

## Research Article

# Comparative Genomic Characterization of Relaxin Peptide Family in Cattle and Buffalo

Muhammad Saif-ur Rehman,<sup>1</sup> Faiz-ul Hassan ,<sup>1</sup> Zia-ur Rehman,<sup>2</sup> Hafiz Noubahar Hussain,<sup>3</sup> Muhammad Adnan Shahid,<sup>1</sup> Muhammad Mushahid ,<sup>1</sup> and Borhan Shokrollahi <sup>4</sup>

<sup>1</sup>Institute of Animal and Dairy Sciences, University of Agriculture, Faisalabad 38040, Pakistan

<sup>2</sup>University of Agriculture Faisalabad-Sub Campus, Toba Tek Singh 36050, Pakistan

<sup>3</sup>Animal Science Division, Nuclear Institute for Agriculture and Biology (NIAB), Faisalabad, Pakistan

<sup>4</sup>Department of Animal Science, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran

Correspondence should be addressed to Borhan Shokrollahi; borhansh@iausdj.ac.ir

Received 30 July 2022; Accepted 9 September 2022; Published 4 October 2022

Academic Editor: Akhtar Ali

Copyright © 2022 Muhammad Saif-ur Rehman et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Relaxin family peptides significantly regulate reproduction, nutrient metabolism, and immune response in mammals. The present study aimed to identify and characterize the relaxin family peptides in cattle and buffalo at the genome level. The genomic and proteomic sequences of cattle, buffalo, goat, sheep, horse, and camel were accessed through the NCBI database, and BLAST was performed. We identified four relaxin peptides genes (*RLN3*, *INSL3*, *INSL5*, and *INSL6*) in *Bos taurus*, whereas three relaxin genes (*RLN3*, *INSL3*, and *INSL6*) in *Bubalus bubalis*. Evolutionary analysis revealed the conserved nature of relaxin family peptides in buffalo and cattle. Physicochemical properties revealed that relaxin proteins were thermostable, hydrophilic, and basic peptides except for *INSL5* which was an acidic peptide. Three nonsynonymous mutations (two in *RLN3* at positions A16>T and P29>A, and one in *INSL6* at position R32>Q) in *Bos taurus*, whereas two nonsynonymous mutations (one in *RLN3* at positions G105>w and one in *INSL3* at position G22>R) in *Bubalus bubalis*, were identified. *INSL3* had one indel (insertion) at position 55 in *Bos taurus*. Gene duplication analysis revealed predominantly segmental duplications (*INSL5/RLN3* and *INSL6/INSL3* gene pairs) that helped expand this gene family, whereas *Bubalus bubalis* showed primarily tandem duplication (*INSL3/RLN3*). Our study concluded that relaxin family peptides remained conserved during the evolution, and gene duplications might help to adapt and enrich specific functions like reproduction, nutrient metabolism, and immune response. Further, the nonsynonymous mutations identified potentially affect these functions in buffalo.

## 1. Introduction

The relaxin peptide family comprises seven peptides with significant structural similarities but low sequence resemblance. It includes seven genes, relaxin like *RLN1*, -2, and -3, and insulin-like *INSL3*, -4, -5, and -6 in most mammals [1], but their number varies. These peptides show a high sequence resemblance with insulin due to the presence of six cysteine residues that provide the 2 interchain and 1 intrachain disulphide linkages. Each relaxin peptide family member is constituted of two chains called A and B chains

[2]. These chains are connected by two disulfide bonds present between them and one disulfide bond within the A chain. Each chain contained the cysteine residues together with distinctive disulfide bonding, which are found conserved across all family members [2].

The *RLN1* and *RLN2* are present in humans and higher primates like apes. Both are referred to as relaxin as human *RLN2* is an orthologue to *RLN1* in other mammals [1, 3]. *RLN3* was first identified in 2002 and is considered the common ancestral gene for all relaxin peptides [2, 4]. *RLN1* and its orthologue *RLN2* play an important role in reproduction

in mammals [5], but in bovines, these genes have been lost during the evolution and have no traces in the genome [6]. *RLN3* has been shown to play a role in nutrient metabolism in cattle [7]. *INSL3* plays a crucial role in testicular descent by promoting the growth and development of the gubernaculum ligament [8]. *INSL4* is highly expressed in the placenta and might be involved in bone development [5, 9]. Its receptor is yet to be identified. *INSL5* with its cognate receptor *RXFP4* has been suggested to play important role in immune response regulation, signal transmission to CNS through the vagus nerve, and autocrine/paracrine function within the intestinal tract [10]. *INSL6* role in sperm production has been determined in buffalo [11].

Relaxin family peptides have important role in ruminant reproduction and nutrient metabolism as mentioned above. The availability of genomic data has provided opportunity to perform genome-wide characterization of protein families using the different available bioinformatics tools. Many studies have been conducted to identify the evolutionary relationships, physicochemical characteristics, comparative amino acid analysis, effects of mutations, and gene duplications in important protein families in ruminants [12–15]. The present study was conducted to characterize relaxin family peptides in cattle (*Bos taurus*) and buffalo (*Bubalus bubalis*) at genome-wide level in order to better understand their evolutionary significance, physicochemical properties, and gene duplications events.

## 2. Materials and Methods

**2.1. Genome-Wide Identification of Relaxin Peptides.** The whole genome and proteomic sequence data of buffalo (UOA WB 1), cattle (ARS-UCD1.2), sheep (Oar rambouillet v1.0), goat (ARS1), camel (CamDro3), and horse (EquCab3.0) were obtained from the National Center for Biotechnology Information (NCBI) database [16]. A genome-wide BLAST and HMM search was performed to look for all putative relaxin peptide genes in *Bos taurus*, *Bubalus bubalis*, and other targeted species [17]. The cattle (*Bos taurus*), buffalo (*Bubalus bubalis*), goat (*Capra hircus*), sheep (*Ovis aries*), horse (*Equus caballus*), and camel (*Camelus dromedarius*) relaxin peptide sequences were also validated through the UniProt database search [18]. All accession numbers of sequences used for this study are presented in Table S1. No information was available for *INSL5* gene annotation of *Bubalus bubalis* in the databases.

**2.2. Phylogenetic Analysis.** Relaxin peptide amino acid sequences from *Bos taurus*, *Bubalus bubalis*, *Capra hircus*, *Ovis aries*, *Equus caballus*, and *Camelus dromedarius* were aligned in ClustalW. Further, the neighbor-joining (NJ) tree was constructed through the MEGA7 software [19].

