mighty Mouse (Yang *et al.,* 2001).

### 2.4.2. Regulation of myoblast proliferation, differentiation, and quiescence

Myostatin appears to prevent myoblast hyperplasia in mammals by inhibiting cell cycle progression (Taylor *et al.,* 2001). Myostatin also inhibits myoblast differentiation (Langley *et al.,* 2002) although the teleological significance of this particular effect may appear controversial because conflicting data suggest that myostatin initiates cell cycle withdrawal, which is a necessary prerequisite for differentiation (Joulia *et al.,* 2003) However, studies with primary myosatellite cells (also known as ñskeletal muscle stem cellsò located below the sarcolemma and basal lamina) from myostatin-null mice suggest that myostatin-stimulated cell cycle withdrawal accompanied by cellular quiescence (McCroskery *et al.,* 2003) rather than differentiation. This explains the apparent discrepancy and supports earlier studies indicating that myostatin is a myoblast survival factor (Rios *et al.,* 2001). A model for myostatin action in mammals suggests that in the absence of other myogenic regulators, myostatin inhibits myoblast hyperplasia by stimulating cell cycle withdrawal and delays differentiation by inducing cellular quiescence. Recent studies further suggest that myostatin-induced cellular quiescence is reversible and is associated with reduced expression of the myogenic regulatory factors Pax-3, Myf-5, and MyoD (Amthor *et al.,* 2006). It is therefore inaccurate to describe myostatinôs actions as solely inhibitory. Indeed, myostatin initiates the first and necessary step in the differentiation process,

cell cycle withdrawal, and prevents apoptosis of the quiescent cells. Its negative effects on myofiber hypertrophy are due to the inhibition of myosatellite cell activation, proliferation, and/or renewal because the fusion of these cells with existing myofibers is largely responsible for postnatal muscle growth (Figeac *et al.,* 2007). These cells are more abundant in skeletal muscle of myostatin-null mice, which proliferate more rapidly than those isolated from wild-type mice (McCroskery *et al.,* 2003). Myostatin also inhibits protein synthesis in differentiated C2C12 myotubes (Taylor *et al.,* 2001). These studies together suggest that myostatin ultimately limits skeletal muscle size by inhibiting the hyperplastic growth of developing myoblasts and thus, the number of cells that eventually differentiate into mature myofibers, and by reducing myofiber protein synthesis and myosatellite cell renewal, both of which inhibit the hypertrophic growth of mature muscle.

### 2.4.3. Receptors and signaling

All TGFb superfamily ligands signal through membrane-bound heteromeric serine-threonine kinase receptor complexes composed of two type 1 and two type 2 receptors (Attisano *et al.,* 2002). Ligand binding to type 2 receptors recruits type 1 and both autophosphorylate via transinteractions with one anotherôs intracellular kinase domains. The signaling pathway to the nucleus is short and quick because specific receptor (R)-Smads are recruited to the complex and are phosphorylated by type 1 receptors. The R-Smads then oligomerize with appropriate co-Smads and translocate into the nucleus. This complex directly binds promoter elements and initiates or represses gene transcription. Myostatin bioactivity is mediated by activin receptors, specifically Acvr2 and Acvr2b. Cross-linking studies and radioreceptor assays indicate that although myostatin binds both, it binds the latter with slightly higher affinity (Lee and McPherron, 2001). Skeletal muscle mass was increased by 125% in transgenic mice overexpressing dominant-negative Acvr2b (Lee and McPherron, 2001) and by 60% just 2 weeks after three ip injections of a soluble form of Acvr2bôs extracellular domain (Lee *et al.,* 2005). Mice homozygous for deactivating mutations in Acvr2 have pectoralis and triceps muscles that are 27ï 40% larger than the same muscles from wild-type mice ((Lee *et al.,* 2005). These muscles are just 20ï 26% larger in mice with mutant Acvr2b receptors, suggesting that Acvr2 may play a more important role in regulating myostatinôs actions, at least in these muscles. The relative distribution of each receptor in different skeletal muscles or even within a specific muscle is not

known. Thus, the contribution of each receptor may be equally relative and may differ between individual muscles or even fiber types. Myostatin activation of either receptor recruits the type I receptors activin like kinase-4 or -5 (Rebbapragada *et al.,* 2003), which in turn phosphorylate Smads 2 and 3. These transcription factors oligomerize with Smad4 and eventually regulate gene transcription, including the expression of Smad7 and c-ski. This particular Smad is an inhibitory Smad because it sequesters Smad4 in the cytoplasm and prevents it from binding to the Smad2/3 complex (Rebbapragada *et al.,* 2003). C-ski is a corepressor that stabilizes the inactive Smad2/3/4 complex on Smad/ski binding elements (Xu *et al.,* 2000). It also appears to inhibit Smad2 and Smad3 signaling in part by directly blocking histone deacetylase activity as well as their association with a transcriptional coactivator(Kobayashi *et al.,* 2007). Nuclear localization of c-ski is required for the differentiation of myoblasts because it enhances myogenin transactivation through direct interactions with MyoD/MEF2 heterodimers (Kobayashi *et al.,* 2007), which is in direct opposition to myostatinôs negative effects on differentiation. These studies together suggest that in skeletal muscle, Smad7 and c-ski serve as intracellular negative feedback mechanisms for myostatin or other TGFb superfamily ligands that signal via Smads 2 and 3.

Myostatin signaling is not limited to canonical Smad pathways because it curiously interacts with mitogenic pathways as well. Myoblast proliferation and cell cycle progression are stimulated by IGF-I, a potent mitogen for many different cell types including myoblasts. IGF-I also stimulates myoblast differentiation and the associated cell cycle arrest (Zapf *et al.,* 1999). The mechanisms of IGF-Iôs ability to stimulate these normally diametrically opposed cellular activities is currently under dispute and may include the local production of IGF binding protein (IGFBP)-3 (Pampusch *et al.,* 2003).. Myostatin also stimulates myoblast cell cycle withdrawal and activates p21 (McCroskery *et al.,* 2003), but unlike IGF-I it inhibits rather than stimulates differentiation. The significance of cross-talk between myostatin and IGF signaling is not known, nor is it known how myostatin activates these pathways while simultaneously arresting the cell cycle. However, myostatin and IGF-I are both survival factors, and thus myostatin activation of these particular aspects of a mitogenic pathway is likely related to its ability to maintain cellular quiescence in growth-arrested cells. Indeed, blocking Erk-1/2 and JNK activation similarly blocks myostatinôs ability to inhibit differentiation (Yang *et al.,* 2006)

## 2.5 Functions and Comparative Genomics of Myostatin

### 2.5.1. Genomic organization

Myostatin genes have been mapped in several vertebrate species, albeit to different degrees of resolution, and are located on chromosomes 2 in humans, cattle, and sheep, 1 in mice, 9 in rats, 15 in pigs, 18 in horse, 37 in dogs, and 7 in chicken (Kambadur *et al.,* 1997) (see also Entrez Gene at www. ncbi. nlm. nih. gov/sites/entrez?db=gene). The fish MSTN-1 and -2 genes have also been cloned in zebrafish (Xu, *et al.,* 2003) and mapped to chromosomes 9 and 22, respectively. They are also located on separate chromosomes in the green-spotted pufferfish (*Tetraodon nigroviridis*), chromosomes 2 and 3, which share more duplicated genes than any other chromosome pair, and are putative paralogs themselves (Jaillon *et al.,* 2004). These data are consistent with the phylogenetic distribution of the two fish myostatin clades and further suggest that MSTN-1 and -2 arose from an early genome duplication event that occurred approximately 350,000,000 years ago (Amores *et al.,* 1998). Ostbye *et al. (* 2007) recently mapped Atlantic salmon MSTN-1a and -1b to separate linkage groups indicating that these paralogs, and likely the MSTN-2a and -2b paralogs as well, arose from the recent (25,000,000ï 100,000,000 yrs ago) tetraploidization of the salmonid genome (Phillips and Rab, 2001), which again is consistent with the well-described phylogenies (Garikipati *et al.,* 2007).

Sequence differences in noncoding regions can also contribute to functional divergence by mechanisms qualitatively different from those in coding regions and are based on two fundamental hypotheses (Wray, 2007): changes in *cis* regulatory elements 1, are more likely to have phenotypic consequences, and 2, are more sensitive to selection pressures. Environmental factors and physiological responses greatly and rapidly influence gene transcription, whereas changes in protein structure occur more slowly. Mutations in *cis* elements can also be codominant (Wittkopp *et al.,* 2004) and are thus more sensitive to natural selection because these changes are often expressed in heterozygotes, whereas similar changes in coding regions are usually recessive (Furney *et al.,* 2006). Thus, polymorphisms in orthologous gene promoters contribute to functional divergence by substantially changing gene expression patterns that ultimately affect fitness. Indeed, differences in promoter activity are inherently indicative of functional divergence because altered expression in different tissues or development stages similarly alters function. Although the precise function for each fish myostatin paralog is not known, they are clearly diverging because the gene expression patterns are very different.

Further analysis of myostatin function and molecular evolution will provide a unique opportunity to better understand fundamental mechanisms of evolutionary change. They may also help understand novel functions for myostatin in other vertebrates, including mammals, because recent studies indicate that myostatin expression in mammals is more diverse, and more similar to fish, than originally presumed.

### 2.5.2. Differential gene expression

Surprisingly little is known about myostatin expression in developing mammalian embryos. It is first detected in the myotome compartment of developing mouse somites and presumably continues in developing myogenic cells. By contrast, many studies have both qualitatively and quantitatively assessed myostatin expression in developing embryos of different fish species (Biga *et al.,* 2005) and in chickens (Castelhano-Barbosa *et al.,* 2005). The most comprehensive of these studies used extensive RNA panels and gene-specific ñreal-timeò assays to correlate expression levels of all myostatin genes in rainbow trout and zebrafish to key ontological events (Garikipati *et al.,* 2006). All transcripts were detected in unfertilized and newly fertilized embryos. This is consistent with maternal deposition and with the expression of MSTN-1 and -2 genes in other fish species and even in chicken embryos (Castelhano-Barbosa *et al.,* 2005). It also suggests that myostatin plays a significant role during early development, but not during gastrulation because all studies to date report a rapid decline in myostatin expression during this stage. Expression levels of all MSTN-1 genes in both fish species progressively increased during somitogenesis, which is consistent with myostatinôs myogenic role in mice. However, myostatin is also expressed in many developing chicken tissues and in most adult fish tissues (see below). Further studies are clearly needed to determine whether this dynamic regulation of myostatin expression occurs in mammalian embryos and whether myostatin is also expressed in developing mammalian tissues other than skeletal muscle. If conserved, the different fish model systems, particularly zebrafish, will prove invaluable to elucidating myostatinôs different developmental functions.

Initial studies suggested that in adult mammals, myostatin expression was limited primarily to skeletal muscle, although subsequent studies identified low levels of expression in mammary gland and heart. By contrast, many non mammalian vertebrates do not share this limited

expression pattern because myostatin mRNA and/or protein is expressed in most fish tissues and in many different developing chicken tissues as well. In fish, the MSTN-1 genes are widely expressed, whereas MSTN-2 expression is more limited and occurs mostly in the brain (Garikipati *et al.,* 2007). Expression of the former is also dynamically regulated during development, whereas MSTN-2 expression changes very little (Garikipati *et al.,* 2007). These data strongly suggest that in fish, myostatinôs actions are not restricted to the negative regulation of skeletal muscle growth and development, but may additionally influence these processes in many other tissues through the differential expression of each paralog. The similarly ubiquitous expression pattern in chickens suggests that myostatinôs more limited expression pattern in mammals evolved more recently. A more thorough analysis of myostatinôs tissue-specific expression pattern in other avians is needed to determine whether limited expression in general is unique to mammals. Nevertheless, the comparative analysis of myostatin expression suggests that myostatin originally functioned as a general differentiation factor, which is consistent with the wide distribution of activin receptors in nonmammalian vertebrates (Garg *et al.,* 1999), and recently adopted a more specialized role in mammals.

### 2.5.3. Alternative processing

Alternative splicing of myostatin transcripts has been described in rainbow trout (Garikipati *et al.,* 2007) and in developing chicken embryos (Castelhano-Barbosa *et al.,* 2005). Although the precise functional significance remains to be determined for both animal models, preliminary evidence suggests that in rainbow trout, it enhances myostatinôs effects in the brain, whereas in the chicken it helps control bioavailability.

Both intact and truncated myostatin transcripts are expressed in chicken embryos (Castelhano-Barbosa *et al.,* 2005). The latter lacks the coding sequence for the C-terminal mature peptide. Translation of this transcript would therefore block myostatin bioavailability because only the LAP domain would be produced. In rainbow trout, MSTN-2a and -2b are lowly expressed in most tissues except for brain where rtMSTN-2a is highly expressed (Garikipati *et al.,* 2007). Nearly all tissues express mostly unprocessed transcripts that retain both introns, which themselves contain several in-frame stop codons. By contrast, rtMSTN-2a, but not -2b, is fully processed in the brain. Tissue-specific alternative processing is rare and is mediated by *cis* recognition of tissue-specific proteins that bind to repeating motifs in the pre-mRNA (Hui and

Bindereif, 2005). It occurs most often in the brain where it is commonly mediated by Nova proteins that act as splice enhancers or silencers, depending on their proximity to the splice site, and is required with transcripts possessing poorly conserved splice sites (Hui *et al.,* 2003). Nova proteins recognize YCAY motifs and are known to enhance the alternative splicing of two neurotransmitter receptors, $GABA_ARg2$ and GlyRa2, and to silence the splicing of some Nova transcripts themselves, specifically Nova-1 (Dredge and Darnell 2003). Both rtMSTN-2 genes lack conserved splice sites, although rtMSTN-2a has almost twice the number of expected YCAY motifs throughout and three times the expected motifs in the second intron alone. Whether Nova regulates the alternative splicing of rtMSTN-2a remains to be determined. Nevertheless, a comparative analysis of rtMSTN-2a and -2b processing provides a perfectly controlled system to investigate the basic mechanisms responsible for tissue-specific pre-mRNA processing.

