



Interferon Regulatory Factor 3 (IRF3) Characterisation in Four Mammalian Species Using Bioinformatics Method

Sanda MOFOYEKE^{1*}, Adenaike ADEYEMI¹, Fafiolu ADEBOYE², Adeyinka SANDA¹, Christian IKEOBI¹

¹ Animal Breeding and Genetics Department, College of Animal Science and Livestock Production, Federal University of Agriculture Abeokuta, Alabata Road, Abeokuta, Ogun State, P.M.B 2240, Nigeria

² Animal Nutrition Department, College of Animal Science and Livestock Production, Federal University of Agriculture Abeokuta, Alabata Road, Abeokuta, Ogun State, P.M.B 2240, Nigeria

* Corresponding author: S. Mofoyeke e-mail: mofoyekesanda@gmail.com

REVIEW

Abstract

IRF3 has gained recognition as a key transcriptional regulator of type1 interferon (IFN) gene expression which induces response to pathogenic infection. The advancement in technology and molecular biology has led to bioinformatics, this field is widely accepted because it is effective and saves time. This study made use of bioinformatics approach to identify and characterize IRF3 in four mammalian species namely: cattle, rabbit, pig and water buffalo. Evolutionary relationships inferred among the IRF3 of four mammalian and other species showed a high level of relatedness. Leucine was the most abundant residue while serine which plays a major role in maintenance of a healthy immune system was present in all the species. The information obtained in this study could serve as a guide when using the IRF3 gene as a marker to select animals for disease resistance because of the critical role they play in antiviral defense.

Keywords: Interferon regulatory factor; bioinformatics; mammals.

INTRODUCTION

The immune system is essentially important for the survival of any species and also the main determinants of productivity. With the current concerns about food security, this highlights the importance of maintaining a productive and disease resistant livestock population. Interferons (IFNs) are antiviral agents and they modulate functions of the immune system, they also limit viral spread by increasing activity, which kills virus-infected cells by promoting apoptosis and the effect of IFN on tumor protein P53 (p53) is also linked to its protective role against certain cancers (Takaoka et al., 2003). Moreover, IFNs are a group of signalling proteins made and released by host cells in response to the presence of pathogens, such as viruses, bacteria, parasites, or tumour cells (De Andrea et al., 2002). In a typical scenario, a virus-infected cell will release interferons causing nearby cells to heighten their anti-viral defense. More than twenty distinct IFN genes and proteins have been identified in animals, including humans. They are typically divided among three classes: Type I IFN, Type II IFN, and Type III IFN. IFNs belonging to all three classes are important for fighting viral infections and for the regulation of the immune system. In general, type I and II interferons are responsible for regulating and activating the immune response (Cohen and Parkin, 2001). IFNs activate signal transducer and activator of transcription (STAT) complexes. STATs are a family of transcription factors that regulate the expression of certain immune system genes. The regulatory elements play

Received: 09 July 2020

Accepted: 08 February 2021

Published: 15 May 2021

DOI:

10.15835/buasvmcn-asb:2020.0013



© 2021 Authors. The papers published in this journal are licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

a major role in transcription and they can be broadly divided into two classes, the proximal or core promoter elements which occur close to the initiation site of transcription and the enhancer or silencer elements which act at distance to regulate basal levels of transcription. In both classes of regulatory elements, there is presence of clusters of transcription factor binding sites (TFBSs) (Arnone and Davidson, 1997).

Interferon regulatory factors (IRFs) are proteins which regulate transcription of interferons likewise they are transcription mediators of virus and bacteria. They play a critical role in antiviral defence, immune response, cell growth regulation and apoptosis (Paun and Pitha, 2007). The IRFs family plays a critical role in the host defence system (Honda and Taniguchi, 2006), these IRFs regulate many protein-encoding genes involved in immune responses, oncogenesis, and cell development and differentiation. IRF3 plays a critical role in antiviral response (Lu and Pitha, 2001). Moreover, in the report of Taniguchi et al. (2018) IRF3 is described as a transcription regulator of cellular responses in many cell types which is known to be essential for innate immunity. Similarly, IRF3 was reported to be bifunctional contributing to tumour suppression via STING pathway and tumour progression by unknown mechanism.