**2.3. Gene Structure, Motif Patterns, and Conserved Domain Analysis of Relaxin Peptide Family.** The conserved motifs in the dataset were analyzed using the MEME suite [20]. As a query, the relaxin protein sequences were submitted in Fasta format, and a site distribution with one occurrence of each site was determined for each sequence. The motifs'

minimum and maximum widths were found to be 6 and 50, respectively. The number of themes was limited to ten. The Gene Structure Display Server (GSDS) [21] was used to import all CDs and genomic sequences. The final gene structure was exhibited and illustrated using the genome annotation data in general feature format utilizing in-house scripts in the TBtools software (GFF).

**2.4. Physicochemical Properties of Relaxin Proteins.** The online available ProtParam tool [22] was used to evaluate the physicochemical characteristics of relaxin peptides including molecular weight (MW), amino acid count (AA), isoelectric point (pI), and aliphatic index (A.I.). Besides, it also included instability index (II) and the grand average of hydropathicity (GRAVY).

**2.5. Multiple Sequence Analysis (MSA).** Multiple align show [23] was used online to explore the mutations and indels in the relaxin peptides using the aligned sequences of *Bos taurus*, *Bubalus bubalis*, *Capra hircus*, and *Ovis aries*.

**2.6. Mutational Analysis.** Further, the mutations found in *Bos taurus* and *Bubalus bubalis* relaxin peptide sequences were subsequently examined using several online tools (PolyPhen-2, MUpro, PROVEAN, IMutant, PhD-SNP, SIFT, SNAP2, PredictSNP, Meta-SNP, and SNAP) to determine their effects on protein structure and functions.

**2.7. Nuclear Hormone Receptor Sites Identification.** The NHR scan [24] was employed to predict nuclear hormone receptor binding sites. Using genomic sequences in Fasta format, NHR scan was performed. The cumulative probability of entering match states was 0.01 using the NHR scan.

**2.8. Synteny Analysis and Gene Duplications.** To find collinear genes, the whole genomes of cattle and buffalo were blasted to each other. The dual-synteny map was constructed using the TBtools after submitting annotation files for both genomes, including information about collinear genes and chromosomal IDs.

Chromosomal locations of relaxin genes were obtained from genomic resources of respective specie. An annotated genome file was saved as a general feature format (GFF) file and was fed into the MCScanX programme [25], which was subsequently used to plot the gene locations on chromosomes and presented in TB tools. In addition, the relaxin peptide gene collinearity plots for *Bos taurus* and *Bubalus bubalis* were generated. Further, for the *Bos taurus* and *Bubalus bubalis* relaxin peptide gene family, pairwise alignment of homologous gene pairs of relaxin peptide genes using MEGA7 [19] with the MUSCLE method was utilized to analyze the frequency of duplications. DnaSP v6.0 [26] was also used to determine pairwise synonymous substitutions per synonymous site (ks) and nonsynonymous substitutions per nonsynonymous site (ka) that were corrected for multiple hits. Synonymous mutations are referred as silent mutations which can result in altered DNA sequence but does not change the encoded amino acid (evolutionary neutral mutations), whereas nonsynonymous mutations cause

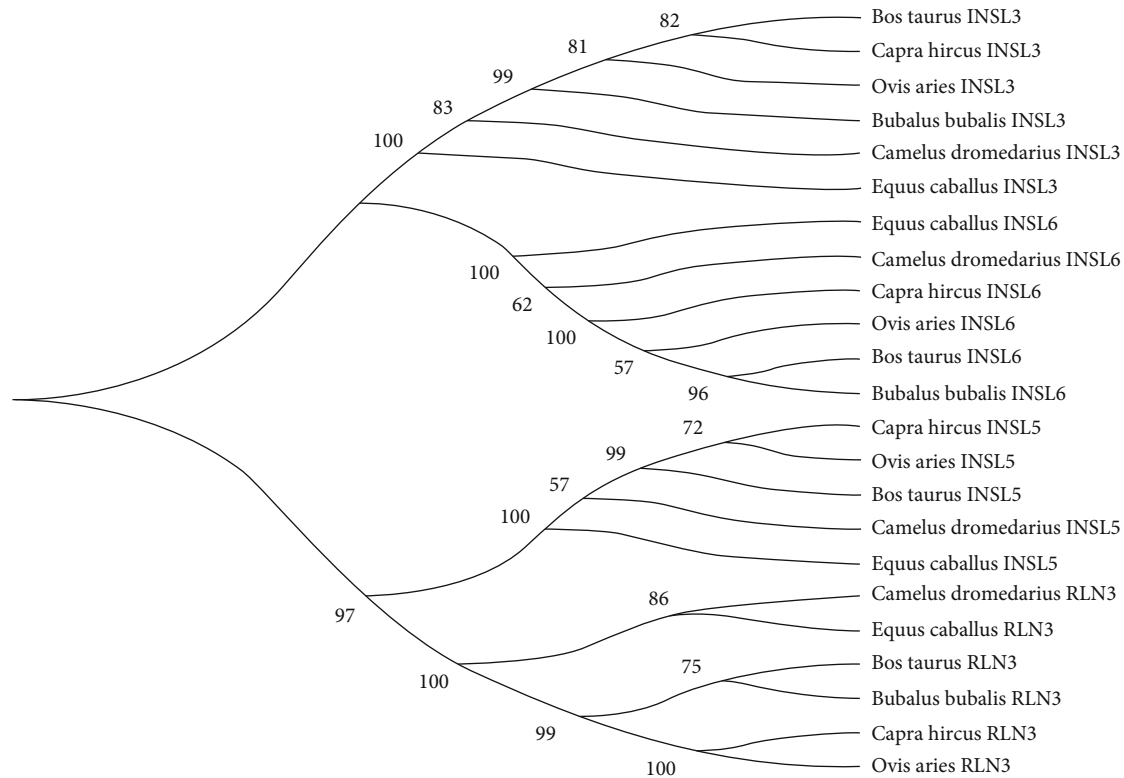


FIGURE 1: Evolutionary relationship of relaxin peptides (*RLN3*, *INSL3*, *INSL5*, *INSL6*) in six mammalian species.

change in both DNA and protein sequence (evolutionary important mutations).

### 3. Results

**3.1. Phylogenetic Analysis.** Results revealed that relaxin peptide family consisted of four genes (*RLN3*, *INSL3*, *INSL5*, and *INSL6*) in all representative species, except for *Bubalus bubalis* which had 3 of them excluding *INSL5* (Figure 1). All four genes were clustered into two sister clades. From top to downward, clade 1 included *INSL3* and *INSL6*, whereas clade 2 included *INSL5* and *RLN3*. Overall phylogenetic analysis of the relaxin peptide family revealed that the “*Bos taurus* and *Bubalus bubalis*,” “*Capra hircus* and *Ovis aries*,” and “*Camelus dromedarius* and *Equus caballus*” had close similarities between them. However, *INSL3* showed more evolutionary similarities between “*Bos taurus* and *Capra hircus*” than between “*Bos taurus* and *Bubalus bubalis*.”