As the mature myostatin peptide is encoded entirely by the third exon, the introduction of alternative transcription start sites in either intron of rtMSTN-2b, or any other salmonid MSTN-2b, could presumably result in the expression of a mature myostatin. Therefore, the inability to remove expressed introns contributes to the pseudogenization of rtMSTN-2b as it prohibits the ñaccidentalò translation of an otherwise silenced gene. This is particularly important as the rtMSTN-2b gene still possess a functionally active promoter. Contributions from noncoding regions to the functional divergence of duplicated genes or to the subfunctionalization of a particular allele are not necessarily limited to gene promoters. Gene expression and pre-mRNA processing are both regulated by protein:DNA interactions that depend upon highly specific binding sites. Thus, alterations in such binding sites over time could not only influence gene function, but ultimately organismal complexity and speciation as well.

## 2.6 Novel Actions

### 2.6.1. Adipose tissue

Myostatin is minimally expressed in adipose tissue, and myostatin-null animals have less total body fat than wild-type animals. Increases in muscle mass have long been known to similarly increase resting energy expenditure (REE), which in turn can reduce fat free mass (Wang *et al.,* 2000) and is inversely correlated to negative outcomes of patients with type 2 diabetes mellitus (Hamilton *et al.,*2007). Thus, the reduced adiposity in myostatin-null animals could simply be due to the caloric draw from enhanced musculature. Circulating levels of leptin, an adipokine and satiety factor that controls body fat, are reduced rather than elevated in these animals. This suggests that the increased REE is indeed responsible. However, myostatin has been shown to directly influence the cellular physiology of three different adipocyte culture systems with conflicting results. It inhibits the differentiation of 3T3-L1 preadipocytes (Stolz *et al.,* 2008) and primary preadipocytes from cattle (Hirai *et al.,* 2007) or humans (Gou *et al.,* 2008). It also down-regulated the expression of several adipogenic markers and transcription factors in these cells, including peroxisome proliferator-activated receptor ɔ, C/EBPŬ, aP2, and leptin.

McPherron and Lee (McPherron and Lee, 2002) directly tested myostatinôs ability to influence the development of obesity by crossing the Mighty Mouse with two murine models of obesity and insulin resistance: the agouti lethal yellow ($A^{y}$) and leptin-deficient obese ($Lep^{ob/ob}$) mice. Introducing a myostatin-null background into either strain increased muscle mass, suppressed adiposity (as indicated by reduced mass of different fat pads), and vastly improved glucose tolerance. Similar results were also observed in transgenic mice overexpressing LAP and fed a high-fat diet (Yang and Zhao 2006). Thus, enhancing muscle mass by blocking myostatin bioactivity and/or bioavailability can prevent the development of obesity and insulin resistance. These results do not, however, prove that myostatinôs effects are mediated entirely by increasing REE because the myostatin-null environment could have influenced the development of preadipocytes *in utero* or even the *de novo* synthesis of triglycerides in already differentiated adipocytes. Several additional questions also remain unanswered. In particular, will disrupting myostatin production and/or bioavailability be similarly beneficial in animals that are already obese and displaying signs of insulin resistance? Regardless, these results are extremely exciting

and potentially introduce a new therapeutic target for treating obesity and type 2 diabetes mellitus.

### 2.6.2. Cardiac muscle

Cardiac expression of myostatin has been documented in sheep (Sharma *et al.,* 1999), chickens (Sundaresan *et al.,* 2008), mice, rats (McKoy *et al.,* 2007), and different fish species (Garikipati *et al.,* 2007). In developing chicken hearts, myostatin expression is detected early and progressively increases until morphogenesis is complete, suggesting a functional role for myostatin in the development of cardiac as well as skeletal muscle. High levels of myostatin expression in primary mouse cardiomyoblasts are correlated with a low proliferative index, whereas recombinant myostatin inhibits the growth of these cells, as well as rat H9C2 cardiomyoblast growth, without inducing apoptosis (McKoy *et al.,* 2007). Myostatin also suppresses cardiomyocyte hypertrophic growth responses, specifically protein synthesis, induced by either phenylepinephrine (McKoy *et al.,* 2007) or IGF-I (Shyu *et al.,* 2005). Recent studies with Akt transgenic mice (Cook *et al.,* 2002), *in vitro* models of cyclic stretch, and IGF-I-stimulated cardiomyocytes (Shyu *et al.,* 2005) all suggest that myostatin not only regulates some cardiac muscle growth process, but that it may function as a cardiac chalone (Gaussin and Depre, 2005) just as it does in skeletal muscle. Indeed, myostatin expression is elevated in all of these models and may therefore provide a negative feedback mechanism to limit cardiac muscle growth, and thus hypertrophy.

In mammals, myostatin is clearly expressed in developing and adult cardiac muscle and is capable of manipulating different cardiac muscle growth processes. It is unknown, therefore, why a significant cardiac phenotype has never been described for any myostatin-null animal. Cardiac tissue lacks myoprogenitor cells analogous to the myosatellite cells found in skeletal muscle and may be less sensitive to myostatinôs inhibitory effects. Thus, blocking myostatinôs actions or removing the gene altogether would have very little effect, except under circumstances when myostatin expression is elevated.

### 2.6.3. Brain

Although myostatin is expressed in different brain regions of different vertebrates, its function in the brain is unknown. However, GDF-11, which is closely related to myostatin and has a nearly

identical bioactive domain (95% similar), is thought to play a role in neurogenesis. It is more widely expressed in mammalian embryos (Nakashima *et al.,* 1999) and is secreted by neuroprogenitor cells and fully differentiated neurons of the olfactory epithelium where it inhibits neurogenesis by inducing cell cycle arrest of the neuroprogenitors (Wu *et al.,* 2003). In the retina, GDF-11 limits the number of ganglion, amacrine, and photoreceptor cells by controlling neuroprogenitor cell competence (Kim *et al.,* 2005). Thus, GDF-11 functions as a negative autoregulator, or chalone, of neural tissue, which mirrors myostatin action in mammalian skeletal muscle. It is therefore possible that myostatin has similar functions in the brain, albeit in different regions and that compensatory changes in GDF-11 expression and/or availability prohibit the development of neural phenotypes in myostatin-null animals. Fish brains maintain a large number of hyperplastic neuroprogenitor cells that can be easily isolated and even cultured in vitro (Evans *et al.,* 2000). *In vivo* models for neuroprogenitor cell proliferation and differentiation have also been developed in goldfish and zebrafish (Otteson *et al.,* 2002). These underutilized comparative model systems may therefore prove valuable in distinguishing the neural functions of myostatin and GDF-11.

## 2.7. Double muscling in different species:

### 2.7.1. Sheep

In the past, double muscling (DM) identification in sheep was based on morphological characteristics such as appearance of intermuscular grooves, and pelvic inclination (Bellinge *et al.,* 2005). However, following myostatin gene characterization by (Mcpherron *et al.,* 1997) and the determination of mutant DM in cattle (Grobet *et al.,* 1997), DM identification is almost achieved via genetic marker testing. Genetic marker testing or the candidate gene approach assumes that a gene involved in the physiology of the trait could harbour a mutation causing variation in that trait.

The *GDF8* allele of Texel sheep is characterized by one G to A transition in the 3'UTR region of myostatin, causing double muscling (Younes *et al.,* 2014). According to Ron and Weller (2007), it seems that genotyping of this SNP could be a good option for the double muscling and muscularity identification in sheep.

Figure3: An example of a Texel double muscle sheep

### 2.7.2. Cattle

Similar phenotypes have also been described in some domestic breeds of cattle including the Piedmontese  Belgian Blue, and Marchigiana, all of which possess mutant alleles for myostatin (Kambadur *et al.,* 1997). The Piedmontese is marketed as the ñMyostatin Breedò because its standard and the North American Piedmontese Cattle Associationôs registry require proof of at least one mutant myostatin allele, which may be the first cattle registry based on a particular genotype (www.piedmontese.org). Meat from Piedmontese scores high in palatability studies and is particularly tender, more so than the other breeds (Page *et al.,* 2002). By contrast, meat from the double-muscled Texel sheep is very tough (Johnson *et al.,* 2005). Thus, enhanced musculature itself does not necessarily impact meat palatability because other genetic factors clearly contribute.

Figure 4: Double muscling in cattle

### 2.7.3. Goat:

At present, many scholars had done numerous researches in pigs and cattle (Walsh and Celeste, 2005; Fan *et al.,* 2010). In sheep, mutations in the *MSTN* gene also had an important role in muscular development (Boman *et al.,* 2009). As a consequence, the *MSTN* gene is primarily responsible for muscle development and could be a potential candidate gene for animal muscle growth. However, similar investigations about the properties of the *MSTN* gene in goat breeds have been limited (Liu *et al.,* 2006).

Results have indicated that the *MSTN* gene is a candidate gene for goat growth performance, and plays a significant role in the body weight of goats . It could be inferred that the *MSTN* gene may be a major gene or linked to the major gene affecting goat growth traits (Zhang *et al.,* 2013). The polymorphic site could be a molecular marker-assisted selection program for body weight (Zhang *et al.,* 2013). In goat breeds, a number of myostatin variants of different phenotypic consequence have been described across a variety of breeds (Li *et al.,* 2006; Javanmard *et al.,* 2010), but there is scarcity of reports on the association analysis of SNPs with growth traits. Polymorphisms have been identified in exon 1 of the *MSTN* gene in goat breeds and could be potential genetic markers for growth traits in goats (An *et al.,* 2011).

Figure 5: Double muscling in goat:

## 2.8. Implications for Biomedical, Agricultural, and Evolutionary Sciences

The most apparent biomedical application for manipulating myostatin action is clearly to enhance muscle growth, although the use of novel ñmyostatin-blockingò technologies need not be limited to the treatment of skeletal muscle pathologiesð injury, sarcopenia, wasting/cachexia, some forms of muscular dystrophy, etc. because exploiting such technologies may prove beneficial in treating survivors of a myocardial infarction as well. Several of these technologies have already been developed, including immunoneutralizing antisera (Liang *et al.,* 2007), a soluble form of Acvr2bôs extracellular domain (Lee *et al.,* 2005) and a quiver of myostatin binding proteins and each successfully stimulates skeletal muscle hypertrophy in wild-type mice and, in some instances, in a murine model for Duchenne muscular dystrophy (Bogdanovich *et al.,* 2002). The indirect effect of neutralizing myostatinôs actions on adiposity suggests that these technologies could also be useful in treating obesity and type 2 diabetes mellitus. A promise of ñmore muscle, less fatò is an attractive marketing campaign and one that these technologies could potentially deliver. Thus, blocking myostatinôs actions for cosmetic purposes is not just probable, but possibly inevitable as the biomedical successes gain visibility among the general public.

Double-muscled cattle like the Belgian Blue and the Piedmontese frequently require cesarean delivery due to large calf size. Piedmontese meat is particularly tender (Wheeler *et al.,* 2001); nevertheless, the high cost of veterinary care associated with parturition has prevented the widespread acceptance of these and other double-muscled breeds in most cattle production markets. Crossing myostatin-null breeds with normal-muscled breeds or even immunoneutralizing cows before breeding, as has been done with mice (Liang *et al.,* 2007), could address issues of growth rate and product yield in many animal production industries. It is unknown, however, whether the potential loss of im fat would potentially offset these commercial gains. These issues are not a concern with egg-laying vertebrates. Thus, myostatin technologies may have a greater commercial impact on domestic fish and fowl production, especially because the mass ratio of muscle to nonmuscle tissues in fish is greater than in any other vertebrate class. Such commercial gains are predicated on the assumption that blocking myostatin actions will have no adverse side effects on nonmuscle tissues, many of which express multiple myostatin genes. A more thorough assessment of myostatinôs nonmuscle actions in fish is therefore needed to determine whether blocking its actions is both feasible and commercially beneficial.

Comparative model systems will also help identify novel functions for both myostatin and GDF-11 in ways pertinent to agriculture and medicine as their more basic and conserved actions are defined. The subset of salmonid myostatin genes is a particularly useful model for investigating the functional divergence of duplicated alleles because differences in coding and noncoding sequences have contributed to their evolution in a manner that influences gene expression, transcript processing, protein structure, and pseudogenization. Determining the underlying mechanisms involved presents a unique opportunity to investigate competing evolutionary modelsð double-recessive *vs.* duplication-degeneration-complementation (Force *et al.,* 1999) that potentially explain the molecular basis of fundamental evolutionary processes. Lack of such mechanistic understanding is an important problem because, without it, we cannot understand how perceivably minor changes in gene structure and function can significantly impact phenotypic differences between species and/or speciation itself. Thus, a growing knowledge of myostatin molecular genetics and bioactivity, in different tissues and in different organisms, has the potential to impact science as well as society.

## 2.9. Protein structure

Protein structure is the three dimensional arrangement of atoms in the protein molecule. Proteins are polymers, specifically polypeptides formed from sequences of monomer amino acids. By convention, a chain under 40 amino acids is considered peptide, rather than protein (Stoker, 2015). To be able to perform their biological functions, proteins fold into one or more specific spatial conformation driven by a number of non-covalent interactions such as hydrogen bonding, ionic interaction, van der waals forces and hydrophobic packing. In order to understand the functions of protein at molecular level, it is often important to predict their three dimensional structures.

Protein structure range in size from tens to several thousands of amino acids (Brocchieri and Karlin, 2005). There are four distinct levels of protein structures which includes; primary, secondary, tertiary and quaternary structures of proteins.

**The primary structure** of protein refers to the linear sequence of amino acids in the polypeptide chain. The primary structure is held together by covalent bonds such as peptide bonds, which are made during the process of protein biosynthesis or translation. Usually, the two ends of the chain are referred to as the amino terminus (N- terminus) and carboxyl terminus (C- terminus) based on the nature of free group on each extremity. The residues are usually counted starting from the N- terminus end (NH2- group), which is the end when amino group is involved in the peptide bond.