According to Tang (2002), biologists have to rely more on computers to organise, store and efficiently make use of data for analyses owing to the fact that vast amounts of data on DNA sequences and proteins has been generated from the Human Genome Project and from laboratories around the world. With the accumulation of nucleotide and protein sequences, computational methods such as bioinformatics can be used to give some indication of the structure and function of folds of protein that are becoming increasingly important (Singh et al., 1999). The most functional unit in the protein sequence is called a motif. Motifs usually correspond to short sequence fragments that reflect some biological role in terms of structure and function (Bujnicki, 2009). Also, in molecular biology the structure or fold of a protein provides the key to understanding the biological function of a particular protein and proteins play a variety of important roles in the body. There is also a need to provide an insight into the evolutionary relationship that exists in IRFs among mammals and other species. This study employed the use of bioinformatics tools to identify, analyse and characterise IRF3 in four mammalian species of animals namely cattle, rabbit, pig and water buffalo.

MATERIALS AND METHODS

Retrieval of nucleotide and amino acid sequences

Nucleotide and amino acid sequences of IRF3 of four mammalian species namely cattle, rabbit, pig and water buffalo together with other species namely goat, sheep, chicken, turkey, Arabian camel Norway rat, wanderfalke, turkey, common quail, chicken, atlantic salmon, domestic ferret, rhesus monkey and African savannah elephant were downloaded from the GenBank. A FASTA format of the nucleotide and amino acid sequences was obtained at the National Centre for Biotechnology Information (NCBI, USA) and the Basic Local Alignment Search Tool (BLASTn) option was used to obtain similar sequences in other organisms. The GenBank accession numbers of the nucleotide and amino acid sequences, the gene name and protein name as well as the sequence length and the name of the organism were retrieved and tabulated.

Identification of putative promoters and determination of GC content

The nucleotide sequences of the IRF3 mammalian species retrieved were used to identify the putative promoter and determine the GC (Guanine Cytosine) content. The promoter regions of the interferon regulatory factor of ten mammalian species were then identified using the Neutral Network Promoter Prediction Algorithm (Reese et al., 1996). These promoter regions were identified from the sequences earlier retrieved by downloading 1000 base pairs upstream of the interferon regulatory factor. The GC was determined using GENSCAN software (Burge and Karlin, 1997) and this software also predicted the coding region of the genomic sequences.

Identification of introns, exons, untranslated regions, start and stop codons

The nucleotide sequences of the IRF3 mammalian species retrieved was also used to identify the introns, exons, untranslated regions, start and stop codon in ten mammalian species. The identification process was carried out using the GENSCAN software (Burge and Karlin, 1997).

Multiple sequence alignment

The multiple sequence alignment was carried out on all the obtained nucleotide and amino acid sequences using CLUSTAL W software (Thompson et al., 1994). In aligning sequences, BLOCK sum 80 Algorithm of CLUSTAL W software was used in order to ensure strict penalty for gap extension and penalty.

Phylogenetic analysis

Phylogenetic tree was constructed to give an insight into the evolutionary relationship between IRF3 of four mammalian species and the IRF3 of other species. The phylogenetic tree was constructed with Molecular Evolution and Genetic Analysis (MEGA6) software (Tamura et al., 2013) using the amino acid sequences of IRF3 of four mammalian species and the IRF3 of other species. The evolutionary history was inferred using the UPGMA method

and the evolutionary distances were computed using the Poisson correction method (Zuckerkindl and Pauling, 1965).

Determination of amino acid composition

The protein sequences of IRF3 of the four mammalian species retrieved were used for the amino acid composition. Their amino acid composition was determined using the FASTA format of amino acid sequence which was submitted into the Network protein sequence analysis, a server in Pole bioinformatique Lyonnais (<http://nps.org>).

Determination of physico-chemical properties

The FASTA format of the protein sequences of IRF3 of four mammalian species was used to determine the various physico-chemical properties namely molecular weight, theoretical pl, aliphatic index, instability index, grand average hydrophobicity (GRAVY), total number of negatively charged residues (aspartate + glutamine) and total number of positively charged residues (arginine + lysine).

The various physico-chemical properties were determined using Swiss Prot and Expert Protein Analysis System (ExPASy) which is the proteomic server of Swiss Institute of Bioinformatics (SIB) (<http://web.expasy.org>). Various tools in the proteomic server such as ProtParam, Protein Calculator, Compute pl/Mw and Protscale were used to deduce the physico-chemical properties.