**3.2. Structural Categorization of the Relaxin Peptide Family.** The examination of gene organization, motif patterns, and conserved domains in the relaxin peptide family of four targeted species, including *Bos taurus*, *Bubalus bubalis*, *Capra hircus*, and *Ovis aries*, were performed to carry out the structural characterization of the relaxin peptide family while taking account of their evolutionary relationships (Figures 2(a)–2(d)). The top ten MEME-conserved motifs were investigated to look for conserved domains (Table 1). No conserved domain was detected in the Pfam search for these motifs. Further, all genes were investigated to look for conserved domains across all targeted species

(Figure 3(c)). The insulin/insulin-like growth factor (IIGF) domain was found conserved across all targeted species, which was further validated through a conserved domain database (CDD) BLAST. A gene structural analysis revealed the evolutionarily conserved nature of relaxin family genes across the studied species (Figure 3(d)). The same gene across different species revealed a similar number of introns and exons.

**3.3. Physico-chemical Properties of the Relaxin Proteins.** The physicochemical properties like location on the chromosome, exon count, molecular weight (Da), number of amino acids (A.A) in each peptide, aliphatic index (A.I.), isoelectric point (pI), instability index (II), and grand average of hydrophobicity index (GRAVY) of the relaxin peptides were evaluated in cattle through the ProtParam tool (Table 2). The molecular weight of relaxin peptides ranged from 14 to 24 kDa. *RLN3* and *INSL3* were located on chromosomes 7, whereas *INSL5* and *INSL6* were located on chromosomes 3 and 8, respectively. All relaxin peptide genes had an exons count of 2 and a variable length of the peptides with amino acid residues. The pI values revealed all relaxin peptides were basic except for *INSL5* which was slightly acidic. The AI values suggested the thermostable nature of all peptides having AI values greater than 65. Moreover, II values greater than 40 revealed that all peptide members of the relaxin family are unstable in vitro. Negative GRAVY values suggested the hydrophilic nature of relaxin peptides.

**3.4. Identification of Mutations in Relaxin Peptides.** Comparative amino acid analysis was performed by aligning the

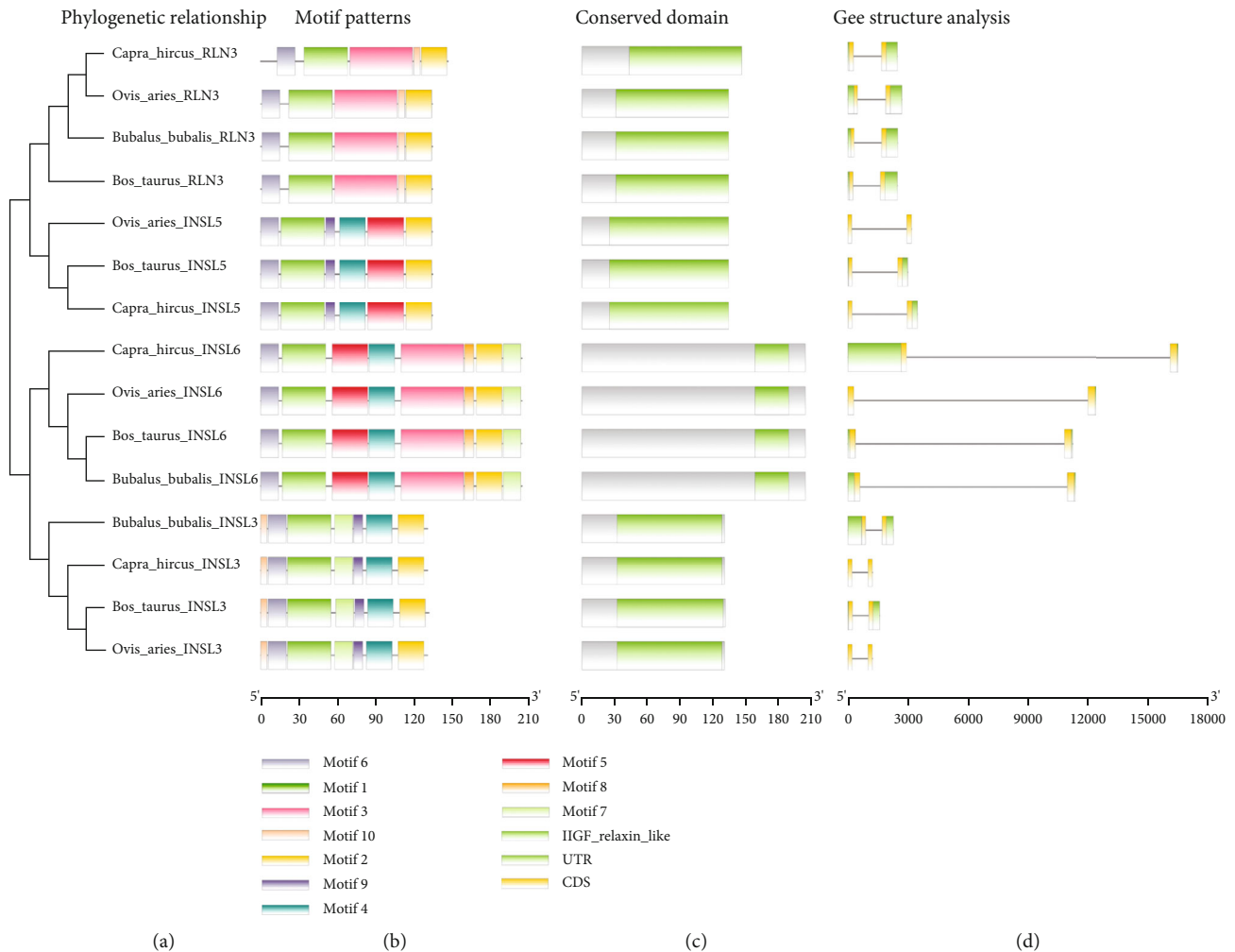


FIGURE 2: Graphical representation of motif patterns (b), conserved domains (c), and gene structure (d) of relaxin peptide family genes corresponding to their phylogenetic relationships (a) in four mammalian species.

TABLE 1: Top ten differentially conserved motifs detected in relaxin peptide family (*RLN3*, *INSL3*, *INSL5*, and *INSL6*).