**The secondary structure** of proteins refers to the highly regular local sub-structures on the actual polypeptide backbone chain. The two main types of secondary structure of protein are the alpha helix and the beta strand or sheet. These were suggested by Linus Pauling and coworkers in 1951. Both Ŭhelix and ɓ sheet represent a way of saturating all the hydrogen bond donors and receptors in the peptide backbone.

**Tertiary structure** of protein is the three dimensional structure of monomeric and multimeric protein molecules. The alpha helices and beta sheets are folded into a compact globular structure. The folding is driven by the non-specific hydrophobic interactions, burial of hydrophobic residues from water but the structure is stable only when the parts of a protein domain are locked into place by specific tertiary interaction.

**Quaternary structure** of protein is the three dimensional structure of a multi subunit protein and how the subunits fit together. The quaternary is stabilized by the same non-covalent interactions and disulfide bonds as the tertiary structure.

## 2.10. Polymorphism

Generally, polymorphism in biology and zoology is the occurrence of two or more morphs or forms, also referred to as alternative phenotypes. But in this context, we are interested in the changes that occur in the gene (genetic polymorphism). Therefore, gene polymorphism is said to occur when more than one allele that geneôs locus within a population (Bio. online). A polymorphic variant of a gene may lead to abnormal expression or to the production of an abnormal form of the gene. For instance, a polymorphic variant of the enzyme CYP411 in which thymidine replaces cytosine at the geneôs nucleotide 8590 position encodes a CYP4A11 protein that substitutes phenylalanine with serine at the proteinôs amino acid position 434. This variant protein has reduced enzyme activity in metabolizing arachidonic acid to blood pressure regulating. Humans bearing this variant in the gene (CYP4A11) have an increased incidence of hypertension, ischemic stroke and coronary artery (Cardiol Rev, 2014).

## 2.11. Genetic diversity

Genetic diversity is the total number of genetic characteristics in a species gene make up. It serves as a way for population to adapt to their changing environment. With more variation, it is likely that some individuals in a population will posses varying alleles that will make them suitable in their environment. These individuals with suitable alleles are more likely to survive produce offspring bearing the same allele and the population will continue for generations because of the success of these individuals (National Bio. Inf., 2011). According to Toro and Caballero (2005), molecular data on within and between genetic diversity are essential for effective management of farm animalsô genetic resources. Also genetic diversity in livestock allows farmers to select stocks or develop new breeds in response to environmental changes, threats of disease, new knowledge of human nutrition requirements, changing market conditions

and societal needs (FAO, 2000). Genetic diversity and species diversity are interdependent and delicate. Changes in species diversity lead to changes in environment leading to adaptation of the remaining species. A change in genetic diversity such as in loss of species leads to a loss of biological diversity (National Bio. Inf., 2011). Loss of genetic diversity in domestic animal population has also been studied and attributed to the extension of markets and economic globalization (Tisdell, 2003: Frankham and Richard, 2005).

### 2.12. Phylogenetics and phylogenetic tree

According to merriam Webster dictionary, phylogenetics is the study of evolutionary history and relationships among individuals or groups of organism (e.g. Species or populations). These relationships are discovered through phylogenetic inference methods that evaluate observed heritable traits such as DNA sequences or morphology under a model of evolution of these traits. The result of this analysis is what we call phylogeny (or phylogenetic tree). Phylogenetic tree is the hypothesis about the history of evolution relationships (www.Bio.nline.org). Phylogenetic analysis has become central to understanding biodiversity or genetic diversity, evolution, ecology and genomes. Phylogenetic tree or evolutionary tree is a branching diagram or tree showing the inferred evolutionary relationships among various biological species- their phylogeny based upon similarities and differences in the physical or genetic characteristics. According to Felsenstein (2004), the phylogenetic tree composed with a nontrival number of sequences are constructed using computational phylogenetic methods. Distance-matrix methods such as neigbour joining or UPGMA which calculate genetic distance from multiple sequence alignment are simplest to implement. Many sequence alignment methods such as clustalW also create trees by using the simpler algorithms (i.e. those based on distance) of tree construction. Another simple method is maximum parsimony, it estimates phylogenetic tree but implies an implicit model of evolution (i.e. parsimony). More advanced methods use the optimality criterion of maximum likelihood, often within a Bayesian framework and apply an explicit model of evolution to phylogenetic tree estimation (Felsenstein, 2004). Identifying the optimal tree using many of these techniques is very hard, heuristic search and optimization method are used in combination with tree-scoring functions to identify a reasonably good tree that fits the data.

## 2.13. Bioinformatics

### 2.13.1. Definition:

Bioinformatics is a broad term covering the use of computer algorithms to analyze biological data. Differs from ñcomputational biologyò in that while computational biology is the use of computer technology to solve a single, hypothesis-based question (Mgbeahuruike, 2015), bioinformatics is the use of computerized statistical analysis to make statistical or comparative inferences

### 2.13.2. Aims of bioinformatics

The aims of bioinformatics are threefold. First, at its simplest bioinformatics organises data in a way that allows researchers to access existing information and to submit new entries as they are produced, e.g. the Protein Data Bank for 3D macromolecular structures (Berman *et al*, 2000). While data-curation is an essential task, the information stored in these databases is essentially useless until analysed. Thus the purpose of bioinformatics extends much further. The second aim is to develop tools and resources that aid in the analysis of data. For example, having sequenced a particular protein, it is of interest to compare it with previously characterized sequences. This needs more than just a simple text-based search and programs such as FASTA (Pearson and Lipman, 1988) and PSI-BLAST (Altschul *et al*, 1997) must consider what comprises a biologically significant match. Development of such resources dictates expertise in computational theory as well as a thorough understanding of biology. The third aim is to use these tools to analyze the data and interpret the results in a biologically meaningful manner. Traditionally, biological studies examined individual systems in detail, and frequently compared those with a few that are related. In bioinformatics, we can now conduct global analyses of all the available data with the aim of uncovering common principles that apply across many systems and highlight novel features.

### 2.13.3. Common uses of bioinformatics

1. Sequence analysis;

a. Geneticists/ molecular biologists analyze genome sequence information to understand disease processes.

b. Crystallographers/ biochemists design drugs using computer-aided tools.

c. Geneticists obtain information about the evolution of organisms by looking for similarities in gene sequences.

1. Ecology and population studies.
2. Bioinformatics is used to handle large amounts of data obtained in population studies.

**Table 1: List of URLs for the databases that are used in bioinformatics**

| Database | URL |
|---|---|
| **Protein sequence (primary)** | |
| SWISS-PROT | www.expasy.ch/sprot/sprot-top.html |
| PIR-International | www.mips.biochem.mpg.de/proj/protseqdb |
| **Protein sequence (composite)** | |
| OWL | www.bioinf.man.ac.uk/dbbrowser/OWL |
| NRDB | www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein |
| **Protein sequence (secondary)** | |
| PROSITE | www.expasy.ch/prosite |
| PRINTS | www.bioinf.man.ac.uk/dbbrowser/PRINTS/PRINTS.html |
| Pfam | www.sanger.ac.uk/Pfam/ |
| **Macromolecular structures** | |
| Protein Data Bank (PDB) | www.rcsb.org/pdb |
| Nucleic Acids Database (NDB) | ndbserver.rutgers.edu/ |
| HIV Protease Database | www.ncifcrf.gov/CRYS/HIVdb/NEW_DATABASE |
| ReLiBase | www2.ebi.ac.uk:8081/home.html |
| PDBsum | www.biochem.ucl.ac.uk/bsm/pdbsum |
| CATH | www.biochem.ucl.ac.uk/bsm/cath |
| SCOP | scop.mrc-lmb.cam.ac.uk/scop |
| FSSP | www2.embl-ebi.ac.uk/dali/fssp |
| **Nucleotide sequences** | |
| GenBank | www.ncbi.nlm.nih.gov/Genbank |
| EMBL | www.ebi.ac.uk/embl |
| DDBJ | www.ddbj.nig.ac.jp |
| **Genome sequences** | |
| Entrez genomes | www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome |
| GeneCensus | bioinfo.mbb.yale.edu/genome |
| COGs | www.ncbi.nlm.nih.gov/COG |
| **Integrated databases** | |
| InterPro | www.ebi.ac.uk/interpro |
| Sequence retrieval system (SRS) | www.expasy.ch/srs5 |
| Entrez | www.ncbi.nlm.nih.gov/Entrez |

# 3.0 CHAPTER THREE: MATERIALS AND METHODS

## 3.1 Sequences of Species

A total of thirty seven (37) myostatin nucleotides with complete coding sequences comprising goats (26), sheep (2) and cattle (9) and their corresponding amino acid sequences were retrieved from the GenBank (NCBI) (www.ncbi.nlm.nih.gov). The Genbank accession numbers of the sequences and sequence variations are shown in Table 2.

**Table 2. Accession number, sequence length and complete coding region of myostatin gene of goats, sheep and cattle**

| Species | Accession No | No of Base Pairs (bp) | Complete Coding |
|---|---|---|---|
| Cattle | AY160688.1 | 1,128 | Complete coding sequence |
| | AF320998.1 | 6,691 | Complete coding sequence |
| | GQ184147.1 | 1,140 | Complete coding sequence |
| | JQ711180.1 | 7,831 | Complete coding sequence |
| | AF019620.1 | 1,128 | Complete coding sequence |
| | AF019761.1 | 1,128 | Complete coding sequence |
| | AB076403.1 | 6,660 | Complete coding sequence |
| | NM_001001525.2 | 2,767 | Complete coding sequence |
| | AY794986.1 | 6,686 | Complete coding sequence |
| Sheep | AF019622.1 | 1,128 | Complete coding sequence |
| | NM_001009428.1 | 1,128 | Complete coding sequence |
| Goat | JN662463.1 | 1,128 | Complete coding sequence |
| | EF591039.1 | 6,355 | Complete coding sequence |
| | JN012228.1 | 5,211 | Complete coding sequence |
| | DQ167575.2 | 5,211 | Complete coding sequence |
| | AY436347.1 | 1,128 | Complete coding sequence |
| | EF588035.1 | 5,217 | Complete coding sequence |
| | EF588034.1 | 5,217 | Complete coding sequence |
| | EF588033.1 | 5,217 | Complete coding sequence |
| | EF588032.1 | 5,217 | Complete coding sequence |
| | EF588031.1 | 5,211 | Complete coding sequence |
| | EF588030.1 | 5,211 | Complete coding sequence |
| | EF588029.1 | 5,211 | Complete coding sequence |
| | EF588028.1 | 5,211 | Complete coding sequence |
| | EF588027.1 | 5,211 | Complete coding sequence |
| | EF588026.1 | 5,211 | Complete coding sequence |
| | EF588025.1 | 5,211 | Complete coding sequence |
| | EF588024.1 | 5,211 | Complete coding sequence |
| | EF588023.1 | 5,211 | Complete coding sequence |
| | EF588022.1 | 5,211 | Complete coding sequence |
| | EF588021.1 | 5,211 | Complete coding sequence |
| | EF588020.1 | 5,211 | Complete coding sequence |
| | EF588019.1 | 5,211 | Complete coding sequence |
| | EF588018.1 | 5,211 | Complete coding sequence |
| | EF588017.1 | 5,211 | Complete coding sequence |
| | GU377303.1 | 1,128 | Complete coding sequence |
| | JX840482.1 | 1,128 | Complete coding sequence |

The two sheep sequences were the only available MSTN genes with complete coding sequence.

### 3.2. Sequence Alignment and Translation

Sequence alignment, translation and comparison of the myostatin gene of the various species were done with ClustalW as described by (Larkin *et al*., 2007) using IUB substitution matrix, gap open penalty of 15 and gap extension penalty of 6.66.

### 3.3. Functional Analysis

The single nucleotide polymorphisms (SNPs) and their related protein sequences of myostatin gene of goats, sheep and cattle were retrieved from the National Center for Biotechnology Information (NCBI) database, public domain (Tay *et al*., 2004; Zhou *et al*., 2008) and amino acid substitutions of goats, sheep and cattle of the present study to determine the beneficial or otherwise of the amino acid mutations. *In silico* functional analysis of missense mutations was obtained using protein variation effect analyzer (PROVEAN) with threshold value of -2.5. PROVEAN collects a set of homologous and distantly related sequences from the NCBI protein database (released August 2011) using basic local alignment search tool (BLASTP) (ver.2.2.25) with an E-value threshold of 0.1. The sequences are clustered based on a sequence identity of 80% to remove redundancy using the clustal database at high identity with tolerance (CD-HIT) program (ver.4.5.5) (Li and Godzik, 2006). If the PROVEAN score is smaller than or equal to a given threshold, the variation is predicted as deleterious (Choi *et al*., 2012).

### 3. 4. Phylogenetic trees analysis

Neighbor-Joining (NJ) trees were constructed using P-distance method, complete deletion and gap/missing data treatment. The construction was done on the basis of genetic distances, depicting phylogenetic relationships among the myostatin nucleotide sequences of the goats, sheep and cattle. The reliability of the trees was calculated by bootstrap confidence values with

1000 bootstrap iterations using molecular evolutionary genetics analysis version 6.0 (MEGA 6.0 software) (Tamura *et al*., 2011). However, the Poisson method was used for the amino acid sequences of the three species. UPGMA trees for the myostatin gene were constructed with consensus nucleotide and amino acid sequences of the three species. All sequences were trimmed to equal length of 1128 (nucleotides) and 375 (amino acids) corresponding to the same region of myostatin gene before generating the neigbour joining (NJ) and unweighted pair group method with averages (UPGMA) trees, respectively.

**3.5. Physico-chemical properties of myostatin gene**

ProtParam Tool developed by Gasteiger was used for the computation of various physical and chemical properties of the myostatin gene using amino acid sequences. The computational parameters included molecular weight, theoretical pI (isoelectric point), amino acid composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathy (GRAVY) (Gasteiger, 2005).