Prediction of motifs

The protein sequences were used to predict the motifs present in the sequences, this was carried out using the PROSITE software (De Castro et al., 2006).

RESULTS AND DISCUSSIONS

Identification of Promoters, Introns, exons and untranslated regions of IRF3 in four mammalian species

Nucleotide sequences information in several mammalian species largely varies (Table 1) but also the physico-chemical properties of IRF3 in four mammalian species we are discussing here (Table 2).

Table 1. Retrieved nucleotide and amino acid sequences of IRF3 in several mammalian species

Genbank accession number	Protein accession number	Species	Gene name	Protein name	Nucleotide sequence length	Amino acid Sequence length
NM_001029845.3	NP_001025016.1	Cattle (<i>Bos tarus</i>)	<i>Bos tarus</i> interferon regulatory factor 3 (IRF3) mRNA	Interferon regulatory factor 3	1564	417
XM_006080088.1	XP_006080150.1	Water buffalo (<i>Bubalus bubalis</i>)	<i>Bubalus bubalis</i> interferon regulatory factor 3(IRF3) transcript variant XI mRNA	Interferon regulatory factor 3(IRF3) isoform XI	1865	456
XM_008252813.1	XP_008251035.1	Rabbit (<i>Oryctolagus cuniculus</i>)	<i>Oryctolagus cuniculus</i> interferon regulatory factor 3 (IRF3) mRNA	Interferon regulatory factor 3	1438	435
KC 860781.1	AGS09135.1	Pig (<i>Sus scrofa</i>)	<i>Sus scrofa</i> interferon regulatory factor 3(IRF3) mRNA, complete cds	Interferon regulatory factor 3	1260	491
XM_004767292.1	XP_004767349.1	Domestic ferret (<i>Mustela putorius furo</i>)	<i>Mustela putorius furo</i> interferon regulatory factor 3(IRF3) transcript variant X3 mRNA	Interferon regulatory factor 3 (IRF3) isoform X3	1930	294
EU204919.1	ABY64977.1	Rhesus monkey (<i>Macaca mulatta</i>)	<i>Macaca mulatta</i> interferon regulatory factor 3(IRF3) mRNA Complete cds	Interferon regulatory factor 3	1328	422
BT048453.1	AC168254.1	Atlantic salmo	<i>Salmo salar</i> clone ssal-rgb2-603-278 interferon regulatory factor 3	Interferon regulatory factor 3	845	237

		(<i>Salmo salar</i>)	putative mRNA complete cds			
HQ22991	AD052204	Goldfish (<i>Carassius auratus</i>)	<i>Carassius auratus</i> interferon regulatory factor 3 mRNA complete cds	Interferon regulatory factor 3	1839	458
NM_001006969	NP_001006970	Norway rat (<i>Rattus norvegicus</i>)	<i>Rattus norvegicus</i> interferon regulatory factor 3(IRF3) mRNA	Interferon regulatory factor 3	1471	421
AB102887	BA089416	Human (<i>Homo sapiens</i>)	<i>Homo sapiens</i> interferon regulatory factor 3(IRF3) mRNA variant 3, complete cds	Interferon regulatory factor 3 variant 3	903	300

Table 2. Physico-chemical properties of IRF3 in four mammalian species

Characteristics	Cattle	Rabbit	Pig	Water buffalo
Molecular weight	46 645.7	47 851.1	46 614.4	50 910.5
Theoretical pI	4.99	5.12	4.99	4.94
Aliphatic index	79.26	76.46	79.14	78.03
Instability index	49.79	47.25	44.11	47.22
GRAVY	-0.448	-0.422	-0.433	0.416
Total number of negative charged residues (aspartate + glutamine)	41	39	41	43
Total number of positively charged residues (arginine + lysine)	6485	6644	6470	7059

Table 3 shows the result on identification of promoters, introns, exons and untranslated regions of IRF3 in four mammalian species. One promoter sequence was found on region 931-981 of cattle with a high cut-off score of 0.96 and it was picked as a putative promoter sequence because of the TATA sequence in this region. The terminal exon of 1262 bp on region 269-1530 bp was predicted so also the poly A signal region 1571-1576 bp, 3'UTR of 40 bp and GC content of 58.41%. The gene structure prediction analysis revealed a CDS of 1245 bp and a predicted peptide sequence of 417 amino acids.