Motif	Protein sequence	Length	Pfam domain
MEME-1	REAAATEAARKLCGRHFIRAVVKLCGGSRWSSREEG	35	—
MEME-2	RGLSEKCKKGGCTKSELLTLC	21	—
MEME-3	PEYQYPEVBLPFESELEEAVASSEILPLTKEPIEFYGNKTBKIGTSPNLF	50	—
MEME-4	DPALNPAPQPLSQEEAIHNMK	21	—
MEME-5	TQLLSZASEKVESFIPDRSESSQTTFPVW	29	—
MEME-6	MRALVLLLLALAVLL	15	—
MEME-7	LPGGDYELLRKLZGL	15	—
MEME-8	WGNHPQRK	8	—
MEME-9	HLLHGLMA	8	—
MEME-10	GDRDPL	6	—

protein sequences of the relaxin peptides of buffalo, cattle, goat, sheep, camel, and horse in multiple align show to look for indels and single amino acid variations in *Bos taurus* and *Bubalus bubalis* (Figures 3–3(d)).

In *RLN3* protein, mutations were observed in *Bos taurus* at positions A16>T, P29>A, and A62>T and *Bubalus*

*bubalis* at position G105>W (Figure 3). *INSL3* had one indel (insertion) at position 55 in *Bos taurus* (Figure 3(b)). In *INSL3*, mutations were observed in *Bubalus bubalis* at positions G22>R, V86>M, and V88>I, whereas no mutation was observed in *Bos taurus* (Figure 3(b)). *INSL5* sequence of *Bubalus bubalis* was not found in the database.





FIGURE 3: (a–d) Comparative amino acid analysis of the relaxin peptide family (RLN3, INSL3, INSL5, and INSL6) in *Bos taurus*, *Bubalus bubalis*, *Capra hircus*, and *Ovis aries*.

TABLE 2: Physicochemical properties of the relaxin family peptides in *Bos taurus*.

Gene	Chromosome	Exon count	MW (kDa)	A.A	pI	AI	II	GRAVY
RLN3	7	2	14.75	135	7.60	86.00	55.82	-0.162
INSL3	7	2	14.38	132	8.69	96.21	67.33	-0.145
INSL5	3	2	15.36	135	6.89	73.04	76.93	-0.389
INSL6	8	2	24.00	205	9.14	77.51	66.54	-0.717

MW: molecular weight; A.A: number of amino acids; pI: isoelectric point; AI: aliphatic index; II: instability index; GRAVY: grand average of hydrophaticity index.

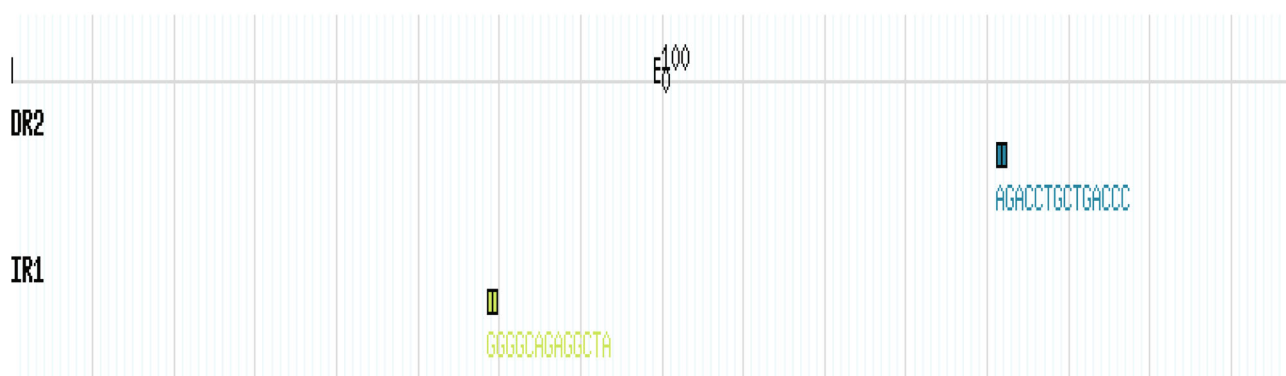
INSL6 had a mutation in *Bos taurus* at position R32 > Q, whereas no mutation was detected in *Bubalus bubalis*.

Additionally, the mutations observed in *Bos taurus* and *Bubalus bubalis* through a comparative amino acid analysis were further analyzed through different online available mutational analysis tools to predict the functional effects of these

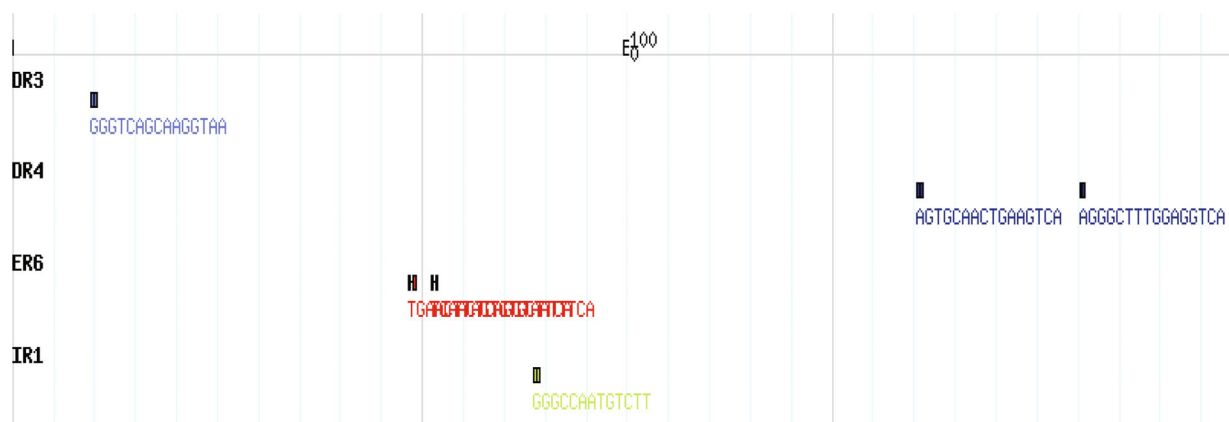
mutations (Table S2). In *Bos taurus*, a total of three nonsynonymous mutations were predicted, two in *RLN3* at positions A16 > T and P29 > A, and one in *INSL6* at position R32 > Q. In *Bubalus bubalis*, a total of two nonsynonymous mutations were predicted, one in *RLN3* at positions G105 > w, G22 > R, and one in *INSL3* at position G22 > R.



(a)



(b)



(c)

FIGURE 4: Continued.



(d)

FIGURE 4: NHR scans patterns of *RLN3* (a), *INSL3* (b), *INSL5* (c), and *INSL6* (d) in *Bos taurus*.