**3.6. The prediction of secondary structure of myostatin gene**

The amino acid sequences of myostatin gene were further subjected to secondary structure prediction using the expert protein analysis system ExPASyôs SOPMA tool. SOPMA (self optimized prediction from multiple alignment) **i**s an improved SOP method. It predicts 69.5% of amino acids for a 3 state description of the secondary structure (a helix, b sheets and coil). It predicts the secondary structure by consensus prediction from multiple alignments.

**3.7. The prediction of transmembrane domain of myostatin gene**

The myostatin sequences of goats, sheep and cattle were also subjected to transmembrane domain identification using TMbase - A Database of Membrane Spanning Protein Segments

(Hofmann and Stoffel, 1993). TMbase is mainly based on SwissProt, but contains information from other sources as well.

### 3.8. The prediction of tertiary structure of myostatin gene

The Phyre and Phyre2 servers were used to predict the 3Dimentional structure of myostatin gene of the three species (cattle, sheep and goat) in the present study. These servers predict the three-dimensional structure of a protein sequence using the principles and techniques of homology modeling (Kelley and Sternberg, 2009). Currently, the most powerful and accurate methods for detecting and aligning remotely related sequences rely on profiles or the hidden Markov models (HMMs). 3DligandSite was used to predict the binding site of the 3D structure of the myostatin gene. Phyre2 is coupled to the 3DligandSite server for protein binding site prediction (Wass *et al*., 2010).

# 4.0 CHAPTER FOUR: RESULTS AND DISCUSSION

## 4.1 The effects of genetic polymorphism

The functional analysis of the effect of amino acid substitutions in goats, sheep and cattle is shown in Tables 3, 4 and 5 respectively.

**Table 3: Functional analysis of coding nsSNP of the myostatin gene of goats using PROVEAN**

| Variant | PROVEAN Score | Prediction |
|---|---|---|
| K153F | -4.692 | Deleterious |
| M180L | 0.117 | Neutral/beneficial |
| T240K | -4.028 | Deleterious |
| L270Q | -4.277 | Deleterious |
| S276N | -2.094 | Neutral/beneficial |
| S279K | -1.893 | Neutral/beneficial |

. Default threshold is -2.5, that is
-Variants with a score equal to or below -2.5 are considered óôdeleteriousôõ
-Variants with a score above -2.5 are considered óôbeneficialôõ

**Table 4: Functional analysis of coding nsSNP of the myostatin gene of sheep using PROVEAN**

| Variant | PROVEAN Score | Prediction |
|---|---|---|
| K178N | -1.472 | Neutral/beneficial |
| V151L | -1.594 | Neutral/beneficial |
| S191P | -2.604 | Deleterious |
| W203L | -10.440 | Deleterious |
| S205C | -3.278 | Deleterious |
| N222Q | -4.900 | Deleterious |
| D231C | -5.942 | Deleterious |
| E247D | -1.253 | Neutral |
| R303G | -4.226 | Deleterious |
| Q329L | -2.472 | Neutral/beneficial |
| G355I | -1.759 | Neutral/beneficial |

. Default threshold is -2.5, that is
-Variants with a score equal to or below -2.5 are considered óôdeleteriousôõ
-Variants with a score above -2.5 are considered óôbeneficialôõ

**Table 5: Functional analysis of coding nsSNP of the myostatin gene of cattle using PROVEAN**

| Variant | PROVEAN Score | Prediction |
|---|---|---|
| R98H | -0.625 | Neutral/beneficial |
| D110L | -4.514 | Deleterious |
| I119T | -2.484 | Neutral/beneficial |
| S125M | -0.814 | Neutral/beneficial |
| G133S | -1.080 | Neutral/beneficial |
| I158A | -2.860 | Deleterious |
| T165N | 0.322 | Neutral/beneficial |
| R175V | -4.330 | Deleterious |
| K193V | -3.628 | Deleterious |
| S205G | -2.613 | Deleterious |
| P301L | -9.240 | Deleterious |
| H328T | -1.298 | Neutral/beneficial |
| F353N | -7.398 | Deleterious |

. Default threshold is -2.5, that is
-Variants with a score equal to or below -2.5 are considered ôdeleteriousõ
-Variants with a score above -2.5 are considered ôbeneficialõ

For goats, three amino acid substitutions (M180L, S276N and S279K) were obtained and they all appeared to be beneficial. The rest three (K153F, T240K and L270Q) were found to be harmful to protein function. Five amino acid mutations from the deduced amino acid sequences of sheep have their score below the default threshold (K178N, V151L, E247D, Q329L and G355I) appeared beneficial while the remaining six (S191P, W203L, S205C, N222Q, D231C and R303G) appeared harmful. In cattle, six substitutions (R98H, I119T, S125M, G133S, T165N and H328T) appeared not to impair the function while the rest seven (D110L, I158A, R175V, K193V, S205G, P301L and F353N) were deleterious.

Myostatin gene (GDF8) is important in the physiology of stock animals because its product produces a direct effect on muscle development and consequently also on meat production (Mota *et al*., 2006). It carries this out by acting as a negative regulation of skeletal muscle growth and keeps the skeletal musculature within appropriate proportions (Tay *et al*., 2004). Gene polymorphism between species as observed in form of varying amino acid non-synonymous substitutions in the present study may play an important role in the gene translation and

regulation, thereby having a positive impact on selection and consequently on muscular development (Crisa *et al*., 2003).

The beneficial amino acids obtained in this study may be exploited as possible markers for growth and development in goats, sheep and cattle. The mutations associated with MSTN are typically recent (Tellam *et al*., 2012). The immediate future of the meat industry will see the increasing exploitation of natural genetic variation contributing to muscling. Enhanced muscling is often associated with morphology change of the animal. This suggests a common change in developmental programming that scales body shape to accommodate enhanced skeletal muscle structure. Changed developmental trajectories for skeletal muscle beginning in early life are also common and these may be linked with morphology changes. Large increases in muscularity are typically associated with a shift toward fast twitch glycolytic fibers, leanness, and poorer eating quality attributes. The reasons for these changes are unclear, although both myoblasts and preadipocytes have a common progenitor cell, suggesting a developmental link (Tellam *et al*., 2012). The incorporation of causal genetic variations into genomic selection strategies will enhance their accuracy and robustness, while allowing targeted selection to achieve more rapid genetic improvement. In the medium term, the discovery of developmental and biochemical pathways contributing to enhanced muscling will open new opportunities for the use of novel and acceptable biochemical and immunological interventions that may play a significant and complementary role to genetic selection in the meat industry.

### 4.2 Genetic diversity (evolutionary history and amino acid composition).

The phylogeny based on nucleotide and amino acid sequences of myostatin gene revealed mainly species-wise clustering (Figures 6 and 7).

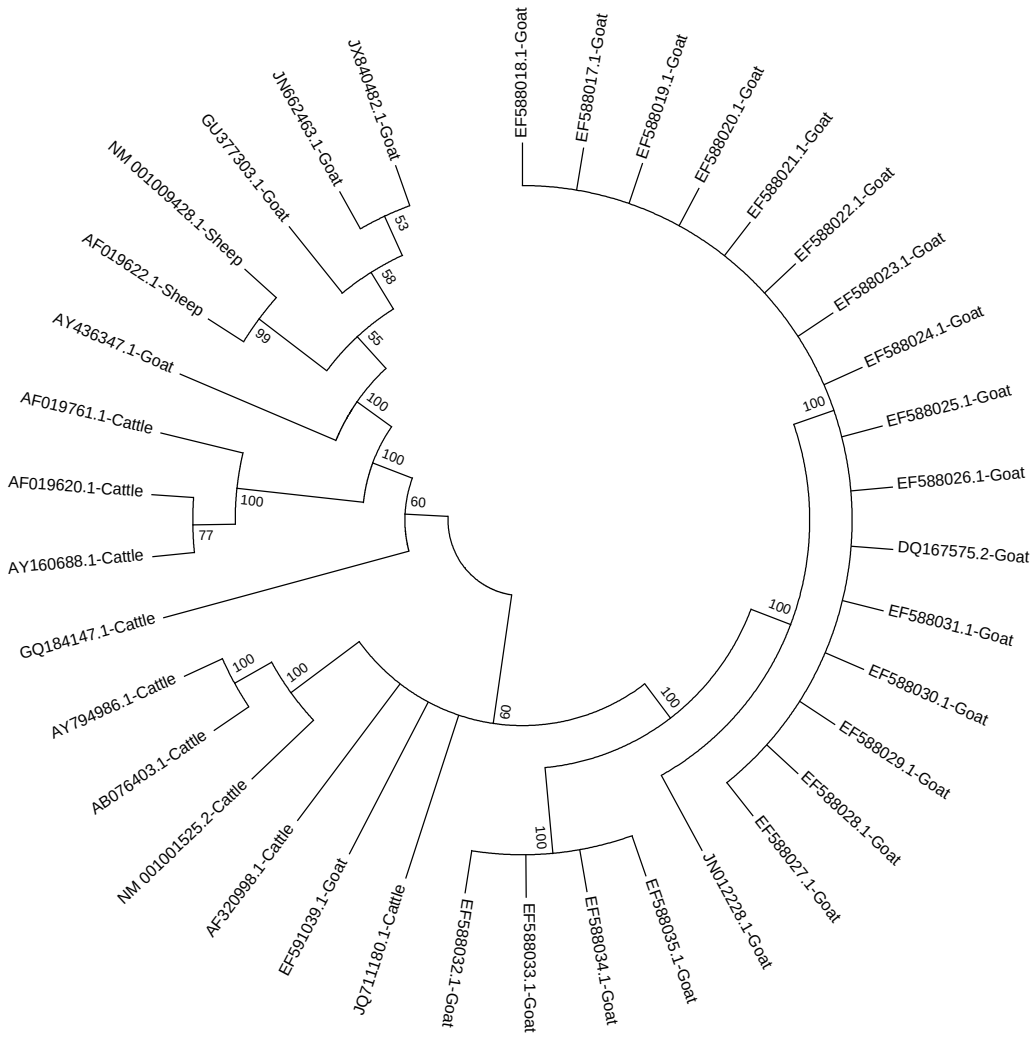



**Figure 6: The evolutionary history indicating phylogeny involving 37 Nucleotide sequences of the myostatin gene in three Bovid species.** The optimal tree with the sum of branch length = 3.24511719 is shown. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. The analysis involved **37 nucleotide sequences**. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated.

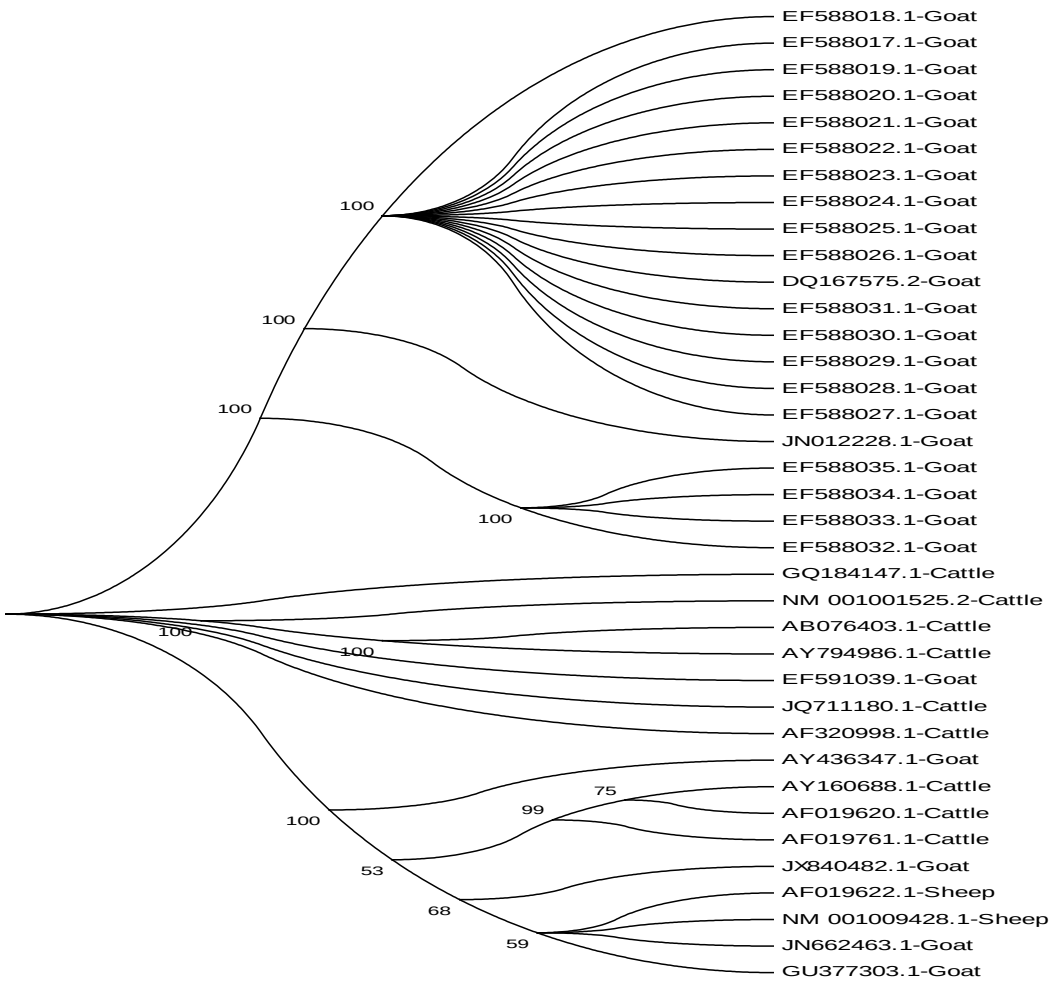



**Figure 7: The evolutionary history indicating phylogeny involving 37 Amino acid sequences of the myostatin gene in three Bovid species.** The optimal tree with the sum of branch length = 10.76916840 is shown. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved **37 amino acid sequences**. All positions containing gaps and missing data were eliminated.

The genetic relationships of myostatin Bovidae subfamily members of goats, sheep and cattle shown in the UPGMA trees derived from consensus sequences revealed that goats and sheep were closer at the myostatin locus compared to cattle (Figures 8 and 9).