Table 3. Promoter sequences predicted by Neutral Network promoter prediction algorithm with a score cut-off of 0.80 (transcription starts are in bold form)

Species	Start	End	Score	Promoter Sequences
Cattle	931	981	0.96	CTGGGCAGCCTATACGACTGCCGGACCCTGCGACGTCCTGACAGACAAG
Rabbit	95	145	0.81	TGTGGCCGGAAGACCGAAAAGGGGCGAGACGGGCGGGAGCATGGGGACGC
	313	363	0.96	CCAGCGGCGCCTATACTCCCGGGGATAGGCCCGACCTGCCAACCTGGAAG
Pig	1112	1162	0.90	GGGCCCTGGTAGACATGGCGGGGACGGGGGCGCCTCCTCCTGGAGAAC
Water	349	399	0.90	AGGAGGCATCTATATTGACGGATCTCGGGAGAGAAGGATCACTCTGCCAA
Buffalo	406	456	0.86	CAGTCCACTCTAGAAAATGGGGATATGCTCTGGGGGAGGTTGTTTGAGAG

In the case of rabbit two promoter sequences were found, the promoter on region 313-363 with high cut-off score of 0.96 was picked as a putative promoter sequence because of the presence of the TATA sequence. Poly A signal on region 1484-1489 bp and GC content of 66.94% was predicted while the gene structure prediction analysis revealed a CDS of 1263 bp and a predicted peptide sequence of 420 amino acids.

Pig also had one promoter sequence, found in region 1112-1162 with a high cut-off score of 0.90 but the TATA sequence was absent in this region. Only the GC content of 63.24%, CDS of 1260 bp and peptide sequence of 419 amino acids was predicted. While for water buffalo, two promoter sequences were found, the promoter on region 349-399 with high cut-off score of 0.86 was picked as a putative promoter sequence because of the presence of TATA sequence. Poly A signal on region 1922-1927 bp and GC content of 57.45% was predicted while the gene structure prediction analysis revealed a CDS of 1305 bp and a predicted peptide sequence of 434 amino acids.

Evolutionary relationship of IRF3

The evolutionary relationship between IRF3 of four mammals and other species is shown in Figure 1. The phylogenetic analysis of IRF3 based on amino acid sequence showed how the various species evolved on branches with species that are highly related and they clustered together on the same branch.

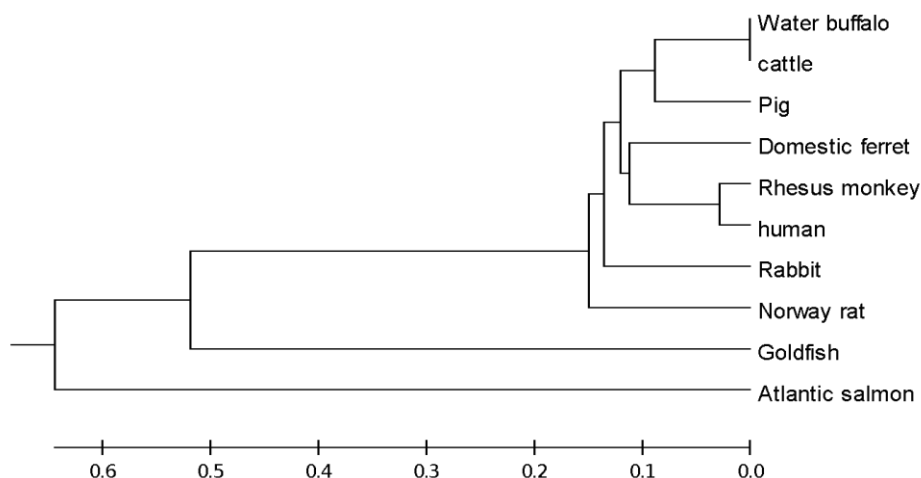


Figure 1. The evolutionary relationship of IRF3 in some species.

Amino acid composition of IRF3 in four mammalian species

Amino acid composition of IRF3 in four mammalian species is presented in Table 4 with cattle having protein containing more leucine (11.27%) and the least amino acid composition (1.68%) was predicted in tyrosine. Similarly, rabbit had more leucine (11.03%) in its protein and the least amino acid composition (1.38%) was predicted in asparagine.