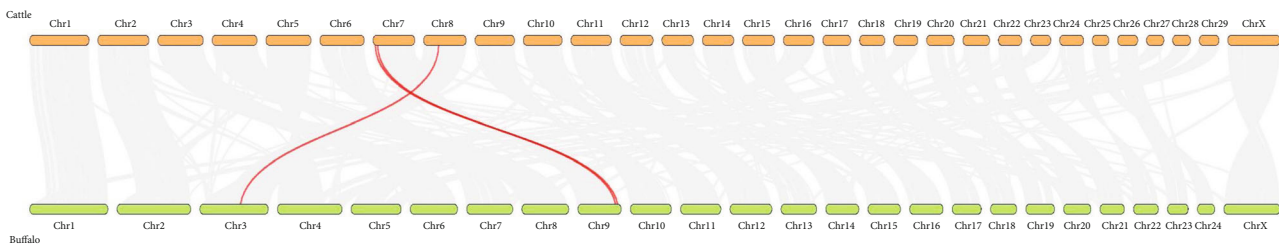


FIGURE 5: Synteny plot between *Bos taurus* and *Bubalus bubalis* genomes.

3.5. *NHR Patterns in Relaxin Peptides.* The *Bos taurus* nuclear hormone receptor sites (NHRs) were searched for all four relaxin peptides (*RLN3*, *INSL3*, *INSL5*, and *INSL6*) (Figures 4(a)–4(d)). A total of 23 NHRs were found of which *RLN3* had 3, *INSL3* had 2, *INSL5* had 6, and *INSL6* had 12. The number of direct repeats (DR) identified in *RLN3*, *INSL3*, *INSL5*, and *INSL6* were 2, 1, 3, and 6, respectively. The number of everted repeats (ER) identified in *RLN3*, *INSL3*, *INSL5*, and *INSL6* were 1, 0, 2, and 3, respectively. The number of inverted repeats (IR) identified in *RLN3*, *INSL3*, *INSL5*, and *INSL6* were 0, 1, 1, and 3, respectively.

3.6. *Synteny Analysis and Gene Duplications.* Collinearity analysis showed that relaxin family genes were randomly distributed over 2 chromosomes in both cattle and buffalo (Figure 5). In *Bos taurus*, relaxin peptide genes were present on chromosomes 7 and 8, whereas in *Bubalus bubalis*, these genes were located on chromosomes 3 and 9.

Further, the gene duplication analysis was performed to look for segmental or tandem duplication gene pairs in the relaxin peptide family of *Bos taurus* and *Bubalus bubalis* (Table 3). In *Bos taurus*, two segmental duplication events were observed between *INSL5/RLN3* and *INSL6/INSL3* gene

TABLE 3: Analysis of duplicated gene pairs and their  $k_a/k_s$  values of relaxin peptide family in *Bos taurus* and *Bubalus bubalis*.

<i>Bos taurus</i> Gene pair	Chromosome	Duplication	$k_a$	$k_s$	$k_a/k_s$
<i>INSL5/RLN3</i>	3/7	SD	0.32	0.47	0.68
<i>INSL6/INSL3</i>	8/7	SD	0.33	0.56	0.60
<i>Bubalus bubalis</i> Gene pair	Chromosome	Duplication	$k_a$	$k_s$	$k_a/k_s$
<i>INSL3/RLN3</i>	9/9	TD	0.37	0.50	0.74

$k_a$ : number of nonsynonymous substitutions per nonsynonymous site;  $k_s$ : number of synonymous substitutions per synonymous site; SD: segmental duplication; TD: tandem duplication.

pairs, whereas in *Bubalus bubalis*, one tandem duplication was detected between *INSL3/RLN3* gene pair. The number of nonsynonymous substitutions per nonsynonymous site/number of synonymous substitutions per synonymous site ( $k_a/k_s$ ) ratios was determined for this duplicated event. *Bos taurus* segmental duplicated pairs *INSL5/RLN3* and *INSL6/INSL3* showed 0.68 and 0.60  $k_a/k_s$  ratios, respectively,

whereas *Bubalus bubalis* tandem duplication pair *INSL3/RNL3* showed 0.74 ka/ks ratio.

#### 4. Discussion

**4.1. Phylogenetic Analysis.** In recent years, genomic sequencing technology, particularly next-generation sequencing, has advanced significantly, resulting in the accessibility of sequenced genomes for many important organisms, opening up a new path for understanding the genomic architecture at the molecular level of diverse animal species [27]. Comparative genomics allows for the discovery of new genes and functional components [28, 29]. Advances in bioinformatics have enabled the utilization of genomic data and look into the protein family evolutionary history, comparative amino acid analysis, gene duplications and prediction of mutations, and their functional and structural effects [12–15].

The relaxin peptide family has been found to contain seven members in most mammals, including relaxin-like genes *RLN1*, *RLN2*, and *RLN3* and insulin-like genes *INSL3*, *INSL4*, *INSL5*, and *INSL6* [30]. In our analysis, we have found four genes (*RNL3*, *INSL3*, *INSL5*, and *INSL6*) in *Bos taurus*, *Capra hircus*, *Ovis aries*, *Camelus dromedarius*, and *Equus caballus*, whereas three genes (*RNL3*, *INSL3*, and *INSL6*) in *Bubalus bubalis* from the sequenced genome of these species. Further, these genes were grouped into two sister major clades, clade 1 included *INSL3* and *INSL6*, whereas clade 2 included *INSL5* and *RLN3*. A variable number of *RLN/INSL* peptides were also observed in different vertebrates [31–33]. These variations could be explained on the basis of gene loss and fixation during the evolution and adaptation to specific niches. Overall phylogenetic analysis of the relaxin peptide family revealed that the “*Bos taurus* and *Bubalus bubalis*,” “*Capra hircus* and *Ovis aries*,” and “*Camelus dromedarius* and *Equus caballus*” had close similarities between them. Previous studies also revealed evolutionary similarities between these species [12, 13]. However, *INSL3* peptide showed more evolutionary similarity between “*Bos taurus* and *Capra hircus*” than between “*Bos taurus* and *Bubalus bubalis*”. Researchers have been fascinated by relaxin evolution for decades. Relaxins are renowned for their high sequence variability across closely related species, although unexpected parallels have been found between quite different species like pigs and whales [34].

**4.2. Structural Features of Relaxin Peptides.** The examination of gene organization, motif patterns, and conserved domains of the relaxin peptide family of four targeted species including *Bos taurus*, *Bubalus bubalis*, *Capra hircus*, and *Ovis aries* revealed the conserved nature of relaxin genes across the targeted species. The insulin/insulin-like growth factor (IIGF) domain was found conserved in all relaxin family genes across all targeted species. The insulin/IGF system (IIGFs) regulates a wide range of physiological processes, including development, linear growth, and aging, as well as metabolism, homeostasis, and central nervous system activities [35, 36]. This domain is important for proper function, and any dysregulation in this domain can result in abnormal

growth, increased development and progression of numerous cancers, and pathologic ailments associated with chronic inflammation and fibrosis [37, 38].