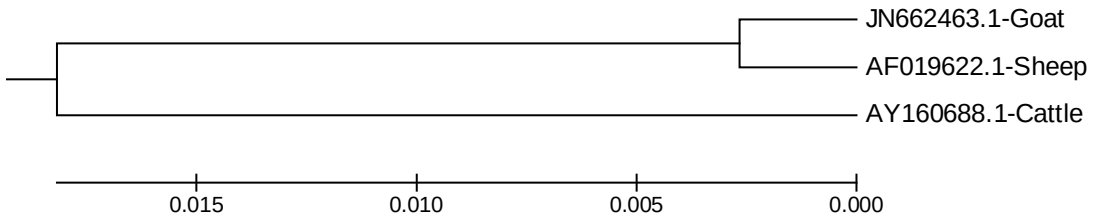



**Figure 8: UPGMA tree indicating the consensus Nucleotide sequences of the myostatin gene in three Bovid species.** The optimal tree with the sum of branch length = 0.03900709 is shown. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. All positions containing gaps and missing data were eliminated.

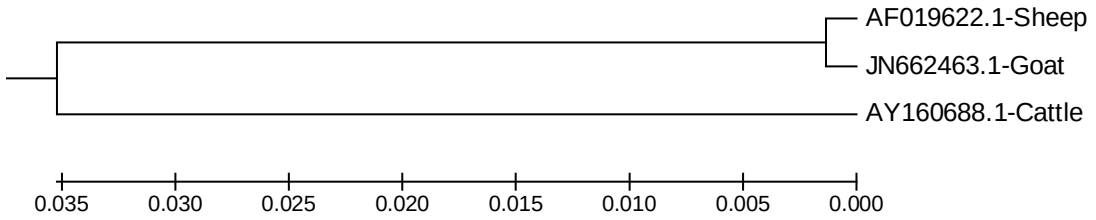



**Figure 9: UPGMA tree indicating the consensus Amino acid sequences of the myostatin gene in three Bovid species.** The optimal tree with the sum of branch length = 0.07175860 is shown. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated.

The close similarity of a gene among ruminants may be ascribed to recent separation in evolutionary process and/or similar selection pressure which the ruminants have suffered during evolution. According to Aleriza *et al.* (2014), this result shows that this gene has a high degree of conservation during evolution of various species, which implies that MSTN, is an essential factor in biological muscle control. The species wise clustering might be due to species specific residues (Takahashi and Nei, 2000; Yakubu *et al*., 2013) and such patterns of the sequences may be explained by gene conversion and balancing selection. The genetic relationships of myostatin gene of Bovidae subfamily members of goat, sheep, and cattle shown in the UPGMA phylogenetic trees were in accordance with the well-known evolutionary history of Bovidae subfamily speciation.

The amino acid composition of the myostatin proteins of goats, sheep and cattle is shown in (Table 6.).

**Table 6: Amino acid percentage composition of the myostatin proteins of goats, sheep and cattle.**

| Amino acid % | Goats | Sheep | Cattle |
|---|---|---|---|
| Alanine | 4.0 | 4.0 | 4.8 |
| Arginine | 4.8 | 4.5 | 4.8 |
| Aspargine | 4.5 | 4.5 | 4.0 |
| Aspartate | 6.1 | 6.1 | 6.1 |
| Cysteine | 3.5 | 3.5 | 3.5 |
| Glutamine | 4.8 | 4.8 | 4.5 |
| Glutamate | 6.7 | 6.7 | 6.7 |
| Glycine | 5.3 | 5.3 | 5.6 |
| Histidine | 1.9 | 1.9 | 1.6 |
| Isoleucine | 6.4 | 6.4 | 6.7 |
| Leucine | 9.9 | 9.9 | 9.9 |
| Lysine | 8.3 | 8.3 | 7.2 |
| Methionine | 2.1 | 2.1 | 2.1 |
| Phenylalanine | 3.7 | 3.7 | 3.7 |
| Proline | 6.4 | 6.4 | 6.4 |
| Serine | 5.6 | 5.6 | 5.9 |
| Threonine | 5.6 | 5.6 | 6.1 |
| Tryptophan | 1.3 | 1.6 | 1.6 |
| Tyrosine | 3.2 | 3.2 | 3.2 |
| Valine | 5.9 | 5.9 | 5.6 |

Goats and sheep appear more similar in their amino acid contents compared to cattle. However, the distribution pattern was the same for the three species in respect of aspartate, cysteine, glutamate, leucine, methionine, phenylalanine, proline and tyrosine.

It was observed that the three species investigated were rich in leucine amino acid. As Leucine is an aliphatic, hydrophobic amino acid, it prefers substitution with other amino acids of the same type. Therefore, MSTN amino acid sequences of goat, sheep and cattle are more hydrophobic and less hydrophilic (Aleriza *et al.,* 2014). Being hydrophobic, Leucine prefers to be buried in protein hydrophobic cores. It also shows a preference for being within alpha helices more so than in beta strands. The Leucine side chain is very non-reactive, and is thus rarely directly involved in protein function, though it can play a role in substrate recognition. In particular, hydrophobic amino acids can be involved in binding/recognition of hydrophobic ligands such as lipids (Betts, 2003).

### 4.3 The physical and chemical properties of myostatin protein.

The physico-chemical characteristics of myostatin proteins of goats, sheep and cattle predicted by ProtParam are shown in (Table 7).

**Table 7: Physico-chemical characteristics of myostatin proteins of goats, sheep and cattle predictedby protparam**

| Specie | No of AA | Mol. Wt | pI | Q | EC | Half life | II | AI | GRAVY |
|---|---|---|---|---|---|---|---|---|---|
| Goats | 375 | 42797.3 | 7.47 | +ve | 46130 | 30 | 44.82 | 84.45 | -0.421 |
| Sheep | 375 | 42827.3 | 7.01 | neutral | 51630 | 30 | 44.87 | 84.45 | -0.411 |
| Cattle | 375 | 42550.9 | 6.14 | -ve | 51630 | 30 | 40.48 | 85.52 | -0.337 |

AA=amino acid, Mol. Wt= molecular weight, pI=isoelectric point, Q=net charge, EC= extinction coefficient, II=instability index, AI= aliphatic index, GRAVY=grand average of hydropathicity.

Apart from extinction coefficient (sheep and cattle), half life (goats, sheep and cattle) and aliphatic index (goats and sheep), other parameters varied from one species to another.

Isoelectric point is the pH at which a protein carries no net charge. The isoelectric point is of significance in protein purification because it is the pH at which solubility is often minimal and at which mobility in an electro focusing system is zero (and therefore the point at which the protein will accumulate). This measure indicates how much light is absorbed by a protein at a particular wavelength. The extinction coefficient of a protein at 280 nm depends almost exclusively on the number of aromatic residues, particularly tryptophan. The half life of a protein is the time it takes before only half of the protein pool for that particular protein is left. The half life of proteins is highly dependent on the presence of the N-terminal amino acid, thus overall protein stability (Ugbo *et al*., 2015).

### 4.4. The tertiary and secondary protein structures of MSTN

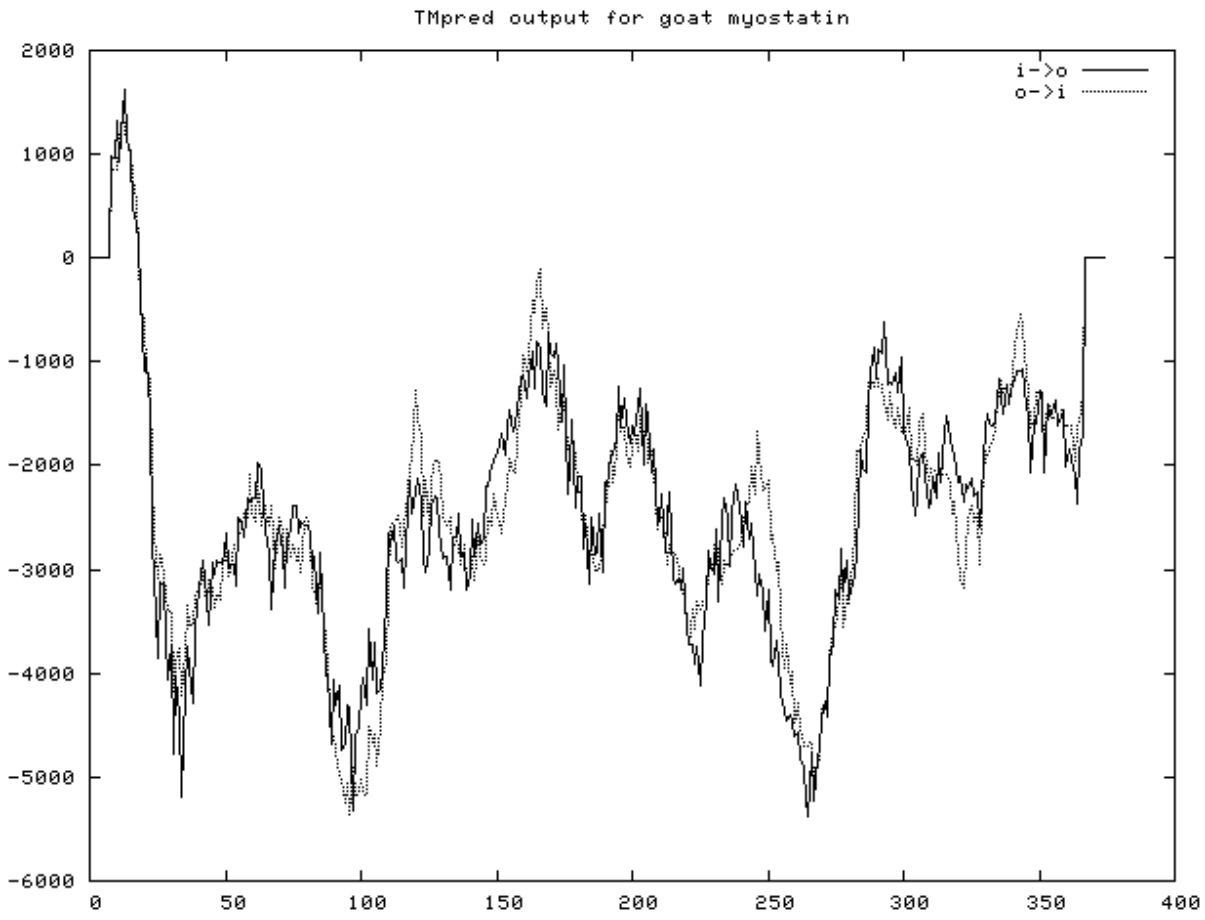
The secondary structure prediction of myostatin protein represented in percentages are in (table 8).

**Table 8: Prediction of secondary structures of myostatin proteins of goats, sheep and cattle.**

| Species | Alpha helix (%) | Extended strand (%) | Beta turn (%) | Random coil (%) |
|---|---|---|---|---|
| Caprine | 22.67 | 25.87 | 8.27 | 43.20 |
| Ovine | 22.67 | 25.87 | 8.27 | 43.20 |
| Bovine | 23.20 | 25.33 | 7.47 | 44.00 |

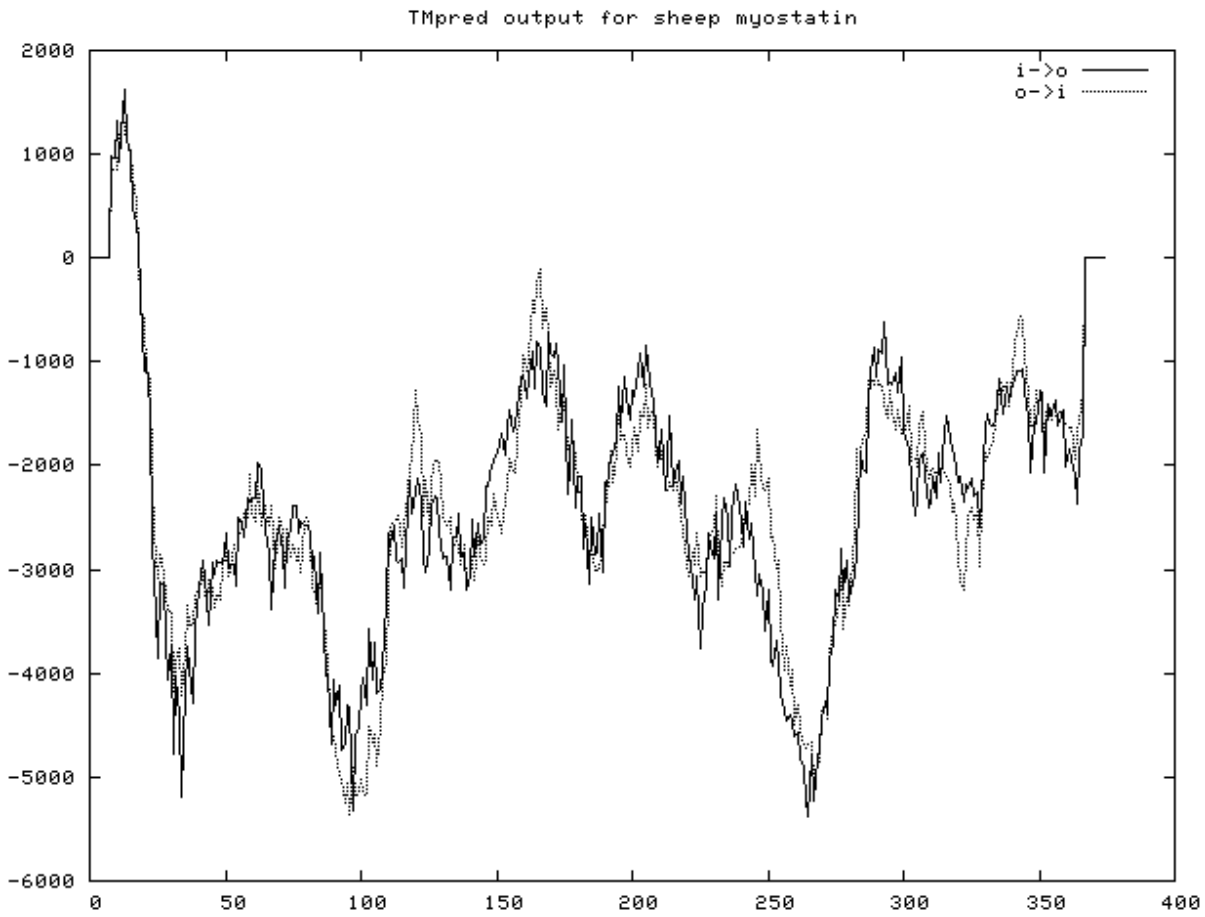
Parameters: Window width= 17, Similarity threshold=8, Number of states= 4.
The bovine myostatin protein showed highest alpha helix (23.20%) and random coil (44.00%). However, the extended strand (25.87) and beta turn (8.27) predictions were higher in both caprine and ovine species (Table 8).