Generally, the amino acid (leucine) composition was higher in the four mammals considered when compared with other amino acids. Pig and water buffalo had higher leucine composition of 11.69% and 10.75% respectively, as against the least amino acid composition of 1.91% each predicted in tyrosine, methionine, asparagine and cysteine for pig, while water buffalo had the least amino acid composition of 1.54% predicted in tyrosine.

Table 4. Amino acid composition of IRF3 protein in four mammalian species

Amino acid (%)	Cattle	Rabbit	Pig	Water buffalo
A (Ala)	5.28	7.59	6.92	5.26
B (Asx)	0.00	0.00	0.00	0.00
C (Cys)	2.40	2.30	1.91	2.41
D (Asp)	7.91	6.21	8.59	7.46
E (Glu)	5.76	6.90	5.73	6.14
F (Phe)	3.60	2.99	3.34	3.51
G (Gly)	8.15	9.89	8.59	9.21
H (His)	1.92	2.76	2.63	1.97
I (Ile)	3.60	1.84	2.39	3.51
K (Lys)	3.84	2.99	3.58	3.29
L (Leu)	11.27	11.03	11.69	10.75
M (Met)	1.92	2.07	1.91	2.41
N (Asn)	2.88	1.38	1.91	2.85
P (Pro)	7.67	9.66	7.40	7.02
Q (Gln)	6.00	6.21	5.01	5.70
R (Arg)	6.00	5.98	6.21	6.14
S (Ser)	7.19	5.98	7.40	7.02
T (Thr)	4.56	3.22	3.82	4.61
V (Val)	5.52	6.44	5.97	5.92
W (Trp)	2.88	2.76	3.10	3.29

X (Unk)	0.00	0.00	0.00	0.00
Y (Tyr)	1.68	1.84	1.91	1.54
Z (Glx)	0.00	0.00	0.00	0.00
Residue MEAN (Daltons)	111.86	110.04	111.25	111.65

Alanine (Ala), aspartic acid (Asx), cysteine (Cys), aspartate (asp), glutamate (Glu), phenylalanine (Phe), glycine (Gly), histidine (His), isoleucine (ile), lysine (Lys), leucine (Leu), methionine (Met), asparagine (Asn), proline (Pro), glutamine (Gln), arginine (Arg), serine (Ser), threonine (Thr), valine (Val), tryptophan (Trp), unknown amino acid (Unk), tyrosine (Tyr), glutamic acid (Glx).

Predicted motifs of IRF3 in four mammalian species

In all the four mammalian species different motifs were predicted. Cattle and water buffalo had the following motifs; N-glycosylation site, Protein Kinase C Phosphorylation site, Casein Kinase II Phosphorylation site and N-myristoylation site. Rabbit and pig's had Protein Kinase C Phosphorylation site, Casein Kinase II Phosphorylation site, N-myristoylation site and IRF tryptophan pentad repeat DNA binding domain signature.

The nucleotide sequences length varied from 903 bp to 1930 base pairs while the amino acid residues length varied from 237 to 491 amino acid residues. This variation observed in the nucleotide and amino acid sequence length of IRF3 can be attributed to evolution and differentiation. Kang et al. (2008) reported that sequence length variations are caused by insertions and deletions. Many length variations are caused by insertions and deletions that further result into amino acid variation within species and are found by comparison with known sequences (Yakubu et al., 2014).

The predicted putative promoter sequences of the domestic mammalian livestock contained transcription start site and general transcription binding sites (TATA); these are the minimal portion of the promoter required to properly initiate transcription of interferon regulatory factor. The location of the TATA box close to the transcription start site is in line with the findings of Smale and Kadonaga (2003) that TATA are located close to the transcription site. Although in pig the promoter observed had no TATA box which implied that the promoters are TATA-less. When TATA box is absent the corresponding promoters are referred to as TATA-less promoters (Solovyev and Shahmuradov, 2003). The highest GC content of 66.94 % was observed in the IRF3 of rabbits and this implies that the gene is stable, owing to stacking interaction between guanine and adenine (Yakovchuk *et al.*, 2006). Subsequently the GC content of a gene correlates positively with the melting temperature of the DNA double (Yakovchuk *et al.*, 2006). Hence the melting temperature of the IRF3 of rabbits will be high.