**4.3. Physico-chemical Properties of Relaxin Peptides.** The physicochemical properties of the relaxin peptide family proteins were evaluated in *Bos taurus* through the ProtParam tool (Table 2). The molecular weight of relaxin peptides ranged from 14 to 24 kDa. The aliphatic index (AI) tells about the thermostability of globular proteins, and values greater than 65 show greater thermostability [39]. In our study, all relaxin family peptides were found thermostable. In vitro stability of proteins can be inferred through the instability index (II), and the II value lower than 40 indicates the in vitro stability of proteins [40], as in our case, all relaxin family peptides showed in vitro instability having values greater than 40. The GRAVY values tell about the hydrophobicity of protein, the negative GRAVY values show hydrophilic nature, whereas the positive GRAVY values show hydrophobic nature of proteins [41], as in our case, all relaxin family peptides showed hydrophilic nature having negative GRAVY values.

**4.4. Comparative Mutational Analysis.** Comparative genomics is a large-scale, integrated technology for the comparison of two or more genomes. Comparative studies at various levels of the genomes may be conducted to obtain distinct perspectives on the organisms [35, 42]. We aligned the sequences of four species *Bos taurus*, *Bubalus bubalis*, *Capra hircus*, and *Ovis aries* in multiple align show to look for indels and single amino acid variations in *Bos taurus* and *Bubalus bubalis*. All relaxin family peptides were found well conserved with few amino acid variations in *Bos taurus* and *Bubalus bubalis*. *INSL3* had one indel (insertion) at position 55 in *Bos taurus*. Indels and mutations have all played a part in the divergence of gene family members from their progenitors [43]. Further, the mutational analysis of observed single amino acid variations predicted three nonsynonymous mutations (two in *RLN3* at positions A16 > T and P29 > A and one in *INSL6* at position R32 > Q) in *Bos taurus*, whereas two nonsynonymous mutations (one in *RLN3* at positions G105 > w, G22 > R, and one in *INSL3* at position G22 > R) in *Bubalus bubalis*. *RLN3* gene was observed to play role in feed efficiency in cattle [7]. *INSL3* is a gender-specific gene that is produced in Leydig cells of male adult and fetus and plays a key role in testicular descent [44]. Higher level of *INSL3* gene was observed in female ruminant blood with male fetus. Mutations in the *INSL3* gene resulted in failure of testicular descent (cryptorchidism) [45, 46]. *INSL6* was detected to play a role during spermatogenesis [11]. The deficiency of *INL6* in mice resulted in a decline in sperm production and immotility [47]. Mutations in these genes can interfere with functions like feed metabolism, testicular descent, and spermatogenesis in bovines.

**4.5. NHR Patterns in Relaxin Peptides.** Diverse biochemical mechanisms are involved in gene regulation and information flow from the DNA to the protein that is transcription and translation, and understanding of these mechanisms is



necessary to explore the cell dynamics [48]. Nuclear receptors bind to target genes at sites referred as hormone response elements (HREs) and help to regulate the transcription. These HREs are usually located in the 5-flanking region of target genes. Even though HREs are primarily found near the primary promoter, they can sometimes be found several kilobases upstream away from the start of the transcription site in enhancer regions [49]. Most of the time, a single NHR has been found to impact many genes, and sometimes, many NHRs have competition for one target gene and result in overlapping networks for the target genes [50]. This competition for the same target gene sometimes results in reduced expression of the gene. The expression of the gene can also be reduced if NHR bind with negative HREs [49]. The pattern of NHR sites in the relaxin peptide family in *Bos taurus* was investigated. A total of 23 NHR sites were detected. In total, 12 direct repeats (DR), 6 everted repeats (ER), and 5 inverted repeats (IR) were found in the *Bos taurus* relaxin genes.

**4.6. Synteny Analysis and Gene Duplications.** Chromosomal regions common between two genomes with the same homologous genes order as in common ancestor sites are called synteny blocks [51]. Different species originating from the common ancestor in the same tree of life can be compared using syntenic relationships, which will give an idea about the chromosomal structure and number variation between species [52, 53]. Synteny analysis revealed that relaxin peptide genes were randomly located over 2 chromosomes in both *Bos taurus* and *Bubalus bubalis*. In *Bos taurus*, relaxin genes were present on chromosomes 7 and 8, whereas in *Bubalus bubalis*, these genes were located on chromosomes 3 and 9. Further, the gene duplication events were examined for *Bos taurus* and *Bubalus bubalis*. Gene duplications have evolutionary significance as it is believed that during the evolution, whole genome duplications occurred and only 5 to 10% of duplicated genomes got fixed to perform specific functions, while others were lost in the process [54, 55]. These duplication events helped in the expansion of genome size and increased complexities to perform specific functions as indicated by two rounds of duplications hypothesis (2R hypothesis) [56, 57]. In our study, the *Bos taurus* relaxin peptide family showed predominantly segmental duplications (*INSL5/RLN3* and *INSL6/INSL3* gene pairs) that helped in the expansion of this gene family, whereas *Bubalus bubalis* showed predominantly tandem duplication (*INSL3/RLN3*). Our results are in agreement with Liu et al. [58], and they explained that segmental duplications are predominant in cattle genome and these duplications in bovine genomes are enriched with specific biological processes related to digestion, lactation, immunity, and reproduction. Further, the  $k_a/k_s$  ratios were lower than 1 for all these observed duplications, indicating the purifying pressure for these duplication events [59].

## 5. Conclusions

Our study revealed four relaxin peptide family genes (*RLN3*, *INSL3*, *INSL5*, and *INSL6*) in *Bos taurus*, whereas three

relaxin peptide genes (*RLN3*, *INSL3*, and *INSL6*) in *Bubalus bubalis* in contrast to seven genes in most of the mammals. The loss of genes might be the result of adaptation to specific niches during evolution. Relaxin family peptides remained conserved during evolution. Nonsynonymous mutations in *RLN3*, *INSL3*, and *INSL6* can interfere with biological functions like spermatogenesis, testicular descent, and feed metabolism in bovines. The segmental duplication in *Bos taurus* and the tandem duplication in *Bubalus bubalis* of relaxin family peptides helped in enrichment to specific functions like reproduction and feed metabolism during evolution.

## Data Availability

All data are shown within the manuscript.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Authors' Contributions

Muhammad Saif-ur Rehman and Faiz-ul Hassan contributed equally to this manuscript.

## Acknowledgments

This study was funded by the "National Centre for Livestock Breeding, Genetics and Genomics (NCLBGG)", Subcentre-UAF, Pakistan.