Prediction of transmembrane proteins of myostatin gene of goats, sheep and cattle indicated that the transmembrane segments used were significant (graph 1, 2 and 3). The scores for the respective species were above 500 the cut-off point (for goats, Inside to outside helices= 1621, Outside to inside helices= 1394; for sheep, Inside to outside helices= 1621, Outside to inside helices= 1394; for cattle, Inside to outside helices= 1422, Outside to inside helices= 1145).



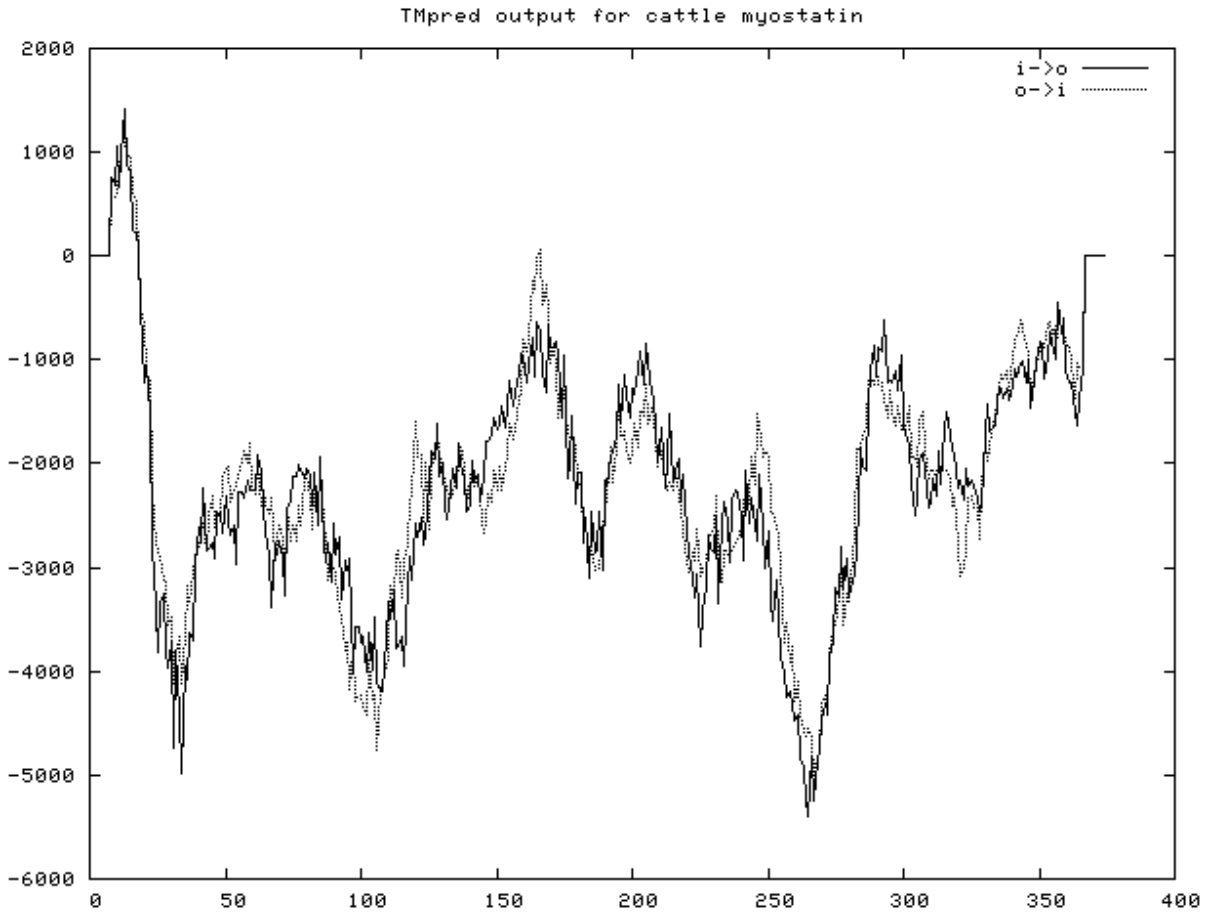

**graph 1: Prediction plot of transmembrane topology of goat myostatin protein**

io=inside to outside, oi=the opposite
inside' means normally the cytoplasmic face
outside' the lumenal face of the membrane depending on the organelle




**graph 2: Prediction plot of transmembrane topology of sheep myostatin protein**

io=inside to outside, oi=the opposite
inside' means normally the cytoplasmic face
outside' the lumenal face of the membrane depending on the organelle




**graph 3: Prediction plot of transmembrane topology of cattle myostatin protein**

io=inside to outside, oi=the opposite
inside' means normally the cytoplasmic face
outside' the lumenal face of the membrane depending on the organelle

Three hundred and twenty four (324) residues (86% of goat sequence) have been modelled with 100.0% confidence by the single highest scoring template. The percentage identity with the template myostatin protein was 86.0%. 324 residues (86% of sheep sequence) have been modelled with 100.0% confidence by the single highest scoring template. The percentage identity with the template myostatin protein was equally 86.0% while 323 residues (86% of cattle sequence) have been modelled with 100.0% confidence by the single highest scoring template. The percentage identity with the template myostatin protein was also 86.0% (plates 1, 2 and 3). The amino acids residues predicted to form part of the binding site of goat myostatin protein 3D structure as well as their residue numbers were Ser (105), Asp (267), Phe (268) Gly (269), Asp (273), Cys (281), Cys (281), Arg (371), Cys (372) and Gly (373), respectively. Those for sheep were Ser (105), Asp (267), Phe (268) Gly (269), Asp (273), Cys (281), Cys (282), Arg (371), Cys (372) and Gly (373), respectively. The amino acids residues predicted to form part of the binding site of cattle myostatin protein 3D structure as well as their residue numbers were Leu (106), Asp (267), Phe (268) Gly (269), Asp (273), Cys (281), Cys (281), Arg (371), Cys (372) and Gly (373), respectively.

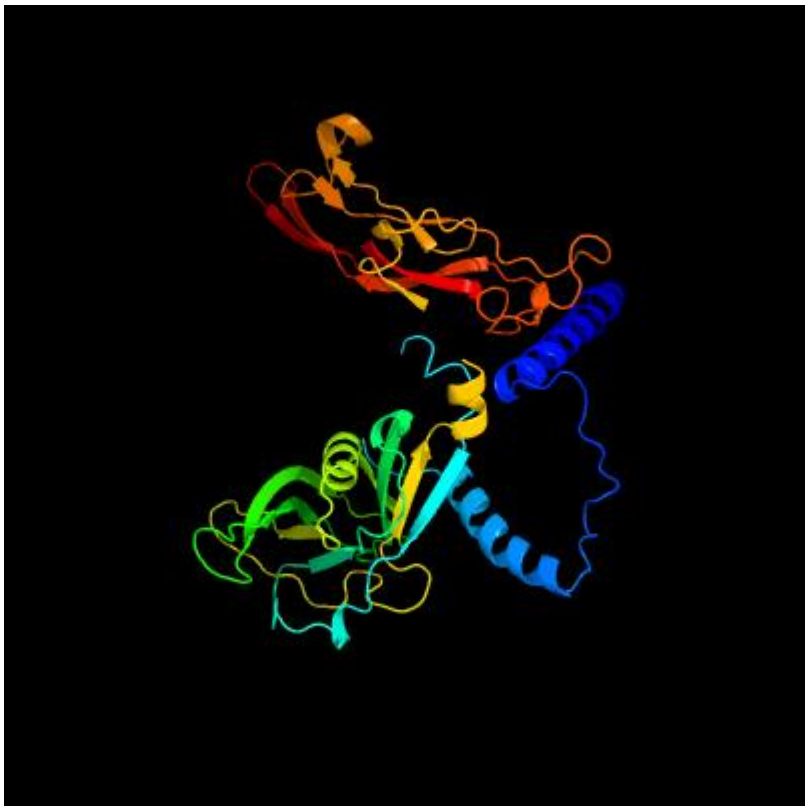


**plate 1: Schematic 3D structure of goat myostatin protein**
Image coloured by rainbow N Ÿ C terminus

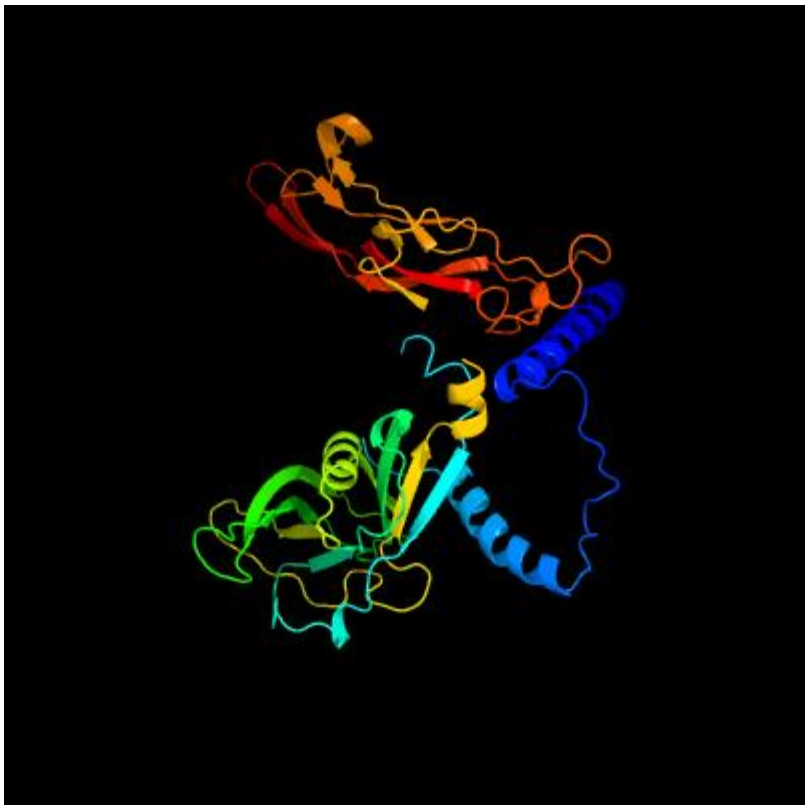


**Plate 2: Schematic 3D structure of sheep myostatin protein**
Image coloured by rainbow N Ÿ C terminus

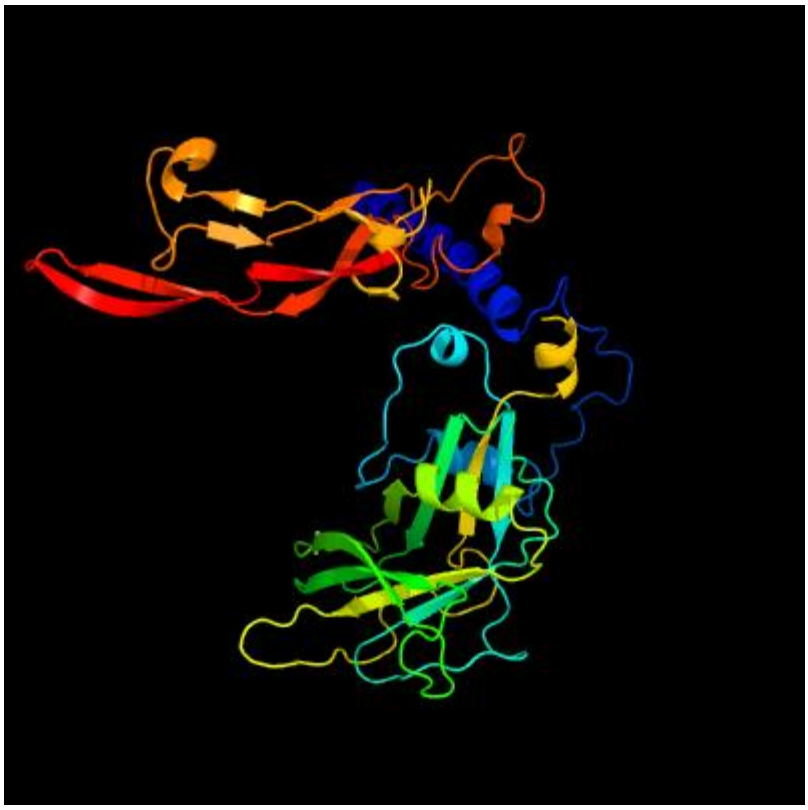


**Plate 3: Schematic 3D structure of cattle myostatin protein**
Image coloured by rainbow N Ÿ C terminus

Many important biological processes such as cell signaling, transport of membrane-impermeable molecules, cellï cell communication, cell recognition and cell adhesion are mediated by membrane proteins (Jones, 2007). Although there has been some recent progress in predicting the full 3-D structure of transmembrane proteins (e.g. Yarov-Yarovoy*et al*., 2006), the most widely applied prediction technique for these proteins is to determine the transmembrane topology, i.e. the insideï outside location of the N and C termini relative to the cytoplasm, along with the number and sequence locations of the membrane spanning regions. Knowing myostatin membrane proteinôs topology can be a significant step toward inferring both its structure and function.