In eukaryotes the expected start codon is (AUG) and for IRF3 in the four mammalian species the observed start codon was AUG. There was variation in the usage of stop codon as the use of TAG as stop codon was observed in IRF3 of cattle, pig and water buffalo while the use of TGA as stop codon was observed in IRF3 of rabbit. This is in agreement with the findings of Li et al. (2007) who also found stop codon variation of the MSHR gene within and among different families. There might be a stop codon usage bias among species and for different genes (Ghosh, 2000). These variances among patterns of stop codon usage are possibly due to environmental adaptation (Wong et al., 2008). The non-usage of TAG as stop codon in IRF3 of rabbit could be attributed to high GC although, the usage of TAG as stop codons in IRF3 of cattle, pig and water buffalo with relatively high GC content is a deviation from the report of Wong et al. (2008) that the higher the GC content, the fewer TAG codons that could be found in the gene.

From the evolutionary findings it was revealed that there was a close relationship between the IRF3 of cattle and water buffalo. With the exception of IRF3 of Atlantic salmon which evolved along a separate clade being the least evolved amongst the species used in the study. Close relationship was also observed in the IRF3 of human and rhesus monkeys. Leucine was the most abundant amino acid residue in the IRF3 of the species examined and this is significant as leucine regulates several cellular processes such as protein synthesis, tissue regeneration and metabolism (Donato et al., 2015). Moreover, serine was also present in all the species examined; this relates to tissue growth and the immune system as it assists in the production of immunoglobulins and antibodies as confirmed in literature. Furthermore, serine is involved in the proper function of RNA, DNA, fat and fatty acid metabolism, muscle formation and the maintenance of a healthy immune system. Tryptophan, histidine, methionine and cysteine are amino acid residues that were less abundant in this study and this is confirmed in an earlier report by Fountoulakis and Lahm (1998). Based on the amino acid composition and the level at which the various amino acid residues are present, the IRF3 protein of cattle, rabbit, water buffalo and pig belong to aliphatic amino acid class this is as a result of the high aliphatic index values deduced. The IRF3 protein of water buffalo had the highest molecular weight while in the other species studied their molecular weights were relatively high. The theoretical pI of 5.12 in IRF3 protein of rabbit was the highest deduced and this indicates that the protein has no net charge at pH of 5.12 below this pH the protein transmits a positive charge while above the pH the protein transmits a negative charge. The lowest theoretical pI of 4.94 was deduced in IRF3 protein of water buffalo and this also implies that at the pH value of 4.94 the protein has no net charge; below this pH the protein transmits a positive charge while above this pH the protein transmits a negative charge. The isoelectric potential value deduced can be used to separate the interferon

regulatory factor protein from other proteins using isoelectric focusing when running electrophoretic gels in the laboratory.

In lieu of all the domestic mammalian livestock used in this study, the GRAVY score obtained was negative and this is an indication that the interferon regulatory factor protein is relatively a hydrophilic protein. Negative GRAVY values indicate a relatively hydrophilic protein, GRAVY values greater than zero indicate a relatively hydrophobic protein (Kyte and Doolittle, 1982).

Negatively charged residues were higher than positively charged residues in the interferon regulatory factor protein of domestic mammalian livestock. This is an indication that the IRF3 are intracellular proteins as these types of proteins have a higher fraction of negatively charged residues. Negatively charged residues like the ones from aspartate and glutamate play a key role in transmembrane ion transport (Lutsenko and Kaplan, 1995).

Relatively high values were deduced for the instability index and it can be implied that interferon regulatory factor protein is likely going to be unstable *in vivo*. Instability index is a measure of proteins used to determine whether a protein will be stable *in vivo* and values greater than 40 specify that the protein may be unstable *in vivo* while values less than 40 is an indication that protein may be stable *in vivo* (Guruprasad et al., 1990).

The presence of N-glycosylation sites in the IRF3 protein of cattle and water buffalo suggests that the protein is highly glycosylated in all these animals and this is in line with the report of Messner (1997) that glycosylated proteins are found in almost all living organisms that have been studied. Furthermore, the N-glycosylation of protein in eukaryotes is one of the most abundant post-translational modification reactions in approximately part of all known proteins in eukaryotes. N-glycosylation modifies protein properties such as stability and solubility make protein more stable to high temperature and pH (Sinclair and Elliott, 2005).