## Supplementary Materials

Table S1: Relaxin peptides gene family members' nucleotide and protein sequences accessions from NCBI database (Genbank accessions) for *Bos taurus*, *Bubalus bubalis*, *Capra hircus*, *Ovis aries*, *Camelus dromedarius* and *Equus caballus*. Table S2: Mutational effects predicted through different online software in relaxin peptides of *Bos taurus* and *Bubalus bubalis*. (*Supplementary Materials*)

## References

- [1] T. N. Wilkinson, T. P. Speed, G. W. Tregear, and R. A. D. Bathgate, "Evolution of the relaxin-like peptide family," *BMC Evolutionary Biology*, vol. 5, no. 1, pp. 14–17, 2005.
- [2] L. Jiaying Chan, M. Akhter Hossain, C. S. Samuel, F. Separovic, and J. D. Wade, "The relaxin peptide family – structure, function and clinical applications," *Protein and Peptide Letters*, vol. 18, no. 3, pp. 220–229, 2011.
- [3] B. A. Evans, P. Fu, and G. W. Tregear, "Characterization of two relaxin genes in the chimpanzee," *The Journal of Endocrinology*, vol. 140, no. 3, pp. 385–392, 1994.
- [4] S. Y. T. Hsu, "New insights into the evolution of the relaxin-LGR signaling system," *Trends in Endocrinology and Metabolism*, vol. 14, no. 7, pp. 303–309, 2003.
- [5] R. A. D. Bathgate, M. L. Halls, E. T. van der Westhuizen, G. E. Callander, M. Kocan, and R. J. Summers, "Relaxin family

- peptides and their receptors," *Physiological Reviews*, vol. 93, no. 1, pp. 405–480, 2013.
- [6] L. Malone, J. C. Opazo, P. L. Ryan, and F. G. Hoffmann, "Progressive erosion of the Relaxin1 gene in bovinds," *General and Comparative Endocrinology*, vol. 252, pp. 12–17, 2017.
- [7] S. D. Perkins, C. N. Key, M. N. Marvin et al., "Effect of residual feed intake on hypothalamic gene expression and meat quality in Angus-sired cattle grown during the hot season1,2," *Journal of Animal Science*, vol. 92, no. 4, pp. 1451–1461, 2014.
- [8] R. Ivell, S. Hartung, and R. A. Anand-Ivell, "Insulin-like factor 3: where are we now?," *Annals of the New York Academy of Sciences*, vol. 1041, no. 1, pp. 486–496, 2005.
- [9] D. Chassin, A. Laurent, J.-L. Janneau, R. Berger, and D. Bellet, "Cloning of a new member of the insulin gene superfamily (INSL4) expressed in human placenta," *Genomics*, vol. 29, no. 2, pp. 465–470, 1995.
- [10] D. Hechter, B. Vahkal, T. Tiede, and S. V. Good, "Reviewing the physiological roles of the novel hormone-receptor pair INSL5-RXFP4: a protective energy sensor?," *Journal of Molecular Endocrinology*, vol. 69, no. 1, pp. R45–R62, 2022.
- [11] Y. Huang, Q. Fu, H. Pan et al., "Spermatogenesis-associated proteins at different developmental stages of buffalo testicular seminiferous tubules identified by comparative proteomic analysis," *Proteomics*, vol. 16, no. 14, pp. 2005–2018, 2016.
- [12] S. U. Rehman, T. Feng, S. Wu et al., "Comparative genomics, evolutionary and gene regulatory regions analysis of casein gene family in *Bubalus bubalis*," *Frontiers in Genetics*, vol. 12, article 662609, 2021.
- [13] M. S. Rehman, F. U. Hassan, Z. U. Rehman, I. Ishtiaq, S. U. Rehman, and Q. Liu, "Molecular characterization of TGF-beta gene family in buffalo to identify gene duplication and functional mutations," *Genes*, vol. 13, no. 8, p. 1302, 2022.
- [14] M. S. Rehman, M. Mushtaq, F. Hassan, Z. Rehman, M. Mushahid, and B. Shokrollahi, "Comparative genomic characterization of insulin-like growth factor binding proteins in cattle and Buffalo," *BioMed Research International*, vol. 2022, Article ID 5893825, 15 pages, 2022.
- [15] S. U. Rehman, A. Nadeem, M. Javed et al., "Genomic identification, evolution and sequence analysis of the heat-shock protein gene family in buffalo," *Genes*, vol. 11, no. 11, p. 1388, 2020.
- [16] NCBI: National Center for Biotechnology Information April 2022, <https://www.ncbi.nlm.nih.gov/>.
- [17] R. D. Finn, J. Clements, and S. R. Eddy, "HMMER web server: interactive sequence similarity searching," *Nucleic Acids Research*, vol. 39, pp. W29–W37, 2011.
- [18] Uniprot, "Protein databse," April 2022, <https://www.uniprot.org/>.
- [19] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.
- [20] T. L. Bailey, M. Boden, F. A. Buske et al., "MEME SUITE: tools for motif discovery and searching," *Nucleic Acids Research*, vol. 37, no. Web Server, pp. W202–W208, 2009.
- [21] B. Hu, J. Jin, A.-Y. Guo, H. Zhang, J. Luo, and G. Gao, "GSDS 2.0: an upgraded gene feature visualization server," *Bioinformatics*, vol. 31, no. 8, pp. 1296–1297, 2015.
- [22] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, and A. Bairoch, "Protein identification and analysis tools on the ExpASY server," *The proteomics protocols handbook*, pp. 571–607, Humana Totowa, NJ, 2005.
- [23] "MSA: multiple align show," April 2022, [https://www.bioinformatics.org/sms/multi\\_align.html](https://www.bioinformatics.org/sms/multi_align.html).
- [24] "NHR-SCAN: Nuclear hormone receptor scan," April 2022, [http://nhrscan.genereg.net/cgi-bin/nhr\\_scan.cgi](http://nhrscan.genereg.net/cgi-bin/nhr_scan.cgi).
- [25] Y. Wang, H. Tang, J. D. DeBarry et al., "MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity," *Nucleic Acids Research*, vol. 40, no. 7, pp. e49–e49, 2012.
- [26] J. Rozas, A. Ferrer-Mata, J. C. Sánchez-DelBarrio et al., "DnaSP 6: DNA sequence polymorphism analysis of large data sets," *Molecular Biology and Evolution*, vol. 34, no. 12, pp. 3299–3302, 2017.
- [27] M. S. Tantia, R. K. Vjih, V. Bhasin et al., "Whole-genome sequence assembly of the water buffalo (*Bubalus bubalis*)," *The Indian Journal of Animal Sciences*, vol. 81, no. 5, p. 38, 2011.
- [28] S. U. Rehman, F. Hassan, X. Luo, Z. Li, and Q. Liu, "Whole-genome sequencing and characterization of buffalo genetic resources: recent advances and future challenges," *Animals*, vol. 11, no. 3, p. 904, 2021.
- [29] W. Li, D. M. Bickhart, L. Ramunno, D. Iamartino, J. L. Williams, and G. E. Liu, "Genomic structural differences between cattle and river buffalo identified through comparative genomic and transcriptomic analysis," *Data in brief*, vol. 19, pp. 236–239, 2018.
- [30] F. G. Hoffmann and J. C. Opazo, "Evolution of the relaxin/insulin-like gene family in placental mammals: implications for its early evolution," *Journal of Molecular Evolution*, vol. 72, no. 1, pp. 72–79, 2011.
- [31] S. V. Good-Avila, S. Yegorov, S. Harron et al., "Relaxin gene family in teleosts: phylogeny, syntenic mapping, selective constraint, and expression analysis," *BMC Evolutionary Biology*, vol. 9, no. 293, pp. 1–19, 2009.
- [32] J.-I. Park, J. Semyonov, C. L. Chang, W. Yi, W. Warren, and S. Y. T. Hsu, "Origin of INSL3-mediated testicular descent in therian mammals," *Genome Research*, vol. 18, no. 6, pp. 974–985, 2008.
- [33] J.-I. Park, J. Semyonov, W. Yi, C. L. Chang, and S. Y. T. Hsu, "Regulation of receptor signaling by relaxin A chain motifs," *The Journal of Biological Chemistry*, vol. 283, no. 46, pp. 32099–32109, 2008.
- [34] C. Schwabe, E. E. Böllesbach, H. Heyn, and M. Yoshioka, "Cetacean relaxin," *The Journal of Biological Chemistry*, vol. 264, no. 2, pp. 940–943, 1989.
- [35] A. Belfiore, F. Frasca, G. Pandini, L. Sciacca, and R. Vigneri, "Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease," *Endocrine Reviews*, vol. 30, no. 6, pp. 586–623, 2009.
- [36] A. Belfiore, R. Malaguarnera, V. Vella et al., "Insulin receptor isoforms in physiology and disease: an updated view," *Endocrine Reviews*, vol. 38, no. 5, pp. 379–431, 2017.
- [37] L. W. Bowers, E. L. Rossi, C. H. O'Flanagan, L. A. deGraffenried, and S. D. Hursting, "The role of the insulin/IGF system in cancer: lessons learned from clinical trials and the energy balance-cancer link," *Frontiers in Endocrinology*, vol. 6, p. 77, 2015.
- [38] K. I. Avgerinos, N. Spyrou, C. S. Mantzoros, and M. Dalamaga, "Obesity and cancer risk: emerging biological mechanisms and perspectives," *Metabolism*, vol. 92, pp. 121–135, 2019.