Determining the structure and function of a novel protein is a cornerstone of many aspects of modern biology. The accuracy of protein structure prediction depends critically on sequence similarity between the query and template as observed in the present study. If a template is detected with >30% sequence identity to the query, then usually most or all of the alignment will be accurate and the resulting relative positions of structural elements in the model will be reliable (Kelley *et al*., 2015). The practical applications of myostatin protein structure prediction in goats, sheep and cattle are of many and varied, including guiding the development of functional hypotheses about hypothetical proteins (Watson *et al*., 2005), improving phasing signals in crystallography (Qian *et al*., 2007) and selecting sites for mutagenesis (Rava and Hussain, 2007). Proteins often perform their function on ligands (e.g. enzyme substrates) or are regulated by them. The explosion of protein sequences from genome sequencing projects makes it essential for automated methods to predict ligand-binding sites. Further, protein structures are often solved in the absence of ligands, making it important to identify binding sites for such proteins (Wass *et al*., 2010). Residue properties are also commonly used to establish different specificity in similar proteins (Capra and Singh, 2008). Therefore, the identification of ligand-binding sites as observed in goats, sheep and cattle of the present study is important to make functional predictions especially for newly discovered myostatin sequences.

# 5.0 CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

## 5.1 Conclusion

There was a great genetic variation and polymorphism in the aligned sequences of myostatin gene within and across species. Computational analysis of non-synonymous mutations revealed some beneficial amino acid substitutions. The dendrograms obtained showed some form of proximity and differentiation in myostatin sequences within and among the mammalian species investigated. Thus, the present genetic information on the myostatin gene of goats, sheep and cattle will guide subsequent studies using in vivo experimental protocol to associate the observed beneficial SNPs in the three species with performance traits such as growth and carcass quality in goats, sheep and cattle in Nigeria.

## 5.2 Recommendation.

During the course of the study, some amino acid substitutions of myostatin gene of the Bovidae family were beneficial. Therefore, it is recommendable to use the beneficial amino acids to enhance the physiology of stock animals in breeding programs. Also, the beneficial amino acids variants obtained from the non-synonymous amino acid substitutions can be can be exploited as possible markers for growth and development in cattle, sheep and goat. Following the similarities seen in Bovidae family with ovine and caprine which showed great propinquity in evolutionary process; the amino acid percentage composition were the same in majority of the amino acid constituent of the MSTN gene while secondary, tertiary structures as well as membrane topology were difficult to distinguish in sheep and goat, thus, myostatin gene of goat and sheep can be substituted in breeding practice. MSTN gene sequences available in the Genebank were from exotic breeds, therefore, it is important to carry out research on local breeds of cattle, sheep and goat to make the sequences available in the Genebank so that comparative analysis may be possible with the already available sequences.

## REFERENCES


Aleriza E, Mostata M and Morteza S. (2014). Bioinformatics analysis of myostatin gene and protein in different species. International Journal of Bio Sciences (IJB), 5( 7); 5-11.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389-3402.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998). Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711ï 1714.

Amthor H, Otto A, Macharia R, McKinnell I, Patel K (2006) Myostatin imposes reversible quiescence on embryonic muscle precursors. Dev Dyn 235:672ï 680.

An XP, Wang JG, Hou JX, Zhao HB, Bai L, Li G, Wang LX, Liu XQ, Xiao WP, Song YX and Cao BY (2011). Polymorphism identification in the goat *MSTN* gene and association analysis with growth traits. Czech J. Anim. Sci. 56(12) 529-539.

Arif M, Kucuktas KH, Dunlam RA and Liu Z. (2002). Molecular Characterization and differential expression of the mysotation gene in channel catfish (Kt-alurus punctatus). Bioch. Biophys. Acta, 157 (5); 99-107.

Attisano L, Wrana JL (2002). Signal transduction by the TGF-ß superfamily. Science 296:1646ï 1647.

Bellinge RHS, Liberles DA, Iaschi SPA, O'Brien A, Tay GK. (2005). Myostatin and its implications on animal breeding: a review. Anim Genet;36:1-6.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. (2000). The Protein Data Bank. Nucleic Acids Res; 28(1):235-42.

Betts MJ. and Russell, RB. (2003).Amino acid properties and consequences of subsitutions. In *Bioinformatics for geneticists* edition MR Barnes and IC Gray Chichester: Wiley.

Biga PR, Roberts SB, Iliev DB, McCauley LA, Moon JS, Collodi P, Goetz FW (2005). The isolation, characterization, and expression of a novel GDF11 gene and a second myostatin form in zebrafish, *Danio rerio*. Comp Biochem Physiol B Biochem Mol Biol 141:218ï 230.

Bogdanovich S, Krag TO, Barton ER, Morris LD, Whittemore LA, Ahima RS, Khurana TS (2002). Functional improvement of dystrophic muscle by myostatin blockade. Nature 420:418ï 421.

Boman IA, Klemetsdal G, Blichfeldt T, Nafstad O, et al. (2009). A frameshift mutation in the coding region of the myostatin gene (MSTN) affects carcass conformation and fatness in Norwegian White Sheep (*Ovis aries*). Anim. Genet. 40: 418-422.

Broccheiri L and Karlin S. (2005). Protein length in eukaryotic and prokaryotic proteome. Nucleic Acid Research 33(10): 3390-3400.

Bromberg Y and Rost B. (2007) SNAP: predict effect of non-synonymous polymorphisms on function, Nucleic Acids Res, 35, 3823-3835.

Capra JA. and Singh, M. (2008). Characterization and prediction of residues determining protein functional specificity. *Bioinformatics*, 24 (13): 1473-1480.

Cardiol R. (2014). Pharmacogenomics in heart failure: where are we now and how can we reach clinical application? Doi: 10.1097/CRD.0000000000000028. 22(5)193-8.

Castelhano-Barbosa EC, Gabriel JE, Alvares LE, Monteiro-Vitorello CB, Coutinho LL (2005). Temporal and spatial expression of the myostatin gene during chicken embryo development. Growth Dev Aging 69:3ï 12.

Choi Y, Sims, GE, Murphy, S, Miller, JR. and Chan, AP. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688.

Cook SA, Matsui T, Li L, Rosenzweig A (2002). Transcriptional effects of chronic Akt activation in the heart. J Biol Chem 277:22528ï 22533.

Crisà A, Marchitelli, C, Savarese, MC. and Valentini, A. (2003). Sequence analysis of myostatin promoter in cattle. *Cytogenet Genome Res*., 102: 48ï 52.

Diel P, Schiffer T, Geisler S, Hertramp FT, Mosler S, Schulz S, Wintgens KF and Adler M. (2010). Analysis of the effects of androgens and training on mysotatin propeptide and follistation concentrations in blood and skeletal muscle using highly sensitive immuno PCR. Molecular and cellular endocrinology 330; 1-9.

Dredge BK and Darnell RB (2003). Nova regulates GABA(A) receptor ɔ2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. Mol Cell Biol 23:4687ï 4700.

Evans JJ, Shoemaker CA, Klesius PH (2000). In vivo and in vitro effects of benzothiazole on sheepshead minnow (*Cyprinodon variegatus*). Mar Environ Res 50:257ï 261

Fan B, Lkhagvadorj S, Cai W, Young J, et al. (2010). Identification of genetic markers associated with residual feed intake and meat quality traits in the pig. *Meat Sci.* 84: 645-650.

FAO (2000). World watch list for domestic animal diversity. Rome, Italy (Available at http://www.fao.org/docrep/009/x8750e/8750e00.GTTM). P 726 3rd ed.

Felsenstein J. (2004). Inferring phylogenies, Sinauer Assoc., pp.xx +664.

Figeac N, Daczewska M, Marcelle C, Jagla K (2007). Muscle stem cells and model systems for their investigation. Dev Dyn 236:3332ï 3342.

Filipo P. (2014). Myostatin, a key protein in muscle growth. www.flipper.diff.org.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531ï 1545.

Frankham and Richard (2005). Genetics and Extiction. Biological Conservation. 126(2): 131-140.

Furney SJ, Alba MM, Lopez-Bigas N (2006). Differences in the evolutionary history of disease genes affected by dominant or recessive mutations. BMC Genomics 7:165.

Garg RR, Bally-Cuif L, Lee SE, Gong Z, Ni X, Hew CL, Peng C (1999). Cloning of zebrafish activin type IIB receptor (ActRIIB) cDNA and mRNA expression of ActRIIB in embryos and adult tissues. Mol Cell Endocrinol 153:169ï 181.

Garikipati DK, Gahr SA, Roalson EH, Rodgers BD (2007). Characterization of rainbow trout myostatin-2 genes (rtMSTN-2a and -2b): genomic organization, differential expression, and pseudogenization. Endocrinology 148:2106ï 2115.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A. (2005). Protein identification and analysis tools on the ExPASY server. The proteomics Protocols Handbook, Human Press. 571-607.

Gasteiger E. (2005). The Proteomics Protocols Handbook. (In) John M Walker (edition): Humana Press. 571-607.

Gaussin V and Depre C (2005). Myostatin, the cardiac chalone of insulin-like growth factor-1. Cardiovasc Res 68:347ï 349.

Gonzalez-cadavid NF and Bahsin S. (2004). Role of mysotatin in metabolism. Curr. Opin. Chin, Nutr. Metab. Care, (7); 451-457.

Gonzalez-Cadavid NF, Taylor WE, Yarasheski K, SinhaHikim I, Ma K, Ezzat S, Shen R, Lalani R, Asa S, Mamita M, Nair G, Arver S, Bhasin S. (1998). Organization of the human myostatin gene and expression in healthy men and HIV-infected men with muscle wasting. Proc Natl Acad Sci USA;95:14938-14943.

Grobet L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, Riquet J, Schoeberlein A, Dunner S, Menissier F, Massabanda J, Fries R, Hanset R, Georges M (1997). A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle [see comments]. Nat Genet 17:71ï 74.

Guo W, Flanagan J, Jasuja R, Kirkland J, Jiang L, Bhasin S (2008). The effects of myostatin on adipogenic differentiation of human bone marrow-derived mesenchymal stem cells are mediated through cross-communication between Smad3 and Wnt/ß-catenin signaling pathways. J Biol Chem 283:9136ï 9145.

Hamilton MT, Hamilton DG, Zderic TW (2007). Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. Diabetes 56:2655ï 2667.

Hill JJ, Davies MV, Pearson AA, Wang JH, Hewick RM, Wolfman NM, Qiu Y (2002). The myostatin propeptide and follistatin-related gene are inhibitory binding proteins of myostatin in normal serum. J Biol Chem 277:40735ï 40741.

Hill JJ, Qiu Y, Hewick RM, Wolfman NM (2003). Regulation of myostatin *in vivo* by growth and differentiation factor-associated serum protein-1: a novel protein with protease inhibitor and follistatin domains. Mol Endocrinol 17:1144ï 1154.

Hirai S, Matsumoto H, Hino N, Kawachi H, Matsui T, Yano H (2007). Myostatin inhibits differentiation of bovine preadipocyte. Domest Anim Endocrinol 32:1ï 14.

Hofmann, K. and Stoffel, W. (1993). TMBASE - A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler*, 374: 166.

http//www.biology-online.org/dictionary/Genetic polymorphism.

Hui J and Bindereif A (2005). Alternative pre-mRNA splicing in the human system: unexpected role of repetitive sequences as regulatory elements. Biol Chem 386:1265ï 1271.

Hui J, Reither G, Bindereif A (2003). Novel functional role of CA repeats and hnRNP L in RNA stability. RNA 9:931ï 936.

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946ï 957.

Javanmard A, Panandam JM, Sugnaseelan S, Yusoff K. (2010): Allele frequencies at six candidate genes associated with growth and carcass quality traits in the Boer goats. African J. Bio. 9, 7236ï 7238.

Johnson PL, McEwan JC, Dodds KG, Purchas RW, Blair HT (2005). Meat quality traits were unaffected by a quantitative trait locus affecting leg composition traits in Texel sheep. J Anim Sci 83:2729ï 2735.

Jones DT. (2007). Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics*, 3 (5): 538ï 544.

Joulia D, Bernardi H, Garandel V, Rabenoelina F, Vernus B, Cabello G (2003). Mechanisms involved in the inhibition of myoblast proliferation and differentiation by myostatin. Exp Cell Res 286:263ï 275.

Joulia- Ekaza D and Cabello G. (2007). The myostatin gene. Physiology and pharmacological relevance. Curr. Opin pharmacol, 7 (3); 310-315.

Joulia-Ekaza D, and Cabello G. (2006). Myostatin regulation of muscle development: Molecular basis, natural mutations, physiopathological aspects. Exp Cell Res;312:2401-2414.

Kambadur R, Sharma M, Smith TP, Bass JJ (1997). Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. Genome Res 7:910ï 916.

Kelley L. A. *et al*. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10: 845-858.

Kelley L.A. and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols,* 4 (3): 363ï 371.

Kim J, Wu HH, Lander AD, Lyons KM, Matzuk MM, Calof AL (2005). GDF11 controls the timing of progenitor cell competence in developing retina. Science 308:1927ï 1930.

Kobayashi N, Goto K, Horiguchi K, Nagata M, Kawata M, Miyazawa K, Saitoh M, Miyazono K (2007) c-Ski activates MyoD in the nucleus of myoblastic cells through suppression of histone deacetylases. Genes Cells 12:375ï 385.

Koli K, Saharinen J, Hyytiainen M, Penttinen C, Keski-Oja J (2001). Latency, activation, and binding proteins of TGF-ß. Microsc Res Tech 52:354ï 362.

Langley B, Thomas M, Bishop A, Sharma M, Gilmour S, Kambadur R (2002). Myostatin inhibits myoblast differentiation by down regulating MyoD expression. J Biol Chem 277:49831ï 49840.

Larkin MA, Blackshields G, Brown, NP., Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, Gibson, TJ. and Higgins, DG. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics,* 23 (21): 2947-2948.

Lee SJ and McPherron AC (2001). Regulation of myostatin activity and muscle growth. Proc Natl Acad Sci USA 98:9306ï 9311.