In eukaryotic proteins, N-myristoylation is essential to cell growth, survival (Rajala et al., 2000; Resh, 2004) also development and host defense (Seaton and Smith, 2008; Boisson et al., 2003). The presence of N-myristoylation sites in the IRF3 protein of all the species examined with glycine being predominant implies that these species will respond quickly against certain diseases. This result is in line with the findings of Morita et al. (2011) which identified N-myristoylated viral peptides as a novel class of cytotoxic lymphocytes.

Protein phosphorylation plays a critical role in the regulation of synapse function (Moss and Smart, 1996). Protein kinase casein phosphorylation is a family of phospholipid dependent serine/ threonine kinase that play key roles in many of the signaling pathways that control cellular growth, proliferation, differentiation and cell death (Steinberg, 2008; Newton, 2010). Protein kinase casein phosphorylation sites presence in the IRF3 protein serve a vital purpose of limiting viral spread by increasing p53 activity which kills virus infected cells by promoting apoptosis (Takaoka et al., 2003) thereby making them more resistant to diseases.

Casein kinase II phosphorylation site was present in all the IRF3 of the four mammalian species. Phosphorylation by casein kinase II promotes p53 dependent stimulation or repression of transcription of genes involved in growth control which leads to the suppression of cell growth (tumor) (Milne et al., 1992) this further shows that immune system cells are viable and growth is controlled.

CONCLUSIONS

The promoter sequence of the IRF3 gene in pigs had no TATA box and they are referred to as TATA-less promoters. Leucine was the most abundant amino acid residue while serine was present in all the four mammalian species which play a vital role of maintaining a healthy immune system. Negative GRAVY score was predicted in all the mammalian species which indicated that the protein is hydrophilic. The molecular weight of IRF3 in the four mammalian species was relatively high. As a result of the close evolutionary relationship in the IRF3 of cattle and water buffalo, this marker could assist in selection programmes since they are closely related and where an improvement in a species can be beneficial to the other species. Moreover, the presence of N-myristoylation, protein kinase casein phosphorylation and casein kinase II phosphorylation sites in IRF3 in all the four mammalian species is an indication that the gene can serve as a marker for high disease resistance in breeding programmes.

Funding Source: This research work was not funded by any grants.

Conflicts of Interest

The authors declare that they do not have any conflict of interest.

REFERENCES

1. Arnone MI, Davidson EH. The hardwiring of development: Organization and function of genomic regulatory systems. *Dev.* 1997; 124:1851-64.
2. Boisson B, Giglione C, Meinnel T. Unexpected protein families including cell defense components feature in the N-Myristoylome of a higher eukaryote. *J Biol Chem.* 2003; 278(44):43418-229.