- [39] I. Atsushi, "Thermostability and aliphatic index of globular proteins," *Journal of Biochemistry*, vol. 88, no. 6, pp. 1895–1898, 1980.
- [40] K. Guruprasad, B. V. B. Reddy, and M. W. Pandit, "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence," *Protein Engineering, Design & Selection*, vol. 4, no. 2, pp. 155–161, 1990.
- [41] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.
- [42] L. Wei, Y. Liu, I. Dubchak, J. Shon, and J. Park, "Comparative genomics approaches to study organism similarities and differences," *Journal of Biomedical Informatics*, vol. 35, no. 2, pp. 142–150, 2002.
- [43] B. Spiller, A. Gershenson, F. H. Arnold, and R. C. Stevens, "A structural view of evolutionary divergence," *Proceedings of the National Academy of Sciences*, vol. 96, no. 22, pp. 12305–12310, 1999.
- [44] R. Anand-Ivell, S. Hiendleder, C. Viñoles et al., "INSL3 in the ruminant: a powerful indicator of Gender- and genetic-specific feto-maternal dialogue," *PLoS One*, vol. 6, no. 5, article e19821, 2011.
- [45] P. Canto, I. Escudero, D. Söderlund et al., "A novel mutation of the insulin-like 3 gene in patients with cryptorchidism," *Journal of Human Genetics*, vol. 48, no. 2, pp. 0086–0090, 2003.
- [46] S. Feng, V. K. Cortessis, A. Hwang et al., "Mutation analysis of INSL3 and GREAT/LGR8 genes in familial cryptorchidism," *Urology*, vol. 64, no. 5, pp. 1032–1036, 2004.
- [47] O. Burnicka-Turek, K. Shirneshan, I. Paprotta et al., "Inactivation of insulin-like factor 6 disrupts the progression of spermatogenesis at late meiotic prophase," *Endocrinology*, vol. 150, no. 9, pp. 4348–4357, 2009.
- [48] E. H. Davidson, *Genomic regulatory systems: in development and evolution*, Elsevier, 2001.
- [49] A. Aranda and A. Pascual, "Nuclear hormone receptors and gene expression," *Physiological Reviews*, vol. 81, no. 3, pp. 1269–1304, 2001.
- [50] D. Cotnoir-White, D. Laperrière, and S. Mader, "Evolution of the repertoire of nuclear receptor binding sites in genomes," *Molecular and Cellular Endocrinology*, vol. 334, no. 1–2, pp. 76–82, 2011.
- [51] I. A. Vergara and N. Chen, "Large synteny blocks revealed between *Caenorhabditis elegans* and *Caenorhabditis briggsae* genomes using OrthoCluster," *BMC Genomics*, vol. 11, no. 1, p. 516, 2010.
- [52] G. Zhang, B. Li, C. Li, M. T. P. Gilbert, E. D. Jarvis, and J. Wang, "Comparative genomic data of the Avian Phylogenomics Project," *Gigascience*, vol. 3, no. 1, p. 2047, 2014.
- [53] K. L. Howe, B. J. Bolt, S. Cain et al., "WormBase 2016: expanding to enable helminth genomic research," *Nucleic Acids Research*, vol. 44, no. D1, pp. D774–D780, 2016.
- [54] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, pp. 860–921, 2001.
- [55] J. C. Venter, M. D. Adams, E. W. Myers et al., "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.
- [56] P. W. H. Holland, J. Garcia-Fernández, N. A. Williams, and A. Sidow, "Gene duplications and the origins of vertebrate development," *Development*, vol. 1994, pp. 125–133, 1994.
- [57] K. H. Wolfe, "Yesterday's polyploids and the mystery of diploidization," *Nature Reviews. Genetics*, vol. 2, no. 5, pp. 333–341, 2001.
- [58] G. E. Liu, M. Ventura, A. Cellamare et al., "Analysis of recent segmental duplications in the bovine genome," *BMC Genomics*, vol. 10, no. 571, pp. 1–16, 2009.
- [59] L. D. Hurst, "The Ka/Ks ratio: diagnosing the form of sequence evolution," *Trends in genetics: TIG*, vol. 18, no. 9, pp. 486–487, 2002.