Lee SJ, Reed LA, Davies MV, Girgenrath S, Goad ME, Tomkinson KN, Wright JF, Barker C, Ehrmantraut G, Holmstrom J, Trowell B, Gertz B, Jiang MS, Sebald SM, Matzuk M, Li E, Liang LF, Quattlebaum E, Stotish RL, Wolfman NM (2005). Regulation of muscle growth by multiple ligands signaling through activin type II receptors. Proc Natl Acad Sci USA 102:18117ï 18122.

Li W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics,* 22: 1658ï 1659.

Li XL, Wu ZL, Liu ZZ, Gong YF, Zhou RY, Zheng GR. (2006): SNP identification and analysis in part of intron 2 of goat *MSTN* gene and variation within and among species. J. Heredity, 97, 285ï 289.

Liang YC, Yeh JY, Ou BR (2007). Effect of maternal myostatin antibody on offspring growth performance and body composition in mice. J Exp Biol 210:477ï 483.

Liu ZT, Li XL, Gong YF and Jin XM (2006). Relationship between polymorphism of goat MSTN gene intron 2 and body weight. Acta Vet. Zootech. Sin. 37: 745-748.

Marchitelli C, Savarese MC, Crisa A, Nardone A, Marsan PA and Valentini A. (2003). Double muscling in Marchigiana beef breed is caused by a stop codon in the third exon of myostatin gene. Mammalian Genome, 14 (6); 392-395.

McCroskery S, Thomas M, Maxwell L, Sharma M, Kambadur R (2003). Myostatin negatively regulates satellite cell activation and self-renewal. J Cell Biol 162:1135ï 1147

McKoy G, Bicknell KA, Patel K, Brooks G (2007). Developmental expression of myostatin in cardiomyocytes and its effect on foetal and neonatal rat cardiomyocyte proliferation. Cardiovasc Res 74:304ï 312.

Mcpherron AC and Lee SJ.( 1997). Double muscling in cattle due to mutations in the myostatin gene. Proceedings of National Academy of Sciences, 94.

McPherron AC, Lawler AM, Lee SJ (1997). Regulation of skeletal muscle mass in mice by a new TGF-ƃ superfamily member. Nature 387:83ï 90.

McPherron AC, Lee SJ (2002). Suppression of body fat accumulation in myostatin-deficient mice. J Clin Invest 109:595ï 601.

Mgbeahuruike A, (2015). Generating data for high impact factor paper through the use of computer algorithms. Practical Bioinformatics workshop organized by Facu;ty of Veterinary Medicine, University of Nigeria Nsukka. 22nd-25th Nov. Not published.

Mota LSLS, Curi, RA, Palmieri, DA, Borges, AS, Lopes, CR, Barbosa, JD. and Gimenes, M.A. (2006). Sequence characterization of coding regions of the myostatin gene (*GDF8*) from Brazilian Murrah buffaloes (*Bubalus bubalis*) and comparison with the *Bos taurus* sequence. *Genetics and Molecular Biology*, 29 (1): 79-82.

Nakashima M, Toyono T, Akamine A, Joyner A 1999 Expression of growth/differentiation factor 11, a new member of the BMP/TGFб superfamily during mouse embryogenesis. Mech Dev 80:185ï 189.

National Biological Information Infrastructure. (2011). Introduction to genetic diversity. U.S. Geological Survey. Archived from the original. Feb. 25.

Ng PC and Henikoff S. (2003). SIFT: Predicting amino acid changes that affect protein function, Nucleic Acids Res, 31, 3812-3814.

Nicholas G, Thomas M, Langley B, Somers W, Patel K, Kemp CF, Sharma M, Kambadur R. (2002). Titin-cap associates with, and regulates secretion of, myostatin. J Cell Physiol 193:120ï 131.

Ostbye TK, Wetten OF, Tooming-Klunderud A, Jakobsen KS, Yafe A, Etzioni S, Moen T, Andersen O (2007). Myostatin (MSTN) gene duplications in Atlantic salmon (*Salmo salar*): evidence for different selective pressure on teleost MSTN-1 and -2. Gene 403:159ï 169.

Otteson DC, Cirenza PF, Hitchcock PF (2002). Persistent neurogenesis in the teleost retina: evidence for regulation by the growth-hormone/insulin-like growth factor-I axis. Mech Dev 117:137ï 149.

Page BT, Casas E, Heaton MP, Cullen NG, Hyndman DL, Morris CA, Crawford AM, Wheeler TL, Koohmaraie M, Keele JW, Smith TP (2002). Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. J Anim Sci 80:3077ï 3085.

Pampusch MS, Kamanga-Sollo E, White ME, Hathaway MR, Dayton WR (2003). Effect of recombinant porcine IGF-binding protein-3 on proliferation of embryonic porcine myogenic cell cultures in the presence and absence of IGF-I. J Endocrinol 176:227ï 235

Phillips R and Rab P (2001). Chromosome evolution in the *Salmonidae* (Pisces): an update. Biol Rev Camb Philos Soc 76:1ï 25.

Phylogeny (2013). Biology online, Retrieved 2013-02-15.

Potts JK, Echeternkamp SE, Smith TP and Reecy JM. (2003). Characterization of gene expression in double-muscled and normal-muscled bovine embryos. Anim Genet, 34 (6); 438-444.

Qian, B, Raman, S, Das, R, Bradley, P, McCoy, AJ, Read, RJ. and Baker, D. (2007). High-resolution structure prediction and the crystallographic phase problem. *Nature,* 450: 259ï 264.

Rava, P. and Hussain, MM. (2007). Acquisition of triacylglycerol transfer activity by microsomal triglyceride transfer protein during evolution. *Biochemistry*, 46: 12263ï 12274.

Rebbapragada A, Benchabane H, Wrana JL, Celeste AJ, Attisano L (2003). Myostatin signals through a transforming growth factor ϐ-like signaling pathway to block adipogenesis. Mol Cell Biol 23:7230ï 7242.

Rios R, Carneiro I, Arce VM and Devesa J (2001). Myostatin regulates cell survival during C2C12 myogenesis. Biochem Biophys Res Commun 280:561ï 566.

Ron M and Weller JI. (2007). From QTL to QTN identification in livestock-winning by points rather than knock-out: a review. Anim Genet; 38:429-439.

Shah MG, Gureshi A, Reismann and Schwartz HJ. (2006). Sequencing and sequence analysis of myostatin gene in the Exon 1 of the camel (Camelus promedarius). Pakistan Vet. J., 26 (4); 176-178.

Sharma M, Kambadur R, Matthews KG, Somers WG, Devlin GP, Conaglen JV, Fowke PJ, Bass JJ (1999). Myostatin, a transforming growth factor-ϐ superfamily member, is expressed in heart muscle and is upregulated in cardiomyocytes after infarct. J Cell Physiol 180:1ï 9.

Shyu KG, Lu MJ, Wang BW, Sun HY, Chang H (2006). Myostatin expression in ventricular myocardium in a rat model of volume-overload heart failure. Eur J Clin Invest 36:713ï 719

Smith TPL, Lopez- Carrles NL, Kappes SM and Sonstegard TS. (1997). Myostatin maps to the interval containing the bovine mh locus. Mamalian Genome, 8: 742-744.

Stoker HS. (2015). Organic and biological chemistry: cengage Learning, p 371.

Stolz LE, Li D, Qadri A, Jalenak M, Klaman LD, Tobin JF (2008). Administration of myostatin does not alter fat mass in adult mice. Diabetes Obes Metab 10:135ï 142.

Sundaresan NR, Saxena VK, Singh R, Jain P, Singh KP, Anish D, Singh N, Saxena M, Ahmed KA (2008). Expression profile of myostatin mRNA during the embryonic organogenesis of domestic chicken (*Gallus gallus domesticus*). Res Vet Sci 85:86ï 91.

Takahashi, K. and Nei, M. (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution,* 17: 1251-1258.

Tamura K, Peterson D, Peterson N, Stecher D, Nei M and Kumar S. (2011). MEGA5: molecular evolutionary genetic analysis using maximum likelihood,evolutionary distance, and maximum parsimony methods. Mol Biol Evol., (28); 2731-2739.

Tay, GK, Iaschi, SPA., Bellinge, RHS., Chong, FN. and Hui, J. (2004). The development of sequence-based-typing of myostatin (GDF-8) to identify the double muscling phenotype in the goat. *Small Ruminant Research*, 52: 1ï 12.

Taylor WE, Bhasin S, Artaza J, Byhower F, Azam M, Willard Jr DH, Kull Jr FC, Gonzalez-Cadavid N (2001). Myostatin inhibits cell proliferation and protein synthesis in C(2)C(12) muscle cells. Am J Physiol Endocrinol Metab 280:E221ï E228.

Tellam, RL, Cockett, NE, Vuocolo, T. and Bidwell, CA. (2012). Genes contributing to genetic variation of muscling in sheep. *Frontiers in Genetics*, 3: doi: 10.3389/fgene.2012.00164.

Thomas M, Langley B, Berry C, Sharma M, Kirk S, Boss J and Kambadaur R. (2000). Myostantin, a negative regulator of muscle Growth, functions by inhibiting myoblast proliferation. Journal of Biological Chemistry, (275); 40235-40243.

Tisdell C. (2003). Socioeconomic causes of loss of animal genetic diversity: Analysis and assessment. Ecological Economics 45(3): 365-376.

Toro M and Caballero A. (2005). Characterization and conservation of diversity in subdivided populations. Philosophical Transactions of Royal society of London, Series B: Biological Science, 360: 1367-1378.

Ugbo, SB, **Yakubu, A**., Bibinu, BS, Yakubu, S. and Dim, NI. (2015). Molecular analysis of proteins of DQA2 gene in goats, sheep and cattle. *Proceedings of the Genetics Society of Nigeria 39$^{th}$ Annual conference* held from 27$^{th}$-31$^{st}$ of October, 2015 at Abubakar Tafawa Balewa University, Bauchi.

Walsh FS and Celeste AJ (2005). Myostatin: a modulator of skeletal-muscle stem cells. Biochem. Soc. Trans. 33: 1513- 1517.

Wang Z, Heshka S, Gallagher D, Boozer CN, Kotler DP, Heymsfield SB (2000). Resting energy expenditure-fat-free mass relationship: new insights provided by body composition modeling. Am J Physiol Endocrinol Metab 279:E539ð E545.

Wass, MN, Kelley, LA. and Sternberg, MJ. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Research,* 38: 469-473.

Watson, JD, Laskowski, RA. and Thornton, JM. (2005). Predicting protein function from sequence and structural data. *Curr. Opin. Struct. Biol*., 15: 275ï 284.

Wheeler TL, Cundiff LV, Shackelford SD, Koohmaraie M (2001). Characterization of biological types of cattle (cycle V): carcass traits and longissimus palatability. J Anim Sci 79:1209ï 1222.

Wittkopp PJ, Haerum BK, Clark AG (2004). Evolutionary changes in cis and trans gene regulation. Nature 430:85ï 88.

Wray GA (2007). The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206ï 216.

Wu HH, Ivkovic S, Murray RC, Jaramillo S, Lyons KM, Johnson JE, Calof AL (2003). Autoregulation of neurogenesis by GDF11. Neuron 37:197ï 207.

www.biology-online.org/dictionary/pylogeny. Retrieved. 2013-02-15.

Xu W, Angelis K, Danielpour D, Haddad MM, Bischof O, Campisi J, Stavnezer E, Medrano EE (2000). Ski acts as a co-repressor with Smad2 and Smad3 to regulate the response to type ß transforming growth factor. Proc Natl Acad Sci USA 97:5924ï 5929.

Yakubu, A, Salako, AE, De Donato, M, Takeet, MI, Peters, SO, Adefenwa, MA, Okpeku, M, Wheto, M, Agaviezor, BO, Sanni, TM, Ajayi, OO, Onasanya, GO, Ekundayo, OJ, Ilori, BM, Amusan, SA. and Imumorin, IG. (2013).Genetic diversity in Exon 2 at the major histocompatibility complex DQB1 locus in Nigerian indigenous goats.*Biochemical Genetics*, 51: 954-966.

Yang J and Zhao B (2006). Postnatal expression of myostatin propeptide cDNA maintained high muscle growth and normal adipose tissue mass in transgenic mice fed a high-fat diet. Mol Reprod Dev 73:462ï 469.

Yang J, Ratovitski T, Brady JP, Solomon MB, Wells KD, Wall RJ (2001). Expression of myostatin pro domain results in muscular transgenic mice. Mol Reprod Dev 60:351ï 361.

Yang W, Chen Y, Zhang Y, Wang X, Yang N, Zhu D (2006). Extracellular signal-regulated kinase 1/2 mitogen-activated protein kinase pathway is involved in myostatin-regulated differentiation repression. Cancer Res 66:1320ï 1326.

Yarov-Yarovoy,V, Schonbrun, J. and Baker, D. (2006). Multipass membrane protein structure prediction using Rosetta. Proteins, 62: 1010ï 1025.

Younes M, Abdolreza S, Davood K and Seyed AA. (2014). Application of myostatin in sheep breeding programs: A review Mol. Bio. Res. Comm. 3(1): 33-43.

Zapf J and Froesch ER (1999). Insulin-like growth factor-I actions on somatic growth. In: Kostyo JL, Goodman HM, eds. Handbook of physiology, the endocrine system: hormonal control of growth. New York: Oxford University Press; 663ï 700.

Zhang ZJ, Ling YH, Wang LJ, Hang YF, Guo XF, Zhang YH, Ding JP and Zhang XR. (2013). Polymorphisms of the myostatin gene (*MSTN*) and its relationship with growth traits in goat breeds. Genet. Mol. Res. 12(2): 965-971.

Zhou, H, Hickford, JGH. and Fang, Q. (2008). Variation in the coding region of the myostatin (GDF8) gene in sheep. *Molecular and Cellular Probes*, 22: 67ï 68.

Zhu X, Topouzis S, Liang LF, Stotish RL (2004) Myostatin signaling through Smad2, Smad3 and Smad4 is regulated by the inhibitory Smad7 by a negative feedback mechanism. Cytokine 26:262ï 272.