3. Bujnicki JM. Prediction of protein Structures, Functions and Interactions. 1st ed. John Wiley& Sons Ltd Great Britain; 2009. p. 6.
4. Burge C Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol.* 1997; 268:78-94.
5. Cohen BPJ. An overview of the immune system. *Lancet* 2001; 357(9270):1777-89.
6. De-Andrea M, Gariglio M, Gioia D, Landolfo SRR. The interferon system: an overview. *Eur J Paediatr Neurol.* 2002; 6:41-58.
7. De-Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A Hulo N. ScanProsite: Detection of PROSITE Signature matches and ProRule associated functional and structural residues in proteins. *Nucleic Acids Res.* 2006; 34:W362-W65.
8. Donato J, Pedroso JAB, Zampini TT. Reviewing the effects of L-Leucine supplementation in the regulation of food intake, energy balance and glucose homeostasis. *Nutrients* 2015; 7(5): 3914-37.
9. Fountoulakis MLH. Hydrolysis and amino acid composition analysis of proteins. *J Chromatogr A.* 1998; 826:109-34.
10. Ghosh T. Studies on codon usage in *Entamoeba histolytica*. *Int J Parasitol.* 2000; 30:715-22.
11. Guruprasad K, Reddy BV, Pandit MW. 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 Eng.* 1990; 4(2):155-61.
12. Honda K, Taniguchi T. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat Rev Immunol.* 2006; 6:644-58.
13. Kang JF, Li XL, Zhou RY, Li LH, Feng FJ Guo XL. Bioinformatics analysis of lactoferrin genes for several species. *Biochem Genet.* 2008; 46:312-22.
14. Kyte J Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982; 157:105-32.
15. Li XL, Zheng GR, Zhou RY. Evolution and differentiation of MSHR genes in different species. *J Hered.* 2007; 3:165-68.
16. Lu R Pitha PM. Monocyte differentiation to macrophage requires interferon regulatory factor 7. *J Biol Chem.* 2001; 276:45491-6.
17. Lutsenko S, Kaplan JH. Organisation of P- type ATPases significance of structural diversity. *Biochem.* 1995; 34(48):15607-13.
18. Messner P. Bacterial glycoprotein. *Glycoconj J.* 1997; 14:3-11.
19. Milne DM, Palmer RH, Campbell DG, Meek DW. Phosphorylation of the p53 tumor suppressor protein at three N- terminal sites by a novel Casein Kinase like enzyme. *Oncogene.* 1992; 7:1361-9.
20. Morita D, Igarashi T, Honike M, Mori N, Sugita M. T cells monitor N-myristoylation of the Nef protein in Simian immunodeficiency virus infected monkeys. *J Immunol.* 2011; 187:608-612.
21. Moss SJ, Smart TG. Modulation of amino acid-gated ion channels by protein phosphorylation. *Int Rev Neurobiol;* 1996; 39:1-52.
22. Newton AC (2010). Protein Kinase C: poised to signal. *Am J Physiol Endocrinol Metab.* 2010; 298(3):E395-402.
23. Paun A, Pitha PM. The IRF family, revisited. *Biochimie;* 2007; 89(6-7):744-53.
24. Rajala RV, Datta RS, Moyana TN, Kakkar R, Carlsen SA Sharma RK. N Myristoyl Transferase. *Mol Cell Biochem.* 2000; 204:135-56.
25. Reese MG, Harris NL, Eeckman FH. Large scale sequencing specific neutral networks for promoter and splice site recognition. *Proceedings of the Biocomputing Pacific Symposium.* 1996; 345-9.
26. Resh MD. A myristoyl switch regulates membrane binding of HIV 1 Gag. *Proc Natl Acad Sci.* 2004; 101(2):417.
27. Seaton KE, Smith CD. N-myristoyltransferase isoenzymes exhibit differential specificity for human immunodeficiency virus type 1 Gag and Nef. *J Gen Virol.* 2008; 89(Pt 1):288-296.
28. Sinclair AM, Elliott S. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. *J Pharm Sci.* 2005; 94(8):1626-35.
29. Singh M, Berger B, Kim PS. LearnCoil-VMF: Computational evidence for coiled coil-like motifs in many viral membrane-fusion proteins. *J Mol Biol.* 1999; 290:1031-41.
30. Smale T, Kadonaga T. "The RNA polymerase II core promoter". *Annu Rev Biochem.* 2003; 72:449-79.

31. Solovyev VV, Shahmuradov IA. PromH: promoters identification using orthologous genomic sequences. *Nucleic Acids Res.* 2003; 31(13): 3540-5.
32. Steinberg SF. Structural bases of Protein kinase C isoform function. *Physiol Rev.* 2008; 88(4):1341-78.
33. Tang C. Bioinformatics: Life Science Research In Silico. *J Young Investig.* 2002; 6(4):1-3.
34. Taniguchi T, Yanai H, Chiba S, Hangai S, Kometani K, Inoue A, Kimura Y, Abe T, Kiyonari H, Nishio J, Taguchi-Atarashi N, Mizushima Y, Negishi H, Grosschedl R. Revisiting the role of IRF3 in inflammation in immunity by conditional and specifically targeted gene ablation in mice. *PNAS.* 2018; 115(20): 5253-8.
35. Takaoka A, Hayakawa S, Yanai H, Stoiber D, Negishi H, Kikuchi H. Integration of interferon-alpha/beta signalling to p53 responses in tumour suppression and antiviral defence. *Nature.* 2003; 424 (6948): 516-523.
36. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6. *Mol Biol Evol.* 2013; 30(12): 2725-9.
37. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22:4673-80.
38. Wong T, Fernandes S, Sankhon N, Leong PP, Kuo J, Liu J. Role of Premature Stop Codons in Bacterial Evolution. *J Bacteriol.* 2008; 190(20):6718-25.
39. Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* 2006; 34(2):564-74.
40. Yakubu A, Alade DE, Dim NI. Molecular analysis of solute carrier family 11 member a1 (SLC11a1) gene in ruminants and non-ruminants using Computational method. *GENETIKA.* 2014; 46 (3):925-34.
41. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. *Evolving genes and Proteins.* Academic Press New York, 1956. p. 97-166.